13 Principal Components and Factor Analysis

13.1 General Purpose and Description

Principal components analysis (PCA) and factor analysis (FA) are statistical techniques applied to a single set of variables when the researcher is interested in discovering which variables in the set form coherent subsets that are relatively independent of one another. Variables that are correlated with one another but largely independent of other subsets of variables are combined into factors.[1] Factors are thought to reflect underlying processes that have created the correlations among variables.

Suppose, for instance, a researcher is interested in studying characteristics of graduate students. The researcher measures a large sample of graduate students on personality characteristics, motivation, intellectual ability, scholastic history, familial history, health and physical characteristics, etc. Each of these areas is assessed by numerous variables; the variables all enter the analysis individually at one time, and correlations among them are studied. The analysis reveals patterns of correlation among the variables that are thought to reflect underlying processes affecting the behavior of graduate students. For instance, several individual variables from the personality measures combine with some variables from the motivation and scholastic history measures to form a factor measuring the degree to which a person prefers to work independently—an independence factor. Several variables from the intellectual ability measures combine with some others from scholastic history to suggest an intelligence factor.

A major use of PCA and FA in psychology is in development of objective tests for measurement of personality and intelligence and the like. The researcher starts out with a very large number of items reflecting a first guess about the items that may eventually prove useful. The items are given to randomly selected subjects, and factors are derived. As a result of the first factor analysis, items are added and deleted, a second test is devised, and that test is given to other randomly selected subjects. The process continues until the researcher has a test with numerous items forming several factors that represent the area to be measured. The validity of the factors is tested in research where predictions are made regarding differences in the behavior of persons who score high or low on a factor.

The specific goals of PCA or FA are to summarize patterns of correlations among observed variables, to reduce a large number of observed variables to a smaller number of factors, to provide an operational definition (a regression equation) for an underlying process by using observed

[1]PCA produces components while FA produces factors, but it is less confusing in this section to call the results of both analyses factors.

variables, or to test a theory about the nature of underlying processes. Some or all of these goals may be the focus of a particular research project.

PCA and FA have considerable utility in reducing numerous variables down to a few factors. Mathematically, PCA and FA produce several linear combinations of observed variables, where each linear combination is a factor. The factors summarize the patterns of correlations in the observed correlation matrix and can be used, with varying degrees of success, to reproduce the observed correlation matrix. But since the number of factors is usually far fewer than the number of observed variables, there is considerable parsimony in using the factor analysis. Further, when scores on factors are estimated for each subject, they are often more reliable than scores on individual observed variables.

Steps in PCA or FA include selecting and measuring a set of variables, preparing the correlation matrix (to perform either PCA or FA), extracting a set of factors from the correlation matrix, determining the number of factors, (probably) rotating the factors to increase interpretability, and, finally, interpreting the results. Although there are relevant statistical considerations to most of these steps, an important test of the analysis is its interpretability.

A good PCA or FA "makes sense"; a bad one does not. Interpretation and naming of factors depend on the meaning of the particular combination of observed variables that correlate highly with each factor. A factor is more easily interpreted when several observed variables correlate highly with it and those variables do not correlate with other factors.

Once interpretability is adequate, the last, and very large, step is to verify the factor structure by establishing the construct validity of the factors. The researcher seeks to demonstrate that scores on the latent variables (factors) covary with scores on other variables, or that scores on latent variables change with experimental conditions as predicted by theory.

One of the problems with PCA and FA is that there are no readily available criteria against which to test the solution. In regression analysis, for instance, the DV is a criterion and the correlation between observed and predicted DV scores serves as a test of the solution—similarly for the two sets of variables in canonical correlation. In discriminant function analysis, logistic regression, profile analysis, and multivariate analysis of variance, the solution is judged by how well it predicts group membership. But in PCA and FA, there is no external criterion such as group membership against which to test the solution.

A second problem with FA or PCA is that, after extraction, there is an infinite number of rotations available, all accounting for the same amount of variance in the original data, but with the factors defined slightly differently. The final choice among alternatives depends on the researcher's assessment of its interpretability and scientific utility. In the presence of an infinite number of mathematically identical solutions, researchers are bound to differ regarding which is best. Because the differences cannot be resolved by appeal to objective criteria, arguments over the best solution sometimes become vociferous. However, those who expect a certain amount of ambiguity with respect to choice of the best FA solution will not be surprised when other researchers choose a different one. Nor will they be surprised when results are not replicated exactly, if different decisions are made at one, or more, of the steps in performing FA.

A third problem is that FA is frequently used in an attempt to "save" poorly conceived research. If no other statistical procedure is applicable, at least data can usually be factor analyzed. Thus, in the minds of many, the various forms of FA are associated with sloppy research. The very power of PCA and FA to create apparent order from real chaos contributes to their somewhat tarnished reputations as scientific tools.

There are two major types of FA: exploratory and confirmatory. In exploratory FA, one seeks to describe and summarize data by grouping together variables that are correlated. The variables themselves may or may not have been chosen with potential underlying processes in mind. Exploratory FA is usually performed in the early stages of research, when it provides a tool for consolidating variables and for generating hypotheses about underlying processes. Confirmatory FA is a much more sophisticated technique used in the advanced stages of the research process to test a theory about latent processes. Variables are carefully and specifically chosen to reveal underlying processes. Currently, confirmatory FA is most often performed through structural equation modeling (Chapter 14).

Before we go on, it is helpful to define a few terms. The first terms involve correlation matrices. The correlation matrix produced by the observed variables is called the *observed correlation matrix*. The correlation matrix produced from factors, that is, correlation matrix implied by the factor solution, is called the *reproduced correlation matrix*. The difference between observed and reproduced correlation matrices is the *residual correlation matrix*. In a good FA, correlations in the residual matrix are small, indicating a close fit between the observed and reproduced matrices.

A second set of terms refers to matrices produced and interpreted as part of the solution. Rotation of factors is a process by which the solution is made more interpretable without changing its underlying mathematical properties. There are two general classes of rotation: orthogonal and oblique. If rotation is *orthogonal* (so that all the factors are uncorrelated with each other), a *loading* matrix is produced. The loading matrix is a matrix of correlations between observed variables and factors. The sizes of the loadings reflect the extent of relationship between each observed variable and each factor. Orthogonal FA is interpreted from the loading matrix by looking at which observed variables correlate with each factor.

If rotation is *oblique* (so that the factors themselves are correlated), several additional matrices are produced. The *factor correlation* matrix contains the correlations among the factors. The loading matrix from orthogonal rotation splits into two matrices for oblique rotation: a *structure* matrix of correlations between factors and variables and a *pattern* matrix of unique relationships (uncontaminated by overlap among factors) between each factor and each observed variable. Following oblique rotation, the meaning of factors is ascertained from the pattern matrix.

Lastly, for both types of rotations, there is a *factor-score* coefficients matrix—a matrix of coefficients used in several regression-like equations to predict scores on factors from scores on observed variables for each individual.

FA produces *factors*, while PCA produces *components*. However, the processes are similar except in preparation of the observed correlation matrix for extraction and in the underlying theory. Mathematically, the difference between PCA and FA is in the variance that is analyzed. In PCA, all the variances in the observed variables are analyzed. In FA, only shared variance is analyzed; attempts are made to estimate and eliminate variance due to error and variance that is unique to each variable. The term *factor* is used here to refer to both components and factors unless the distinction is critical, in which case the appropriate term is used.

Theoretically, the difference between FA and PCA lies in the reason that variables are associated with a factor or component. Factors are thought to "cause" variables—the underlying construct (the factor) is what produces scores on the variables. Thus, exploratory FA is associated with theory development and confirmatory FA is associated with theory testing. The question in exploratory FA is: What are the underlying processes that could have produced correlations among these variables? The question in confirmatory FA is: Are the correlations among variables consistent with a hypothesized factor structure? Components are simply aggregates of correlated variables. In that sense, the variables "cause"—or produce—the component. There is no underlying theory about which variables should be associated with which factors; they are simply empirically associated. It is understood that any labels applied to derived components are merely convenient descriptions of the combination of variables associated with them, and do not necessarily reflect some underlying process.

Tiffin, Kaplan, and Place (2011) gathered responses from 673 adolescents (age range 12–18) for an exploratory FA of a 75-item test of perceptions of family functioning. Items were derived from previous work that seemed to indicate a five-factor structure and from feedback from both professionals and other adolescents. The exploratory FA with varimax rotation revealed three significant factors accounting for 73% of the variance. However, the first factor seemed to be a composite of three themes; the 75 items were pruned and a five-factor structure was accepted with between five and seven items per factor. The five factors were labeled Nurture, Problem Solving, Expressed Emotion, Behavioral Boundaries, and Responsibility.

LaVeist, Isaac, and Williams (2009) used principle components analysis to reduce a 17-item Medical Mistrust Index to 7 items. They used a nicely constructed telephone survey of 401 persons to identify the first principle component and then a follow-up interview of 327 of them three weeks later to investigate utilization of health services. Those who scored higher on the 7-item Medical Mistrust Index were more likely to fail to take medical advice, fail to keep a follow-up appointment, postpone receiving needed care, and fail to fill a prescription; they were not, however, more likely to fail to get needed medical care.

Kinoshita and Miyashita (2011) used maximum likelihood extraction and promax rotation to study difficulties felt by ICU nurses in providing end-of-life care. A total of 224 ICU nurses from the Kanto region of Japan participated in the first part of the study. The researchers hypothesized that the nurses would have difficulties in nine different areas, and generated 75 items to assess those areas. However, the FA revealed only five factors with a total of 28 items. Although promax rotation might have revealed correlated factors, the highest actual correlation between two factors was .51. The final five factors were called "the purpose of the ICU is recovery and survival"; "nursing system and model nurse for end-of-life care"; "building confidence in end-of-life care"; "caring for patients and families at end-of-life"; and "converting from curative care to end-of-life care".

13.2 Kinds of Research Questions

The goal of research using PCA or FA is to reduce a large number of variables to a smaller number of factors, to concisely describe (and perhaps understand) the relationships among observed variables, or to test theory about underlying processes. Some of the specific questions that are frequently asked are presented in Sections 13.2.1 through 13.2.5.

13.2.1 Number of Factors

How many reliable and interpretable factors are there in the data set? How many factors are needed to summarize the pattern of correlations in the correlation matrix? In the graduate student example, two factors are discussed: Are these both reliable? Are there any more factors that are reliable? Strategies for choosing an appropriate number of factors and for assessing the correspondence between observed and reproduced correlation matrices are discussed in Section 13.6.2.

13.2.2 Nature of Factors

What is the meaning of the factors? How are the factors to be interpreted? Factors are interpreted by the variables that correlate with them. Rotation to improve interpretability is discussed in Section 13.6.3; interpretation itself is discussed in Section 13.6.5.

13.2.3 Importance of Solutions and Factors

How much variance in a data set is accounted for by the factors? Which factors account for the most variance? In the graduate student example, does the independence or intellectual ability factor account for more of the variance in the measured variables? How much variance does each account for? In a good factor analysis, a high percentage of the variance in the observed variables is accounted for by the first few factors. And, because factors are computed in descending order of magnitude, the first factor accounts for the most variance, with later factors accounting for less and less of the variance until they are no longer reliable. Methods for assessing the importance of solutions and factors are in Section 13.6.4.

13.2.4 Testing Theory in FA

How well does the obtained factor solution fit an expected factor solution? If the researcher had generated hypotheses regarding both the number and the nature of the factors expected of graduate students, comparisons between the hypothesized factors and the factor solution provide a test of the hypotheses. Tests of theory in FA are addressed, in preliminary form, in Sections 13.6.2 and 13.6.7.

More highly developed techniques are available for testing theory in complex data sets in the form of structural equation modeling, which can also be used to test theory regarding factor structure. These techniques are sometimes known by the names of the most popular programs for doing them—EQS and LISREL. Structural equation modeling is the focus of Chapter 14. Confirmatory FA is demonstrated in Section 14.7.

13.2.5 Estimating Scores on Factors

Had factors been measured directly, what scores would subjects have received on each of them? For instance, if each graduate student were measured directly on independence and intelligence, what scores would each student receive for each of them? Estimation of factor scores is the topic of Section 13.6.6.

13.3 Limitations

13.3.1 Theoretical Issues

Most applications of PCA or FA are exploratory in nature; FA is used primarily as a tool for reducing the number of variables or examining patterns of correlations among variables. Under these circumstances, both the theoretical and the practical limitations to FA are relaxed in favor of a frank exploration of the data. Decisions about number of factors and rotational scheme are based on pragmatic rather than theoretical criteria.

The research project that is designed specifically to be factor analyzed, however, differs from other projects in several important respects. Among the best detailed discussions of the differences is the one found in Comrey and Lee (1992), from which some of the following discussion is taken.

The first task of the researcher is to generate hypotheses about factors believed to underlie the domain of interest. Statistically, it is important to make the research inquiry broad enough to include five or six hypothesized factors so that the solution is stable. Logically, in order to reveal the processes underlying a research area, all relevant factors have to be included. Failure to measure some important factor may distort the apparent relationships among measured factors. Inclusion of all relevant factors poses a logical, but not statistical, problem to the researcher.

Next, one selects variables to observe. For each hypothesized factor, five or six variables, each thought to be a relatively pure measure of the factor, are included. Pure measures are called *marker variables.* Marker variables are highly correlated with one and only one factor and load on it regardless of extraction or rotation technique. Marker variables are useful because they define clearly the nature of a factor; adding potential variables to a factor to round it out is much more meaningful if the factor is unambiguously defined by marker variables to begin with.

The complexity of the variables is also considered. Complexity is indicated by the number of factors with which a variable correlates. A pure variable, which is preferred, is correlated with only one factor, whereas a complex variable is correlated with several. If variables differing in complexity are all included in an analysis, those with similar complexity levels may "catch" each other in factors that have little to do with underlying processes. Variables with similar complexity may correlate with each other because of their complexity and not because they relate to the same factor. Estimating (or avoiding) the complexity of variables is part of generating hypotheses about factors and selecting variables to measure them.

Several other considerations are required of the researcher planning a factor analytic study. It is important, for instance, that the sample chosen exhibits spread in scores with respect to the variables and the factors they measure. If all subjects achieve about the same score on some factor, correlations among the observed variables are low and the factor may not emerge in analysis. Selection of subjects expected to differ on the observed variables and underlying factors is an important design consideration.

One should also be wary of pooling the results of several samples, or the same sample with measures repeated in time, for factor analytic purposes. First, samples that are known to be different with respect to some criterion (e.g., socioeconomic status) may also have different factors. Examination of group differences is often quite revealing. Second, underlying factor structure may shift in time for the same subjects with learning or with experience in an experimental setting and these differences may also be quite revealing. Pooling results from diverse groups in FA may obscure differences rather than illuminate them. On the other hand, if different samples do produce the same factors, pooling them is desirable because of increase in sample size. For example, if men and women produce the same factors, the samples should be combined and the results of the single FA reported.

13.3.2 Practical Issues

Because FA and PCA are exquisitely sensitive to the sizes of correlations, it is critical that honest correlations be employed. Sensitivity to outlying cases, problems created by missing data, and degradation of correlations between poorly distributed variables all plague FA and PCA. A review of these issues in Chapter 4 is important to FA and PCA. Thoughtful solutions to some of the problems, including variable transformations, may markedly enhance FA, whether performed for exploratory or confirmatory purposes. However, the limitations apply with greater force to confirmatory FA if done through FA rather than SEM programs.

13.3.2.1 Sample Size and Missing Data

Correlation coefficients tend to be less reliable when estimated from small samples. Therefore, it is important that sample size be large enough that correlations are reliably estimated. The required sample size also depends on magnitude of population correlations and number of factors: if there are strong correlations and a few, distinct factors, a smaller sample size is adequate.

MacCallum, Widaman, Zhang, and Hong (1999) show that samples in the range of 100–200 are acceptable with well-determined factors (i.e., most factors defined by many indicators, i.e., marker variables with loadings $> .80$) and communalities (squared multiple correlations among variables) in the range of .5. *At least 300 cases are needed with low communalities, a small number of factors, and just three or four indicators for each factor.* Sample sizes well over 500 are required under the worst conditions of low communalities and a larger number of weakly determined factors. Impact of sample size is reduced with consistently high communalities (all greater than .6) and well-determined factors. In such cases, samples well below 100 are acceptable, although such small samples run the computational risk of failure of the solution to converge.

McCallum et al. recommend designing studies in which variables are selected to provide as high a level of communalities as possible (a mean level of at least .7) with a small range of variation, a small to moderate number of factors (say, three to five) and several indicators per factor (say, five to seven).

If cases have missing data, either the missing values are estimated or the cases deleted. Consult Chapter 4 for methods of finding and estimating missing values. Consider the distribution of missing values (is it random?) and remaining sample size when deciding between estimation and deletion. If cases are missing values in a nonrandom pattern or if sample size becomes too small, estimation is in order. However, beware of using estimation procedures (such as regression) that are likely to overfit the data and cause correlations to be too high. These procedures may “create” factors.

13.3.2.2 Normality

As long as PCA and FA are used descriptively as convenient ways to summarize the relationships in a large set of observed variables, assumptions regarding the distributions of variables are not in force. If variables are normally distributed, the solution is enhanced. To the extent that normality fails, the solution is degraded but may still be worthwhile.

However, multivariate normality is assumed when statistical inference is used to determine the number of factors. Multivariate normality is the assumption that all variables, and all linear combinations of variables, are normally distributed. Although tests of multivariate normality are overly sensitive, *normality among single variables is assessed by skewness and kurtosis* (see Chapter 4 and Section 13.7.1.2). If a variable has substantial skewness and kurtosis, variable transformation is considered. Some SEM (Chapter 14) programs (e.g., Mplus and EQS) permit PCA/FA with nonnormal variables.

13.3.2.3 Linearity

Multivariate normality also implies that relationships among pairs of variables are linear. The analysis is degraded when linearity fails, because correlation measures linear relationship and does not reflect nonlinear relationship. *Linearity among pairs of variables is assessed through inspection of* *scatterplots.* Consult Chapter 4 and Section 13.7.1.3 for methods of screening for linearity. If nonlinearity is found, transformation of variables is considered.

13.3.2.4 Absence of Outliers Among Cases

As in all multivariate techniques, cases may be outliers either on individual variables (univariate) or on combinations of variables (multivariate). Such cases have more influence on the factor solution than other cases. Consult Chapter 4 and Section 13.7.1.4 for methods of detecting and reducing the influence of both univariate and multivariate outliers.

13.3.2.5 Absence of Multicollinearity and Singularity

In PCA, multicollinearity is not a problem because there is no need to invert a matrix. For most forms of FA and for estimation of factor scores in any form of FA, singularity or extreme multicollinearity is a problem. For FA, if the determinant of **R** and eigenvalues associated with some factors approach 0, multicollinearity or singularity may be present.

To investigate further, look at the SMCs for each variable where it serves as DV with all other variables as IVs. If any of the SMCs is one, singularity is present; if any of the SMCs is very large (near one), multicollinearity is present. Delete the variable with multicollinearity or singularity. Chapter 4 and Section 13.7.1.5 provide examples of screening for and dealing with multicollinearity and singularity.

13.3.2.6 Factorability of R

A matrix that is factorable should include several sizable correlations. The expected size depends, to some extent, on *N* (larger sample sizes tend to produce smaller correlations), but if no correlation exceeds .30, use of FA is questionable because there is probably nothing to factor analyze. *Inspect **R** for correlations in excess of .30, and, if none is found, reconsider use of FA.*

High bivariate correlations, however, are not ironclad proof that the correlation matrix contains factors. It is possible that the correlations are between only two variables and do not reflect the underlying processes that are simultaneously affecting several variables. For this reason, it is helpful to examine matrices of partial correlations where pairwise correlations are adjusted for effects of all other variables. If there are factors present, then high bivariate correlations become very low partial correlations. IBM SPSS and SAS produce partial correlation matrices.

Bartlett's (1954) test of sphericity is a notoriously sensitive test of the hypothesis that the correlations in a correlation matrix are zero. The test is available in IBM SPSS FACTOR but because of its sensitivity and its dependence on *N*, the test is likely to be significant with samples of substantial size even if correlations are very low. Therefore, use of the test is recommended only if there are fewer than, say, five cases per variable.

Several more sophisticated tests of the factorability of **R** are available through IBM SPSS and SAS. Both programs give significance tests of correlations, the anti-image correlation matrix, and Kaiser's (1970, 1974) measure of sampling adequacy. Significance tests of correlations in the correlation matrix provide an indication of the reliability of the relationships between pairs of variables. If **R** is factorable, numerous pairs are significant. The anti-image correlation matrix contains the negatives of partial correlations between pairs of variables with effects of other variables removed. If **R** is factorable, there are mostly small values among the off-diagonal elements of the anti-image matrix. Finally, Kaiser's measure of sampling adequacy is a ratio of the sum of squared correlations to the sum of squared correlations plus sum of squared partial correlations. The value approaches 1 if partial correlations are small. Values of .6 and above are required for good FA.

13.3.2.7 Absence of Outliers Among Variables

After FA, in both exploratory and confirmatory FA, variables that are unrelated to others in the set are identified. These variables are usually not correlated with the first few factors although they often correlate with factors extracted later. These factors are usually unreliable, both because they account for very little variance and because factors that are defined by just one or two variables are not stable. Therefore, one never knows whether these factors are "real." Suggestions for determining reliability of factors defined by one or two variables are in Section 13.6.2.

If the variance accounted for by a factor defined by only one or two variables is high enough, the factor is interpreted with great caution or is ignored, as pragmatic considerations dictate. In confirmatory FA done through FA rather than SEM programs, the factor represents either a promising lead for future work or (probably) error variance, but its interpretation awaits clarification by more research.

A variable with a low squared multiple correlation with all other variables and low correlations with all important factors is an outlier among the variables. The variable is usually ignored in the current FA and either deleted or given friends in future research. Screening for outliers among variables is illustrated in Section 13.7.1.7.

13.4 Fundamental Equations for Factor Analysis

Because of the variety and complexity of the calculations involved in preparing the correlation matrix, extracting factors, and rotating them, and because, in our judgment, little insight is produced by demonstrations of some of these procedures, this section does not show them all. Instead, the relationships between some of the more important matrices are shown, with an assist from IBM SPSS FACTOR for underlying calculations.

Table 13.1 lists many of the important matrices in FA and PCA. Although the list is lengthy, it is composed mostly of *matrices of correlations* (between variables, between factors, and between variables and factors), *matrices of standard scores* (on variables and on factors), *matrices of regression weights* (for producing scores on factors from scores on variables), and the *pattern matrix* of unique relationships between factors and variables after oblique rotation.

Also in the table are the matrix of eigenvalues and the matrix of their corresponding eigenvectors. Eigenvalues and eigenvectors are discussed here and in Appendix A, albeit scantily, because of their importance in factor extraction, the frequency with which one encounters the terminology, and the close association between eigenvalues and variance in statistical applications.

A data set appropriate for FA consists of numerous subjects each measured on several variables. A grossly inadequate data set appropriate for FA is in Table 13.2. Five subjects who were trying on ski boots late on a Friday night in January were asked about the importance of each of four variables to their selection of a ski resort. The variables were cost of ski ticket (COST), speed of ski lift (LIFT), depth of snow (DEPTH), and moisture of snow (POWDER). Larger numbers indicate greater importance. The researcher wanted to investigate the pattern of relationships among the variables in an effort to understand better the dimensions underlying choice of ski area.

TABLE 13.1 Commonly Encountered Matrices in Factor Analyses

Label	Name	Rotation	Size[a]	Description
R	Correlation matrix	Both orthogonal and oblique	$p \times p$	Matrix of correlations between variables
Z	Variable matrix	Both orthogonal and oblique	$N \times p$	Matrix of standardized observed variable scores
F	Factor-score matrix	Both orthogonal and oblique	$N \times m$	Matrix of standardized scores on factors or components
A	Factor loading matrix Pattern matrix	Orthogonal Oblique	$p \times m$	Matrix of regression-like weights used to estimate the unique contribution of each factor to the variance in a variable. If orthogonal, also correlations between variables and factors
B	Factor-score coefficients matrix	Both orthogonal and oblique	$p \times m$	Matrix of regression-like weights used to generate factor scores from variables
C	Structure matrix[b]	Oblique	$p \times m$	Matrix of correlations between variables and (correlated) factors
Φ	Factor correlation matrix	Oblique	$m \times m$	Matrix of correlations among factors
L	Eigenvalue matrix[c]	Both orthogonal and oblique	$m \times m$	Diagonal matrix of eigenvalues, one per factor[e]
V	Eigenvector matrix[d]	Both orthogonal and oblique	$p \times m$	Matrix of eigenvectors, one vector per eigenvalue

[a]Row by column dimensions where

p = number of variables

N = number of subjects

m = number of factors or components.

[b]In most textbooks, the structure matrix is labeled **S**. However, we have used **S** to represent the sum-of-squares and cross-products matrix elsewhere and will use **C** for the structure matrix here.

[c]Also called characteristic roots or latent roots.

[d]Also called characteristic vectors.

[e]If the matrix is of full rank, there are actually p rather than m eigenvalues and eigenvectors. Only m are of interest, however, so the remaining $p - m$ are not displayed.

Notice the pattern of correlations in the correlation matrix as set off by the vertical and horizontal lines. The strong correlations in the upper left and lower right quadrants show that scores on COST and LIFT are related, as are scores on DEPTH and POWDER. The other two quadrants show that scores on DEPTH and LIFT are unrelated, as are scores on POWDER and LIFT, and so on. With luck, FA will find this pattern of correlations, easy to see in a small correlation matrix but not in a very large one.

TABLE 13.2 Small Sample of Hypothetical Data for Illustration of Factor Analysis

	Variables			
Skiers	*COST*	*LIFT*	*DEPTH*	*POWDER*
S_1	32	64	65	67
S_2	61	37	62	65
S_3	59	40	45	43
S_4	36	62	34	35
S_5	62	46	43	40

Correlation Matrix

	COST	**LIFT**	**DEPTH**	**POWDER**
COST	1.000	−.953	−.055	−.130
LIFT	−.953	1.000	−.091	−.036
DEPTH	−.055	−.091	1.000	.990
POWDER	−.130	−.036	.990	1.000

13.4.1 Extraction

An important theorem from matrix algebra indicates that, under certain conditions, matrices can be diagonalized. Correlation and covariance matrices are among those that often can be diagonalized. When a matrix is diagonalized, it is transformed into a matrix with numbers in the positive diagonal[2] and zeros everywhere else. In this application, the numbers in the positive diagonal represent variance from the correlation matrix that has been repackaged as follows:

$$\mathbf{L} = \mathbf{V}'\mathbf{R}\mathbf{V} \tag{13.1}$$

Diagonalization of **R** is accomplished by post- and pre-multiplying it by the matrix **V** and its transpose.

The columns in **V** are called eigenvectors, and the values in the main diagonal of **L** are called eigenvalues. The first eigenvector corresponds to the first eigenvalue, and so forth.

Because there are four variables in the example, there are four eigenvalues with their corresponding eigenvectors. However, because the goal of FA is to summarize a pattern of correlations with as few factors as possible, and because each eigenvalue corresponds to a different potential factor, usually only factors with large eigenvalues are retained. In a good FA, these few factors almost duplicate the correlation matrix.

In this example, when no limit is placed on the number of factors, eigenvalues of 2.02, 1.94, .04, and .00 are computed for each of the four possible factors. Only the first two factors, with

[2]The positive diagonal runs from upper left to lower right in a matrix.

values over 1.00, are large enough to be retained in subsequent analyses. FA is rerun specifying extraction of just the first two factors; they have eigenvalues of 2.00 and 1.91, respectively, as indicated in Table 13.3.

Using Equation 13.1 and inserting the values from the example, we obtain

$$\mathbf{L} = \begin{bmatrix} -.283 & .177 & .658 & .675 \\ .651 & -.685 & .252 & .207 \end{bmatrix} \begin{bmatrix} 1.000 & -.953 & -.055 & -.130 \\ -.953 & 1.000 & -.091 & -.036 \\ -.055 & -.091 & 1.000 & .990 \\ -.130 & -.036 & .990 & 1.000 \end{bmatrix} \begin{bmatrix} -.283 & .651 \\ .177 & -.685 \\ -.658 & .252 \\ .675 & .207 \end{bmatrix}$$

$$= \begin{bmatrix} 2.00 & .00 \\ .00 & 1.91 \end{bmatrix}$$

(All values agree with computer output. Hand calculation may produce discrepancies due to rounding error.)

The matrix of eigenvectors pre-multiplied by its transpose produces the identity matrix with ones in the positive diagonal and zeros elsewhere. Therefore, pre- and post-multiplying the correlation matrix by eigenvectors does not change it so much as repackage it.

$$\mathbf{V'V} = \mathbf{I} \tag{13.2}$$

For the example:

$$\begin{bmatrix} -.283 & .177 & .658 & .675 \\ .651 & -.685 & .252 & .207 \end{bmatrix} \begin{bmatrix} -.283 & .651 \\ .177 & -.685 \\ .658 & .252 \\ .675 & .207 \end{bmatrix} = \begin{bmatrix} 1.000 & .000 \\ .000 & 1.000 \end{bmatrix}$$

The important point is that because correlation matrices often meet requirements for diagonalizability, it is possible to use on them the matrix algebra of eigenvectors and eigenvalues with FA as the result. When a matrix is diagonalized, the information contained in it is repackaged. In

TABLE 13.3 Eigenvectors and Corresponding Eigenvalues for the Example

Eigenvector 1	Eigenvector 2
−.283	.651
.177	−.685
.658	.252
.675	.207
Eigenvalue 1	**Eigenvalue 2**
2.00	1.91

FA, the variance in the correlation matrix is condensed into eigenvalues. The factor with the largest eigenvalue has the most variance and so on, down to factors with small or negative eigenvalues that are usually omitted from solutions.

Calculations for eigenvectors and eigenvalues are extremely laborious and not particularly enlightening (although they are illustrated in Appendix A for a small matrix). They require solving p equations in p unknowns with additional side constraints and are rarely performed by hand. Once the eigenvalues and eigenvectors are known, however, the rest of FA (or PCA) more or less "falls out," as is seen from Equations 13.3 to 13.6.

Equation 13.1 can be reorganized as follows:

$$\mathbf{R} = \mathbf{VLV}' \tag{13.3}$$

The correlation matrix can be considered a product of three matrices—the matrices of eigenvalues and corresponding eigenvectors.

After reorganization, the square root is taken of the matrix of eigenvalues.

$$\mathbf{R} = \mathbf{V}\sqrt{\mathbf{L}}\sqrt{\mathbf{L}}\mathbf{V}' \tag{13.4}$$

or

$$\mathbf{R} = (\mathbf{V}\sqrt{\mathbf{L}})(\sqrt{\mathbf{L}}\mathbf{V}')$$

If $\mathbf{V}\sqrt{\mathbf{L}}$ is called $\mathbf{A}$ and $\sqrt{\mathbf{L}}\mathbf{V}'$ is $\mathbf{A}'$ then

$$\mathbf{R} = \mathbf{AA}' \tag{13.5}$$

The correlation matrix can also be considered a product of two matrices—each a combination of eigenvectors and the square root of eigenvalues.

Equation 13.5 is frequently called the fundamental equation for FA.[3] It represents the assertion that the correlation matrix is a product of the factor loading matrix, **A**, and its transpose.

Equations 13.4 and 13.5 also reveal that the major work of FA (and PCA) is calculation of eigenvalues and eigenvectors. Once they are known, the (unrotated) factor loading matrix is found by straightforward matrix multiplication, as follows.

$$\mathbf{A} = \mathbf{V}\sqrt{\mathbf{L}} \tag{13.6}$$

For the example:

$$\mathbf{A} = \begin{bmatrix} -.283 & .651 \\ .177 & -.685 \\ .658 & .252 \\ .675 & .207 \end{bmatrix} \begin{bmatrix} \sqrt{2.00} & 0 \\ 0 & \sqrt{1.91} \end{bmatrix} = \begin{bmatrix} -.400 & .900 \\ .251 & -.947 \\ .932 & .348 \\ .956 & .286 \end{bmatrix}$$

[3]In order to reproduce the correlation matrix exactly, as indicated in Equations 13.4 and 13.5, all eigenvalues and eigenvectors are necessary, not just the first few of them.

The factor loading matrix is a matrix of correlations between factors and variables. The first column is correlations between the first factor and each variable in turn, COST (−.400), LIFT (.251), DEPTH (.932), and POWDER (.956). The second column is correlations between the second factor and each variable in turn, COST (.900), LIFT (−.947), DEPTH (.348), and POWDER (.286). A factor is interpreted from the variables that are highly correlated with it—that have high loadings on it. Thus, the first factor is primarily a snow conditions factor (DEPTH and POWDER), while the second reflects resort conditions (COST and LIFT). Subjects who score high on the resort conditions factor (Equation 13.11) tend to assign high value to COST and low value to LIFT (the negative correlation); subjects who score low on the resort conditions factor value LIFT more than COST.

Notice, however, that all the variables are correlated with both factors to a considerable extent. Interpretation is fairly clear for this hypothetical example, but most likely would not be for real data. Usually, a factor is most interpretable when a few variables are highly correlated with it and the rest are not.

13.4.2 Orthogonal Rotation

Rotation is ordinarily used after extraction to maximize high correlations between factors and variables and minimize low ones. Numerous methods of rotation are available (see Section 13.5.2) but the most commonly used, and the one illustrated here, is *varimax*. Varimax is a variance-maximizing procedure. The goal of varimax rotation is to maximize the variance of factor loadings by making high loadings higher and low ones lower for each factor.

This goal is accomplished by means of a transformation matrix $\mathbf{\Lambda}$ (as defined in Equation 13.8), where

$$\mathbf{A}_{\text{unrotated}}\mathbf{\Lambda} = \mathbf{A}_{\text{rotated}} \tag{13.7}$$

The unrotated factor loading matrix is multiplied by the transformation matrix to produce the rotated loading matrix.

For the example:

$$\mathbf{A}_{\text{rotated}} = \begin{bmatrix} -.400 & .900 \\ .251 & -.947 \\ .932 & .348 \\ .956 & .286 \end{bmatrix} \begin{bmatrix} .946 & -.325 \\ .325 & .946 \end{bmatrix} = \begin{bmatrix} -.086 & .981 \\ -.071 & -.977 \\ .994 & .026 \\ .997 & -.040 \end{bmatrix}$$

Compare the rotated and unrotated loading matrices. Notice that in the rotated matrix the low correlations are lower and the high ones are higher than in the unrotated loading matrix. Emphasizing differences in loadings facilitates interpretation of a factor by making unambiguous the variables that correlate with it.

The numbers in the transformation matrix have a spatial interpretation.

$$\mathbf{\Lambda} = \begin{bmatrix} \cos\Psi & -\sin\Psi \\ \sin\Psi & \cos\Psi \end{bmatrix} \tag{13.8}$$

The transformation matrix is a matrix of sines and cosines of an angle Ψ.

TABLE 13.4 Relationships Among Loadings, Communalities, SSLs, Variance, and Covariance of Orthogonally Rotated Factors

	Factor 1	Factor 2	Communalities (h^2)
COST	−.086	.981	$\sum a^2 = .970$
LIFT	−.071	−.977	$\sum a^2 = .960$
DEPTH	.994	.026	$\sum a^2 = .989$
POWDER	.997	−.040	$\sum a^2 = .996$
SSLs	$\sum a^2 = 1.994$	$\sum a^2 = 1.919$	3.915
Proportion of variance	.50	.48	.98
Proportion of covariance	.51	.49	

For the example, the angle is approximately 19°. That is, cos 19 ≈ .946 and sin 19 ≈ .325. Geometrically, this corresponds to a 19° swivel of the factor axes about the origin. Greater detail regarding the geometric meaning of rotation is in Section 13.5.2.3.

13.4.3 Communalities, Variance, and Covariance

Once the rotated loading matrix is available, other relationships are found, as in Table 13.4. The communality for a variable is the variance accounted for by the factors. It is the squared multiple correlation of the variable as predicted from the factors. Communality is the sum of squared loadings (SSL) for a variable across factors. In Table 13.4, the communality for COST is $(-.086)^2 + .981^2 = .970$. That is, 97% of the variance in COST is accounted for by Factor 1 plus Factor 2.

The proportion of variance *in the set of variables* accounted for by a factor is the SSL for the factor divided by the number of variables (if rotation is orthogonal).[4] For the first factor, the proportion of variance is $\left[(-.086)^2 + (-.071)^2 + .994^2 + .997^2\right]/4 = 1.994/4 = .50\right]$. Fifty percent of the variance in the variables is accounted for by the first factor. The second factor accounts for 48% of the variance in the variables and, because rotation is orthogonal, the two factors together account for 98% of the variance in the variables.

The proportion of variance *in the solution* accounted for by a factor—the proportion of covariance—is the SSL for the factor divided by the sum of communalities (or, equivalently, the sum of the SSLs). The first factor accounts for 51% of the variance in the solution (1.994/3.915) while the second factor accounts for 49% of the variance in the solution (1.919/3.915). The two factors together account for all of the covariance.

[4]For unrotated factors only, the sum of the squared loadings for a factor is equal to the eigenvalue. Once loadings are rotated, the sum of squared loadings is called SSL and is no longer equal to the eigenvalue.

The reproduced correlation matrix for the example is generated using Equation 13.5:

$$\bar{\mathbf{R}} = \begin{bmatrix} -.086 & .981 \\ -.071 & -.977 \\ .994 & .026 \\ .997 & -.040 \end{bmatrix} \begin{bmatrix} -.086 & -.071 & .994 & .997 \\ .981 & -.977 & .026 & -.040 \end{bmatrix}$$

$$= \begin{bmatrix} .970 & -.953 & -.059 & -.125 \\ -.953 & .962 & -.098 & -.033 \\ -.059 & -.098 & .989 & .990 \\ -.125 & -.033 & .990 & .996 \end{bmatrix}$$

Notice that the reproduced correlation matrix differs slightly from the original correlation matrix. The difference between the original and reproduced correlation matrices is the residual correlation matrix:

$$\mathbf{R}_{\text{res}} = \mathbf{R} - \bar{\mathbf{R}} \tag{13.9}$$

The residual correlation matrix is the difference between the observed correlation matrix and the reproduced correlation matrix.

For the example, with communalities inserted in the positive diagonal of **R**:

$$\mathbf{R}_{\text{res}} = \begin{bmatrix} .970 & -.953 & -.055 & -.130 \\ -.953 & .960 & -.091 & -.036 \\ -.055 & -.091 & .989 & .990 \\ -.130 & -.036 & .990 & .996 \end{bmatrix} - \begin{bmatrix} .970 & -.953 & -.059 & -.125 \\ -.953 & .960 & -.098 & -.033 \\ -.059 & -.098 & .989 & .990 \\ -.125 & -.033 & .990 & .996 \end{bmatrix}$$

$$= \begin{bmatrix} .000 & .000 & .004 & -.005 \\ .000 & .000 & .007 & -.003 \\ .004 & .007 & .000 & .000 \\ -.005 & -.003 & .000 & .000 \end{bmatrix}$$

In a "good" FA, the numbers in the residual correlation matrix are small because there is little difference between the original correlation matrix and the correlation matrix generated from factor loadings.

13.4.4 Factor Scores

Scores on factors can be predicted for each case once the loading matrix is available. Regression-like coefficients are computed for weighting variable scores to produce factor scores. Because $\mathbf{R}^{-1}$ is the inverse of the matrix of correlations among variables and **A** is the matrix of correlations between factors and variables, Equation 13.10 for factor score coefficients is similar to Equation 5.6 for regression coefficients in multiple regression.

$$\mathbf{B} = \mathbf{R}^{-1}\mathbf{A} \tag{13.10}$$

Factor score coefficients for estimating factor scores from variable scores are a product of the inverse of the correlation matrix and the factor loading matrix.

For the example:[5]

$$\mathbf{B} = \begin{bmatrix} 25.485 & 22.689 & -31.655 & 35.479 \\ 22.689 & 21.386 & -24.831 & 28.312 \\ -31.655 & -24.831 & 99.917 & -103.950 \\ 35.479 & 28.312 & -103.950 & 109.567 \end{bmatrix} \begin{bmatrix} -.087 & .981 \\ -.072 & -.978 \\ .994 & .027 \\ .997 & -.040 \end{bmatrix}$$

$$= \begin{bmatrix} 0.082 & 0.537 \\ 0.054 & -0.461 \\ 0.190 & 0.087 \\ 0.822 & -0.074 \end{bmatrix}$$

To estimate a subject's score for the first factor, all of the subject's scores on variables are standardized and then the standardized score on COST is weighted by 0.082, LIFT by 0.054, DEPTH by 0.190, and POWDER by 0.822, and the results are added. In matrix form,

$$\mathbf{F} = \mathbf{ZB} \tag{13.11}$$

Factor scores are a product of standardized scores on variables and factor score coefficients.

For the example:

$$\mathbf{F} = \begin{bmatrix} -1.22 & 1.14 & 1.15 & 1.14 \\ 0.75 & -1.02 & 0.92 & 1.01 \\ 0.61 & -0.78 & -0.36 & -0.47 \\ -0.95 & 0.98 & -1.20 & -1.01 \\ 0.82 & -0.30 & -0.51 & -0.67 \end{bmatrix} \begin{bmatrix} 0.082 & 0.537 \\ 0.054 & -0.461 \\ 0.190 & 0.087 \\ 0.822 & -0.074 \end{bmatrix}$$

$$= \begin{bmatrix} 1.12 & -1.16 \\ 1.01 & 0.88 \\ -0.45 & 0.69 \\ -1.08 & -0.99 \\ -0.60 & 0.58 \end{bmatrix}$$

[5]The numbers in **B** are different from the factor score coefficients generated by computer for the small data set. The difference is due to rounding error following inversion of a multicollinear correlation matrix. Note also that the **A** matrix contains considerable rounding error.

The first subject has an estimated standard score of 1.12 on the first factor and −1.16 on the second factor, and so on for the other four subjects. The first subject strongly values both the snow factor and the resort factor, one positive and the other negative (indicating stronger importance assigned to speed of LIFT). The second subject values both the snow factor and the resort factor (with more value placed on COST than LIFT); the third subject places more value on resort conditions (particularly COST) and less value on snow conditions, and so forth. The sum of standardized factor scores across subjects for a single factor is zero.

Predicting scores on variables from scores on factors is also possible. The equation for doing so is

$$\mathbf{Z} = \mathbf{F}\mathbf{A}' \tag{13.12}$$

Predicted standardized scores on variables are a product of scores on factors weighted by factor loadings.

For example:

$$\mathbf{Z} = \begin{bmatrix} 1.12 & -1.16 \\ 1.01 & 0.88 \\ -0.45 & 0.69 \\ -1.08 & -0.99 \\ -0.60 & 0.58 \end{bmatrix} \begin{bmatrix} -.086 & -.072 & .994 & .997 \\ .981 & -.978 & .027 & -.040 \end{bmatrix}$$

$$= \begin{bmatrix} -1.23 & 1.05 & 1.08 & 1.16 \\ 0.78 & -0.93 & 1.03 & 0.97 \\ 0.72 & -0.64 & -0.43 & -0.48 \\ -0.88 & 1.05 & -1.10 & -1.04 \\ 0.62 & -0.52 & -0.58 & -0.62 \end{bmatrix}$$

That is, the first subject (the first row of **Z**) is predicted to have a standardized score of −1.23 on COST, 1.05 on LIFT, 1.08 on DEPTH, and 1.16 on POWDER. Like the reproduced correlation matrix, these values are similar to the observed values if the FA captures the relationship among the variables.

It is helpful to see these values written out because they provide an insight into how scores on variables are conceptualized in factor analysis. For example, for the first subject,

$$\begin{aligned} -1.23 &= -.086(1.12) + .981(-1.16) \\ 1.05 &= -.072(1.12) - .978(-1.16) \\ 1.08 &= .994(1.12) + .027(-1.16) \\ 1.16 &= .997(1.12) - .040(-1.16) \end{aligned}$$

Or, in algebraic form,

$$z_{\text{COST}} = a_{11}F_1 + a_{12}F_2$$
$$z_{\text{LIFT}} = a_{21}F_1 + a_{22}F_2$$
$$z_{\text{DEPTH}} = a_{31}F_1 + a_{32}F_2$$
$$z_{\text{POWDER}} = a_{41}F_1 + a_{42}F_2$$

A score on an observed variable is conceptualized as a properly weighted and summed combination of the scores on factors that underlie it. The researcher believes that each subject has the same latent factor structure, but different scores on the factors themselves. A particular subject's score on an observed variable is produced as a weighted combination of that subject's scores on the underlying factors.

13.4.5 Oblique Rotation

All the relationships mentioned thus far are for orthogonal rotation. Most of the complexities of orthogonal rotation remain and several others are added when oblique (correlated) rotation is used. Consult Table 13.1 for a listing of additional matrices and a hint of the discussion to follow.

IBM SPSS FACTOR is run on the data from Table 13.2 using the default option for oblique rotation (cf. Section 13.5.2.2) to get values for the pattern matrix, **A**, and factor-score coefficients, **B**.

In oblique rotation, the loading matrix becomes the pattern matrix. Values in the pattern matrix, when squared, represent the unique contribution of each factor to the variance of each variable but do not include segments of variance that come from overlap between correlated factors. For the example, the pattern matrix following oblique rotation is

$$\mathbf{A} = \begin{bmatrix} -.079 & .981 \\ -.078 & -.978 \\ .994 & .033 \\ .977 & -.033 \end{bmatrix}$$

The first factor makes a unique contribution of $-.079^2$ to the variance in COST, $-.078^2$ to LIFT, $.994^2$ to DEPTH, and $.997^2$ to POWDER.

Factor-score coefficients following oblique rotation are also found:

$$\mathbf{B} = \begin{bmatrix} 0.104 & 0.584 \\ 0.081 & -0.421 \\ 0.159 & -0.020 \\ 0.856 & 0.034 \end{bmatrix}$$

Applying Equation 13.11 to produce factor scores results in the following values:

$$\mathbf{F} = \begin{bmatrix} -1.22 & 1.14 & 1.15 & 1.14 \\ 0.75 & -1.02 & 0.92 & 1.01 \\ 0.61 & -0.78 & -0.36 & -0.47 \\ -0.95 & 0.98 & -1.20 & -1.01 \\ 0.82 & -0.30 & -0.51 & -0.67 \end{bmatrix} \begin{bmatrix} 0.104 & 0.584 \\ 0.081 & -0.421 \\ 0.159 & -0.020 \\ 0.856 & 0.034 \end{bmatrix}$$

$$= \begin{bmatrix} 1.12 & -1.18 \\ 1.01 & 0.88 \\ -0.46 & 0.68 \\ -1.07 & -0.98 \\ -0.59 & 0.59 \end{bmatrix}$$

Once the factor scores are determined, correlations among factors can be obtained. Among the equations used for this purpose is

$$\mathbf{\Phi} = \left(\frac{1}{N - 1}\right)\mathbf{F}'\mathbf{F} \tag{13.13}$$

One way to compute correlations among factors is from cross-products of standardized factor scores divided by the number of cases minus one.

The factor correlation matrix is a standard part of computer output following oblique rotation. For the example:

$$\mathbf{\Phi} = \frac{1}{4}\begin{bmatrix} 1.12 & 1.01 & -0.46 & -1.07 & -0.59 \\ -1.18 & 0.88 & 0.68 & -0.98 & 0.59 \end{bmatrix} \begin{bmatrix} 1.12 & -1.16 \\ 1.01 & 0.88 \\ -0.45 & 0.69 \\ -1.08 & -0.99 \\ -0.60 & 0.58 \end{bmatrix}$$

$$= \begin{bmatrix} 1.00 & -0.01 \\ -0.01 & 1.00 \end{bmatrix}$$

The correlation between the first and second factor is quite low, −.01. For this example, there is almost no relationship between the two factors, although considerable correlation could have been produced had it been warranted. Usually, one uses orthogonal rotation in a case like this because complexities introduced by oblique rotation are not warranted by such a low correlation among factors.

However, if oblique rotation is used, the structure matrix, **C**, is the correlations between variables and factors. These correlations assess the unique relationship between the variable and the factor (in the pattern matrix) plus the relationship between the variable and the overlapping variance among the factors. The equation for the structure matrix is

$$\mathbf{C} = \mathbf{A\Phi} \tag{13.14}$$

The structure matrix is a product of the pattern matrix and the factor correlation matrix.

For example:

$$\mathbf{C} = \begin{bmatrix} -.079 & .981 \\ -.078 & -.978 \\ .994 & .033 \\ .997 & -.033 \end{bmatrix} \begin{bmatrix} 1.00 & -.01 \\ -.01 & 1.00 \end{bmatrix} = \begin{bmatrix} -.069 & .982 \\ -.088 & -.977 \\ .994 & .023 \\ .997 & -.043 \end{bmatrix}$$

COST, LIFT, DEPTH, and POWDER correlate −.069, −.088, .994, and .997 with the first factor and .982, −.977, .023, and −.043 with the second factor, respectively.

There is some debate as to whether one should interpret the pattern matrix or the structure matrix following oblique rotation. The structure matrix is appealing because it is readily understood. However, the correlations between variables and factors are inflated by any overlap between factors. The problem becomes more severe as the correlations among factors increase and it may be hard to determine which variables are related to a factor. On the other hand, the pattern matrix contains values representing the unique contributions of each factor to the variance in the variables. Shared variance is omitted (as it is with standard multiple regression), but the set of variables that composes a factor is usually easier to see. If factors are very highly correlated, it may appear that no variables are related to them because there is almost no unique variance once overlap is omitted.

Most researchers interpret and report the pattern matrix rather than the structure matrix. However, if the researcher reports either the structure or the pattern matrix and also Φ, then the interested reader can generate the other using Equation 13.14 as desired.

In oblique rotation, $\bar{\mathbf{R}}$, is produced as follows:

$$\bar{\mathbf{R}} = \mathbf{CA}' \tag{13.15}$$

The reproduced correlation matrix is a product of the structure matrix and the transpose of the pattern matrix.

Once the reproduced correlation matrix is available, Equation 13.9 is used to generate the residual correlation matrix to diagnose adequacy of fit in FA.

13.4.6 Computer Analyses of Small-Sample Example

A two-factor principal factor analysis (PFA) with varimax rotation using the example is shown for IBM SPSS FACTOR and SAS FACTOR in Tables 13.5 and 13.6.

For a PFA with varimax rotation, IBM SPSS FACTOR requires that you specify EXTRACTION PAF and ROTATION VARIMAX.[6] IBM SPSS FACTOR (Table 13.5) begins by printing out SMCs for each variable, labeled Initial in the Communalities portion of the output. In the same table are final (Extraction) communalities. These show the portion of variance in each variable accounted for by the solution (h^2 in Table 13.4).

[6]The defaults for IBM SPSS FACTOR are principal components analysis with no rotation.

TABLE 13.5 Syntax and IBM SPSS FACTOR Output for Factor Analysis on Sample Data of Table 13.2

```
FACTOR
  /VARIABLES COST LIFT DEPTH POWDER /MISSING LISTWISE
  /ANALYSIS COST LIFT DEPTH POWDER
  /PRINT INITIAL EXTRACTION ROTATION
  /CRITERIA MINEIGEN(1) ITERATE(25)
  /EXTRACTION PAF
  /CRITERIA ITERATE(25)
  /ROTATION VARIMAX
  /METHOD=CORRELATION.
```

Communalities

	Initial	Extraction
COST	.961	.970
LIFT	.953	.960
DEPTH	.990	.989
POWDER	.991	.996

Extraction Method: Principal Axis Factoring.

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.016	50.408	50.408	2.005	50.118	50.118	1.995	49.866	49.866
2	1.942	48.538	98.945	1.909	47.733	97.852	1.919	47.986	97.852
3	.038	.945	99.891						
4	.004	.109	100.000						

Extraction Method: Principal Axis Factoring.

(*continued*)

TABLE 13.5 Continued

Factor Matrix[a]

	Factor	
	1	2
COST	−.400	.900
LIFT	.251	−.947
DEPTH	.932	.348
POWDER	.956	.286

Extraction Method: Principal Axis Factoring.

a. 2 factors extracted. 4 iterations required.

Rotated Factor Matrix[a]

	Factor	
	1	2
COST	−.086	−.981
LIFT	−.071	.977
DEPTH	.994	−.026
POWDER	.997	.040

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

Factor Transformation Matrix

Factor	1	2
1	.946	.325
2	.325	−.946

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.

TABLE 13.6 Syntax and Selected SAS FACTOR Output for Factor Analysis of Sample Data of Table 13.2

```
proc factor data=SASUSER.SSFACTOR;
  method=prinit priors=smc nfactors=2 rotate=v;
  var cost lift depth powder;
run;
```

The FACTOR Procedure
Initial Factor Method: Iterated Principal Factor Analysis

Prior Communality Estimates: SMC

COST	LIFT	DEPTH	POWDER
0.96076070	0.95324069	0.98999165	0.99087317

Preliminary Eigenvalues: Total = 3.89486621 Average = 0.97371655

	Eigenvalue	Difference	Proportion	Cumulative
1	2.00234317	0.09960565	0.5141	0.5141
2	1.90273753	1.89753384	0.4885	1.0026
3	0.00520369	0.02062186	0.0013	1.0040
4	-.01541817		-0.0040	1.0000

2 factors will be retained by the NFACTOR criterion.

WARNING: Too many factors for a unique solution.

Eigenvalues of the Reduced Correlation Matrix: Total = 3.91277649 Average = 0.97819412

	Eigenvalue	Difference	Proportion	Cumulative
1	2.00473399	0.09539900	0.5124	0.5124
2	1.90933499	1.90037259	0.4880	1.0003
3	0.00896240	0.01921730	0.0023	1.0026
4	-.01025490		-0.0026	1.0000

(continued)

TABLE 13.6 Continued

Initial Factor Method: Iterated Principal Factor Analysis

Factor Pattern

	Factor1	Factor2
COST	-0.40027	0.89978
LIFT	0.25060	-0.94706
DEPTH	0.93159	0.34773
POWDER	0.95596	0.28615

Variance Explained by Each Factor

Factor1	Factor2
2.0047340	1.9093350

Final Communality Estimates: Total = 3.914069

COST	LIFT	DEPTH	POWDER
0.96982841	0.95972502	0.98877384	0.99574170

Rotation Method: Varimax

Orthogonal Transformation Matrix

	1	2
1	0.94565	-0.32519
2	0.32519	0.94565

Rotated Factor Pattern

	Factor1	Factor2
COST	-0.08591	0.98104
LIFT	-0.07100	-0.97708
DEPTH	0.99403	0.02588
POWDER	0.99706	-0.04028

Variance Explained by Each Factor

Factor1	Factor2
1.9946455	1.9194235

Final Communality Estimates: Total = 3.914069

COST	LIFT	DEPTH	POWDER
0.96982841	0.95972502	0.98877384	0.99574170

The next table shows a great deal of information about variance accounted for by the factors. Initial Eigenvalues, % of Variance, and percent of variance cumulated over the four factors (Cumulative %) are printed out for the four initial factors. (Be careful not to confuse factors with variables.) The remainder of the table shows the percent of variance (sums of squared loadings—see Table 13.3) accounted for by the two factors extracted with eigenvalues greater than 1 (the default value), after extraction and after rotation.

For the two extracted factors, an unrotated Factor (loading) Matrix is then printed. The Rotated Factor Matrix, which matches loadings in Table 13.4, is given along with the Factor Transformation Matrix (Equation 13.8) for orthogonal varimax rotation with Kaiser's normalization.

SAS FACTOR (Table 13.6) requires a bit more instruction to produce a PFA with orthogonal rotation for two factors. You specify the type (`method=prinit`), initial communalities (`priors=smc`), number of factors to be extracted (`nfactors=2`), and the type of rotation (`rotate=v`). `Prior Communality Estimates: SMCs` are given, followed by `Preliminary Eigenvalues` for all four factors; also given is the `Total` of the eigenvalues and their `Average`. The next row shows `Differences` between successive eigenvalues. For example, there is a small difference between the first and second eigenvalues (0.099606) and between the third and fourth eigenvalues (0.020622), but a large difference between the second and third eigenvalues (1.897534). `Proportion` and `Cumulative` proportion of variance are then printed for each factor. This is followed by corresponding information for the `Reduced Correlation Matrix` (after factoring). Information on the iterative process is not shown.

The `Factor Pattern` matrix contains unrotated factor loadings for the first two factors. (Note that the signs of the FACTOR2 loadings are the reverse of those of IBM SPSS.) SSLs for each factor are in the table labeled `Variance explained by each factor.` Both `Final Communality Estimates` (h^2) and the `Total` h^2 are then given. The `Orthogonal Transformation Matrix` for rotation (Equation 13.8) is followed by the rotated factor loadings in the `Rotated Factor Pattern` matrix. SSLs for rotated factors—`Variance explained by each factor`—appear below the loadings. `Final Communality Estimates` are then repeated.

13.5 Major Types of Factor Analyses

Numerous procedures for factor extraction and rotation are available. However, only those procedures available in IBM SPSS and SAS packages are summarized here. Other extraction and rotational techniques are described in Mulaik (1972), Harman (1976), Rummel (1970), Comrey and Lee (1992), and Gorsuch (1983), among others.

13.5.1 Factor Extraction Techniques

Among the extraction techniques available in the packages are principal components (PCA), principal factors, maximum likelihood factoring, image factoring, alpha factoring, and unweighted and generalized (weighted) least squares factoring (see Table 13.7). Of these, PCA and principal factors are the most commonly used.

All the extraction techniques calculate a set of orthogonal components or factors that, in combination, reproduce **R**. Criteria used to establish the solution, such as maximizing variance

TABLE 13.7 Summary of Extraction Procedures

Extraction Technique	Program	Goal of Analysis	Special Features
Principal components	IBM SPSS SAS	Maximize variance extracted by orthogonal components	Mathematically determined, empirical solution with common, unique, and error variance mixed into components
Principal factors	IBM SPSS SAS	Maximize variance extracted by orthogonal factors	Estimates communalities to attempt to eliminate unique and error variance from variables
Image factoring	IBM SPSS SAS (Image and Harris)	Provides an empirical factor analysis	Uses variances based on multiple regression of a variable with all other variables as communalities to generate a mathematically determined solution with error variance and unique variance eliminated
Maximum likelihood factoring	SAS IBM SPSS	Estimate factor loadings for population that maximize the likelihood of sampling the observed correlation matrix	Has significance test for factors; especially useful for confirmatory factor analysis
Alpha factoring	IBM SPSS SAS	Maximize the generalizability of orthogonal factors	
Unweighted least squares	IBM SPSS SAS	Minimize squared residual correlations	
Generalized least squares	IBM SPSS SAS	Weights variables by shared variance before minimizing squared residual correlations	

or minimizing residual correlations, differ from technique to technique. But differences in solutions are small for a data set with a large sample, numerous variables, and similar communality estimates. In fact, one test of the stability of a FA solution is that it appears regardless of which extraction technique is employed. Table 13.8 shows solutions for the same data set after extraction with several different techniques, followed by varimax rotation. Similarities among the solutions are obvious.

None of the extraction techniques routinely provides an interpretable solution without rotation. All types of extractions may be rotated by any of the procedures described in Section 13.5.2.

Lastly, when using FA the researcher should hold in abeyance well-learned proscriptions against data snooping. It is quite common to use PCA or PFA as a preliminary extraction technique, followed by one or more of the other procedures, perhaps varying number of factors, communality estimates, and rotational methods with each run. Analysis terminates when the researcher decides on the preferred solution.

TABLE 13.8 Results of Different Extraction Methods on Same Data Set

	Factor 1				Factor 2			
Variables	*PCA*	*PFA*	*Rao*	*Alpha*	*PCA*	*PFA*	*Rao*	*Alpha*
	Unrotated factor loadings							
1	.58	.63	.70	.54	.68	.68	−.54	.76
2	.51	.48	.56	.42	.66	.53	−.47	.60
3	.40	.38	.48	.29	.71	.55	−.50	.59
4	.69	.63	.55	.69	−.44	−.43	.54	−.33
5	.64	.54	.48	.59	−.37	−.31	.40	−.24
6	.72	.71	.63	.74	−.47	−.49	.59	−.40
7	.63	.51	.50	.53	−.14	−.12	.17	−.07
8	.61	.49	.47	.50	−.09	−.09	.15	−.03
	Rotated factor loadings (varimax)							
1	.15	.15	.15	.16	.89	.91	.87	.92
2	.11	.11	.10	.12	.83	.71	.72	.73
3	−.02	.01	.02	.00	.81	.67	.69	.66
4	.82	.76	.78	.76	−.02	−.01	−.03	.01
5	.74	.62	.62	.63	.01	.04	.03	.04
6	.86	.86	.87	.84	.04	−.02	−.01	−.03
7	.61	.49	.48	.50	.20	.18	.21	.17
8	.57	.46	.45	.46	.23	.20	.20	.19

Note: The largest difference in communality estimates for a single variable between extraction techniques was 0.08.

13.5.1.1 PCA Versus FA

One of the most important decisions is the choice between PCA and FA. Mathematically, the difference involves the contents of the positive diagonal in the correlation matrix (the diagonal that contains the correlation between a variable and itself). In either PCA or FA, the variance that is analyzed is the sum of the values in the positive diagonal. In PCA, ones are in the diagonal and there is as much variance to be analyzed as there are observed variables; each variable contributes a unit of variance by contributing a 1 to the positive diagonal of the correlation matrix. All the variance is distributed to components, including error and unique variance for each observed variable. So if all components are retained, PCA duplicates exactly the observed correlation matrix and the standard scores of the observed variables.

In FA, on the other hand, only the variance that each observed variable shares with other observed variables is available for analysis. Exclusion of error and unique variance from FA is based on the belief that such variance only confuses the picture of underlying processes. Shared variance is estimated by *communalities*, values between 0 and 1 that are inserted in the positive diagonal of the correlation matrix.[7] The solution in FA concentrates on variables with high communality values. The sum of the communalities (sum of the SSLs) is the variance that is distributed among factors and is less than the total variance in the set of observed variables. Because unique and error variances are omitted, a linear combination of factors approximates, but does not duplicate, the observed correlation matrix and scores on observed variables.

PCA analyzes variance; FA analyzes covariance (communality). The goal of PCA is to extract maximum variance from a data set with a few orthogonal components. The goal of FA is to reproduce the correlation matrix with a few orthogonal factors. PCA is a unique mathematical solution, whereas most forms of FA are not unique.

The choice between PCA and FA depends on your assessment of the fit between the models, the data set, and the goals of the research. If you are interested in a theoretical solution uncontaminated by unique and error variability and have designed your study on the basis of underlying constructs that are expected to produce scores on your observed variables, FA is your choice. If, on the other hand, you simply want an empirical summary of the data set, PCA is the better choice.

13.5.1.2 Principal Components

The goal of PCA is to extract maximum variance from the data set with each component. The first principal component is the linear combination of observed variables that maximally separates subjects by maximizing the variance of their component scores. The second component is formed from residual correlations; it is the linear combination of observed variables that extracts maximum variability uncorrelated with the first component. Subsequent components also extract maximum variability from residual correlations and are orthogonal to all previously extracted components.

The principal components are ordered, with the first component extracting the most variance and the last component the least variance. The solution is mathematically unique and, if all components are retained, exactly reproduces the observed correlation matrix. Further, since the components are orthogonal, their use in other analyses (e.g., as DVs in MANOVA) may greatly facilitate interpretation of results.

PCA is the solution of choice for the researcher who is primarily interested in reducing a large number of variables down to a smaller number of components. PCA is also useful as an initial step in FA where it reveals a great deal about maximum number and nature of factors.

13.5.1.3 Principal Factors

Principal factors extraction differs from PCA in that estimates of communality, instead of ones, are in the positive diagonal of the observed correlation matrix. These estimates are derived through an iterative procedure, with SMCs (squared multiple correlations of each variable with all other variables) used as the starting values in the iteration. The goal of analysis, like that for PCA, is to extract maximum orthogonal variance from the data set with each succeeding factor. Advantages to principal factors extraction are that it is widely used (and understood) and that it conforms to the factor analytic model in which common variance is analyzed with unique and error variance removed. Because the goal is to maximize variance extracted, however, principal factors extraction is sometimes not as good as other extraction techniques in reproducing the correlation matrix. Also,

[7]Maximum likelihood extraction manipulates off-diagonal elements rather than values in the diagonal.

communalities must be estimated and the solution is, to some extent, determined by those estimates. PFA is available through both IBM SPSS and SAS.

13.5.1.4 Image Factor Extraction

The technique is called image factoring because the analysis distributes among factors the variance of an observed variable that is *reflected* by the other variables, in a manner similar to the SMC. Image factor extraction provides an interesting compromise between PCA and principal factors. Like PCA, image extraction provides a mathematically unique solution because there are fixed values in the positive diagonal of **R**. Like principal factors, the values in the diagonal are communalities with unique and error variability excluded.

Image scores for each variable are produced by multiple regression, with each variable, in turn, serving as a DV. A covariance matrix is calculated from these image (predicted) scores. The variances from the image score covariance matrix are the communalities for factor extraction. Care is necessary in interpreting the results of image analysis, because loadings represent covariances between variables and factors rather than correlations.

Image factoring is available through IBM SPSS, and SAS FACTOR (with two types—"image" and Harris component analysis).

13.5.1.5 Maximum Likelihood Factor Extraction

The maximum likelihood method of factor extraction was developed originally by Lawley in the 1940s (see Lawley & Maxwell, 1963). Maximum likelihood extraction estimates population values for factor loadings by calculating loadings that maximize the probability of sampling the observed correlation matrix from a population. Within constraints imposed by the correlations among variables, population estimates for factor loadings are calculated that have the greatest probability of yielding a sample with the observed correlation matrix. This method of extraction also maximizes the canonical correlations between the variables and the factors (see Chapter 12).

Maximum likelihood extraction is available through IBM SPSS FACTOR and SAS FACTOR.

13.5.1.6 Unweighted Least Squares Factoring

The goal of unweighted least squares factor extraction is to minimize squared differences between the observed and reproduced correlation matrices. Only off-diagonal differences are considered; communalities are derived from the solution rather than estimated as part of the solution. Thus, unweighted least squares factoring can be seen as a special case of PFA in which communalities are estimated after the solution.

The procedure, originally called minimum residual, was developed by Comrey (1962) and later modified by Harman and Jones (1966). The latter procedure is available through IBM SPSS FACTOR and SAS FACTOR.

13.5.1.7 Generalized (Weighted) Least Squares Factoring

Generalized least squares extraction also seeks to minimize (off-diagonal) squared differences between observed and reproduced correlation matrices, but in this case weights are applied to the variables. Variables that have substantial shared variance with other variables are weighted more heavily than variables that have substantial unique variance. In other words, variables that are not as strongly related to other variables in the set are not as important to the solution. This method of extraction is available through IBM SPSS FACTOR and SAS FACTOR.

13.5.1.8 Alpha Factoring

Alpha factor extraction, available through IBM SPSS FACTOR and SAS FACTOR, grew out of psychometric research, where the interest is in discovering which common factors are found consistently when repeated samples of variables are taken from a population of *variables.* The problem is the same as identifying mean differences that are found consistently among samples of subjects taken from a population of subjects—a question at the heart of most univariate and multivariate statistics.

In alpha factoring, however, the concern is with the reliability of the common factors rather than with the reliability of group differences. Coefficient alpha is a measure derived in psychometrics for the reliability (also called generalizability) of a score taken in a variety of situations. In alpha factoring, communalities that maximize coefficient alpha for the factors are estimated using iterative procedures (and sometimes exceed 1.0).

Probably, the greatest advantage to the procedure is that it focuses the researcher's attention squarely on the problem of sampling variables from the domain of variables of interest. Disadvantages stem from the relative unfamiliarity of most researchers with the procedure and the reason for it.

13.5.2 Rotation

The results of factor extraction, unaccompanied by rotation, are likely to be hard to interpret regardless of which method of extraction is used. After extraction, rotation is used to improve the interpretability and scientific utility of the solution. It is *not* used to improve the quality of the mathematical fit between the observed and reproduced correlation matrices because all orthogonally rotated solutions are mathematically equivalent to one another and to the solution before rotation.

Just as different methods of extraction tend to give similar results with a good data set, different methods of rotation tend to give similar results if the pattern of correlations in the data is fairly clear. In other words, a stable solution tends to appear regardless of the method of rotation used.

A decision is required between orthogonal and oblique rotation. In orthogonal rotation, the factors are uncorrelated. Orthogonal solutions offer ease of interpreting, describing, and reporting results; yet they strain "reality" unless the researcher is convinced that underlying processes are almost independent. The researcher who believes that underlying processes are correlated uses an oblique rotation. In oblique rotation, the factors may be correlated with conceptual advantages but practical disadvantages in interpreting, describing, and reporting results.

Among the dozens of rotational techniques that have been proposed, only those available in both reviewed packages are included in this discussion (see Table 13.9). The reader who wishes to know more about these or other techniques is referred to Gorsuch (1983), Harman (1976), or Mulaik (1972). For the industrious, a presentation of rotation by hand is in Comrey and Lee (1992).

13.5.2.1 Orthogonal Rotation

Varimax, quartimax, and equamax—three orthogonal techniques—are available in both packages. Varimax is easily the most commonly used of all the rotations available.

TABLE 13.9 Summary of Rotational Techniques

Rotational Technique	Program	Type	Goals of Analysis	Comments
Varimax	SAS IBM SPSS	Orthogonal	Minimize complexity of factors (simplify columns of loading matrix) by maximizing variance of loadings on each factor.	Most commonly used rotation; recommended as default option
Quartimax	SAS IBM SPSS	Orthogonal	Minimize complexity of variables (simplify rows of loading matrix) by maximizing variance of loadings on each variable.	First factor tends to be general, with others subclusters of variables.
Equamax	SAS IBM SPSS	Orthogonal	Simplify both variables and factors (rows and columns); compromise between quartimax and varimax.	May behave erratically
Orthogonal with gamma (orthomax)	SAS	Orthogonal	Simplify either factors or variables, depending on the value of gamma (Γ).	Gamma (Γ) continuously variable
Parsimax	SAS	Orthogonal	Simplifies both variables and factors: $(\Gamma) = (p^*(m-1))/p + m - 2$.	
Direct oblimin	IBM SPSS	Oblique	Simplify factors by minimizing cross-products of loadings.	Continuous values of gamma, or delta, δ (SPSS), available; allows wide range of factor intercorrelations
(Direct) quartimin	IBM SPSS	Oblique	Simplify factors by minimizing sum of cross-products of squared loadings in pattern matrix.	Permits fairly high correlations among factors. Achieved in SPSS by setting $\delta = 0$ with direct oblimin.
Orthoblique	SAS (HK) IBM SPSS	Both orthogonal and oblique	Rescale factor loadings to yield orthogonal solution; non-rescaled loadings may be correlated.	
Promax	SAS	Oblique	Orthogonal factors rotated to oblique positions.	Fast
Procrustes	SAS	Oblique	Rotate to target matrix.	Useful in confirmatory FA

Just as the extraction procedures have slightly different statistical goals, so also the rotational procedures maximize or minimize different statistics. The goal of varimax rotation is to simplify factors by maximizing the variance of the loadings within factors, across variables. The spread in loadings is maximized—loadings that are high after extraction become higher after rotation and loadings that are low become lower. Interpreting a factor is easier because it is obvious which variables correlate with it. Varimax also tends to reapportion variance among factors so that they become relatively equal in importance; variance is taken from the first factors extracted and distributed among the later ones.

Quartimax does for variables what varimax does for factors. It simplifies variables by increasing the dispersion of the loadings within variables, across factors. Varimax operates on the columns of the loading matrix; quartimax operates on the rows. Quartimax is not nearly as popular as varimax because one is usually more interested in simple factors than in simple variables.

Equamax is a hybrid between varimax and quartimax that tries simultaneously to simplify the factors and the variables. Mulaik (1972) reports that equamax tends to behave erratically unless the researcher can specify the number of factors with confidence.

Thus, varimax rotation simplifies the factors, quartimax the variables, and equamax both. They do so by setting levels on a simplicity criterion—such as Γ (gamma)—of 1, 0, and $1/2$ respectively. Gamma can also be continuously varied between 0 (variables simplified) and 1 (factors simplified) by using the orthogonal rotation that allows the user to specify Γ level. In SAS FACTOR, this is done through orthomax with Γ. Parsimax in SAS uses a formula incorporating numbers of factors and variables to determine Γ (see Table 13.9).

Varimax is the orthogonal rotation of choice for many applications; it is the default option of packages that have defaults.

13.5.2.2 Oblique Rotation

An *embarrasse de richesse* awaits the researcher who uses oblique rotation (see Table 13.9). Oblique rotations offer a continuous range of correlations between factors. The amount of correlation permitted between factors is determined by a variable called delta (δ) by IBM SPSS FACTOR. The values of delta and gamma determine the maximum amount of correlation permitted among factors. When the value is less than zero, solutions are increasingly orthogonal; at about −4 the solution is orthogonal. When the value is zero, solutions can be fairly highly correlated. Values near 1 can produce factors that are very highly correlated. Although there is a relationship between values of delta or gamma and size of correlation, the maximum correlation at a given size of gamma or delta depends on the data set.

It should be stressed that factors do not necessarily correlate when an oblique rotation is used. Often, in fact, they do not correlate and the researcher reports the simpler orthogonal rotation. The family of procedures used for oblique rotation with varying degrees of correlation in IBM SPSS is direct oblimin. In the special case where Γ or $\delta = 0$ (the default option for the programs), the procedure is called direct quartimin. Values of gamma or delta greater than zero permit high correlations among factors, and the researcher should take care that the correct number of factors is chosen. Otherwise, highly correlated factors may be indistinguishable one from the other. Some trial and error, coupled with inspection of the scatterplots of relationships between pairs of factors, may be required to determine the most useful size of gamma or delta. Or, one might simply trust the default value.

Orthoblique rotation uses the quartimax algorithm to produce an orthogonal solution *on rescaled factor loadings*; therefore, the solution may be oblique with respect to the original factor loadings.

In promax rotation, available through SAS and IBM SPSS, an orthogonally rotated solution (usually varimax) is rotated again to allow correlations among factors. The orthogonal loadings are raised to powers (usually powers of 2, 4, or 6) to drive small and moderate loadings to zero while larger loadings are reduced, but not to zero. Even though factors correlate, simple structure is maximized by clarifying which variables do and do not correlate with each factor. Promax has the additional advantage of being fast.

In Procrustean rotation, available in SAS, a target matrix of loadings (usually zeros and ones) is specified by the researcher, and a transformation matrix is sought to rotate extracted factors to the target, if possible. If the solution can be rotated to the target, then the hypothesized factor structure is said to be confirmed. Unfortunately, as Gorsuch (1983) reports, with Procrustean rotation, factors are often extremely highly correlated and sometimes a correlation matrix generated by random processes is rotated to a target with apparent ease.

13.5.2.3 Geometric Interpretation

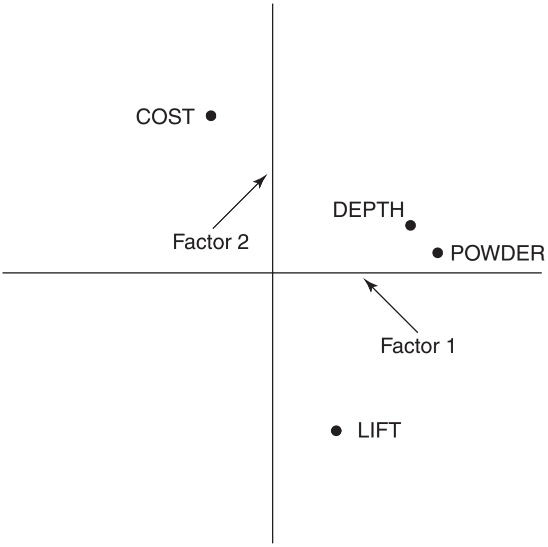
A geometric interpretation of rotation is in Figure 13.1 where 13.1(a) is the unrotated and 13.1(b) the rotated solution to the example in Table 13.2. Points are represented in two-dimensional space by listing their coordinates with respect to X and Y axes. With the first two unrotated factors as axes, unrotated loadings are COST (−.400, .900), LIFT (.251, −.947), DEPTH (.932, .348), and POWDER (.956, .286).

The points for these variables are also located with respect to the first two rotated factors as axes in Figure 13.1(b). The position of points does not change, but their coordinates change in the new axis system. COST is now (−.086, .981), LIFT (−.071, −.977), DEPTH (.994, .026), and POWDER (.997, −.040). Statistically, the effect of rotation is to amplify high loadings and reduce low ones. Spatially, the effect is to rotate the axes so that they "shoot through" the variable clusters more closely.

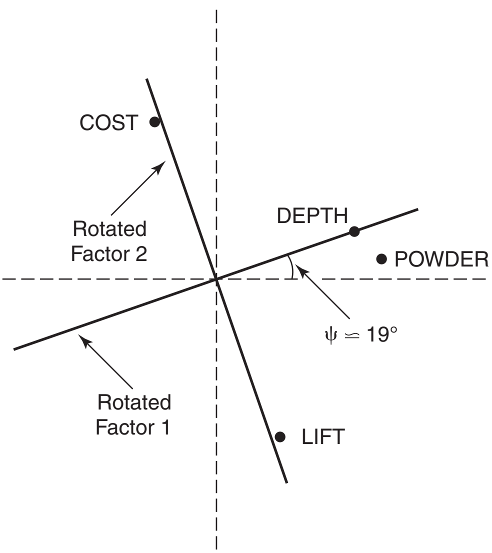
Factor extraction yields a solution in which observed variables are vectors that run from the origin to the points indicated by the coordinate system. The factors serve as axes for the system. The coordinates of each point are the entries from the loading matrix for the variable. If there are three factors, then the space has three axes and three dimensions, and each observed variable is positioned by three coordinates. The length of the vector for each variable is the communality of the variable.

If the factors are orthogonal, the factor axes are all at right angles to one another and the coordinates of the variable points are correlations between the common factors and the observed variables. Correlations (factor loadings) are read directly from these graphs by projecting perpendicular lines from each point to each of the factor axes.

One of the primary goals of PCA or FA, and the motivation behind extraction, is to discover the minimum number of factor axes needed to reliably position variables. A second major goal, and the motivation behind rotation, is to discover the meaning of the factors that underlie responses to observed variables. This goal is met by interpreting the factor axes that are used to define the space. Factor rotation repositions factor axes so as to make them maximally interpretable. Repositioning the axes changes the coordinates of the variable points but not the positions of the points with respect to each other.



(a) Location of COST, LIFT, DEPTH, and POWDER after extraction, before rotation



(b) Location of COST, LIFT, DEPTH, and POWDER vis-à-vis rotated axes

FIGURE 13.1 Illustration of rotation of axes to provide a better definition of factors vis-à-vis the variables with which they correlate.

Factors are usually interpretable when some observed variables load highly on them and the rest do not. And, ideally, each variable loads on one, and only one, factor. In graphic terms, this means that the point representing each variable lies far out along one axis but near the origin on the other axes, that is, that coordinates of the point are large for one axis and near zero for the other axes.

If you have only one observed variable, it is trivial to position the factor axis—variable point and axis overlap in a space of one dimension. However, with many variables and several factor axes, compromises are required in positioning the axes. The variables form a "swarm" in which variables that are correlated with one another form a cluster of points. The goal is to shoot an axis to the swarm of points. With luck, the swarms are about 90° away from one another so that an orthogonal solution is indicated. And with lots of luck, the variables cluster in just a few swarms with empty spaces between them so that the factor axes are nicely defined.

In oblique rotation the situation is slightly more complicated. Because factors may correlate with one another, factor axes are not necessarily at right angles. And, although it is easier to position each axis near a cluster of points, axes may be very near each other (highly correlated), making the solution harder to interpret. See Section 13.6.3 for practical suggestions of ways to use graphic techniques to judge the adequacy of rotation.

13.5.3 Some Practical Recommendations

Although an almost overwhelmingly large number of combinations of extraction and rotation techniques is available, in practice differences among them are often trivial (Velicer & Jackson, 1990; Fava & Velicer, 1992). The results of extraction are similar regardless of the method used when there are a large number of variables with some strong correlations among them, with the same, well-chosen number of factors, and with similar values for communality. Further, differences that are apparent after extraction tend to disappear after rotation.

Most researchers begin their FA by using principal components extraction and varimax rotation. From the results, one estimates the factorability of the correlation matrix (Section 13.3.2.6), the rank of the observed correlation matrix (Sections 13.3.2.5 and 13.7.1.5), the likely number of factors (Section 13.6.2), and variables that might be excluded from subsequent analyses (Sections 13.3.2.7 and 13.7.1.7).

During the next few runs, researchers experiment with different numbers of factors, different extraction techniques, and both orthogonal and oblique rotations. Some number of factors with some combination of extraction and rotation produces the solution with the greatest scientific utility, consistency, and meaning; this is the solution that is interpreted.

13.6 Some Important Issues

Some of the issues raised in this section can be resolved through several different methods. Usually, different methods lead to the same conclusion; occasionally they do not. When they do not, results are judged by the interpretability and scientific utility of the solutions.

13.6.1 Estimates of Communalities

FA differs from PCA in that communality values (numbers between 0 and 1) replace ones in the positive diagonal of **R** before factor extraction. Communality values are used instead of ones to remove the unique and error variance of each observed variable; only the variance a variable shares with the factors is used in the solution. But communality values are estimated, and there is some dispute regarding how that should be done.

The SMC of each variable as DV with the others in the sample as IVs is usually the starting estimate of communality. As the solution develops, communality estimates are adjusted by iterative procedures (which can be directed by the researcher) to fit the reproduced to the observed correlation matrix with the smallest number of factors. Iteration stops when successive communality estimates are very similar.

Final estimates of communality are also SMCs, but now between each variable as DV and the factors as IVs. Final communality values represent the proportion of variance in a variable that is predictable from the factors underlying it. Communality estimates do not change with orthogonal rotation.

Image extraction and maximum likelihood extraction are slightly different. In image extraction, variances from the image covariance matrix are used as the communality values throughout. Image extraction produces a mathematically unique solution because communality values are not changed. In maximum likelihood extraction, number of factors instead of communality values are estimated and off-diagonal correlations are "rigged" to produce the best fit between observed and reproduced matrices.

IBM SPSS and SAS provide several different starting statistics for communality estimation. IBM SPSS FACTOR permits user supplied values for principal factor extraction only, but otherwise uses SMCs. SAS FACTOR offers, for each variable, a choice of SMC: SMC adjusted so that the sum of the communalities is equal to the sum of the maximum absolute correlations, maximum absolute correlation with any other variable, user-specified values, or random numbers between 0 and 1.

The seriousness with which estimates of communality should be regarded depends on the number of observed variables. If there are 20 or more variables, sample SMCs probably provide reasonable estimates of communality. Furthermore, with 20 or more variables, the elements in the positive diagonal are few compared with the total number of elements in R, and their sizes do not influence the solution very much. Actually, if the communality values for all variables in FA are of approximately the same magnitude, results of PCA and FA are very similar (Velicer & Jackson, 1990; Fava & Velicer, 1992).

If communality values equal or exceed 1, problems with the solution are indicated. There is too little data, or starting communality values are wrong, or the number of factors extracted is wrong; addition or deletion of factors may reduce the communality below 1. Very low communality values, on the other hand, indicate that the variables with them are unrelated to other variables in the set (Sections 13.3.2.7 and 13.7.1.7). SAS FACTOR has two alternatives for dealing with communalities greater than 1; HEYWOOD sets them to 1, and ULTRAHEYWOOD allows them to exceed 1, but warns that doing so can cause convergence problems.

13.6.2 Adequacy of Extraction and Number of Factors

Because inclusion of more factors in a solution improves the fit between observed and reproduced correlation matrices, adequacy of extraction is tied to number of factors. The more factors extracted, the better the fit and the greater the percent of variance in the data "explained" by the factor solution. However, the more factors extracted, the less parsimonious the solution. To account for all the variance (PCA) or covariance (FA) in a data set, one would normally have to have as many factors as observed variables. It is clear, then, that a trade-off is required: One wants to retain enough factors for an adequate fit, but not so many that parsimony is lost.

Selection of the number of factors is probably more critical than selection of extraction and rotational techniques or communality values. In confirmatory FA, selection of the number of factors is really selection of the number of theoretical processes underlying a research area. You can partially confirm a hypothesized factor structure by asking if the theoretical number of factors adequately fits the data.

There are several ways to assess adequacy of extraction and number of factors. For a highly readable summary of these methods, not all currently available through the statistical packages, see Gorsuch (1983) and Zwick and Velicer (1986). Reviewed below are methods available through IBM SPSS and SAS.

A first quick estimate of the number of factors is obtained from the sizes of the eigenvalues reported as part of an initial run with principal components extraction. Eigenvalues represent variance. Because the variance that each standardized variable contributes to a principal components extraction is 1, a component with an eigenvalue less than 1 is not as important, from a variance perspective, as an observed variable. The number of components with eigenvalues greater than 1 is usually somewhere between the number of variables divided by 3 and the number of variables divided by 5 (e.g., 20 variables should produce between 7 and 4 components with eigenvalues greater than 1). If this is a reasonable number of factors for the data, if the number of variables is 40 or fewer, and if sample size is large, the number of factors indicated by this criterion is probably about right. In other situations, this criterion is likely to overestimate the number of factors in the data set.

A second criterion is the scree test (Cattell, 1966) of eigenvalues plotted against factors. Factors, in descending order, are arranged along the abscissa with eigenvalue as the ordinate. The plot is appropriately used with principal components or factor analysis at initial and later runs to find the number of factors. The scree plot is available through IBM SPSS and SAS FACTOR.

Usually the scree plot is negatively decreasing—the eigenvalue is highest for the first factor and moderate but decreasing for the next few factors before reaching small values for the last several factors, as illustrated for real data through IBM SPSS in Figure 13.2. You look for the point where a line drawn through the points changes slope. In the example, a single straight line can comfortably fit the first four eigenvalues. After that, another line, with a noticeably different slope, best fits the remaining eight points. Therefore, there appear to be about four factors in the data of Figure 13.2.

Unfortunately, the scree test is not exact; it involves judgment of where the discontinuity in eigenvalues occurs and researchers are not perfectly reliable judges. As Gorsuch (1983) reports, results of the scree test are more obvious (and reliable) when sample size is large, communality values are high, and each factor has several variables with high loadings. Zoski and Jurs (1996) recommend a refinement to the visual scree test that involves computing the standard error of the eigenvalues for the last few components.

Horn (1965) proposed parallel analysis as an alternative to retaining all principal components with eigenvalues larger than 1. This is a three step process. First, a randomly generated data set

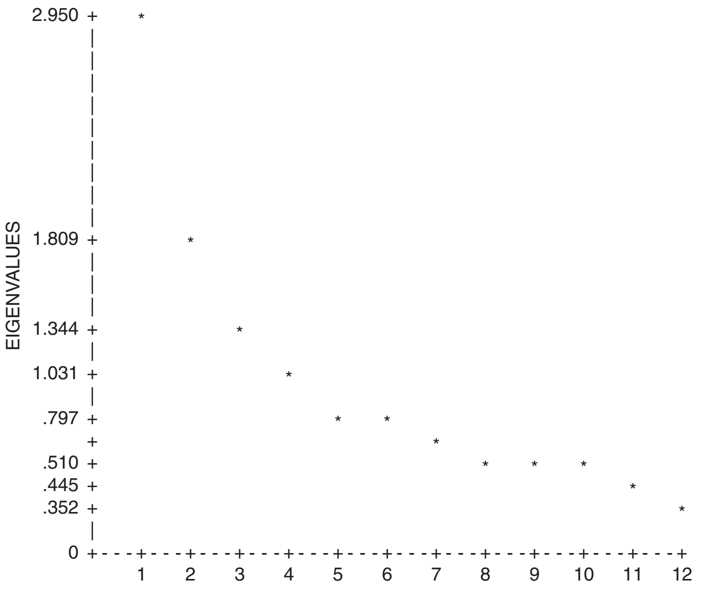


FIGURE 13.2 Scree output produced by IBM SPSS FACTOR. Note break in size of eigenvalues between the fourth and fifth factors.

with the same number of cases and variables is generated. Next, PCA is repeatedly performed on the randomly generated data set and all eigenvalues noted for each analysis. Those eigenvalues are then averaged for each component and compared to the results from the real data set. Only components from the real data set whose eigenvalues exceed the averaged eigenvalue from the randomly generated data set are retained. A major advantage to this procedure is to remind the user that even randomly generated data can have relationships based on chance that produce components with eigenvalues larger than 1, sometimes substantially so.

As an alternative, Velicer (1976) proposed the minimum average partial correlation (MAP) test. The first step is to perform PCA with one component. Partial correlation is used to take the variance of the first component from the variable intercorrelations before the mean squared coefficient of all partial correlations (the values off of the main diagonal) is computed. Then PCA is performed with two components, and the procedure is repeated. Mean squared partial correlations are computed for all solutions until the minimum squared partial correlation is identified. The number of components that produces the minimum mean squared partial correlation is the number of components to retain. Gorsuch (1983) points out that this procedure does not work well when some components have only a few variables that load on them.

Zwick and Velicer (1986) tested the scree test, Horn's parallel test, and Velicer's MAP test (among others) in simulation studies using a data set with a clear factor structure. Both the parallel test and the minimum average partial test seemed to work well. These procedures have been extended successfully to PFA. O'Connor (2000) provides programs for conducting the parallel test and the minimum average partial test through both IBM SPSS and SAS.

Once you have determined the number of factors, it is important to look at the rotated loading matrix to determine the number of variables that load on each factor (see Section 13.6.5). If only one variable loads highly on a factor, the factor is poorly defined. If two variables load on a factor, then whether or not it is reliable depends on the pattern of correlations of these two variables with each other and with other variables in **R**. If the two variables are highly correlated with each other (say, $r > .70$) and relatively uncorrelated with other variables, the factor may be reliable. Interpretation of factors defined by only one or two variables is hazardous, however, under even the most exploratory factor analysis.

For principal components extraction and maximum likelihood extraction in confirmatory factor analysis done through FA rather than SEM programs there are significance tests for the number of factors. Bartlett's test evaluates all factors together and each factor separately against the hypothesis that there are no factors. However, there is some dispute regarding use of these tests. The interested reader is referred to Gorsuch (1983) or one of the other newer FA texts for discussion of significance testing in FA.

There is a debate about whether it is better to retain too many or too few factors if the number is ambiguous. Sometimes a researcher wants to rotate, but not interpret, marginal factors for statistical purposes (e.g., to keep some factors with communality values less than 1). Other times, the last few factors represent the most interesting and unexpected findings in a research area. These are good reasons for retaining factors of marginal reliability. However, if the researcher is interested in using only demonstrably reliable factors, the fewest possible factors are retained.

13.6.3 Adequacy of Rotation and Simple Structure

The decision between orthogonal and oblique rotation is made as soon as the number of reliable factors is apparent. In many factor analytic situations, oblique rotation seems more reasonable on the face of it than orthogonal rotation because it seems more likely that factors are correlated than that they are not. However, reporting the results of oblique rotation requires reporting the elements of the pattern matrix (**A**) and the factor correlation matrix (Φ) whereas reporting orthogonal rotation requires only the loading matrix (**A**). Thus, simplicity of reporting results favors orthogonal rotation. Further, if factor scores or factorlike scores (Section 13.6.6) are to be used as IVs or DVs in other analyses, or if a goal of analysis is comparison of factor structure in groups, then orthogonal rotation has distinct advantages.

Perhaps the best way to decide between orthogonal and oblique rotation is to request oblique rotation with the desired number of factors and look at the correlations among factors. The oblique rotations available by default in IBM SPSS and SAS calculate factors that are fairly highly correlated if necessary to fit the data. However, if factor correlations are not driven by the data, the solution remains nearly orthogonal.

Look at the factor correlation matrix for correlations around .32 and above. If correlations exceed .32, then there is 10% (or more) overlap in variance among factors, enough variance to warrant oblique rotation unless there are compelling reasons for orthogonal rotation. Compelling reasons include a desire to compare structure in groups, a need for orthogonal factors in other analyses, or a theoretical need for orthogonal rotation.

Once the decision is made between orthogonal and oblique rotation, the adequacy of rotation is assessed several ways. Perhaps, the simplest way is to compare the pattern of correlations in the correlation matrix with the factors. Are the patterns represented in the rotated solution? Do highly correlated variables tend to load on the same factor? If you included marker variables, do they load on the predicted factors?

Another criterion is simple structure (Thurstone, 1947). If simple structure is present (and factors are not too highly correlated), several variables correlate highly with each factor and only one factor correlates highly with each variable. In other words, the columns of **A**, which define factors vis-à-vis variables, have several high and many low values while the rows of **A**, which define variables vis-à-vis factors, have only one high value. Rows with more than one high correlation correspond to variables that are said to be complex because they reflect the influence of more than one factor. It is usually best to avoid complex variables because they make interpretation of factors more ambiguous.

Adequacy of rotation is also ascertained through the PLOT instructions of the four programs. In the figures, factors are considered two at a time with a different pair of factors as axes for each plot. Look at the *distance*, *clustering*, and *direction* of the points representing variables relative to the factor axes in the figures.

The *distance* of a variable point from the origin reflects the size of factor loadings; variables highly correlated with a factor are far out on that factor's axis. Ideally, each variable point is far out on one axis and near the origin on all others. *Clustering* of variable points reveals how clearly defined a factor is. One likes to see a cluster of several points near the end of each axis and all other points near the origin. A smattering of points at various distances along the axis indicates a factor that is not clearly defined, while a cluster of points midway between two axes reflects the presence of another factor or the need for oblique rotation. The *direction* of clusters after orthogonal rotation may also indicate the need for oblique rotation. If clusters of points fall between factor axes after orthogonal rotation, if the angle between clusters with respect to the origin is not 90°, then a better fit to the clusters is provided by axes that are not orthogonal. Oblique rotation may reveal substantial correlations among factors. Several of these relationships are depicted in Figure 13.3.

13.6.4 Importance and Internal Consistency of Factors

The importance of a factor (or a set of factors) is evaluated by the proportion of variance or covariance accounted for by the factor after rotation. The proportion of variance attributable to individual factors differs before and after rotation because rotation tends to redistribute variance among factors somewhat. Ease of ascertaining proportions of variance for factors depends on whether rotation was orthogonal or oblique.

After orthogonal rotation, the importance of a factor is related to the size of its SSLs (Sum of Squared Loadings from **A** after rotation). SSLs are converted to proportion of variance for a factor by dividing by p (the number of variables). SSLs are converted to proportion of covariance for a factor by dividing its SSL by the sum of SSLs or, equivalently, sum of communalities. These computations are illustrated in Table 13.4 and Section 13.7 for the example.

The proportion of variance accounted for by a factor is the amount of variance in the original variables (where each has contributed one unit of variance) that has been condensed into the factor. Proportion of *variance* is the variance of a factor relative to the variance in the variables. The proportion of covariance accounted for by a factor indicates the relative importance of the factor to the total covariance accounted for by all factors. Proportion of

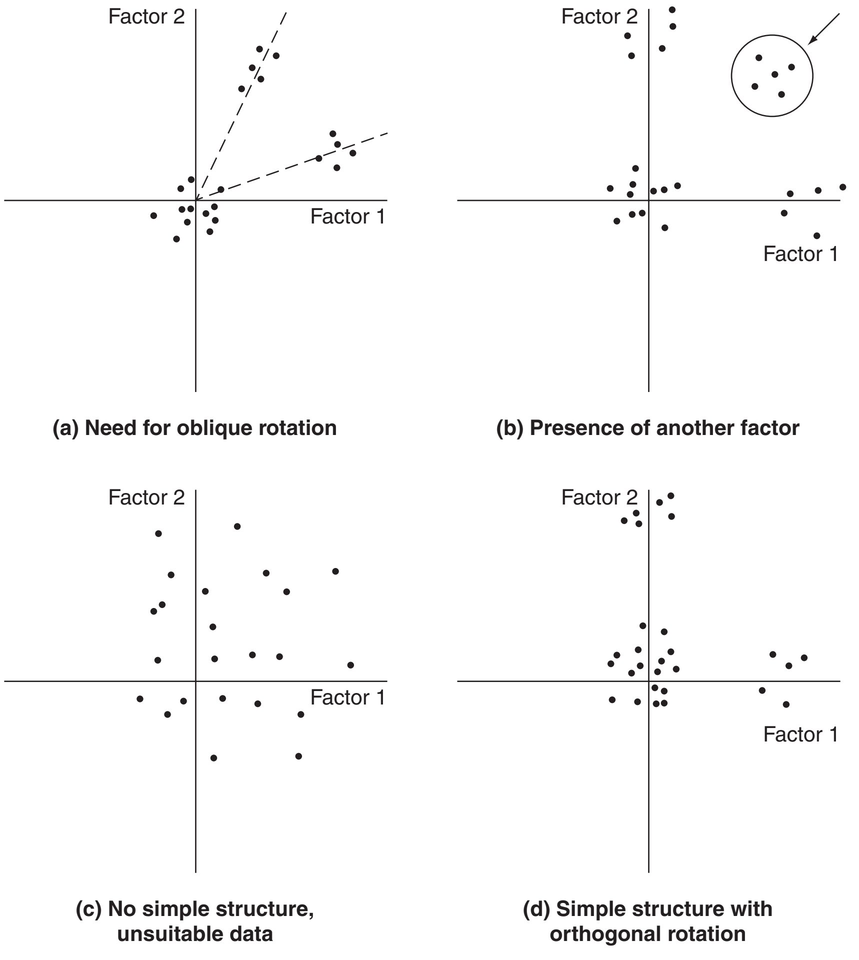


FIGURE 13.3 Pairwise plots of factor loadings following orthogonal rotation and indicating: (a) need for oblique rotation; (b) presence of another factor; (c) unsuitable data; and (d) simple structure.

covariance is the variance of a factor relative to the variance in the solution. The variance in the solution is likely to account for only a fraction of the variance in the original variables.

In oblique rotation, proportions of variance and covariance can be obtained from **A** *before* rotation by the methods just described, but they are only rough indicators of the proportions of variance and covariance of factors after rotation. Because factors are correlated, they share overlapping variability, and assignment of variance to individual factors is ambiguous. After oblique rotation, the size of the SSL associated with a factor is a rough approximation of its importance—factors with bigger SSLs are more important—but proportions of variance and covariance cannot be specified.

An estimate of the internal consistency of the solution—the certainty with which factor axes are fixed in the variable space—is given by the squared multiple correlations of factor scores predicted from scores on observed variables. In a good solution, SMCs range between 0 and 1; the larger the SMCs, the more stable the factors. A high SMC (say, .70 or better) means that the observed variables account for substantial variance in the factor scores. A low SMC means the factors are poorly defined by the observed variables. If an SMC is negative, too many factors have been retained. In oblique rotation, SMCs can exceed 1 and are, therefore, not an indication of factor stability.

IBM SPSS FACTOR prints these SMCs as the diagonal of the covariance matrix for estimated regression factor scores. In SAS FACTOR, SMCs are printed along with factor score coefficients by the SCORE option.

13.6.5 Interpretation of Factors

To interpret a factor, one tries to understand the underlying dimension that unifies the group of variables loading on it. In both orthogonal and oblique rotations, loadings are obtained from the loading matrix, **A**, but the meaning of the loadings is different for the two rotations.

After orthogonal rotation, the values in the loading matrix are correlations between variables and factors. The researcher decides on a criterion for meaningful correlation (usually .32 or larger), collects together the variables with loadings in excess of the criterion, and searches for a concept that unifies them.

After oblique rotation, the process is the same, but the interpretation of the values in **A**, the pattern matrix, is no longer straightforward. The loading is not a correlation but is a measure of the unique relationship between the factor and the variable. Because factors correlate, the correlations between variables and factors (available in the structure matrix, **C**) are inflated by overlap between factors. A variable may correlate with one factor through its correlation with another factor rather than directly. The elements in the pattern matrix have overlapping variance among factors "partialed out," but at the expense of conceptual simplicity.

Actually, the reason for interpretation of the pattern matrix rather than the structure matrix is pragmatic—it's easier. The difference between high and low loadings is more apparent in the pattern matrix than in the structure matrix.

As a rule of thumb, only variables with loadings of .32 and above are interpreted. The greater the loading, the more the variable is a pure measure of the factor. Comrey and Lee (1992) suggest that loadings in excess of .71 (50% overlapping variance) are considered excellent, .63 (40% overlapping variance) very good, .55 (30% overlapping variance) good, .45 (20% overlapping variance) fair, and .32 (10% overlapping variance) poor. Choice of the cutoff for size of loading to be interpreted is a matter of researcher preference. Sometimes there is a gap in loadings across the factors and, if the cutoff is in the gap, it is easy to specify which variables load and which do not. Other times, the cutoff is selected because one can interpret factors with that cutoff but not with a lower cutoff.

The size of loadings is influenced by the homogeneity of scores in the sample. If homogeneity is suspected, interpretation of lower loadings is warranted. That is, if the sample produces similar scores on observed variables, a lower cutoff is used for interpretation of factors.

At some point, a researcher usually tries to characterize a factor by assigning it a name or a label, a process that involves art as well as science. Rummel (1970) provides numerous helpful hints on interpreting and naming factors. Interpretation of factors is facilitated by output of the matrix of sorted loadings where variables are grouped by their correlations with factors. Sorted loadings are produced routinely by REORDER in SAS FACTOR, and SORT in IBM SPSS.

The replicability, utility, and complexity of factors are also considered in interpretation. Is the solution replicable in time and/or with different groups? Is it trivial or is it a useful addition to scientific thinking in a research area? Where do the factors fit in the hierarchy of "explanations" about a phenomenon? Are they complex enough to be intriguing without being so complex that they are uninterpretable?

13.6.6 Factor Scores

Among the potentially more useful outcomes of PCA or FA are factor scores. Factor scores are estimates of the scores that subjects would have received on each of the factors had they been measured directly.

Because there are normally fewer factors than observed variables, and because factor scores are nearly uncorrelated if factors are orthogonal, use of factor scores in other analyses may be very helpful. Multicollinear matrices can be reduced to orthogonal components using PCA, for instance. Or, one could use PCA to reduce a large number of DVs to a smaller number of components for use as DVs in MANOVA. Alternatively, one could reduce a large number of IVs to a small number of factors for purposes of predicting a DV in multiple regression or group membership in discriminant analysis or logistic regression. If factors are few in number, stable, and interpretable, their use enhances subsequent analyses. In the context of a theoretical FA, factor scores are estimates of the values that would be produced if the underlying constructs could be measured directly.

Procedures for estimating factor scores range between simple minded (but frequently adequate) and sophisticated. Comrey and Lee (1992) describe several rather simple-minded techniques for estimating factor scores. Perhaps the simplest is to sum scores on variables that load highly on each factor. Variables with bigger standard deviations contribute more heavily to the factor scores produced by this procedure, a problem that is alleviated if variable scores are standardized first or if the variables have roughly equal standard deviations to begin with. For many research purposes, this "quick and dirty" estimate of factor scores is entirely adequate.

There are several sophisticated statistical approaches to estimating factors. All produce factor scores that are correlated, but not perfectly, with the factors. The correlations between factors and factor scores are higher when communalities are higher and when the ratio of variables to factors is higher. But as long as communalities are estimated, factor scores suffer from indeterminacy because there is an infinite number of possible factor scores that all have the same mathematical characteristics. As long as factor scores are considered only estimates, however, the researcher is not overly beguiled by them.

The method described in Section 13.4 (especially Equations 13.10 and 13.11) is the regression approach to estimating factor scores. This approach results in the highest correlations between factors and factor scores. The distribution of each factor's scores has a mean of 0 and a standard deviation of 1 (after PCA) or equal to the SMC between factors and variables (after FA). However, this regression method, like all others (see Chapter 5), capitalizes on chance relationships among variables so that factor-score estimates are biased (too close to "true" factor scores). Further, there are often correlations among scores for factors even if factors are orthogonal and factor scores sometimes correlate with other factors (in addition to the one they are estimating).

The regression approach to estimating factor scores is available through SAS and IBM SPSS. Both packages write component/factor scores to files for use in other analyses. SAS and IBM SPSS print standardized component/factor score coefficients.

IBM SPSS FACTOR provides two additional methods of estimating factor scores. In the Bartlett method, factor scores correlate only with their own factors and the factor scores are unbiased (i.e., neither systematically too close nor too far away from "true" factor scores). The factor scores correlate with the factors almost as well as in the regression approach and have the same mean and standard deviation as in the regression approach. However, factor scores may still be correlated with each other.

The Anderson–Rubin approach (discussed by Gorsuch, 1983) produces factor scores that are uncorrelated with each other even if factors are correlated. Factor scores have mean 0 and standard deviation 1. Factor scores correlate with their own factors almost as well as in the regression approach, but they sometimes also correlate with other factors (in addition to the one they are estimating) and they are somewhat biased. If you need uncorrelated scores, the Anderson–Rubin approach is best; otherwise, the regression approach is probably best simply because it is best understood and most widely available.

13.6.7 Comparisons Among Solutions and Groups

Frequently, a researcher is interested in deciding whether or not two groups that differ in experience or characteristics have the same factors. Comparisons among factor solutions involve the *pattern* of the correlations between variables and factors, or both the *pattern and magnitude* of the correlations between them. Rummel (1970), Levine (1977), and Gorsuch (1983) have excellent summaries of several comparisons that might be of interest. Some of the simpler of these techniques are described in an earlier version of this book (Tabachnick & Fidell, 1989).

Tests of theory (in which theoretical factor loadings are compared with those derived from a sample) and comparisons among groups are currently the province of structural equation modeling. These techniques are discussed in Chapter 14.

13.7 Complete Example of FA

During the second year of the panel study described in Appendix B, Section B.1, participants completed the Bem Sex Role Inventory (BSRI; Bem, 1974). The sample included 351 middle-class, English-speaking women between the ages of 21 and 60 who were interviewed in person.

Forty-four items from the BSRI were selected for this research, where 20 items measure femininity, 20 masculinity,[8] and 5 social desirability. Respondents attribute traits (e.g., "gentle,"

[8]Due to clerical error, one of the masculine items, "aggression," was omitted from the questionnaires.

"shy," "dominant") to themselves by assigning numbers between 1 ("never or almost never true of me") and 7 ("always or almost always true of me") to each of the items. Responses are summed to produce separate masculine and feminine scores. Masculinity and femininity are conceived as orthogonal dimensions of personality with both, any, or neither descriptive of any given individual. Files are FACTOR.*.

Previous factor analytic work had indicated the presence of between three and five factors underlying the items of the BSRI. Investigation of the factor structure for this sample of women is a goal of this analysis.

13.7.1 Evaluation of Limitations

Because the BSRI was neither developed through nor designed for factor analytic work, it meets only marginally the requirements listed in Section 13.3.1. For instance, marker variables are not included and variables from the feminine scale differ in social desirability as well as in meaning (e.g., "tender" and "gullible"), so some of these variables are likely to be complex.

13.7.1.1 Sample Size and Missing Data

Data are available initially from 369 women, 18 of whom had missing values on one or more variables. With those cases deleted as well as 11 outlying cases (see below), the FA is conducted on responses of 340 women. Using the guidelines of Section 13.3.2.1, over 300 cases provide a good sample size for factor analysis.

13.7.1.2 Normality

Distributions of the 44 variables are examined for skewness through SAS MEANS (cf. Chapter 12). Many of the variables are negatively skewed and a few are positively skewed. However, because the BSRI is already published and is in use, no deletion of variables or transformations of them is performed.

Because the variables fail in normality, significance tests are inappropriate. And because the direction of skewness is different for different variables, we also anticipate a weakened analysis due to lowering of correlations in **R**.

13.7.1.3 Linearity

The differences in skewness for variables suggest the possibility of curvilinearity for some pairs of variables. With 44 variables, however, examination of all pairwise scatterplots (about 1,000 plots) is impractical. Therefore, a spot check on a few plots is run through SAS PLOT. Figure 13.4 shows the plot expected to be among the worst—between LOYAL (with strong negative skewness) and MASCULIN (with strong positive skewness). Although the plot is far from pleasing and shows departure from linearity as well as the possibility of outliers, there is no evidence of true curvilinearity (Section 4.1.5.2). And again, transformations are viewed with disfavor, considering the variable set and the goals of analysis.

```
      Plot of masculin*loyal.  Legend:  A = 1 obs, B = 2 obs, etc.
     |
   6 +                                         A         A         A
     |
     |
     |
     |
   5 +                                                   A         E
     |
     |
m    |
a  4 +                                                   F         P
s    |
c    |
u    |
l    |
i  3 +                                                   N         T
n    |
     |
     |
     |
   2 +                               A         D         U         Z
     |
     |
     |
     |
   1 + A                   C                   J         Z         Z
     |
     --+---------+---------+---------+---------+---------+---------+-
       1         2         3         4         5         6         7
                                   loyal
NOTE: 168 obs hidden.
```

FIGURE 13.4 Spot check for linearity among variables. Syntax and output from SAS PLOT.

13.7.1.4 Outliers

Multivariate outliers are identified using SAS REG (cf. Chapter 12) which adds a leverage variable to the data set, now labeled FACTORLEV. Eleven cases are identified as outliers using the criterion suggested by Lunneborg (1994) of critical $h_{ii} = 2(k/N) = 0.25$. The decision is made to delete the 11 cases, and to run remaining analyses on the data set with 340 cases.

Because of the large number of outliers and variables, a case-by-case analysis (cf. Chapter 4) is not feasible. Instead, a stepwise discriminant analysis is used to identify variables that significantly discriminate between outliers and nonoutliers. First, a variable labeled `DUMMY` is added to the data set, in which each outlier is coded 1 and the remaining cases are labeled 0. Then `DUMMY` is declared the `class` (grouping) variable in the stepwise regression run through SAS STEPDISC, as seen in Table 13.10. Means in each group are requested for all variables. On the last step of the discriminant analysis, five variables (RELIANT, DEFBE, LEADERAB, SELFSUFF, and WARM) discriminate outliers as a group with $p < .001$

A reduced data set that includes only the 340 nonoutlying cases is created, called FACTORR to be used for all subsequent analyses.

TABLE 13.10 Description of Variables Causing Multivariate Outliers Using SAS STEPDISC (Syntax and Selected Output)

```
proc stepdisc data=SASUSER.FACTORLEV;
  class DUMMY;
  var HELPFUL RELIANT DEFBEL YIELDING CHEERFUL INDPT ATHLET SHY ASSERT
  STRPERS FORCEFUL AFFECT FLATTER LOYAL ANALYT FEMININE SYMPATHY MOODY SENSITIV UNDSTAND
  COMPASS LEADERAB SOOTHE RISK DECIDE SELFSUFF CONSCIEN DOMINANT MASCULIN STAND HAPPY
  SOFTSPOK WARM TRUTHFUL TENDER GULLIBLE LEADACT CHILDLIK INDIVID FOULLANG LOVCHIL
  COMPETE AMBITIOU GENTLE;
run;
```

The STEPDISC Procedure
Simple Statistics

DUMMY = 0

Variable	Label	N	Sum	Mean	Variance	Standard Deviation
reliant	reliant	340	2027	5.96176	1.22862	1.1084
defbel	defbel	340	2033	5.97941	1.52465	1.2348
leaderab	leaderab	340	1574	4.62941	2.94780	1.7169
selfsuff	selfsuff	340	1957	5.75588	1.52430	1.2346
warm	warm	340	1928	5.67059	0.97081	0.9853

DUMMY = 1

Variable	Label	N	Sum	Mean	Variance	Standard Deviation
reliant	reliant	11	47.00000	4.27273	6.61818	2.5726
defbel	defbel	11	46.00000	4.18182	5.56364	2.3587
leaderab	leaderab	11	54.00000	4.90909	4.49091	2.1192
selfsuff	selfsuff	11	70.00000	6.36364	0.85455	0.9244
warm	warm	11	66.00000	6.00000	3.40000	1.8439

(continued)

TABLE 13.10 Continued

The STEPDISC Procedure
Stepwise Selection: Step 19

Statistics for Removal, DF = 1, 332

Variable	label	Partial R-Square	F Value	Pr > F
reliant	reliant	0.1152	43.22	<.0001
defbel	defbel	0.0347	11.92	0.0006
yielding	yielding	0.0324	11.11	0.0010
affect	affect	0.0189	6.38	0.0120
loyal	loyal	0.0230	7.81	0.0055
feminine	feminine	0.0148	5.00	0.0260
leaderab	leaderab	0.0716	25.59	<.0001
soothe	soothe	0.0202	6.85	0.0093
risk	risk	0.0130	4.36	0.0374
selfsuff	selfsuff	0.0721	25.81	<.0001
dominant	dominant	0.0095	3.20	0.0747
warm	warm	0.0406	14.05	0.0002
leadact	leadact	0.0683	24.34	<.0001
childlik	childlik	0.0210	7.11	0.0081
individ	individ	0.0063	2.09	0.1488
foullang	foullang	0.0132	4.43	0.0360
lovchil	lovchil	0.0298	10.21	0.0015
ambitiou	ambitiou	0.0108	3.63	0.0576

13.7.1.5 Multicollinearity and Singularity

Nonrotated PCA run through SAS FACTOR reveals that the smallest eigenvalue is 0.126, not dangerously close to 0. The largest SMC between variables where each, in turn, serves as DV for the others is .76, not dangerously close to 1 (Table 13.11). Multicollinearity is not a threat in this data set.

13.7.1.6 Factorability of **R**

The SAS FACTOR correlation matrix (not shown) reveals numerous correlations among the 44 items, well in excess of .30; therefore, patterns in responses to variables are anticipated. Table 13.11 syntax produces Kaiser's measures of sampling adequacy `(msa)`, which are acceptable because all are greater than .6 (not shown). Most of the values in the negative anti-image correlation matrix (also not shown) are small, another requirement for good FA.

13.7.1.7 Outliers Among Variables

SMCs among variables (Table 13.11) are also used to screen for outliers among variables, as discussed in Section 13.3.2.7. The lowest SMC among variables is .11. It is decided to retain all 44 variables although many are largely unrelated to others in the set. (In fact, 45% of the 44 variables in the analysis have loadings too low on all the factors to assist interpretation in the final solution.)

13.7.2 Principal Factors Extraction With Varimax Rotation

Principal components extraction with varimax rotation through SAS FACTOR is used in an initial run to estimate the likely number of factors from eigenvalues.[9] The first 13 eigenvalues are shown in Table 13.12. The maximum number of factors (eigenvalues larger than 1) is 11. However, retention of 11 factors seems unreasonable so sharp breaks in size of eigenvalues are sought using the scree test (Section 13.6.2).

`Eigenvalue`s for the first four factors are all larger than two, and, after the sixth factor, changes in successive eigenvalues are small. This is taken as evidence that there are probably between four and six factors. The scree plot visually suggests breaks between four and six factors. These results are consistent with earlier research suggesting three to five factors on the BSRI.

A common factor extraction model that removes unique and error variability from each variable is used for the next several runs and the final solution. Principal factors are chosen from among methods for common factor extraction. Several PFA runs specifying four to six factors are planned to find the optimal number of factors.

The PFA run with five factors has five eigenvalues larger than 1 among rotated factors and the fifth factor has three loadings larger than .45, the criterion for interpretation chosen for this research. The first seven eigenvalues from the five-factor solution are shown in Table 13.13.

As another test of adequacy of extraction and number of factors, it is noted (but not shown) that most values in the residual correlation matrix for the five-factor orthogonal solution are near zero. This is further confirmation that a reasonable number of factors is five.

[9]Principal components extraction is chosen to estimate the maximum number of factors that might be interesting. PFA, which produces fewer eigenvalues greater than 1, is a reasonable alternative for estimation.

TABLE 13.11 Syntax and Selected SAS Factor Output to Assess Multicollinearity

```
proc factor data=SASUSER.FACTORR prior=smc msa;
   var HELPFUL RELIANT DEFBEL YIELDING CHEERFUL INDPT ATHLET SHY ASSERT
  STRPERS FORCEFUL AFFECT FLATTER LOYAL ANALYT FEMININE SYMPATHY MOODY SENSITIV UNDSTAND
  COMPASS LEADERAB SOOTHE RISK DECIDE SELFSUFF CONSCIEN DOMINANT MASCULIN STAND HAPPY
  SOFTSPOK WARM TRUTHFUL TENDER GULLIBLE LEADACT CHILDLIK INDIVID FOULLANG LOVCHIL
  COMPETE AMBITIOU GENTLE;
run;
```

Prior Communality Estimates: SMC

helpful	reliant	defbel	yielding	cheerful	indpt	athlet	shy
0.37160303	0.48940632	0.40430726	0.22938776	0.49962905	0.54110375	0.26006896	0.31873301
assert	strpers	forceful	affect	flatter	loyal	analyt	feminine
0.56223304	0.61367020	0.56839062	0.55534755	0.28016976	0.38123886	0.23166431	0.38508982
sympathy	moody	sensitiv	undstand	compass	leaderab	soothe	risk
0.43611044	0.36566060	0.47547415	0.61093141	0.64882134	0.76236310	0.43642100	0.41069381
decide	selfsuff	conscien	dominant	masculin	stand	happy	softspok
0.49669066	0.65463203	0.39122820	0.56226427	0.35126516	0.55484492	0.54313590	0.38918408
warm	truthful	tender	gullible	leadact	childlik	individ	foullang
0.61341943	0.34495670	0.58662165	0.28701425	0.75777345	0.29636797	0.37887704	0.11353506
		lovchil	compete	ambitiou	gentle		
		0.28560598	0.47864699	0.46940471	0.59198996		

TABLE 13.12 Eigenvalues and Proportions of Variance for First 13 Components (SAS FACTOR PCA Syntax and Selected Output)

```
proc factor data=SASUSER.FACTORR simple corr scree;
     var HELPFUL RELIANT DEFBEL YIELDING CHEERFUL INDPT ATHLET SHY ASSERT
  STRPERS FORCEFUL AFFECT FLATTER LOYAL ANALYT FEMININE SYMPATHY MOODY SENSITIV UNDSTAND
  COMPASS LEADERAB SOOTHE RISK DECIDE SELFSUFF CONSCIEN DOMINANT MASCULIN STAND HAPPY
  SOFTSPOK WARM TRUTHFUL TENDER GULLIBLE LEADACT CHILDLIK INDIVID FOULLANG LOVCHIL
  COMPETE AMBITIOU GENTLE;
run;
```

Eigenvalues of the Correlation Matrix: Total = 44 Average = 1

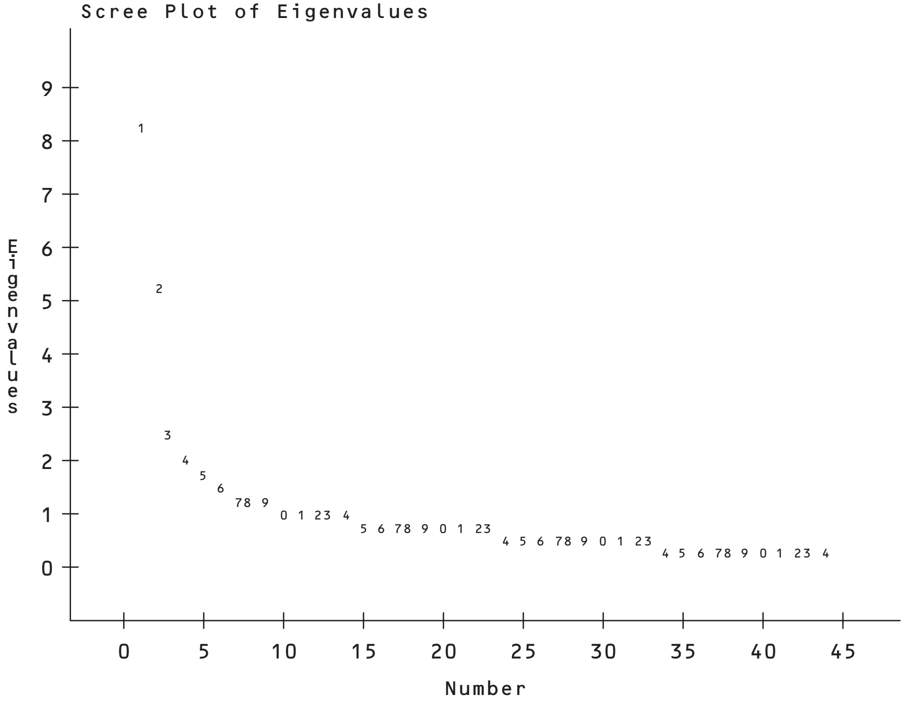
	Eigenvalue	Difference	Proportion	Cumulative
1	8.13953452	2.95691419	0.1850	0.1850
2	5.18262033	2.68422078	0.1178	0.3028
3	2.49839956	0.39333457	0.0568	0.3596
4	2.10506498	0.44826534	0.0478	0.4074
5	1.65679964	0.27255495	0.0377	0.4451
6	1.38424469	0.04304596	0.0315	0.4765
7	1.34119873	0.11660683	0.0305	0.5070
8	1.22459190	0.07627839	0.0278	0.5348
9	1.14831351	0.06537346	0.0261	0.5609
10	1.08294005	0.06111883	0.0246	0.5855
11	1.02182122	0.05627173	0.0232	0.6088
12	0.96554949	0.04701065	0.0219	0.6307
13	0.91853885	0.01390491	0.0209	0.6516

.
.

11 factors will be retained by the MINEIGEN criterion.

(continued)

TABLE 13.12 Continued



The decision between oblique and orthogonal rotation is made by requesting principal factor extraction with oblique rotation of five factors. Promax is the oblique method employed; `power = 2` sets the degree of allowable correlation among factors. The highest correlation (.343) is between factors 2 and 3 (see Table 13.14).

The request for an output data set (`outfile=SASUSER.FACSCORE`) in the syntax produces factor scores, which are plotted in Figure 13.5. The generally oblong shape of the scatterplot of factor scores between these two factors confirms the correlation. This level of correlation can be considered borderline between accepting an orthogonal solution versus dealing with the complexities of interpreting an oblique solution. The simpler, orthogonal, solution is chosen.

The solution that is evaluated, interpreted, and reported is the run with principal factors extraction, varimax rotation, and five factors. In other words, after "trying out" oblique rotation, the decision is made to interpret the earlier run with orthogonal rotation. Syntax for this run is in Table 13.13.

Communalities are inspected to see if the variables are well defined by the solution. Communalities indicate the percent of variance in a variable that overlaps variance in the factors. As seen in Table 13.15, communality values for a number of variables are quite low (e.g., FOULLANG). Seven of the variables have communality values lower than .2 indicating

TABLE 13.13 Eigenvalues and Proportions of Variance for First Six Factors. Principal Factors Extraction and Varimax Rotation (SAS FACTOR Syntax and Selected Output)

```
proc factor data=SASUSER.FACTORR prior=smc nfact=5 method=prinit rotate=varimax plot;
     var HELPFUL RELIANT DEFBEL YIELDING CHEERFUL INDPT ATHLET SHY ASSERT
  STRPERS FORCEFUL AFFECT FLATTER LOYAL ANALYT FEMININE SYMPATHY MOODY SENSITIV UNDSTAND
  COMPASS LEADERAB SOOTHE RISK DECIDE SELFSUFF CONSCIEN DOMINANT MASCULIN STAND HAPPY
  SOFTSPOK WARM TRUTHFUL TENDER GULLIBLE LEADACT CHILDLIK INDIVID FOULLANG LOVCHIL
  COMPETE AMBITIOU GENTLE;
run;
```

Eigenvalues of the Reduced Correlation Matrix: Total = 16.6698011 Average = 0.37885911

	Eigenvalue	Difference	Proportion	Cumulative
1	7.58593121	2.94049536	0.4551	0.4551
2	4.64543585	2.76137423	0.2787	0.7337
3	1.88406162	0.35565059	0.1130	0.8468
4	1.52841103	0.50167641	0.0917	0.9385
5	1.02673462	0.27278166	0.0616	1.0000
6	0.75395296	0.13981395	0.0452	1.0453
7	0.61413901	0.05512438	0.0368	1.0821

TABLE 13.14 **Syntax and Selected SAS FACTOR PFA Output of Correlations Among Factors Following Promax Rotation**

```
proc factor data=SASUSER.FACTORR prior=smc nfact=5 method=prinit rotate=promax power=2
  reorder out=SASUSER.FACSCORE;
     var HELPFUL RELIANT DEFBEL YIELDING CHEERFUL INDPT ATHLET SHY ASSERT
  STRPERS FORCEFUL AFFECT FLATTER LOYAL ANALYT FEMININE SYMPATHY MOODY SENSITIV UNDSTAND
  COMPASS LEADERAB SOOTHE RISK DECIDE SELFSUFF CONSCIEN DOMINANT MASCULIN STAND HAPPY
  SOFTSPOK WARM TRUTHFUL TENDER GULLIBLE LEADACT CHILDLIK INDIVID FOULLANG LOVCHIL
  COMPETE AMBITIOU GENTLE;
run;
```

Inter-Factor Correlations

	Factor1	Factor2	Factor3	Factor4	Factor5
Factor1	1.00000	0.17144	0.03217	0.04955	0.23476
Factor2	0.17144	1.00000	0.34331	0.14280	0.22937
Factor3	0.03217	0.34331	1.00000	0.20455	-0.02264
Factor4	0.04955	0.14280	0.20455	1.00000	-0.03207
Factor5	0.23476	0.22937	-0.02264	-0.03207	1.00000

```
proc plot data=sasuser.facscore;
        plot factor3*factor2;
run;
```

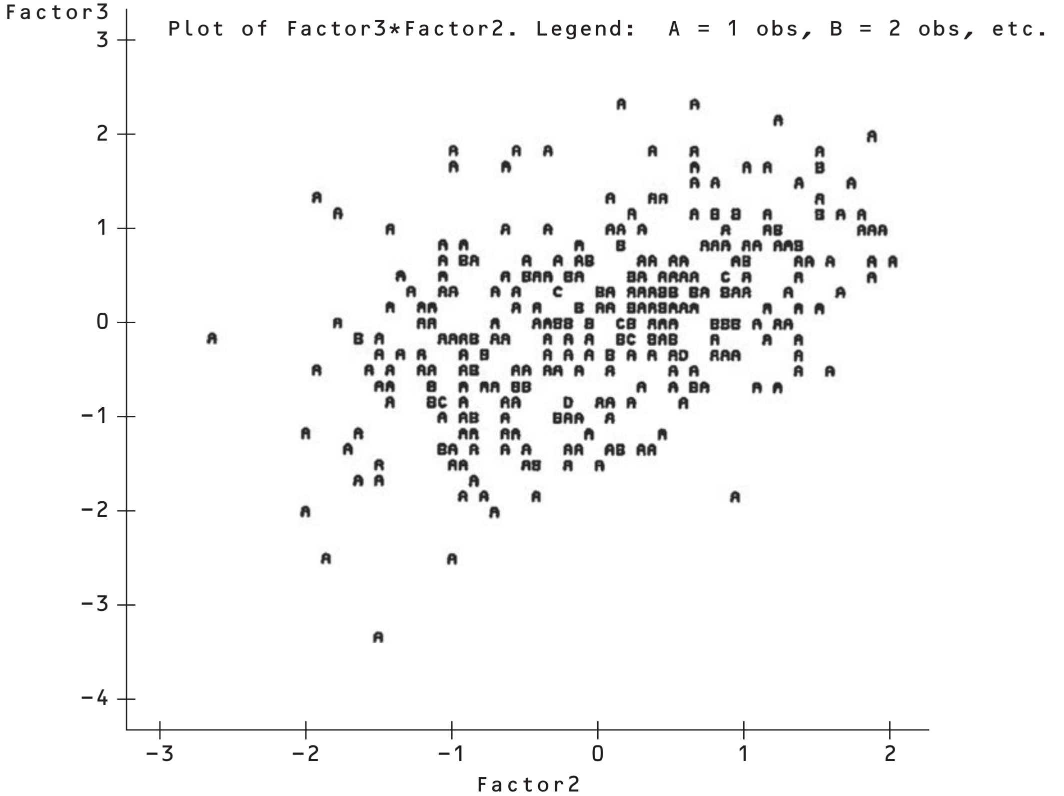


FIGURE 13.5 Scatterplot of factor scores with pairs of factors (2 and 3) as axes following oblique rotation.

considerable heterogeneity among the variables. It should be recalled, however, that factorial purity was not a consideration in development of the BSRI.

Adequacy of rotation (Section 13.6.3) is assessed, in part, by scatterplots with pairs of rotated factors as axes and variables as points, as partially shown in Figure 13.6. Ideally, variable points are at the origin (the unmarked middle of figures) or in clusters at the ends of factor axes. Scatterplots between factor 1 and factor 2 (the only one shown), between factor 2 and factor 4, and between factor 3 and factor 4 seem reasonably clear. The scatterplots between other pairs of factors show evidence of correlation among factors as found during oblique rotation. Otherwise, the scatterplots are disappointing but consistent with the other evidence of heterogeneity among the variables in the BSRI.

Simplicity of structure (Section 13.6.3) in factor loadings following orthogonal rotation is assessed from `Rotated Factor Pattern` table (see Table 13.16). In each column, there are a few high and many low correlations between variables and factors. There are also numerous moderate loadings so several variables will be complex (load on more than one factor) unless a fairly high cutoff for interpreting loadings is established. Complexity of variables (Section 13.6.5)

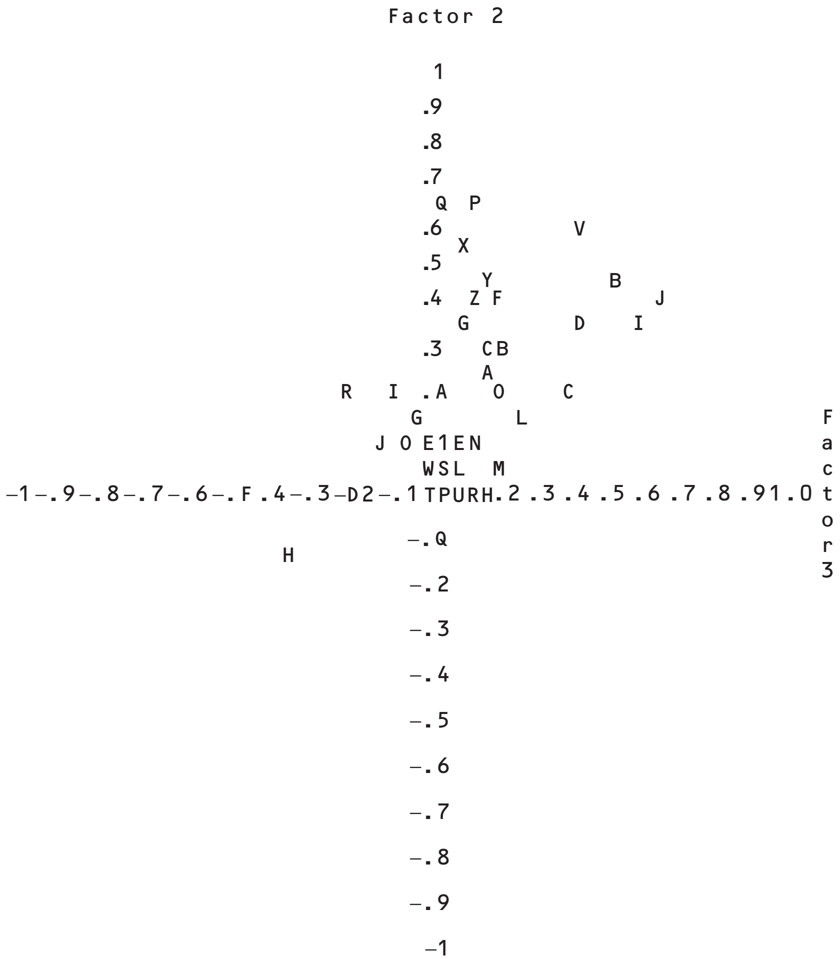
TABLE 13.15 Communality Values (Five Factors). Selected Output From SAS FACTOR PFA (See Table 13.13 for Syntax)

Final Communality Estimates: Total = 16.670574

helpful	reliant	defbel	yielding	cheerful	indpt	athlet	shy
0.27989969	0.40324796	0.27517026	0.14762865	0.46884461	0.45362663	0.18445190	0.25460045
assert	strpers	forceful	affect	flatter	loyal	analyt	feminine
0.47588701	0.57377345	0.48071231	0.48559859	0.21227516	0.29741719	0.15475733	0.16738890
sympathy	moody	sensitiv	undstand	compass	leaderab	soothe	risk
0.43347916	0.31232603	0.42057011	0.55491827	0.67643224	0.58944801	0.39532766	0.32659884
decide	selfsuff	conscien	dominant	masculin	stand	happy	softspok
0.39618000	0.66277946	0.34487215	0.52740913	0.20577933	0.42509594	0.51320831	0.33367049
warm	truthful	tender	gullible	leadact	childlik	individ	foullang
0.62488952	0.17522560	0.51576382	0.27873614	0.54555466	0.20768568	0.24669893	0.05316525

lovchil	compete	ambitiou	gentle
0.13540009	0.46497719	0.43043228	0.55866996

```
Plot of Factor Pattern for Factor2 and Factor3
```



helpful=A	reliant=B	defbel=C	yielding=D	cheerful=E	indpt=F	athlet=G
shy=H	assert=I	strpers=J	forceful=I	affect=L	flatter=M	loyal=N
analyt=O	feminine=P	sympathy=Q	moody=R	sensitiv=S	undstand=T	compass=U
leaderab=V	soothe=W	risk=X	decide=Y	selfsuff=Z	conscien=A	dominant=B
masculin=C	stand=D	happy=E	softspok=F	warm=G	truthful=H	tender=I

FIGURE 13.6 Selected SAS FACTOR PFA output showing scatterplot of variable loadings with factors 1 and 2 as axes. (Syntax in Table 13.13.)

TABLE 13.16 Factor Loadings for Principal Factors Extraction and Varimax Rotation of Five Factors. Selected SAS FACTOR Output (Syntax Appears in Table 13.13)

Rotated Factor Pattern

		Factor1	Factor2	Factor3	Factor4	Factor5
compass	compass	0.81668	-0.01671	0.04398	0.08225	-0.02214
undstand	undstand	0.73038	-0.00708	-0.00988	0.14600	-0.00042
sympathy	sympathy	0.64913	-0.10789	0.01231	0.01740	-0.00460
sensitiv	sensitiv	0.64054	0.02520	0.02927	0.07249	-0.05949
warm	warm	0.60918	0.16824	-0.04652	-0.12163	0.45664
soothe	soothe	0.60373	0.02882	-0.00896	-0.02478	0.17122
gentle	gentle	0.58373	0.19188	-0.25101	-0.10604	0.32690
tender	tender	0.56803	0.19266	-0.11069	-0.12432	0.35817
affect	affect	0.48900	0.14585	0.24264	-0.22048	0.34310
loyal	loyal	0.45093	0.11644	0.13102	0.01516	0.25124
truthful	truthful	0.32812	0.01592	0.11448	0.19525	0.12681
helpful	helpful	0.32381	0.23344	0.13844	0.21628	0.23369
lovchil	lovchil	0.26709	0.09035	-0.08345	-0.13418	0.17588
analyt	analyt	0.22253	0.21688	0.17446	0.12667	-0.10826
compete	compete	-0.02619	0.65931	0.10250	-0.10947	0.08433
ambitiou	ambitiou	0.07739	0.64616	0.00664	0.06766	0.04790
leaderab	leaderab	0.11055	0.59325	0.41303	0.21812	0.08430
leadact	leadact	-0.01292	0.58378	0.40237	0.17552	0.10897
risk	risk	0.14850	0.54482	0.07787	-0.01907	0.03592
decide	decide	0.11026	0.46168	0.15703	0.37565	0.07143
individ	individ	0.10359	0.40925	0.17666	0.18416	0.05797
athlet	athlet	-0.05591	0.34809	0.08364	-0.04645	0.22584
masculin	masculin	-0.20901	0.29630	0.14796	-0.04085	-0.22525
strpers	strpers	0.10866	0.38036	0.63953	0.03088	0.08566
forceful	forceful	0.01823	0.33443	0.59767	0.07952	-0.07078
assert	assert	0.13242	0.34642	0.56971	0.11351	0.02987
dominant	dominant	-0.11815	0.47366	0.51036	0.09073	-0.14282
stand	stand	0.25559	0.35940	0.41605	0.22860	0.07238
defbel	defbel	0.27631	0.20395	0.38164	0.10563	0.02054
foullang	foullang	0.08549	0.11060	-0.16580	0.00139	0.07832
shy	shy	-0.06372	-0.12686	-0.42860	-0.10852	-0.19741
softspok	softspok	0.24469	-0.00175	-0.48667	0.10792	0.15907
selfsuff	selfsuff	0.11528	0.38461	0.12977	0.69376	0.05850
reliant	reliant	0.10525	0.30950	0.16698	0.50590	0.11210
indpt	indpt	0.03392	0.41769	0.17277	0.49740	-0.02752
conscien	conscien	0.32354	0.19952	0.02069	0.43232	0.11426
flatter	flatter	0.14775	0.06476	0.19752	-0.29752	0.24232
childlik	childlik	-0.04596	0.05458	0.03482	-0.43488	-0.11072
gullible	gullible	0.16064	0.10022	-0.15872	-0.46645	0.01086
happy	happy	0.18789	0.11322	-0.01725	0.12670	0.66988
cheerful	cheerful	0.17664	0.08527	0.06174	0.13179	0.63968
feminine	feminine	0.26994	0.02034	0.00163	0.14409	0.27082
yielding	yielding	0.19451	-0.02065	-0.22552	-0.02952	0.24008
moody	moody	0.04974	0.02232	0.08758	-0.34425	-0.42799

(*continued*)

TABLE 13.16 Continued

```
                 Variance Explained by Each Factor

      Factor1      Factor2      Factor3      Factor4      Factor5

    4.9338827    3.9884412    2.9812180    2.4732606    2.2937718
```

is assessed by examining loadings for a variable across factors. With a loading cut of .45, only two variables—WARM and INDPT—load on more than one factor.

The importance of each factor (Section 13.4 and Section 13.6.4) is assessed by the percent of variance and covariance it represents. SSLs, called `Variance Explained by Each Factor` below the loadings in Table 13.16, are used in the calculations. It is important to use SSLs from rotated factors, because the variance is redistributed during rotation. Proportion of variance for a factor is SSL for the factor divided by number of variables. Proportion of covariance is SSL divided by sum of SSLs. Results, converted to percent, are shown in Table 13.17. Each of the factors accounts for between 5% and 14% of the variance in the set of variables, not an outstanding performance.

Internal consistency of the factors (Section 13.6.4) is assessed through SMCs, available in SAS FACTOR when factor scores are requested (`out=SASUSER.FACSCPFA` in the syntax of Table 13.13). These are found in the `Squared Multiple Correlations of the Variables with Each Factor` table, in which factors serve as DVs with variables as IVs. Factors that are well defined by the variables have high SMCs, whereas poorly defined factors have low SMCs. As can be seen in Table 13.18, all factors are internally consistent. (The off-diagonal elements in these matrices are correlations among factor scores. Although uniformly low, the values are not zero. As discussed in Section 13.6.6, low correlations among scores on factors are often obtained even with orthogonal rotation.)

TABLE 13.17 Percents of Variance and Covariance Explained by Each of the Rotated Orthogonal Factors

	Factors				
	1	*2*	*3*	*4*	*5*
SSL	4.93	3.99	2.98	2.47	2.29
Percent of variance	11.20	9.09	6.77	5.61	5.20
Percent of covariance	29.59	23.94	17.89	14.82	13.74

TABLE 13.18 SMCs for Factors With Variables as IVs. Selected Output from SAS FACTOR PFA with Orthogonal (Varimax) Rotation (Syntax in Table 13.13)

```
Squared Multiple Correlations of the Variables with Each Factor

          Factor1     Factor2     Factor3     Factor4     Factor5

       0.88645424  0.79508536  0.76857909  0.79094331  0.75327949
```

TABLE 13.19 Order (by Size of Loadings) in Which Variables Contribute to Factors

Factor 1: ***Empathy***	**Factor 2:** ***Leadership***	**Factor 3:** ***Dominance***	**Factor 4:** ***Independence***	**Factor 5:** ***Positive Affect***
Compassionate	Competitive	Strong personality	Self-sufficient	Happy
Understanding	Ambitious	Forceful	Self-reliant	Cheerful
Sympathetic	Has leadership ability	Assertive	Independent	Warm
Sensitive	Acts as a leader	Dominant	Not gullible	
Warm	Willing to take risks	Not soft spoken		
Eager to soothe hurt feelings	Makes decisions			
Gentle	Dominant			
Tender				
Affectionate				
Loyal				
Not Childlike				

Factors are interpreted through their factor loadings (Section 13.6.5) from Table 13.16. It is decided to use a loading of .45 (20% variance overlap between variable and factor). With the use of the .45 cut, Table 13.19 is generated to further assist interpretation. In more informal presentations of factor analytic results, this table might be reported instead of Table 13.16. Factors are put in columns and variables with the largest loadings are put on top. In interpreting a factor, items near the top of the columns are given somewhat greater weight. Variable names are written out in full detail and labels for the factors (e.g., Dominance) are suggested at the top of each column. Table 13.20 shows a more formal summary table of factor loadings, including communalities as well as percents of variance and covariance.

Table 13.21 provides a checklist for FA. A Results section in journal format follows for the data analyzed in this section.

TABLE 13.20 Factor Loadings, Communalities (h^2), and Percents of Variance and Covariance for Principal Factors Extraction and Varimax Rotation on BSRI Items

Item	F_1[a]	F_2	F_3	F_4	F_5	h^2
Compassionate	.82	.00	.00	.00	.00	.68
Understanding	.73	.00	.00	.00	.00	.55
Sympathetic	.65	.00	.00	.00	.00	.43
Sensitive	.64	.00	.00	.00	.00	.42
Warm	.61	.00	.00	.00	.46	.62
Eager to soothe hurt feelings	.60	.00	.00	.00	.00	.40
Gentle	.58	.00	.00	.00	.00	.56
Tender	.57	.00	.00	.00	.00	.52
Affectionate	.49	.00	.00	.00	.00	.49
Loyal	.45	.00	.00	.00	.00	.30
Competitive	.00	.66	.00	.00	.00	.46
Ambitious	.00	.65	.00	.00	.00	.43
Leadership ability	.00	.59	.00	.00	.00	.59
Acts like a leader	.00	.58	.00	.00	.00	.55
Willing to take risks	.00	.54	.00	.00	.00	.33
Makes decisions	.00	.46	.00	.00	.00	.40
Strong personality	.00	.00	.64	.00	.00	.57
Forceful	.00	.00	.60	.00	.00	.48
Assertive	.00	.00	.57	.00	.00	.48
Dominant	.00	.47	.51	.00	.00	.53
Soft spoken	.00	.00	−.49	.00	.00	.33
Self-sufficient	.00	.00	.00	.69	.00	.66
Self-reliant	.00	.00	.00	.51	.00	.40
Independent	.00	.00	.00	.50	.00	.45
Gullible	.00	.00	.00	−.47	.00	.28
Childlike	.00	.00	.00	−.43	.00	.21
Happy	.00	.00	.00	.00	.67	.51
Cheerful	.00	.00	.00	.00	.64	.47
Truthful	.00	.00	.00	.00	.00	.18
Helpful	.00	.00	.00	.00	.00	.38
Loves children	.00	.00	.00	.00	.00	.14
Analytical	.00	.00	.00	.00	.00	.15
Individualistic	.00	.00	.00	.00	.00	.25
Athletic	.00	.00	.00	.00	.00	.18
Masculine	.00	.00	.00	.00	.00	.21
Takes stand	.00	.00	.00	.00	.00	.27
Defends beliefs	.00	.00	.00	.00	.00	.43
Uses foul language	.00	.00	.00	.00	.00	.05
Shy	.00	.00	.00	.00	.00	.25
Conscientious	.00	.00	.00	.00	.00	.34
Easily flattered	.00	.00	.00	.00	.00	.21
Feminine	.00	.00	.00	.00	.00	.17
Yielding	.00	.00	.00	.00	.00	.15
Moody	.00	.00	.00	.00	.00	.31
Percent of variance	11.20	9.09	6.77	5.61	5.20	
Percent of covariance	29.59	23.94	17.89	14.82	13.74	

[a]Factor labels:

F_1—Empathy;

F_2—Leadership;

F_3—Dominance;

F_4—Independence;

F_5—Positive affect.

TABLE 13.21 Checklist for Factor Analysis

1. Limitations
   a. Outliers among cases
   b. Sample size and missing data
   c. Factorability of **R**
   d. Normality and linearity of variables
   e. Multicollinearity and singularity
   f. Outliers among variables
2. Major analyses
   a. Number of factors
   b. Nature of factors
   c. Type of rotation
   d. Importance of factors
3. Additional analyses
   a. Factor scores
   b. Distinguishability and simplicity of factors
   c. Complexity of variables
   d. Internal consistency of factors
   e. Outlying cases among the factors

Results

Principal factors extraction with varimax rotation was performed through SAS FACTOR on 44 items from the BSRI for a sample of 340 women. Principal components extraction was used prior to principal factors extraction to estimate number of factors, presence of outliers, absence of multicollinearity, and factorability of the correlation matrices. With α = .001 cutoff level, 11 of 351 women produced scores that identified them as outliers; these cases were deleted from principal factors extraction.[10]

[10] Outliers were compared as a group to nonoutliers through discriminant analysis. As a group, at $p < .01$, the 11 women were less self-reliant and likely to defend beliefs and more warm, self-sufficient, and reported more leadership ability than non-outlying women.

Five factors were extracted. As indicated by SMCs, all factors were internally consistent and well defined by the variables; the lowest of the SMCs for factors from variables was .75. [*Information on SMCs is from Table 13.18.*] The reverse was not true, however; variables were, by and large, not well defined by this factor solution. Communality values, as seen in Table 13.15, tended to be low. With a cutoff of .45 for inclusion of a variable in interpretation of a factor, 16 of 44 variables did not load on any factor. Failure of numerous variables to load on a factor reflects heterogeneity of items on the BSRI. However, only two of the variables in the solution, "warm" and "dominant," were complex.

When oblique rotation was requested, factors interpreted as Leadership and Dominant correlated .34. However, because the correlation was modest and limited to one pair of factors, and because remaining correlations were fairly low, orthogonal rotation was chosen.

Loadings of variables on factors, communalities, and percents of variance and covariance are shown in Table 13.20. Variables are ordered and grouped by size of loading to facilitate interpretation. Loadings under .45 (20% of variance) are replaced by zeros. Interpretive labels are suggested for each factor in a footnote.

In sum, the five factors on the BSRI for this group of women are empathy (e.g., compassionate, understanding), leadership (e.g., competitive, ambitious), dominance (e.g., strong personality, forceful), independence (e.g., self-sufficient, self-reliant), and positive affect (e.g., happy, cheerful).

13.8 Comparison of Programs

IBM SPSS, SAS, and SYSTAT each have a single program to handle both FA and PCA. The first two programs have numerous options for extraction and rotation and give the user considerable latitude in directing the progress of the analysis. Features of three programs are described in Table 13.22.

13.8.1 IBM SPSS Package

IBM SPSS FACTOR does a PCA or FA on a correlation matrix or a factor loading matrix, helpful to the researcher who is interested in higher-order factoring (extracting factors from previous FAs). Several extraction methods and a variety of orthogonal rotation methods are available. Oblique rotation is done using direct oblimin, one of the best methods currently available (see Section 13.5.2.2).

Univariate output is limited to means, standard deviations, and number of cases per variable, so that the search for univariate outliers must be conducted through other programs. Similarly, there is no provision for screening for multivariate outliers among cases. But the program is very helpful in assessing factorability of **R**, as discussed in Section 13.3.2.6.

Output of extraction and rotation information is extensive. The residual and reproduced correlation matrices are provided as an aid to diagnosing adequacy of extraction and rotation.

IBM SPSS FACTOR is the only program reviewed that, under conditions requiring matrix inversion, prints out the determinant of the correlation matrix, helpful in signaling the need to check for multicollinearity and singularity (Sections 13.3.2.5 and 4.1.7). Determination of number of factors is aided by an optional printout of a scree plot (Section 13.6.2). Several estimation procedures for factor scores (Section 13.6.6) are available as output to a file.

13.8.2 SAS System

SAS FACTOR is another highly flexible, full-featured program for FA and PCA. About the only weakness is in screening for outliers. SAS FACTOR accepts rotated loading matrices, as long as factor correlations are provided, and can analyze a partial correlation or covariance matrix (with specification of variables to partial out). There are several options for extraction, as well as orthogonal and oblique rotation. Maximum-likelihood estimation provides a χ^2 test for number of factors. Standard errors may be requested for factor loadings with maximum-likelihood estimation and promax rotation. A target pattern matrix can be specified as a criterion for oblique rotation in confirmatory FA. Additional options include specification of proportion of variance to be accounted for in determining the number of factors to retain and the option to allow communalities to be greater than 1.0. The correlation matrix can be weighted to allow the generalized least squares method of extraction.

Factor scores can be written to a data file. SMCs of factors as DVs with variables as IVs are given, to evaluate the reliability of factors.

TABLE 13.22 Comparison of Factor Analysis Programs

Feature	IBM SPSS FACTOR	SAS FACTOR	SYSTAT FACTOR
Input			
Correlation matrix	Yes	Yes	Yes
About origin	No	Yes	No
Covariance matrix	Yes	Yes	Yes
About origin	No	Yes	No
SSCP matrix	No	No	Yes
Factor loadings (unrotated pattern)	Yes	Yes	Yes
Factor-score coefficients	No	Yes	Data file
Factor loadings (rotated pattern) and factor correlations	No	Yes	Yes
Options for missing data	Yes	No	Yes
Analyze partial correlation or covariance matrix	No	Yes	No
Specify maximum number of factors	FACTORS	NFACT	NUMBER
Extraction method (see Table 13.7)			
PCA	PC	PRIN	PCA
PFA	PAF	PRINIT	IPA
Image (Little Jiffy, Harris)	IMAGE	Yes[a]	No
Maximum likelihood	ML	ML	MLA
Alpha	ALPHA	ALPHA	No
Unweighted least squares	ULS	ULS	No
Generalized least squares	GLS	Yes	No
Specify communalities	Yes	Yes	No
Specify minimum eigenvalues	MINEIGEN	MIN	EIGEN
Specify proportion of variance to be accounted for	No	PROPORTION	No
Specify maximum number of iterations	ITERATE	MAXITER	ITER
Option to allow communalities 1	No	HEYWOOD	No
Specify tolerance	No	SING	No
SPecify convergence criterion for extraction	ECONVERGE	CONV	CONV
Specify convergence criterion for rotation	RCONVERGE	No	No
Rotation method (see Table 13.9)			
Varimax	Yes	Yes	Yes
Quartimax	Yes	Yes	Yes
Equamax	Yes	Yes	Yes
Orthogonal with gamma	No	ORTHOMAX	ORTHOMAX
Parsimax	No	Yes	No
Direct oblimin	Yes	No	Yes

(*continued*)

TABLE 13.22 Continued

Feature	IBM SPSS FACTOR	SAS FACTOR	SYSTAT FACTOR
Input *(continued)*			
Rotation method (see Table 13.9) *(continued)*			
Direct quartimin	DELTA = 0	No	No
Orthoblique	No	HK	No
Promax	No	Yes	No
Procrustes	No	Yes	No
Prerotation criteria	No	Yes	No
Optional kaiser's normalization	Yes	Yes	Normalized only
Optional weighting by Cureton-Mulaik technique	No	Yes	No
Optional rescaling of pattern matrix to covariances	No	Yes	No
Weighted correlation matrix	No	WEIGHT	No
Alternate methods for computing factor scores	Yes	No	No
Output			
Means and standard deviations	Yes	Yes	No
Number of cases per variable (missing data)	Yes	No	No
Significance of correlations	Yes	No	No
Covariance matrix	Yes	Yes	Yes
Initial communalities	Yes	Yes	Yes
Final communalities	Yes	Yes	Yes
Eigenvalues	Yes	Yes	Yes
Difference between successive eigenvalues	No	Yes	No
Standard error for each eigenvector element	No	No	Yes
Percent of variance total variance explained by factors	Yes	No	Yes
Cumulative percent of variance	Yes	No	No
Percent of covariance	No	No	Yes
Unrotated factor loadings	Yes	Yes	Yes
Variance explained by factors for all loading matrices	No	Yes	Yes
Simplicity criterion, each rotation iteration	δ [b]	No	No
Rotated factor loadings (pattern)	Yes	Yes	Yes
Rotated factor loadings (structure)	Yes	Yes	Yes
Eigenvectors	No	Yes	Yes
Standard error for each eigenvector element	No	No	Yes

TABLE 13.22 Continued

Feature	IBM SPSS FACTOR	SAS FACTOR	SYSTAT FACTOR
Transformation matrix	Yes	Yes	No
Factor-score coefficients	Yes	Yes	Data file[c]
Standardized factor scores	Data file	Data file	Data file[c]
Residual component scores	No	No	Data file[c]
Sum of squared residuals (Q)	No	No	Data file[c]
Probability for Q	No	No	Data file[c]
Scree plot	Yes	Yes	Yes
Plots of unrotated factor loadings	No	Yes	No
Plots of rotated factor loadings	Yes	Yes	Yes
Sorted rotated factor loadings	Yes	Yes	Yes
χ^2 test for number of factors (with maximum likelihood estimation)	No	Yes	No
χ^2 test that all eigenvalues are equal	No	No	Yes
χ^2 test that last n eigenvalues are equal	No	No	Yes
Standard errors of factor loadings (with maximum likelihood estimation and promax solutions)	No	Yes	No
Inverse of correlation matrix	Yes	Yes	No
Determinant of correlation matrix	Yes	No	No
Partial correlations (anti-image matrix)	AIC	MSA	No
Measure of sampling adequacy	AIC, KMO	MSA	No
Anti-image covariance matrix	AIC	No	No
Bartlett's test of sphericity	KMO	No	No
Residual correlation matrix	Yes	Yes	Yes
Reproduced correlation matrix	Yes	No	No
Correlations among factors	Yes	Yes	Yes

[a]Two types.

[b]Oblique only.

[c]PCA only.

13.8.3 SYSTAT System

The current SYSTAT FACTOR program is less limited than earlier versions. Wilkinson (1990) advocated the use of PCA rather than FA because of the indeterminacy problem (Section 13.6.6). However, the program now does PFA (called IPA) as well as PCA and maximum likelihood (MLA) extraction. Four common methods of orthogonal rotation are provided, as well as provision for oblique rotation. SYSTAT FACTOR can accept correlation or covariance matrices as well as raw data.

The SYSTAT FACTOR program provides scree plots and plots of factor loadings and will optionally sort the loading matrix by size of loading to aid interpretation. Additional information is available by requesting that standardized component scores, their coefficients, and loadings be sent to a data file. Factor scores (from PFA or MLA) cannot be saved. Residual scores (actual minus predicted z-scores) also can be saved, as well as the sum of the squared residuals and a probability value for it.