n was 45% and 36% for Caucasian and Japanese women, respectively. After nent for effects of age, the only predictor retained in the stepwise analysis for aces was "surgery related to a female disorder." For Caucasian women only, ditional predictor, medication, was selected in the stepwise analysis. For age, juared, surgery, and medication, discriminant function coefficients and their rd errors were given.[13] Finally, predictors common to both races were selected eterogeneity of coefficients between Japanese and Caucasian samples were .[14] No significant heterogeneity was found—that is, the two races did not produce cantly different discriminant functions. An alternative strategy for this test is orial discriminant function analysis, with race as one IV and menopausal vs. enopausal as the other.

[13] Because this two-group analysis was run through a multiple regression program, discriminant ction coefficients were produced as β weights (cf. Chapter 5) and standard errors were available.

[14] The test for heterogeneity reported was that of Rao (1952) as applied by Goodman et al. (1974).

Chapter 1

Principal Components and Factor Analysis

12.1 GENERAL PURPOSE AND DESCRIPTION

Principal components analysis (PCA) and factor analysis (FA) are statistical techniques applied to a single set of variables where 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 motivation and scholastic history to suggest a person who 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 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

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

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 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 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, each linear combination a factor. The factors summarize the patterns of the correlations in the observed correlation matrix and can, in fact, be used to reproduce the observed correlation matrix. But the number of factors is usually far fewer than the number of observed variables so there is considerable parsimony in a 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, the final test of the analysis is usually 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.

One of the problems with PCA and FA is that there is no criterion beyond interpretability 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, 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 are an infinite number of rotations available, all accounting for the same amount of variance in the original data, but with 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 the best FA solution will not be surprised when other researchers select a different one. Nor will they be surprised when results are not replicated 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, FA is 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 or to investigate hypothesized differences in latent processes between groups of subjects. Variables are carefully and specifically chosen to reveal underlying processes.

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 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 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: 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 rotation, there is a *factor-score* coefficients matrix, a matrix of coefficients used to estimate 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. The difference between PCA and FA is in the variance that is analyzed. In PCA, all the variance in the observed variables is 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.

12.2 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 12.2.1 through 12.2.6.

12.2.1 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 any more reliable factors present? Strategies for choosing an appropriate number of factors and for assessing the correspondence between observed and reproduced correlation matrices are discussed in Section 12.6.2.

12.2.2 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 12.6.3; interpretation itself is discussed in Section 12.6.5.

12.2.3 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 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 12.6.4.

12.2.4 Testing Theory in FA

How well does the obtained factor solution fit an expected factor solution? The researcher generates 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 12.6.2 and 12.6.7.

12.2.5 Comparing Factor Solutions for Different Groups

How similar are the factors for persons with different characteristics or different experiences? For instance, are the factors for graduate students the same as the factors for undergraduate business majors or other groups that have not gone on to graduate school? Similarity in factors between two groups is assessed, in preliminary fashion, using techniques described in Section 12.6.7.

12.2.6 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 12.6.6.

12.3 LIMITATIONS

12.3.1 Theoretical Issues

Most applications of PCA or FA are exploratory in nature; FA is used as a tool for reducing the number of variables or examining patterns of correlations among variables without a serious intent to test theory. 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 (1973, pp. 189–211), 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 extractional or rotational 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 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 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 exhibit 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 leery about 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. Strategies for evaluating differences in factors among groups are discussed in Section 12.6.7.

12.3.2 Practical Issues

Because FA and PCA are exquisitely sensitive to the sizes of correlations, it is critical that honest, reliable 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.

12.3.2.1 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. Comrey (1973) gives as a guide sample sizes of 50 as very poor, 100 as poor, 200 as fair, 300 as good, 500 as very good, and 1000 as excellent. Others suggest that a sample size of 100 to 200 is good enough for most purposes, particularly when factors are strong and distinct and number of variables is not too large. *As a general rule of thumb, it is comforting to have at least five cases for each observed variable.*

The required sample size depends also on magnitude of population correlations and number of factors. If there are strong, reliable correlations and a few, distinct factors, a sample size of 50 may even be adequate, as long as there are notably more cases than factors.

If cases have missing data, either the missing values are estimated or the cases are 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.

12.3.2.2 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, when statistical inference is used to determine the number of factors, multivariate normality is assumed. Multivariate normality is the assumption that all variables, and all linear combinations of variables, are normally distributed. Although normality of all linear combinations of variables is not testable, *normality among single variables is assessed by skewness and kurtosis* (see Chapter 4 and Section 12.8.1.2). If a variable has substantial skewness and kurtosis, variable transformation is considered.

12.3.2.3 Linearity Multivariate normality also implies that relationships among pairs of variables are linear. Because correlation measures linear relationship and does not reflect nonlinear relationship, the analysis is degraded when linearity fails. *Linearity among pairs of variables is assessed through inspection of scatterplots.* Consult Chapter 4 and Section 12.8.1.3 for methods of screening for linearity. If nonlinearity is found, transformation of variables is considered.

12.3.2.4 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 12.8.1.4 for methods of detecting and reducing the influence of both univariate and multivariate outliers.

12.3.2.5 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 12.8.1.5 provide examples of screening for and dealing with multicollinearity and singularity.

12.3.2.6 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,* except in its most exploratory and pragmatic sense.

High bivariate correlations are, however, not ironclad proof that the correlation matrix contains factors. It is possible that the correlations are between only two variables and do not reflect 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. BMDP, SPSS^x, and SAS produce partial correlation matrices.

Bartlett's test of sphericity (1954) is a notoriously sensitive test of the hypothesis that the correlations in a correlation matrix are zero. The test is available in SPSS^x 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 SPSS^x and SAS. Both programs give significance tests of correlations, the anti-image correlation matrix, and the Kaiser's (1970, 1974) measure of sampling adequacy. Significance tests of pairs 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.[2] Values of .6 and above are required for good FA.

[2] BMDP4M prints partial correlations between pairs of variables with effects of other variables removed through the PARTIAL option so Kaiser's measure of sampling adequacy could (with some pain) be hand-calculated.

12.3.2.7 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 or not these factors are "real." Suggestions for determining reliability of factors defined by one or two variables are in Section 12.6.2.

In exploratory FA, 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 ignored, as pragmatic considerations dictate. In confirmatory FA, 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 12.8.1.7.

12.3.2.8 Outlying Cases among the Factors In FA and PCA, cases may be unusual with respect to their scores on the factors. These are cases that have unusually large or small scores on the factors as estimated from factor score coefficients. The deviant scores are from cases for which the factor solution is inadequate. Examination of these cases for consistency is informative if it reveals the kinds of cases for which the FA is *not* appropriate.

If BMDP4M is used, *outlying cases among the factors are cases with large Mahalanobis distances, estimated as chi square values, from the location of the case in the space defined by the factors to the centroid of all cases in the same space.* If scatterplots between pairs of factors are requested, these cases appear along the borders. Screening for these outlying cases is in Section 12.8.1.8.

12.4 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 SPSS^x FACTOR for underlying calculations.

Table 12.1 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.

TABLE 12.1 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 standard scores on factors or components
A	Factor loading matrix Pattern matrix	Orthogonal Oblique	$p \times m$	Matrix of regressionlike 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 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.

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 12.2. 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.

TABLE 12.2 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

The variables were cost of ski ticket (COST), kind of ski lift (LIFT), depth of snow (DEPTH), and kind 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.

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.

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[3] 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:

$$\mathbf{L} = \mathbf{V}'\mathbf{R}\mathbf{V} \qquad (12.1)$$

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

TABLE 12.3 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

Diagonalization of **R** is accomplished by post- and premultiplying 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. These few factors duplicate the correlation matrix as faithfully as possible.

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 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 the Table 12.3.

Using Equation 12.1 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 premultiplied by its transpose produces the identity matrix with ones in the positive diagonal and zeros elsewhere. Therefore, pre- and postmultiplying the correlation matrix by eigenvectors does not change it so much as repackage it.

$$\mathbf{V'V} = \mathbf{I} \tag{12.2}$$

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 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 12.3 to 12.6.

Equation 12.1 can be reorganized as follows:

$$\mathbf{R} = \mathbf{VLV'} \tag{12.3}$$

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.

$$\mathbf{R} = \mathbf{V}\sqrt{\mathbf{L}}\sqrt{\mathbf{L}}\,\mathbf{V'} \tag{12.4}$$

or

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

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

$$\mathbf{R} = \mathbf{AA'} \tag{12.5}$$

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

Equation 12.5 is frequently called the fundamental equation for FA.[4] It represents the assertion that the correlation matrix is a product of the factor loading matrix, **A**, and its transpose.

[4] In order to reproduce the correlation matrix exactly, as indicated in Equations 12.4 and 12.5, all eigenvalues and eigenvectors are necessary, not just the first few of them.

Equations 12.4 and 12.5 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.

$$\mathbf{A} = \mathbf{V}\sqrt{\mathbf{L}} \tag{12.6}$$

For the example:

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

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). (The negative correlation indicates that more attractive lifts are also more costly.)

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.

Rotation is ordinarily used after extraction to maximize high correlations and minimize low ones. Numerous methods of rotation are available (see Section 12.5.2) but the most commonly used, and the one illustrated here, is varimax. Varimax is a *var*iance *max*imizing 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 12.8), where

$$\mathbf{A}_{\text{unrotated}}\,\mathbf{\Lambda} = \mathbf{A}_{\text{rotated}} \tag{12.7}$$

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

For the example:

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

TABLE 12.4 RELATIONSHIPS AMONG LOADINGS, COMMUNALITIES, SSLs, VARIANCE, AND COVARIANCE OF ORTHOGONALLY ROTATED FACTORS

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

97% of the variance of this variable accounted for by the factor

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

The numbers in the transformation matrix have a spatial interpretation.

$$\mathbf{\Lambda} = \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix} \tag{12.8}$$

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

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 12.5.2.3.

Once the rotated loading matrix is available, other relationships are found, as in Table 12.4. 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 12.4, 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).[5] * For the first factor, the proportion of variance is $[(-.086)^2 + (-.071)^2 + .994^2 + .997^2]/4 = 1.994/4 = .50$. Fifty percent of the variance in the variables is accounted for by the first factor. The second factor accounts for 48% of the variance

[5] 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.

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.

The reproduced correlation matrix for the example is generated using Equation 12.5:

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

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

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:

$$\mathbf{R}_{res} = \mathbf{R} - \bar{\mathbf{R}} \tag{12.9}$$

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**:

$$\mathbf{R}_{res} = \begin{bmatrix} .970 & -.953 & -.055 & -.130 \\ -.953 & .960 & -.091 & -.036 \\ -.055 & -.091 & .989 & .990 \\ -.130 & -.036 & .990 & .996 \end{bmatrix} - \begin{bmatrix} .000 & .000 & .004 & -.005 \\ .000 & -.002 & .007 & -.003 \\ .004 & .007 & .000 & .000 \\ -.005 & -.003 & .000 & .000 \end{bmatrix}$$

$$= \begin{bmatrix} .000 & .000 & .004 & -.005 \\ .000 & -.002 & .007 & -.003 \\ .004 & .007 & .000 & .000 \\ -.005 & -.003 & .000 & .000 \end{bmatrix}$$

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.

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

scu
example

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 12.10 for factor score coefficients is similar to Eq. 5.6 for regression coefficients in multiple regression.

$$\mathbf{B} = \mathbf{R}^{-1}\mathbf{A} \tag{12.10}$$

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:[6]

$$\mathbf{B} = \begin{bmatrix} 25.485 & 22.689 & -31.655 & 35.479 \\ 22.689 & 21.386 & -24.831 & 28.312 \\ -31.655 & -24.831 & 99.917 & -103.950 \\ 35.479 & 28.312 & -103.950 & 109.567 \end{bmatrix} \begin{bmatrix} -.086 & .981 \\ -.072 & -.978 \\ .994 & .027 \\ .997 & -.040 \end{bmatrix}$$

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

To estimate a subject's score for the first factor, then, 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,

$$\mathbf{F} = \mathbf{ZB} \tag{12.11}$$

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

For the example:

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

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

[6] 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.

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 values the snow factor highly and the cost factor low. The second subject values both the snow factor and the cost factor; the third subject values the cost factor more than the snow factor, 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

$$\mathbf{Z} = \mathbf{FA}' \tag{12.12}$$

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,

$$\begin{aligned} z_{\text{COST}} &= a_{11}F_1 + a_{12}F_2 \\ z_{\text{LIFT}} &= a_{21}F_1 + a_{22}F_2 \end{aligned}$$

$$\begin{aligned} z_{\text{DEPTH}} &= a_{31}F_1 + a_{32}F_2 \\ z_{\text{POWDER}} &= a_{41}F_1 + a_{42}F_2 \end{aligned}$$

A score on an observed variable is conceptualized as a properly weighted and summed combination of the scores on factors that underlie it. The underlying processes are factors; the outward manifestations are scores on variables, driven by underlying processes.

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 12.1 for a listing of additional matrices and a hint of the discussion to follow.

SPSS^x FACTOR is run on the data from Table 12.2 using the default option for oblique rotation (cf. Section 12.5.2.2) 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

$$\mathbf{A} = \begin{bmatrix} -.079 & .981 \\ -.078 & -.978 \\ .994 & .033 \\ .997 & -.033 \end{bmatrix}$$

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 12.11 to produce factor scores results in the following values:

$$\mathbf{F} = \begin{bmatrix} 1.12 & 1.14 & 1.15 & 1.14 \\ 0.75 & -1.02 & 0.92 & 1.01 \\ 0.61 & -0.78 & -0.36 & -0.47 \\ -0.95 & 0.98 & -1.20 & -1.01 \\ 0.82 & -0.30 & -0.51 & -0.67 \end{bmatrix} \begin{bmatrix} 0.104 & 0.584 \\ 0.081 & -0.421 \\ 0.159 & -0.020 \\ 0.856 & 0.034 \end{bmatrix}$$

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

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

$$\mathbf{\Phi} = \left(\frac{1}{N-1}\right)\mathbf{F'F} \qquad (12.13)$$

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:

$$\mathbf{\Phi} = \frac{1}{4}\begin{bmatrix} 1.12 & 1.01 & -0.46 & -1.07 & -0.59 \\ -1.18 & 0.88 & 0.68 & -0.98 & 0.59 \end{bmatrix}\begin{bmatrix} 1.12 & -1.18 \\ 1.01 & 0.88 \\ -0.46 & 0.68 \\ -1.07 & -0.98 \\ -0.59 & 0.59 \end{bmatrix}$$

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

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. Ordinarily 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 represent both the unique relationship between the variable and the factor (shown in the pattern matrix) and the relationship between the variable and the variance a factor shares with other factors. The equation for the structure matrix is

$$\mathbf{C} = \mathbf{A\Phi} \qquad (12.14)$$

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 12.14 as desired.

why you use the pat matrix patter than structu for obli solutio

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

$$\bar{\mathbf{R}} = \mathbf{CA'} \qquad (12.15)$$

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 12.9 is used to generate the residual correlation matrix to diagnose adequacy of fit in FA.

Examples of setup and output for the example using SPSS^x, BMDP, SAS, and SYSTAT are shown in Tables 12.5 through 12.8. Principal factor analysis with varimax rotation is shown for SPSS^x FACTOR, BMDP4M, and SAS FACTOR. For SYSTAT FACTOR, version 3.0, principal components is the only available method of extraction.

SPSS^x FACTOR (Table 12.5) begins by printing out SMCs for each variable, labeled COMMUNALITY in the INITIAL STATISTICS portion of the output. In a parallel but unrelated table, EIGENVALUEs, PCT OF VARiance, and percent of variance cumulated over the four factors (CUM PCT) are printed out for the four initial factors. (Be careful not to confuse factors with variables.) The program then indicates the number of factors extracted with eigenvalues greater than 1 (the default value).

Explains of SPSS out-put

For the two extracted factors, an unrotated FACTOR loading MATRIX is then printed. In the output labeled FINAL STATISTICS are final COMMUNALITY values for each variable (h^2 in Table 12.4), and, parallel with them, EIGENVALUEs for the two retained factors (see Table 12.3), the PCT OF VARiance for each factor and CUM PCT of variance accounted for by successive factors. The ROTATED FACTOR loading MATRIX, which matches loadings in Table 12.4 is given along with the FACTOR TRANSFORMATION MATRIX (Equation 12.8) for orthogonal varimax rotation with Kaiser normalization.

BMDP4M (Table 12.6) first prints out the CORRELATION MATRIX for the variables, and then the COMMUNALITIES (h^2) for each variable, OBTAINED FROM 2 FACTORS with eigenvalues greater than 1. In the next table, eigenvalues are given for all 4 factors along with two types of cumulative proportion of variance, neither

TABLE 12.5 SETUP AND SPSS[x] FACTOR OUTPUT FOR FACTOR ANALYSIS ON SAMPLE DATA OF TABLE 12.2

```
        TITLE       SMALL SAMPLE FACTOR ANALYSIS
        FILE HANDLE TAPE17
        DATA LIST   FILE=TAPE17        FREE
                    /SUBJNO,COST,LIFT,DEPTH,POWDER
        FACTOR      VARIABLES=COST,LIFT,DEPTH,POWDER/
                    EXTRACTION=PAF/


ANALYSIS NUMBER  1  LISTWISE DELETION OF CASES WITH MISSING VALUES


EXTRACTION  1  FOR ANALYSIS  1, PRINCIPAL AXIS FACTORING (PAF)


INITIAL STATISTICS:

VARIABLE      COMMUNALITY     FACTOR    EIGENVALUE    PCT OF VAR    CUM PCT

COST              .96076         1        2.01631        50.4         50.4
LIFT              .95324         2        1.94151        48.5         98.9
DEPTH             .98999         3         .03781          .9         99.9
POWDER            .99087         4         .00437          .1        100.0

  PAF EXTRACTED   2 FACTORS.    4 ITERATIONS REQUIRED.


FACTOR MATRIX:

            FACTOR  1     FACTOR  2

COST          -.40027       .89978
LIFT           .25060     -.94706
DEPTH          .93159       .34773
POWDER         .95596       .28615


FINAL STATISTICS:

VARIABLE      COMMUNALITY     FACTOR    EIGENVALUE    PCT OF VAR    CUM PCT

COST              .96983         1        2.00473        50.1         50.1
LIFT              .95973         2        1.90933        47.7         97.9
DEPTH             .98877
POWDER            .99574

- - - - - - - - - - - -   F A C T O R   A N A L Y S I S   - - - - - - - - - - - -

VARIMAX   ROTATION  1  FOR EXTRACTION  1  IN ANALYSIS  1 - KAISER NORMALIZATION.

  VARIMAX CONVERGED IN    3 ITERATIONS.

ROTATED FACTOR MATRIX:

            FACTOR  1     FACTOR  2

COST          -.08591      -.98104
LIFT          -.07100       .97708
DEPTH          .99403      -.02588
POWDER         .99706       .04028


FACTOR TRANSFORMATION MATRIX:

            FACTOR  1     FACTOR  2

FACTOR  1      .94565       .32519
FACTOR  2      .32519      -.94565
```

TABLE 12.6 SETUP AND SELECTED BMDP4M OUTPUT FOR FACTOR ANALYSIS ON SAMPLE DATA OF TABLE 12.2

```
        /PROBLEM     TITLE IS 'SMALL SAMPLE FACTOR ANALYSIS THROUGH BMDP4M'.
        /INPUT       VARIABLES ARE 5.  FORMAT IS '(A2,4F4.0)'.
        /VARIABLE    NAMES ARE SUBJNO, COST, LIFT, DEPTH, POWDER.
                     LABEL IS SUBJNO.
        /FACTOR      METHOD=PFA.
        /END

CORRELATION MATRIX
------------------

            COST      LIFT        DEPTH     POWDER

               2         3          4          5
COST     2   1.000
LIFT     3  -0.953     1.000
DEPTH    4  -0.055    -0.091      1.000
POWDER   5  -0.130    -0.036      0.990      1.000

COMMUNALITIES OBTAINED FROM  2 FACTORS AFTER    4 ITERATIONS.
-------------------------------------------------------------

THE COMMUNALITY OF A VARIABLE IS ITS SQUARED MULTIPLE
CORRELATION WITH THE FACTORS.

   2 COST        0.9698
   3 LIFT        0.9597
   4 DEPTH       0.9888
   5 POWDER      0.9957

FACTOR  VARIANCE    CUMULATIVE PROPORTION OF VARIANCE   CARMINES
        EXPLAINED    IN DATA SPACE   IN FACTOR SPACE      THETA
------  ---------   ---------------------------------    ------
   1     2.0047         0.5110           0.5122          0.6810
   2     1.9093         0.9977           1.0000
   3     0.0090         1.0000
   4    -0.0103

TOTAL VARIANCE IS DEFINED AS THE SUM OF THE POSITIVE EIGEN VALUES OF THE
CORRELATION MATRIX.

NEGATIVE VALUES FOR VARIANCE EXPLAINED INDICATE THE DEGREE
TO WHICH THE COMMUNALITIES HAVE BEEN UNDERESTIMATED.
LARGE NEGATIVE VALUES FOR VARIANCE EXPLAINED INDICATE THAT
THE COVARIANCE OR CORRELATION MATRIX IS POORLY ESTIMATED.


UNROTATED FACTOR LOADINGS (PATTERN)
-----------------------------------

                FACTOR      FACTOR
                  1           2

COST      2     -0.400      -0.900
LIFT      3      0.251       0.947
DEPTH     4      0.932      -0.348
POWDER    5      0.956      -0.286

         VP      2.005       1.909

THE VP FOR EACH FACTOR IS THE SUM OF THE SQUARES OF THE
ELEMENTS OF THE COLUMN OF THE FACTOR LOADING MATRIX
CORRESPONDING TO THAT FACTOR.  THE VP IS THE VARIANCE
EXPLAINED BY THE FACTOR.

ORTHOGONAL ROTATION, GAMMA =      1.0000
```

TABLE 12.6 (Continued)

```
ROTATED FACTOR LOADINGS (PATTERN)
---------------------------------

                  FACTOR     FACTOR
                     1          2

COST       2     -0.086      0.981
LIFT       3     -0.071     -0.977
DEPTH      4      0.994      0.026
POWDER     5      0.997     -0.040

          VP      1.995      1.919

THE VP FOR EACH FACTOR IS THE SUM OF THE SQUARES OF THE
ELEMENTS OF THE COLUMN OF THE FACTOR PATTERN MATRIX
CORRESPONDING TO THAT FACTOR.  WHEN THE ROTATION IS
ORTHOGONAL, THE VP IS THE VARIANCE EXPLAINED BY THE FACTOR.

FACTOR SCORE COVARIANCE (COMPUTED FROM FACTOR SCORES)
-----------------------------------------------------

                 FACTOR    FACTOR

                    1         2
  FACTOR   1      0.997
  FACTOR   2     -0.000     0.982
```

of which exactly matches that of SPSS^x. CARMINES THETA is a measure of the reliability of the first principal component that is closely related to Cronbach's alpha.

UNROTATED FACTOR LOADINGS are given for each variable on the two factors, along with the SSLs for each factor, labeled VP. This is followed by ROTATED FACTOR LOADINGS, with SSLs again shown as VP. Finally, BMDP4M prints out the FACTOR SCORE COVARIANCE matrix, in which the off-diagonal elements are set to zero by the nature of orthogonal rotation. The diagonal elements are the SMCs of each factor with the variables.

SAS FACTOR (Table 12.7) requires a bit more instruction to produce a principal factor analysis with orthogonal rotation for two factors. You specify the initial communalities (PRIOR = SMC), number of factors to be extracted (NFACTOR = 2) and the type of rotation (ROTATE = V). PRIOR COMMUNALITY ESTIMATES—SMCs—are given, followed by 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.

The FACTOR PATTERN matrix contains unrotated factor loadings for the first two factors. 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 transformation matrix for orthogonal rotation (Equation 12.8) 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.

SYSTAT FACTOR (Table 12.8) currently provides only principal components extraction and requires instructions for both number of factors (EIGEN = 2) and type

TABLE 12.7 SETUP AND SAS FACTOR OUTPUT FOR FACTOR ANALYSIS OF SAMPLE DATA OF TABLE 12.2

```
DATA SSAMPLE;
INFILE TAPE17;
INPUT SUBJNO COST LIFT DEPTH POWDER;
PROC FACTOR
     PRIOR=SMC
     NFACTORS=2
     ROTATE=V;
VAR  COST LIFT DEPTH POWDER;


INITIAL FACTOR METHOD: PRINCIPAL FACTORS

               PRIOR COMMUNALITY ESTIMATES: SMC

           COST       LIFT      DEPTH     POWDER
       0.960761   0.953241   0.989992   0.990873

EIGENVALUES OF THE REDUCED CORRELATION MATRIX:  TOTAL =  3.89487   AVERAGE =  0.973717

                        2          3          4
  EIGENVALUE     2.002343   1.902738   0.005204  -0.015418
  DIFFERENCE     0.099606   1.897534   0.020622
  PROPORTION       0.5141     0.4885     0.0013    -0.0040
  CUMULATIVE       0.5141     1.0026     1.0040     1.0000

        2 FACTORS WILL BE RETAINED BY THE NFACTOR CRITERION

                         FACTOR PATTERN

  FACTOR1    FACTOR2

             COST      -0.38290    0.90476
             LIFT       0.23317   -0.95009
             DEPTH      0.93844    0.33099
             POWDER     0.95953    0.26820

             VARIANCE EXPLAINED BY EACH FACTOR

                   FACTOR1    FACTOR2
                  2.002343   1.902738

  FINAL COMMUNALITY ESTIMATES: TOTAL =    3.905081

             COST       LIFT      DEPTH     POWDER
         0.965198   0.957031   0.990217   0.992634

ROTATION METHOD: VARIMAX

              ORTHOGONAL TRANSFORMATION MATRIX

                          1          2

                 1     0.95140   -0.30795
                 2     0.30795    0.95140

                ROTATED FACTOR PATTERN

                      FACTOR1    FACTOR2

             COST     -0.08566    0.97870
             LIFT     -0.07075   -0.97572
             DEPTH     0.99476    0.02591
             POWDER    0.99549   -0.04033

             VARIANCE EXPLAINED BY EACH FACTOR

                     FACTOR1    FACTOR2
                    1.992897   1.912184

  FINAL COMMUNALITY ESTIMATES: TOTAL =    3.905081

             COST       LIFT      DEPTH     POWDER
         0.965198   0.957031   0.990217   0.992634
```

TABLE 12.8 SETUP AND SYSTAT FACTOR OUTPUT FOR PRINCIPAL COMPONENTS ANALYSIS OF SAMPLE DATA OF TABLE 12.2

```
        USE TAPE17
        OUTPUT=OUTFILE
        EIGEN=2
        ROTATE=VARIMAX
        FACTOR COST LIFT DEPTH POWDER

MATRIX TO BE FACTORED

                 COST       LIFT       DEPTH      POWDER

      COST       1.000
      LIFT      -0.953      1.000
     DEPTH      -0.055     -0.091      1.000
    POWDER      -0.130     -0.036      0.990      1.000

LATENT ROOTS (EIGENVALUES)

                 1          2          3          4

                 2.016      1.942      0.038      0.004

COMPONENT LOADINGS

                 1          2

      COST      -0.500      0.856
      LIFT       0.357     -0.925
     DEPTH       0.891      0.449
    POWDER       0.919      0.390

VARIANCE EXPLAINED BY COMPONENTS

                 1          2

                 2.016      1.942

PERCENT OF TOTAL VARIANCE EXPLAINED

                 1          2

                50.408     48.536
ROTATED LOADINGS

                 1          2

      COST      -0.087      0.988
      LIFT      -0.072     -0.989
     DEPTH       0.997      0.026
    POWDER       0.998     -0.040

VARIANCE EXPLAINED BY ROTATED COMPONENTS

                 1          2

                 2.003      1.955

PERCENT OF TOTAL VARIANCE EXPLAINED

                 1          2

                50.067     48.879
```


of rotation. The output includes the correlation matrix—MATRIX TO BE FACTORED—and the eigenvalues for all four factors. The unrotated loading matrix (COMPONENT LOADINGS) is printed, followed by the SSLs (VARIANCE EXPLAINED BY COMPONENTS) and the percent of variance for each component. That information is repeated for the rotated solution after the matrix of ROTATED LOADINGS.

12.5 MAJOR TYPES OF FACTOR ANALYSIS

Numerous procedures for factor extraction and rotation are available. However, only those procedures available in the BMDP, SPSS[x], SAS, and SYSTAT packages are summarized here. Other extraction and rotational techniques are described in Mulaik (1972), Harman (1976, 1983), and Rummel (1970).

12.5.1 Factor Extraction Techniques

Among the extraction techniques available in the four packages are principal components (PCA), principal factors, maximum likelihood factoring, Rao's canonical factoring, image factoring, alpha factoring, and unweighted and generalized least squares factoring (see Table 12.9). 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 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 12.10 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 extraction may be rotated by any of the procedures described in Section 12.5.2, except Kaiser's Second Little Jiffy Extraction, which has its own rotational procedure.

Lastly, when using FA the researcher should hold in abeyance well-learned proscriptions against data snooping. It is quite common to use PCA 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.

12.5.1.1 PCA vs. 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

TABLE 12.9 SUMMARY OF EXTRACTION PROCEDURES

Extraction technique	Program	Goal of analysis	Special features
Principal components	SPSS[x] BMDP4M SAS SYSTAT	Maximize variance extracted by orthogonal components	Mathematically determined, empirical solution with common, unique, and error variance mixed into components
Principal factors	SPSS[x] BMDP4M SAS	Maximize variance extracted by orthogonal factors	Estimates communalities to attempt to eliminate unique and error variance from factors
Image factoring	SPSS[x] BMDP4M (Second Little Jiffy) SAS (Image and Harris)		Uses SMCs between each variable and all others as communalities to generate a mathematically determined solution with error variance and unique variance eliminated
Maximum likelihood factoring	BMDP4M SPSS[x] SAS	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	SPSS[x] SAS	Maximize the generalizability of orthogonal factors	Somewhat likely to produce communalities larger than 1
Unweighted least squares	SPSS[x] SAS	Minimize squared residual correlations	
Generalized least squares	SPSS[x] SAS	Weights variables by shared variance before minimizing squared residual correlations	

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

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

TABLE 12.10 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.

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, FA is your choice. If, on the other hand, you want an empirical summary of the data set, PCA is the better choice.

12.5.1.2 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 recommended as the first step in FA where it reveals a great deal about probable number and nature of factors. PCA is available through all four computer packages.

12.5.1.3 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 extraction, however, principal factors is sometimes not as good as other extraction techniques in reproducing the correlation matrix. Also, communalities must be estimated and the solution is, to some extent, determined by those estimates. Principal factor analysis is available through SPSS, BMDP, and SAS.

12.5.1.4 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, 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. The compromise is struck by using the squared multiple correlation (SMC or R^2) of each variable as DV with all others serving as IVs as the communality for that variable. With a hefty sample size and more than 10 to 15 observed variables, SMCs are stable and, depending on the adequacy of sampling of *variables*, provide a decent estimate of communality.

Image factoring is available through SPSS^x FACTOR, BMDP4M (as Kaiser's Second Little Jiffy), and SAS FACTOR (with two types—"image" and Harris component analysis).

12.5.1.5 Maximum Likelihood Factor Extraction The maximum likelihood method of factor extraction was developed originally by Lawley in the 1940s (see Lawley and 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 6).

Maximum likelihood extraction is available through BMDP4M, SPSS^x FACTOR, and SAS FACTOR. It is the extraction procedure recommended by BMDP when the common factor model is appropriate, the number of variables is fewer than 60, and the correlation matrix is not singular.

12.5.1.6 Unweighted Least Squares (Minres) Factoring The goal of unweighted least squares (minimum residual) 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. This procedure gives the same results as principal factors if communalities are the same. The procedure was developed by Comrey (1962) and Harman and Jones (1966) and is available through SPSS^x FACTOR and SAS FACTOR.

12.5.1.7 Generalized 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. Differences for variables that have substantial shared variance with other variables are weighted more heavily than differences for variables that have substantial unique variance. In other words, differences for variables that are not as strongly related to other variables in the set are not as important to the solution. This relatively new method of extraction is available through SPSS^x FACTOR and SAS FACTOR.

12.5.1.8 Alpha Factoring Alpha factor extraction, available through SPSS^x 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 are estimated, using iterative procedures, that maximize coefficient alpha for the factors.

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.

12.5.2 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 the different methods of extraction tend to give similar results with a good data set, so also the 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 the four reviewed packages are included in this discussion (see Table 12.11). 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 (1973, pp. 109—145).

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

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 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.

TABLE 12.11 SUMMARY OF ROTATIONAL TECHNIQUES

Rotational technique	Program	Type	Goals of analysis	Comments
Varimax	BMDP4M SAS SPSS^x SYSTAT	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 (in BMDP, Γ [gamma] = 1)
Quartimax	BMDP4M SAS SPSS^x SYSTAT	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 (in BMDP, Γ = 0)
Equamax	BMDP4M SAS SPSS^x SYSTAT	Orthogonal	Simplify both variables and factors (rows and columns); compromise between quartimax and varimax	May behave erratically (Γ = 1/2)
Orthogonal with gamma	BMDP4M SAS	Orthogonal	Simplify either factors or variables, depending on the value of gamma (Γ)	Gamma (Γ) continuously variable
Direct oblimin	BMDP4M SPSS^x	Oblique	Simplify factors by minimizing cross products of loadings	Continuous values of gamma, Γ (BMDP) or delta, δ (SPSS^x), available; allows wide range of factor intercorrelations
(Direct) quartimin	BMDP4M SPSS^x	Oblique	Simplify factors by minimizing sum of cross products of squared loadings in pattern matrix	Permits fairly high correlations among factors. Recommended oblique rotation by BMDP series; in BMDP, Γ = 0. Achieved in SPSS^x by setting δ = 0
Orthoblique	BMDP4M SAS (HK)	Both orthogonal and oblique	Rescale factor loadings to yield orthogonal solution; nonrescaled loadings may be correlated	Accompanies Kaiser's Second Little Jiffy (image) extraction in BMDP4M
Promax	SAS	Oblique	Orthogonal factors rotated to oblique positions	Fast and inexpensive
Procrustes	SAS	Oblique	Rotate to target matrix	Useful in confirmatory FA

Although varimax rotation simplifies the factors, quartimax the variables, and equamax both, they do so in BMDP4M and SAS FACTOR by setting levels on a simplicity criterion—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.

For many applications, varimax is the rotation of choice; it is the default option of all four packages.

12.5.2.2 Oblique Rotation An *embarrasse de richesse* awaits the researcher who uses oblique rotation (see Table 12.11). 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 SPSS^x FACTOR and gamma (Γ) by BMDP4M.[8] 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 SPSS and BMDP is direct oblimin. In the special case where Γ or $\delta = 0$ (the default option for both SPSS and BMDP), 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 to the default value.

Orthoblique rotation is designed to accompany Kaiser's Second Little Jiffy (Image) Factor extraction, and does so automatically through BMDP4M. 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.

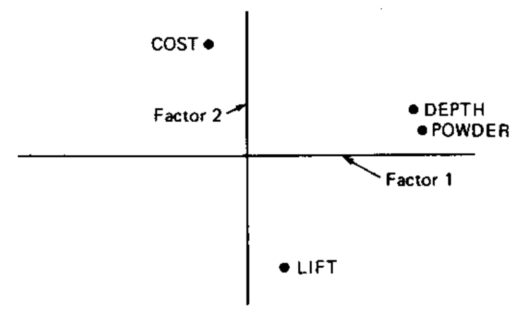
Promax and Procrustes are available through SAS. In promax rotation, an orthogonally rotated solution (usually varimax) is rotated again to allow correlations among factors. The orthogonal loadings are raised to powers (usually powers of 2, 4, or 6) to drive small and moderate loadings to zero while 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 and inexpensive.

[8] In BMDP, gamma is used to indicate the nature of simplicity in orthogonal rotation and the amount of obliqueness in oblique rotation.

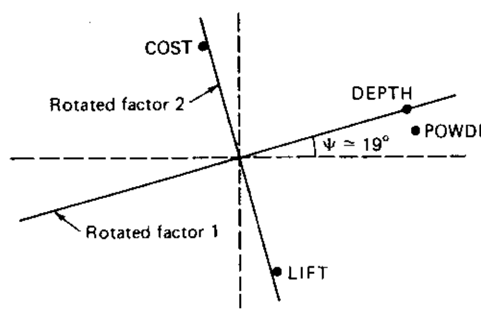
In Procrustes rotation, a target matrix of loadings (usually zeros and ones) is specified by the researcher and a transformation matrix is sought to rotate extracted factors to the target, if possible. If the solution can be rotated to the target, then the hypothesized factor structure is said to be 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.

12.5.2.3 Geometric Interpretation A geometric interpretation of rotation is in Figure 12.1 where 12.1(a) is the unrotated and 12.1(b) the rotated solution to the example in Table 12.2. 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 12.1(b). The position of points does not change, but their coordinates change in the new axis system. COST is now (-0.86, .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.



(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 12.1 Illustration of rotation of axes to provide a better definition of factors vis à vis the variables with which they correlate.

Factor extraction yields a solution in which observed variables are vectors that terminate at 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.

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, though 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 12.6.3 for practical suggestions of ways to use graphic techniques to judge the adequacy of rotation.

12.5.3 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 slight. With a large number of variables, strong correlations among them, with the same, well-chosen number of factors, and with similar values for communality, the results of extraction are similar regardless of which method is used. 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 (Sections 12.3.2.6 and 12.8.1.6), the rank of the observed correlation matrix (Sections 12.3.2.5 and 12.8.1.5), the number of factors (Sections 12.6.2), and variables that might be excluded from subsequent analyses (Sections 12.3.2.7 and 12.8.1.7).

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.

12.6 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.

12.6.1 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, SMCs of each variable with all other variables 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.

BMDP, SPSS[x], and SAS provide several different starting statistics for communality estimation. BMDP4M offers SMCs, user-specified values, or maximum absolute correlation with any other variable as initial communality estimates. SPSS[x] 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. Fewer iterations are usually required when starting from SMCs.

The seriousness with which estimates of communality should be regarded depends on the number of observed variables. If the number of variables exceeds, say, 20, 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.

If communality values equal or exceed 1, problems with the solution are indicated. There are too few 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 outliers (Sections 12.3.2.7 and 12.8.1.7).

12.6.2 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). Reviewed below are methods available through SPSS[x], BMDP, SAS, and SYSTAT. SYSTAT provides the first two and SPSS[x], BMDP, and SAS all the methods described here.

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 may either over- or underestimate 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 SPSS[x] FACTOR and SAS FACTOR, but it is easy to produce from the list of factors and eigenvalues available through BMDP and SYSTAT.

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 SPSS[x] in Figure 12.2. What you are looking for is the point where a line drawn through the points changes direction. 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 12.2.

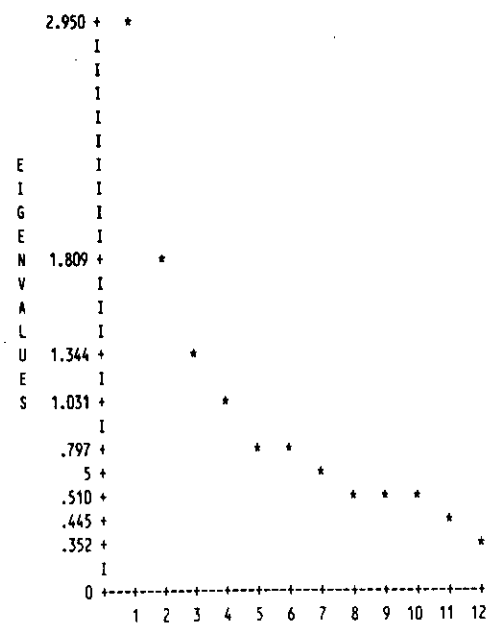


Figure 12.2 Screen output for sample data produced by SPSS[x] FACTOR. Note break in size of eigenvalues between the fourth and fifth factors.

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. Under less than optimal conditions, the scree test is still usually accurate to within one or two factors. If you are unsure of the number of factors, perform several factor analyses, each time specifying a different number of factors, repeating the scree test, and examining the residual correlation matrix.

The residual correlation matrix is available through SPSS^x, BMDP, and SAS. As discussed in Section 12.4, the residual correlation matrix is obtained by subtracting the reproduced correlation matrix from the observed correlation matrix. The numbers in the residual matrix are actually partial correlations between pairs of variables with effects of factors removed. If the analysis is good, the residuals are small. Several moderate residuals (say, .05 to .10) or a few large residuals (say >.10) suggest the presence of another factor.

Once you have determined the number of factors by these criteria, it is important to look at the rotated loading matrix to determine the number of variables that load on each factor (see Section 12.6.5). 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 cautious, however, under even the most exploratory factor analysis.

For principal components extraction and maximum likelihood extraction in confirmatory factor analysis there are significance tests for 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 factor analysis texts for discussion of significance testing in FA.

There is 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 all communality values < 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.

12.6.3 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 12.6.6) are to be used as IVs or DVs in other analyses, or if a goal of analysis is comparison of factor structure in groups (Section 12.6.7), 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 SPSS^x, BMDP, 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.

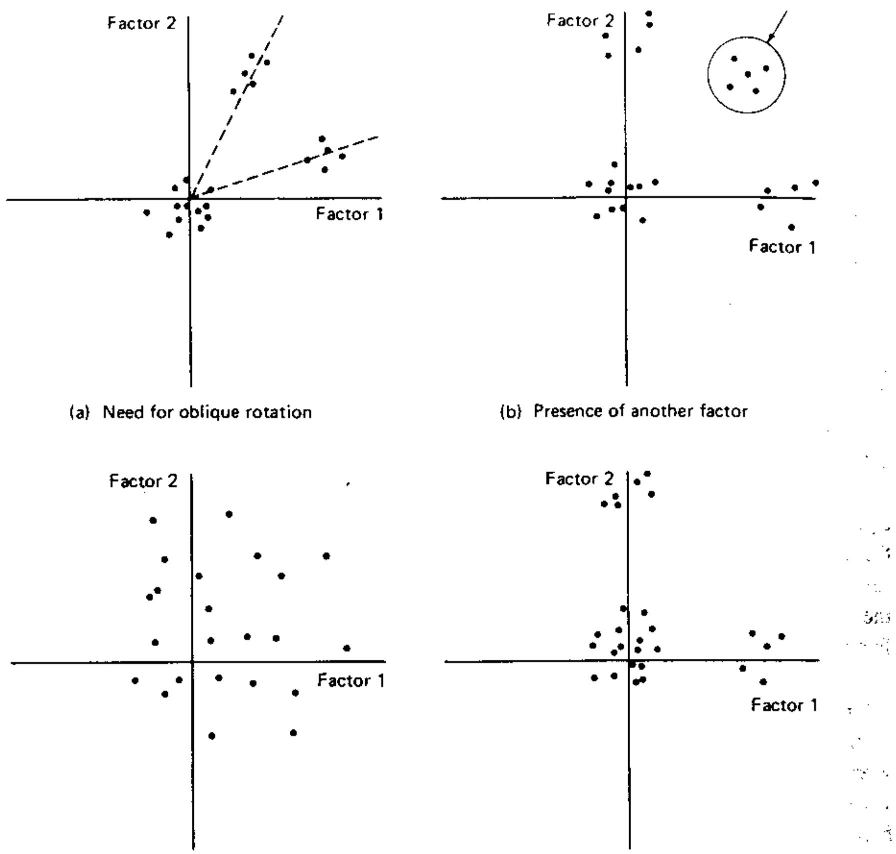
Look at the factor correlation matrix for correlations of .30 and above. If correlations exceed .30, then there is 10% (or more) overlap in variance among factors, enough variance to warrant oblique rotation unless there are compelling reasons for orthogonal rotation. Compelling reasons include a desire 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, 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 commands of SPSS^x FACTOR, SAS FACTOR, and BMDP4M. 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 the respect to the origin is not 90°, then a better fit to the clusters is



(a) Need for oblique rotation (b) Presence of another factor

(c) No simple structure, unsuitable data (d) Simple structure with orthogonal rotation

Figure 12.3 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.

provided by axes that are not orthogonal. Oblique rotation may reveal substantial correlations among factors. Several of these relationships are depicted in Figure 12.3.

12.6.4 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 associated with 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 individual factors is related to the sizes of their SSLs (*S*um of *S*quared *L*oadings from **A** after rotation). SSLs are converted to proportion of variance for a factor by dividing its SSL 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 12.4 and Section 12.5 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 *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. If an SMC is above 1, the entire solution needs to be reevaluated.

BMDP4M prints these SMCs as the positive diagonal of the factor-score covariance matrix. SPSS^x FACTOR prints them 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.

12.6.5 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 .30 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 "partialled 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 .30 and above are interpreted. The greater the loading, the more the variable is a pure measure of the factor. Comrey (1973) suggests 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 BMDP4M, by REORDER in SAS FACTOR, and SORT in SPSS^x FACTOR.

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?

12.6.6 Factor Scores

Among the potentially more useful outcomes of PCA or FA are factor scores. Factor scores are estimates of the scores 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 FA to reduce a large number of DVs to a smaller number of factors 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. If factors are few in number, stable, and interpretable, their use enhances subsequent analyses.

Procedures for estimating factor scores range between simple-minded (but frequently adequate) and sophisticated. Comrey (1973) describes 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 and there is no way to decide among them. As long as factor scores are considered only estimates, however, the researcher is not overly beguiled by them.

The method described in Section 12.4 (especially Equations 12.10 and 12.11) 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 zero 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 SYSTAT and the other three packages. All four packages write factor scores to files for use in other analyses and all four print standardized factor score coefficients—routinely through SYSTAT and BMDP, and through FSCORE in SPSS^x and SCORE in SAS.

SPSS^x 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 (that is, 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 zero, 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.

12.6.7 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. Only the easier comparison techniques are mentioned here.

It is important to note that theory can be tested in FA using these procedures. Theory regarding factor structure is used to generate one set of loadings that is compared with loadings derived from a sample. Estimation of the magnitude of factor loadings for variables from theory does not have to be very precise: ones can be used as loadings for variables that are expected to load on a factor, while zeros are used as loadings for the other variables. Comparisons between the *pattern* of loadings from theory and loadings from sample data are then conducted in confirmatory FA.

The first step in comparing factors from two different samples is to generate them. When comparison is the goal, similar procedures are employed at the various stages of analysis with the two data sets. Similar variables and, if possible, similar marker variables are measured during data collection. Similar procedures for handling missing data and outliers are employed. The same considerations are used to produce transformations of variables, if desired. Extractional and rotational techniques are the same, as is the criterion for determining number of factors. If factor scores are to be compared, they are generated by the same procedures.

Once data sets are factor analyzed, one decides which pairs of factors to compare. Comparing all possible pairs of factors can result in spuriously significant results by capitalizing on chance relationships. Presence of marker variables simplifies choosing pairs.

Before going on, careful inspection of the loading matrices for both groups may reveal similarities or differences in factor structure sufficiently clear as to obviate the need for more formal procedures. Did both groups generate the same number of factors? Do almost the same variables load highly on the different factors for the two groups? Could you reasonably use the same labels to name factors for both groups? If all three questions are answered in the affirmative, it is unnecessary to proceed to statistical comparisons.

If formal procedures are needed, an important decision is whether to compare just the pattern of loadings or both the pattern and magnitude of loadings in the data sets. Comparisons involving both pattern and magnitude are more stringent than those involving just pattern. Magnitude of loadings is influenced by extraneous features of data collection such as homogeneity of a sample for factors being compared, so magnitude is considered when the researcher believes that these influences are absent.

Cattell's salient similarly index, s (Cattell and Baggaley, 1960; Cattell, 1957), is used to compare patterns of loadings. The Pearson product-moment correlation coefficient, r, is used to compare both pattern and magnitude of loadings.

To illustrate these two methods, the two loading matrices in Table 12.12 are used. Both are products of overactive imagination but do illustrate a typical problem in factor comparison, namely, that factor 1 in Set 1 is similar to factor 2 (rather than

TABLE 12.12 LOADING MATRICES FROM TWO HYPOTHETICAL DATA SETS

	Set 1		Set 2	
	Factor 1	Factor 2	Factor 1	Factor 2
COST	−.086	.981	.732	.265
LIFT	−.072	−.978	.649	.537
DEPTH	.994	.027	.211	.874
POWDER	.997	−.040	.189	.796

to factor 1) in Set 2. Sometimes it is hard to decide which factors to compare. But if the decision is that difficult, perhaps you have your answer.

In calculating s, the first step is to construct a two-way frequency table, like that in Table 12.13, with pairs of loadings for each variable on each factor contributing a single tally to the table according to whether the loadings are positively salient (PS), negatively salient (NS), or neither (hyperplane or HP) on the two factors being compared. Cattell used a cut of .10 for determining salience; loadings at or above .10 were salient while lower ones were not. But a cut of .32 or higher is better for this example and is employed here.

For the COST variable, the loading of −.086 in Set 1 is in the hyperplane as

TABLE 12.13 CALCULATION OF CATTELL'S SALIENT SIMILARITY INDEX s

		Set 1		
		PS	HP	NS
Set 2	PS	c_{11}	c_{12}	c_{13}
	HP	c_{21}	c_{22}	c_{23}
	NS	c_{31}	c_{32}	c_{33}

For the example:

		Set 1		
		PS	HP	NS
Set 2	PS	2	1	
	HP		1	
	NS			

$$s = \frac{2 + 0 - 0 - 0}{2 + 0 + 0 + 0 + .5(1 + 0 + 0 + 0)}$$
$$= \frac{2}{2.5}$$
$$= .80$$

is the loading of .265 in Set 2. Therefore, a tally is placed in the c_{22} cell of the table. For the LIFT variable, $-.072$ is HP while .537 is PS, resulting in a tally in cell c_{12}. For DEPTH, .994 and .874 are both PS as, for POWDER, are both .997 and .796. These give the two tallies in cell c_{11}.

Once the frequency table is constructed, s is calculated as follows:

$$s = \frac{c_{11} + c_{33} - c_{13} - c_{31}}{c_{11} + c_{33} + c_{13} + c_{31} + .5(c_{12} + c_{21} + c_{23} + c_{32})} \tag{12.16}$$

The c values in the equation are replaced by frequency counts from cells in the frequency table. Application of the equation for comparison of factor 1 in Set 1 and factor 2 in Set 2 results in an s value of .80.

Estimates of probability values for s are provided by Cattell and colleagues (1969) and reproduced in Appendix C, Table C.7. Probabilities are assessed considering both the number of variables, p, and the percentage of cases that fall into the hyperplane for the pair of factors being compared: 60%, 70%, 80%, or 90%. If a value of s exceeds that of v_s for some hyperplane percentage and number of variables, then the factors are reliably similar. For instance, if the hyperplane count is 60% and 40 variables are compared, an s value in excess of .26 indicates similarity of factors at the .041 significance level. (Significance for the example cannot be determined because only 25% of the loadings are in the hyperplane and only four variables are included.)

The pattern and magnitude of loadings are compared for two factors and two groups by computing Pearson's r (see Equation 3.29). The loadings for factor 1 in Set 1 and factor 2 in Set 2 from Table 12.12 correlate .91. Although calculating r for loadings is a straightforward procedure and the meaning of r is widely understood, this method of comparing factors has drawbacks. If there are numerous variables, it is possible for r to be large even though no variables with large loadings are the same for the two factors. The correlation is large because of the numerous variables with small loadings that are not loaded on either factor. Thus, caution is urged in interpretation of r used to compare factors.

Another method, experimental at this stage, involves generating pairs of factor scores for a group by using the factor-score coefficients for that group and then generating factor scores by using factor-score coefficients for the other group. The pairs of factor scores are correlated. If correlation is high, it implies that there is good correspondence between factor scores generated by the two different groups and that factor structure is therefore similar.

12.7 COMPARISON OF PROGRAMS

BMDP, SPSS^x, and SAS each have a single program to handle both FA and PCA. All three programs have numerous options for extraction and rotation and give the user considerable latitude in directing the progress of the analysis. The programs are all flexible and rich in information. The SYSTAT FACTOR 3.0 program is limited to PCA and orthogonal rotation. Features of all four programs are described in Table 12.14.

TABLE 12.14 COMPARISON OF FACTOR ANALYSIS PROGRAMS

Feature	SPSS^x FACTOR	BMDP4M	SAS FACTOR	SYSTAT FACTOR
Input				
Correlation matrix	Yes	Yes	Yes	Yes
About origin	No	Yes	Yes	No
Covariance matrix	No	Yes	Yes	Yes
About origin	No	Yes	Yes	No
Factor loadings (unrotated pattern)	Yes	Yes	Yes	No
Factor-score coefficients	No	Yes	Yes	No
Factor loadings (rotated pattern) and factor correlations	Yes	No	Yes	No
Specify maximum number of factors	Yes	Yes	Yes	Yes
Extraction method (see Table 12.9)				
PCA	Yes	Yes	Yes	Yes
PFA	Yes	Yes	Yes	No
Image (Little Jiffy, Harris)	Yes	Yes	Yes[a]	No
Maximum likelihood (Rao's canonical)	Yes	Yes	Yes	No
Alpha	Yes	No	Yes	No
Unweighted least squares	Yes[b]	No	Yes	No
Generalized least squares	Yes[b]	No	Yes	No
Specify communalities	Yes	Yes	Yes	N.A.
Specify minimum eigenvalue	Yes	Yes	Yes	Yes
Specify proportion of variance to be accounted for	No	No	Yes	No
Specify maximum number of iterations	Yes	Yes	Yes	N.A.
Option to allow communalities > 1	No	No	Yes	N.A.
Specify tolerance	No	Yes	Yes	No
Specify convergence criterion	Yes	Yes	Yes	N.A.
Rotation method (see Table 12.11)				
Varimax	Yes	Yes	Yes	Yes
Quartimax	Yes	Yes	Yes	Yes
Equamax	Yes	Yes	Yes	Yes
Direct oblimin	Yes	Yes	No	No
Direct quartimin	$\delta = 0$	Yes	No	No
Indirect oblimin	No	Yes	No	No
Orthoblique	No	Yes	HK	No
Promax	No	No	Yes	No
Procrustes	No	No	Yes	No
Prerotation criteria	No	No	Yes	No
Optional Kaiser's normalization	Yes[c]	Yes	Yes	Normalized only
Optional weighting by Cureton-Mulaik technique	No	No	Yes	No
Optional rescaling of pattern matrix to covariances	No	No	Yes	No
Maximum numbers of factors	No limit	10[d]	No limit	No limit
Differential case weighting	No	Yes	No	No
Differential variable weighting	No	No	Yes	No
Output				
Means and standard deviations	Yes	Yes	Yes	No
Number of cases per variable (missing data)	Yes	No	No	No