Meyers, L.A., Gamst, G., & Guarino, A.J. (2006). Applied multivariate research: Design and interpretation. Thousand Oaks, CA: Sage.

CHAPTER 12A

Principal Components and Factor Analysis

How Factor Analysis Is Used in Psychological Research

The term *data crunching* generally refers to a process in which a considerable quantity of data is reduced down to a more manageable or consolidated whole. Principal components analysis and factor analysis are the quintessential data-crunching procedures. Their general purpose is to identify a relatively small number of themes, dimensions, components, or factors underlying a relatively large set of variables. The way they do this is by distinguishing sets of variables that have more in common with each other than with the other variables in the analysis. What the subsets of variables have in common are the underlying components or factors.

Principal components analysis and factor analysis are very closely related data reduction techniques differing at both the statistical and the conceptual levels. Because many researchers use the term *factor analysis* in a generic way whether they are referring to principal components analysis or to one of the variations of factor analysis (Gorsuch, 1983), we will adopt this generic language in the beginning of this chapter just for introductory purposes, but we will address the differences between component and factor analysis both later on in this chapter as well as in Chapter 13A because it is important to distinguish between these techniques.

One of the most common applications of this data-crunching procedure is in test development, and the examples we present here are drawn from this domain. Using factor analysis in this context, researchers analyze the responses of participants to a relatively large number of test items. Their

goals are to choose the items that will appear on the final version of the survey and to determine if the construct assessed by the inventory can best be measured by reasonable and viable subscales. The factors will indicate which items should be assigned to which subscale.

Another potential application of factor analysis might be to help researchers organize or conceptualize a set of measures that they have obtained in the context of a research program. For example, researchers may have administered a large battery of cognitive tests to a sample of participants to determine which ones might be assessing a similar mental process. Here, each factor will represent a set of measures that are relatively strongly related to each other. Further analyses can then be conducted based on the factors rather than on the individual measures themselves; for example, we can examine group differences in a MANOVA using the factors as dependent variables, we can predict group membership through logistic regression or discriminant analysis using the factors as predictors, or we can predict the value of a quantitative variable in multiple regression with the factors as the independent variables.

In the prior sections of this book, we have talked about independent variable and dependent variable variates (weighted linear combinations of variables). Factor analysis still involves the construction of variates, but we do not ordinarily need to distinguish between independent variables and dependent variables in this context. The key requirement is that we have a set of quantitatively measured variables whose underlying structure we wish to explore. Factors resulting from the analysis are the variates, the weighted combinations of variables. The statistical procedure is designed, in part, to produce the weights of those variables in the linear combination.

Origins of Factor Analysis

A very readable and informative history of the development of factor analysis is provided in the beginning of Harry Harmon's (1960, 1967) classic text. Harmon tells us that the origins of this statistical technique can be traced to Charles Spearman (1904), who is considered to be the father of factor analysis, although Anastasi (Anastasi & Urbana, 1997) suggests that an earlier paper by Karl Pearson (1901) might have established a beginning on which Spearman could build. In any case, Spearman proposed a two-factor theory of intelligence by identifying a general intelligence factor (which came to be symbolized as *g*) underlying all cognitive abilities as well as factors specific to particular abilities (e.g., memory for number sequences, ability to perceive spatial relations).

Harmon (1960) goes on to say that

> it had become quite apparent that Spearman's Two-Factor Theory was not always adequate to describe a battery of psychological tests. So group factors found their way into factor analysis; although the experimenters, at first, were very reluctant to admit such deviation from the base theory and restricted the group factors to as small a number as possible. What actually happened was that the *theory* of a general and specific factors in Spearman's original form was superseded by theories of many group factors, but the early *method* [of Spearman] continued to be employed to determine these many factors. Then it naturally followed that some workers explored the possibility of extracting several factors directly from a matrix of correlations among tests, and thus arose the concept of multiple-factor analysis in the work of Garnett [1919]. (pp. 3–4)

Garnett (1919) had given us the freedom to extract more than two factors directly from the correlation matrix, and Harmon (1960) talks about a surge of interest in these kinds of techniques in the mid-1920s and 1930s leading to the continued advancement of factor analysis. As we will see later in this chapter, another extremely productive flurry of activity in the late 1940s and early 1950s resulted in the development of factor rotation.

Sample Size Issues

Factor analysis is one of the large-sample statistical procedures that we discuss in this book. Guidelines for judging the adequacy of the sample size for such an analysis are available from several sources (e.g., Bryant & Yarnold, 1995; Comrey & Lee, 1992; Gorsuch, 1983; Hutcheson & Sofroniou, 1999). Comrey and Lee (1992), for example, have provided the following very general evaluations of the adequacy of various sample sizes for factor analysis:

50	very poor
100	poor
200	fair
300	good
500	very good
1,000	excellent

It is useful to think about the sample size for your analysis not only in general but also in terms of the ratio of participants to variables in the analysis (e.g., number of survey items). Our general recommendation is this: *Do not drop below an N of 200. Start with a target ratio of 10 participants for every variable. With a larger number of variables, you can back off in this sample-size-to-number-of-variables ratio.*

Here are some examples that may help to illustrate our recommendations. If you are analyzing a 10-item inventory, plan to collect data on about 200 participants. If you are analyzing a 25-item inventory, run about 250 participants. If you are analyzing a 90-item inventory, run about 400 participants. If you are analyzing a 500-item inventory, you can probably run between 700 and 1,000 participants and expect to find a stable solution.

A Simplified Example: Correlations Near 1 and 0

Factor analysis attempts to summarize and synthesize the relationships between the variables contained in the analysis. It starts at the same place that many of the other multivariate procedures begin—with the correlation matrix showing the Pearson *r*s between pairs of variables. Bivariate correlations assess the degree to which the variables are associated or related. A correlation matrix contains these pairwise correlations for all the variables. To make the relationships of the variables that much easier to see, in the following discussion, we will talk about correlations that are either extremely high or extremely low.

A Three-Variable Example

As a prelude to our discussion of factor analysis, consider the very small correlation matrix of the three hypothetical survey items labeled as A, B, and C shown in Table 12a.1. These are correlations of the responses to particular questions on a hypothetical survey or inventory. That is, for the A-B correlation, respondents' answers to Item A are evaluated with respect to their responses to Item B; for the A-C correlation, respondents' answers to Item

Table 12a.1 Correlation Matrix for Variables A Through C

	B	*C*
Variable A	.87	.12
Variable B		.07

A are evaluated with respect to their responses to Item C; and for the B-C correlation, respondents' answers to Item B are evaluated with respect to their responses to Item C. With only three variables in the matrix it is relatively easy to see, without any elaborate statistical procedure such as factor analysis, that Items A and B are very strongly correlated but that neither is related much to Item C.

The obtained Pearson r value of .87 assesses the association between A and B. From this value, we know that responses to Item B are very predictable from a knowledge of the responses to Item A. For Pearson r values of .12 and .07 to be found between C and A and C and B, respectively, we know that responses to Item C are not at all predictable from the responses participants gave to either Item A or Item B.

Translated into the language of the test developer, these correlations suggest to us that the content addressed in Item A is presumably related to the content addressed in Item B because the two questions tend to evoke corresponding (i.e., correlated, covarying) responses; stronger endorsement of A matches up with stronger endorsement of B, and weaker endorsement of A matches up with weaker endorsement of B. By the same reasoning, these correlations also tell us that the content of Item C is presumably not related to that of Items A and B because the responses of the questionnaire respondents to C are, by and large, independent of their responses to A and B. We can thus go on to say that Variables A and B have quite a bit in common with each other (i.e., they exhibit quite a bit of shared variance) but do not have very much in common with Variable C.

A Four-Variable Example

Now consider a slightly larger set of variables in an expansion of the preceding example, which we will use to broach the topic of factor analysis. In this somewhat larger study, we have added Item D to the original three survey questions. The correlation matrix that includes this newly added variable is displayed in Table 12a.2. Here, we see that D is pretty strongly related to C but is not related to A and B. At this juncture, we can then draw the

Table 12a.2 Correlation Matrix for Variables A Through D

	B	*C*	*D*
Variable A	.87	.12	.11
Variable B		.07	.17
Variable C			.86

inference that A and B have much in common, that C and D have much in common, and that A and B have little in common with C and D.

Another way to verbalize what we can infer from the relationships between the survey questions is to say that the set of four variables shown in Table 12a.2 really represents two sets of two variables each. Because factors distinguish sets of variables that have more in common with each other than with the other variables in the analysis, we may then also say that the four variables represent two underlying factors. From the correlations of these variables, Items A and B appear to be the primary indicators of one factor, and Items C and D appear to be the primary indicators of the other factor. In essence, what we have just done from our examination of the correlation matrix is to perform a conceptually based factor analysis.

These factors that we have identified are not themselves directly measured variables. Instead, they are latent variables (*variates* or *weighted linear composites* as we have called them in the prior chapters). Each factor is composed of the four variables in the analysis, but the variables have different weights in the two factors. In one variate or factor, A and B should be weighted quite highly, whereas C and D should have negligible weights. We therefore say that A and B are strong indicators of that factor (and that C and D are weak indicators of that factor). The situation is reversed in the other variate. There, C and D should be weighted quite highly, whereas A and B should have negligible weights. In this case, C and D would be the strong indicators of that factor (and A and B would be weak indicators).

As we have seen in other multivariate designs (e.g., regression analysis, discriminant analysis), it is also useful to examine the relationship (correlation) between each of the variables contained in the linear composite and the variate as a whole. You probably recall that such correlations are called *structure coefficients.* The weights assigned to the variables in the linear composite and the structure coefficients both provide us with a perspective for interpreting the variate or factor.

The weights of the variables as well as the structure coefficients associated with the variables are ordinarily shown in a table called a *factor matrix.* Such tables are a standard feature in the computer output of a factor analysis procedure. We show such a factor matrix in Table 12a.3 for Variables A through D. In this table, the variables in the analysis occupy the rows and the factors occupy the columns. These factors were obtained from the results of the rotation phase of the analysis (we will talk about rotation in detail later); we therefore refer to this factor matrix as a *rotated factor matrix.*

Because an orthogonal strategy was used for the rotation (we will explain this in our discussion of rotation), the numerical entries in the table represent at the same time the weights of the variables in the linear composite as

Table 12a.3 Two-Factor Rotated Matrix for Variables A Through D

	Factor 1	*Factor 2*
B	.96563	.06356
A	.96439	.05828
C	.03566	.96422
D	.08635	.96065

well as the structure coefficients. Either way of viewing these values is acceptable in the process of interpreting the results of the factor analysis. For ease of presentation, we have organized the results in Table 12a.3 in a fashion that groups or sorts the correlations (the weights) by factor and orders them within each factor. All statistical programs such as SPSS are capable of generating this kind of output format.

The rotated factor matrix is one of the main ingredients composing a statistical solution to the factor analysis procedure. A convenient way to examine the rotated factor matrix is to first focus on the rows of the table to understand how the variables are behaving. Once we have digested this information, we can then concentrate on the columns in an effort to interpret the factors.

Using this strategy, we first examine the rows of Table 12a.3. Remember that we can view the table entries as structure coefficients. From this perspective, the top two rows of this rotated factor matrix show the correlations of Variables A and B with each of the two factors. Variable B is (slightly) more strongly correlated with Factor 1 than is Variable A and so is listed first in the top grouping. Reading across these first two rows reveals a consistent story: A and B are both strongly correlated with Factor 1 and are not correlated with Factor 2. The last two rows of Table 12a.3 show the correlations of C and D with each of the two factors. Reading across these rows, we can see that Variables C and D are not correlated with Factor 1 but are strongly correlated with Factor 2.

We can now focus on the columns (the factors) of this rotated factor matrix. In this matrix, each factor is composed of a weighted combination of the four variables, and we can view the numbers in the table as representing those weights. The key to dealing with the information contained in the rotated factor matrix is to notice the difference in the pattern of the weights in the two factors. Looking first at Factor 1, we find that it is strongly exemplified by Variables A and B and weakly represented by C and D. That is, A and B have been weighted quite strongly, and C and D have been weighted quite weakly in the establishment of Factor 1. An analogous finding is obtained for Factor 2 in which C and D have received the very substantial weights, whereas the weights assigned to A and B are negligible.

Table 12a.4 Correlation Matrix for Variables A Through F

	B	*C*	*D*	*E*	*F*
Variable A	.87	.12	.11	.92	.08
Variable B		.07	.17	.89	.13
Variable C			.86	.10	.88
Variable D				.14	.91
Variable E					.06

The process of examining the columns or factors of the rotated factor matrix is the way that the factors are "interpreted" by researchers—knowing the content of Items A and B in effect unveils the essence of Factor 1 because these variables are the primary indicators of this factor, and knowing the content of Items C and D discloses the essence of Factor 2 because they are the primary indicators of that factor.

A Six-Variable Example

Let's expand the example further by adding two additional variables, E and F, to the correlation matrix. This is shown in Table 12a.4. There are many more relationships to deal with now given the inclusion of these two additional variables. But because we are building on what we presented in the previous table and because we are still working with correlations that are relatively near 1 and 0, it is still possible to discern the interrelationships between the six variables contained in Table 12a.4. Because we have constructed this example with unambiguous interpretation in mind, we have made E strongly related to A and B, whereas we have made F strongly related to C and D. Given this pattern, it is not surprising that we have built this example such that E and F have no substantial relationship.

Although there is more information in Table 12a.4 than in the prior correlation matrices, we can still identify two factors based on the correlations between the pairs of variables; that is, we can still identify two relatively distinct sets of survey items. One set of variables (one factor) is composed of Inventory Questions A, B, and E, and the other set of questions is composed of C, D, and F. Analogous to what we stated above, we can say here that A, B, and E have a great deal in common and that C, D, and F, likewise, have a great deal in common. That is, it appears that the variables within each set share a great deal of variance and that there is not much shared variance (not much of a relationship) between the two sets.

Just as in the prior example, what we have essentially done here for this more complex correlation matrix is to again perform a conceptually based

Table 12a.5 Two-Factor Rotated Matrix for Variables A Through F

	Factor 1	*Factor 2*
E	.97136	.04630
A	.96344	.05026
B	.95153	.07340
F	.03577	.96865
D	.09035	.95795
C	.04405	.94933

factor analysis of the Variables A, B, C, D, E, and F. We have determined that there are two factors underlying this set of variables. One of the factors was represented by A, B, and E, and the other factor was represented by C, D, and F. Each set of variables shares a great deal of variance. The factor is, at a conceptual level, that shared variance, and the factor would ordinarily be labeled as whatever it is that the items in the set have in common.

Our conceptually based factor analysis coincides very well with the results of the statistical analysis that we performed on these correlations. We show in Table 12a.5 the rotated factor matrix from that analysis. In looking at the rows of this matrix, we can see a very general pattern—variables are pretty strongly correlated with one of the factors while showing no substantial correlation with the other factor. Note that even though the pattern is quite similar, the values of these correlations are different from those shown in the prior rotated factor matrix for Variables A through D. This is to be expected because the two variates (factors) represented in Table 12a.5 are linear composites of six rather than four variables and therefore depict entities that are at least partially different from those in Table 12a.3.

Examining the columns (the factors) in Table 12a.5 reaffirms and extends the story that emerged in the previous example. Factor 1 here is best indicated by Survey Questions A, B, and E (with C, D, and F playing little role in that composite variate). In contrast, Factor 2 is best signified by Questions C, D, and F (with A, B, and E playing little role in that composite variate).

Introducing a Little Fuzziness Into the Example: Modest Relationship Strengths

We will now expand our example one final time to illustrate another point. We have added Items G and H to our set of variables, and we show the entire correlation matrix in Table 12a.6. We acknowledge that the table is beginning to get very cluttered, but stay with us for just a little longer.

Table 12a.6 Correlation Matrix for Variables A Through H

	B	*C*	*D*	*E*	*F*	*G*	*H*
Variable A	.87	.12	.11	.92	.08	.61	.08
Variable B		.07	.17	.89	.13	.57	.15
Variable C			.86	.10	.88	.14	.56
Variable D				.14	.91	.09	.60
Variable E					.06	.55	.07
Variable F						.08	.58
Variable G							.07

We have written in correlations for G that tie it to A, B, and E, but not to the other variables. Similarly, we have written in correlations for H that tie it to C, D, and F, but not to the other variables. Thus, we still have two factors represented by the variables in this matrix—A, B, E, and G on the one hand and C, D, F, and H on the other hand. And we can still say that the variables within each factor have a great deal more in common with each other than they have with those in the other factor.

But here we have much more diversity in the strengths of the relationships of the variables within each factor. Consider the factor represented by A, B, E, and G. If you look at the correlations for these variables, you will see that the first three of them are tied together quite strongly. Given that the correlation for A-B is .87, for A-E is .92, and for B-E is .89, we understand that there is a large amount of shared variance among this trio. Although G is related to these three, the ties are not nearly that strong. The correlations for G-A, G-B, and G-E are .61, .57, and .55, respectively. A similar situation exists for the second factor.

When we think in terms of factors, we would say that one factor is represented by variables A, B, E, and G and another factor is indexed by variables C, D, F, and H. But given the ties that bind them together, we hope you can see that G is less related to its factor than are A, B, and E, and H is less related to its factor than are C, D, and F.

This conceptual treatment is reinforced by the statistical outcome of the analysis. We present the rotated factor matrix for this two-factor solution in Table 12a.7. Once again, the results are unambiguous in that variables correlate substantially with one factor and do not correlate with the other factor. Variables G and H, because they are less related to the other variables in their respective factors as we have described above, show a relatively lower (but still substantial) correlation with their respective variate.

Examining the columns of the rotated factor matrix in Table 12a.7, we see that the factors are clearly differentiated. Factor 1 is best represented by

Table 12a.7 Two-Factor Rotated Matrix for Variables A Through H

	Factor 1	*Factor 2*
A	.95558	.04616
E	.94828	.04119
B	.93463	.08318
G	.73099	.06213
F	.03521	.95125
D	.08330	.94641
C	.05697	.92905
H	.05845	.73754

A, E, B, and G and Factor 2 is best represented by F, D, C, and H. Yet in these two sets of variables, G and H are less potent indicators of their respective variate than the other three variables in the set. When we discuss how you would interpret a factor, we will have you focus on the content of those variables that are more strongly related to the factor. Here, for example, we would focus on A, E, and B to interpret Factor 1 and on F, D, and C to interpret Factor 2.

In the research you perform, you can expect that the fuzziness we introduced into this last correlation matrix will be closer to what you will obtain than the exaggerated examples we used earlier. That is, you may find that although a number of variables are correlated with a factor to a certain extent, not all will be strongly correlated with it. Some variables will have a stronger relationship with (overlap more with, be better indicators of) the factor than other variables. Generally, we expect the variables to have different weights when we obtain the linear function that defines the factor. In the process of choosing a name for (interpreting) a factor, something that the researcher almost always has to (and wants to) do, we should allow ourselves to be more influenced by variables that are more strongly related to (correlated with) the factor than those that are more weakly related to it.

Acquiring Perspective on Factor Analysis

This final example with Variables A through H is really an oversimplified, almost-best-case illustration of the factor analysis process. We started with a set of eight variables—it would be unlikely that we would have such a small number of items if we were actually developing a test, survey, or inventory—and reduced them down to two pretty clear factors. Because this was a contrived example, it worked out to be a wonderfully clean solution. That is, we

lost very little information in the process of this data crunching. Even given the simplicity of our example, however, looking at some of the issues and implications of what went on will provide a good foundation for appreciating what factor analysis is all about.

The Importance of Individual Variables

The importance of the individual variables in a factor analysis presents us with one of the great ironies of behavioral research. It can be best illustrated in the context of test development. As you have no doubt already learned, developing good-quality items that assess the proper content of a construct is an absolute requirement of proper test development. The presence of poorly written items and items that are not germane to the topic can adversely affect the assessment process by substantially lowering the validity and reliability of the instrument.

Faced with a well-understood and well-documented content domain, the number of different items that can be generated is, if not infinite, then very, very large. Yet of the zillions of items that can potentially be written to measure a construct, only four or five dozen are likely to ever be produced, and only about two or three dozen may ever be included on the preliminary versions of the inventory. These few dozen items will be used to represent a content domain that could be composed of thousands and thousands of potential items.

We do not write thousands of items for a very good reason. There would be too much redundancy in the set to justify the effort—it is just too much work for no real gain. But it is naive to believe that the two or three dozen items we did write are completely independent of each other. By virtue of the fact that the items are all related to the same global construct, we would expect that they bear some relation to each other. Now, common sense tells us that for many of the constructs we wish to assess, there are going to be sets of items that are more related to each other than to other items in the inventory, something that we have already illustrated and that occurs time and time again.

For the sake of this illustration, assume that A through H completely represent the content domain that the inventory is designed to measure (pretend that it is a small domain). But these are eight of thousands of potential items that could have been written. Based on the content domain, Item A could just as easily have been written as AA, B as BB, and so on, where AA is a variation of A, BB is a variation of B, and so forth.

Assume that the relationships of these items are represented in Table 12a.6. If that is true, and if we had written AA instead of A, BB instead of B,

and so on through HH instead of H, then it would be reasonable to expect that AA, BB, EE, and GG would form one factor and that CC, DD, FF, and HH would form a second factor.

We could carry out this example further using AAA through HHH, AAAA through HHHH, and so on for hundreds of variations. And in each case, we would expect the same sort of factor structure to emerge. As long as all these items were of good quality, it should not matter if we used A or AA or AAA or AAAA in our inventory. As long as that part of the content domain was covered, and written well, we should be happy.

So here is the irony. Although we sweat to carefully write specific test items that are clear, unambiguous, and unitary, and although test takers are responding to these specific items, in a larger sense, the items are not important in and of themselves but are important only insofar as they represent a particular part of the content domain. Whether we had used A or AA or AAA or AAAA in combination with B or BB or BBB or BBBB and C or CC . . . , we should still find that the As, Bs, Es, and Gs would emerge as one factor and that the Cs, Ds, Fs, and Hs would emerge as a second factor.

What is important at the level of interpreting what the inventory is measuring is not the individual items (most of the time) but, rather, the fewer number of factors that underlie these items. For although we could envision writing thousands or millions or zillions of items for this content domain, they would still represent only the two general factors that describe this construct. In some sense, then, although the items are the concrete entities of the measurement enterprise, what are more "real," or certainly more useful, are the more abstract (latent) factors underlying these items.

How Much Information Is Lost in the Crunch

One of the hopes of researchers using factor analysis is that not much information is lost in the process of reducing a large number of variables down to a few number of factors. The outcome of this hope will be reflected in the quality of the solution. To the extent that variables are strongly related to their factors and not strongly related to the other factors, we say that the factor solution is a good one where not much information is lost. To the extent that the results depart from this ideal, we say that the solution is less good. We will show you how to make this assessment after covering the basics of factor analysis.

When Visual Inspection No Longer Works

We have used contrived and simplified small examples to illustrate the essence of factor analysis. If everything was always this easy, we would have

very limited use for the actual statistical procedures available on SPSS and other statistical programs. But the real world of research often uses more than eight variables, and it is always the case that the interrelationships of the variables are more complicated than presented even in our most extended example. And so it happens that we can rarely rely exclusively on visual inspection of the correlation matrix to tell us what our factors are. Fortunately, we can direct a computerized statistical package such as SPSS to process the correlations by performing mathematical routines in place of our visual inspection. We must always bear in mind, however, that at the end of this mathematical processing, we still need to interpret the results as researchers; that is, we need to interpret the results at the human level.

Factor Analysis as an Umbrella Label

A Hint of Complexity

We have been using the term *factor analysis* as a generic label to stand for a process that allows us to identify a few components, factors, or dimensions that underlie a larger set of variables. As was true for some of the other procedures that we have covered in earlier chapters, such as multiple regression and discriminant analysis, factor analysis subsumes a set of related statistical techniques; therefore, some distinctions (decisions in some of the procedures to be used) must be made when actually running such an analysis.

We need to make certain distinctions here because factor analysis is ordinarily conducted in two successive phases—the extraction phase is done first and is then followed by the rotation phase. Each phase can be accomplished using different analytic methods. In the extraction phase, one can "extract" factors via several alternative processes, or one can extract components (a simpler kind of factor that we will actually use to illustrate the extraction process). From a technical perspective, we would call what we were doing either a factor analysis or a principal components analysis depending on the analytic technique used in the extraction phase. Once components or factors have been extracted, most researchers rotate them before interpreting the results. Not surprisingly, different rotation procedures may be applied as well.

Our Approach

The most straightforward way we have found to describe the extraction process is to focus on a principal components analysis because it is a

bit simpler to explicate than a factor analysis (simpler does not imply better—many researchers prefer one of the factor analysis procedures over principal components). In this discussion, we will use the term *component* rather than *factor* as a label for the variate. Following that, we will conceptually distinguish between principal components and factor analysis, and we will briefly characterize some of the extraction methods that fall under the category of factor analysis. We will then address the rotation process, discussing both orthogonal and oblique strategies.

The First Phase: Component Extraction

The Concept of Extraction

As has been true for the other multivariate procedures that we have covered, the goal of the extraction process is to account for variance. At the start of the extraction process, all the variance is unaccounted for; when it finishes, it will have accounted for all the variance. A solution will be selected by researchers someplace between these extremes but presumably with more rather than less variance accounted for.

What is being "extracted" is variance. Think of the notion of extraction as accounting for variance. Each successively identified component will account for (extract) a certain amount of the total original variance. For this process to make sense, two questions need to be answered: (a) What is a component? (b) From what is variance being extracted?

What Is a Component?

A component is a weighted linear combination of the variables being analyzed—a variate. The mathematical function describing this straight line contains all the variables in the analysis. Each variable is weighted based on its contribution or relationship to the principal component in much the same way as is done in multiple regression. This variate is analogous to the dependent variable in a multiple regression analysis (Pedhazur & Schmelkin, 1991).

From What Is Variance Being Extracted?

Variance is being extracted from the total variance of the variables in the analysis. In principal components analysis (unlike the other multivariate procedures we have described), each variable can be thought of as contributing one unit of variance. If there are eight variables in the analysis, for example, then the total amount of variance is equal to eight.

In addition to equaling the total variance, the number of variables also equals the total number of components that it is possible to extract from the set. When all eight components are extracted from a set of eight variables, we say that the analysis has gone to completion in the sense that it has accounted for all the variance. Although SPSS will take the analysis to completion, we will wish to stop far short of that in identifying a principal components solution that we wish to interpret.

The Extraction Process

Although the extraction process is done at a mathematical level by the computer, we will explain what is happening with a pictorial illustration. This picture that we construct is, admittedly, oversimplified. Despite that limitation, however, it can capture the important features of the first few steps in component extraction. We will first put the picture with all its props in place, then tell you the rules of extraction, and finally describe the process itself.

A Pictorial Illustration

Imagine a miniature cuboid galaxy hanging in the air in front of you punctuated by a few bright "stars" here and there as drawn in Figure 12a.1. The "stars" in the picture are the variables in the analysis. The cuboid galaxy represents at least the visible part of the factor or multidimensional space defined by the positions of the variables. We have asked you to imagine a three-dimensional universe (a cube) because that is a shape that people can see in their mind's eye and a shape that we can draw. There are actually more than three dimensions in this space. In fact, there are as many dimensions as there are variables in the analysis. This multidimensional space (sometimes referred to as hyperspace) is the site of the component extraction process.

In this image of a cuboid factor space, the positions of the variables (representing the variables) relative to each other are determined by their correlations. Variables that are more strongly correlated are closer to each other. We depict this in Figure 12a.2, where the two data points relatively close to each other are shown as being moderately correlated, whereas the data points at opposite sides of the space are shown as being very weakly related to each other. This idea that correlation is a gauge of distance is important and will be elaborated upon later in our discussion.

Now include in this image a set of laser beams, each a little bit longer than our little factor space. You have complete control over where you place these lasers. These laser beams are straight lines, and they will be the components once they are positioned in this space. You have as many of these components as there are variables in this multidimensional space because the extraction process will run to completion.

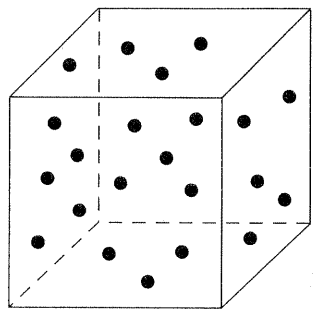

Figure 12a.1 Three-Dimensional View of Multidimensional Space

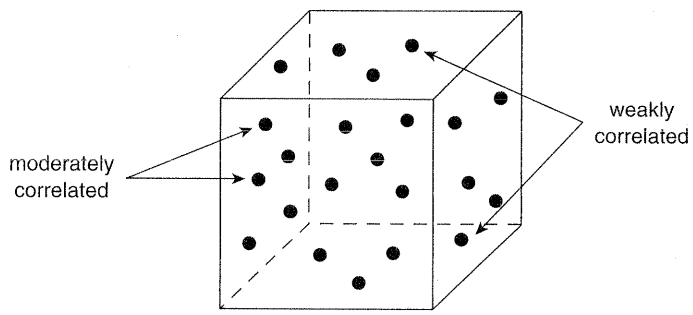

Figure 12a.2 Variables Located Closer Together Are More Correlated Than Those Located Farther Apart

The Rules of Component Extraction

Here are the rules governing the extraction of components from this multidimensional space:

1. Extraction, the fitting of lines into this multidimensional space in our pictorial illustration, is done one component at a time.
2. Every component must be thrust through the gravitational center of that space.
3. Once the first component has been thrust into the space, all succeeding components must intersect the one or ones already in the space at right (90°) angles.

4. When a component is inserted into this multidimensional space, it must account for all the variance that it is possible for it to account for, that is, it must occupy the one orientation that meets the criterion for best fit.

5. If there are already components in the space, a newly extracted component must account for variance not already accounted for by the ones already there.

Extracting Components

Do not let the terminology get in the way of your understanding of the extraction process. Principal components are fitted into the multidimensional space and represent the lines of best fit. Best fit is defined in terms of accounting for the most variance and, as we will see shortly, meets the least squares rule for fitting straight lines to data arrays.

The term *extraction* is used to convey the idea that these components, these fitted straight lines or variates, are accounting for or extracting variance from the system. Thus, we start with all the variance unaccounted for. When we fit the first component to the variables, we will account for a certain amount of variance. It is in this sense that we have extracted variance. Ultimately, when all the components have been fitted, all the variance will have been extracted—that is, accounted for.

Let's extract our first component. Using our pictorial illustration, you must position one of the components through the center of this factor space at any angle such that you come as close as possible to *all* the variables. The position that your beam should occupy is solvable once you identify the rule governing the choice of this position—namely, that there is only one place (i.e., only one linear function) where the sum of squares of the distances of the variables from the line are a minimum—all you need to do is apply the least squares rule (or permit the computer to do it for you). We have illustrated this in Figure 12a.3. Thus successful, you now have a factor space with a laser beam through its gravitational center.

This is the first extracted component and will have accounted for a certain amount of variance. In fact, because this is the very first component, it will have been positioned in the best place in the entire space. This is given to us in Rule 4 above. It will thus have accounted for (extracted) more variance than any subsequently fitted line.

We will next extract (fit) the second component (Rule 1 states that we must do this process one component at a time). Extracted components are to intersect each other at 90° according to Rule 3 above. This means that the components during this extraction phase are orthogonal to each other; that

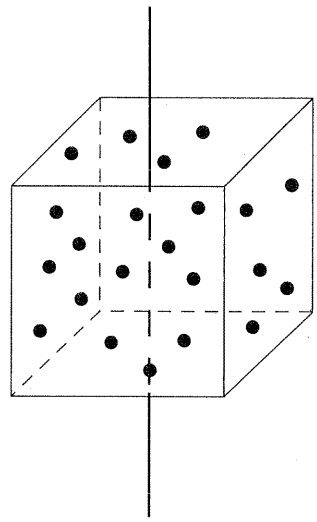

Figure 12a.3 The First Component Is Placed at the Optimal Location as Defined by the Least Squares Rule

is, they are independent of (uncorrelated with) each other. In our imaginary factor space, we will take a second laser beam and position it perpendicular to the one already in place (Rule 3) as shown in Figure 12a.4. It, too, must intersect the factor space's center (Rule 2), and it, too, must occupy the best place left to it given that the first component has the very best absolute place (Rule 4). Applying the least squares rule will yield the information necessary to place this second component appropriately.

The third component to be extracted must also intersect the factor space's center and must be perpendicular to the components (two thus far) that have been identified to this point. This component, too, must occupy the best place left to it given that the first and second components have taken the very best and next best absolute places already. This is pictured in Figure 12a.5. With the third component extracted from this component space, our imaginary factor space now has three components sticking out of it.

Unfortunately, it is impossible for us to provide you with a visual representation of the imaginary factor space once we extract more than three components because we will have run out of tangible dimensions for the universe in which we live. Through the wonders of mathematics, however, any number of dimensions beyond three can be put in place, but if the image with three straight lines made sense to you, it is not necessary to stretch your visual imaging beyond that.

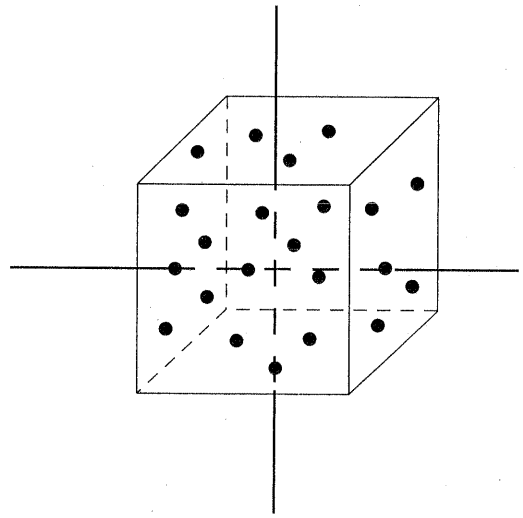

Figure 12a.4 The Second Component Is Placed in the Next Most Optimal Location Orthogonal to the First

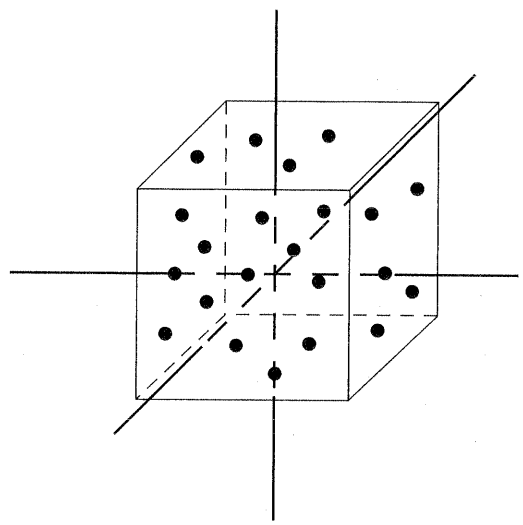

Figure 12a.5 The Third Component Is Placed in the Third Most Optimal Location Orthogonal to the First Two

Point c

One
compon
few com
compon
ance), a
is called
is to stop

Distanc
Compo

We l
of distan
we say t
are stron
idea that
ciation.
of relatic
Pictu
was done
stronger
to take t
we would
toward e
way, it ca
they bec
the corre
verbalizat

- We
(th
- Str
(th

Compon

Comp
analysis,
regressior
each com
score on

Point of Diminishing Returns

One can continue on in the manner just described, extracting as many components as there are variables. But it will also be true that once the first few components have been fit, there will not be any decent places for the other components to be positioned (i.e., they will not account for very much variance), and their existence will be superfluous. Thus, a *complete solution* as it is called is not very interesting; the real issue in principal components analysis is to stop the process when "enough" components have been extracted.

Distances of Variables From a Component Correlation as a Distance Measure

We have already indicated that correlation can be used as a measure of distance between variables. It is time to more fully discuss this idea. When we say that two variables are highly correlated, we are also saying that they are strongly related to or are associated with each other. So important is this idea that researchers have gone so far as to quantify this strength of association. As discussed in an earlier chapter, for the Pearson r, the strength of relationship is indexed by examining the r^2.

Picturing the correlation between two variables as overlapping circles as was done in Chapter 4A, the area common to them (the overlap) is r^2. The stronger the relationship, the more overlap we would draw. If we were able to take two variables and continuously make them more and more related, we would see the two circles start off with little overlap and gradually move toward each other as the strength of relation increased. In this diagrammatic way, it can be seen that the variables are indeed getting physically closer as they become increasingly more related. Thus, it is that correlation (actually the correlation squared) can be used as a measure of distance, although the verbalization of it may at first sound backward (it's not):

- Weaker correlations represent more distance between the variables (they are farther apart).
- Stronger correlations represent less distance between the variables (they are closer together).

Component-Variable Distances

Components are variates—linear combinations of the variables in the analysis, very much analogous to what we have already seen in multiple regression. We could compute a component score for each participant on each component if we so wished. This would be done by taking the person's score on the first variable and multiplying by its coefficient, taking the

person's score on the second variable and multiplying that by its coefficient, and so on, and then adding all those values together to achieve a single value for the component.

Once we had the numeric value of each component for each participant, we could then correlate the component score with the score for each variable. To compute a correlation, all we need are two quantitative values. In this situation, one value would be the score on the component (variate), and the other would be the score on the variable under consideration. Listing these pairs of numbers for each participant gives us the setup to compute a Pearson r and thereby allows us to generate the r^2 for each component-variable pair. These correlations, which we have talked about earlier, are structure coefficients.

The squared correlations—these r^2 values—are directly translatable into distance indexes. Keeping with our pictorial illustration, we have these components passing near some variables and farther from other variables. We can use the r^2 values to let us know how close each variable is to the beam (component). A large r^2 tells us that the two entities—here, the component and the variable—are quite close to each other, whereas a small r^2 tells us that they are relatively far apart.

This idea is schematized in Figure 12a.6. For simplicity, we show five data points at different distances from the component or variate (assume for this example that we have pictured the situation after the rotation phase has been completed). These distances are measured in terms of the strength of association (correlation) between the variate and the variable. As can be seen, the smallest distances represent the strongest correlations between the component and the variable, the middle distance represents a more moderate correlation, and the largest distance represents a relatively weak correlation.

Eigenvalue

Distances of the variables to the component, measured in terms of r^2, are quite important. We (well, the computer) can use these correlations to apply the least squares rule, and thus determine the proper location of the component. That is, the sum of squares of the distances of the variables to the line is minimal at the position of the component. If distance is measured in terms of r^2, then we can simply add all the r^2 values to obtain the sum of these squares for each component. The sum of these squared correlations for each component over the full set of variables is called an *eigenvalue.* In terms of the least squares rule, this optimal position where the squared distances are minimal represents the best or closest fit of the line to the data points. Thus, this position represents the place where the eigenvalue of the component is maximal.

We have illustrated the computation of an eigenvalue in Figure 12a.7, where we show again the distances between the five variables and a component

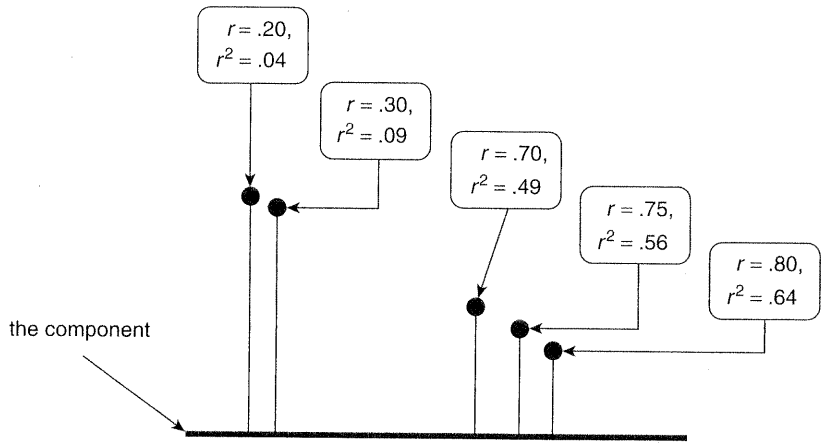

Figure 12a.6 The Squared Variable-Component Correlation Coefficient Is an Indicator of the Distance Between the Variable and the Factor

from the prior figure. In this situation, the eigenvalue of the component is equal to 1.82 and is obtained from determining the sum of the r^2 values (distances from the variate) of the variables.

Each component has an associated eigenvalue. This makes sense in that it is possible to assess the distance (again measured in terms of correlation) between any given data point and any given component. We have illustrated this in Table 12a.8, which shows the two-factor rotated factor matrix for the analysis pictured in Figure 12a.7. Assume that we have determined that there are two factors underlying these five variables. Further assume that the distances of the variables we provided to you earlier were between the variables and Factor 1 and are reproduced in the first numerical column of Table 12a.8. But because there is a second factor in the solution (intersecting the first at 90°), one can also examine the distances of the variables from that second factor as well.

The positioning of the variables on these factors is shown in Figure 12a.8. Each variable occupies a coordinate coinciding to the distance it lies from each factor. For example, Variable V correlates .80 with Factor 1 and thus lies relatively close to it, but it correlates only .05 with Factor 2 and so lies quite far from that factor. In Figure 12a.8, Variable V is placed at the (Factor 1, Factor 2) coordinate of (.80, .05).

The eigenvalue for each factor is the sum of the squared correlations of the variables in the analysis. As we have seen, the eigenvalue for Factor 1 is 1.82. Because we have supplied the correlations of the variables with the second factor in Table 12a.7, it is now possible to also calculate the

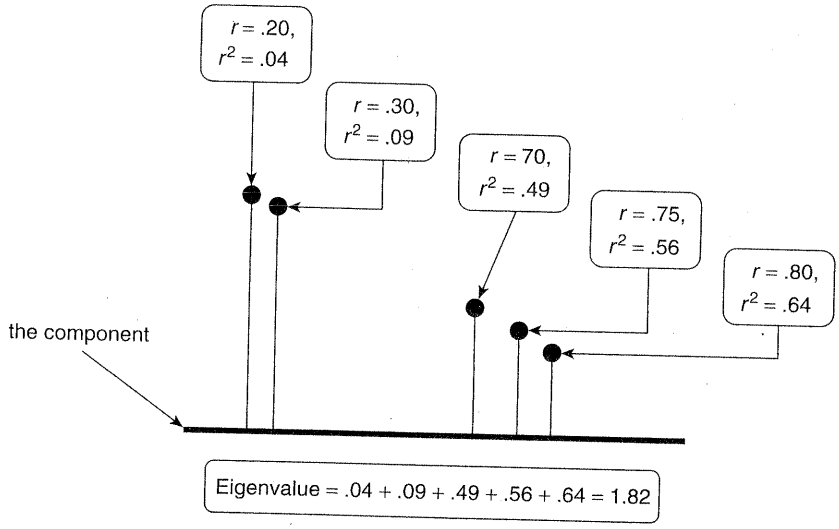

Figure 12a.7 The Sum of the Squared Correlations Is the Eigenvalue of the Component

Table 12a.8 Two-Factor Rotated Matrix for Variables V Through Z

	Factor 1	*Factor 2*
V	.80	.05
W	.75	.10
X	.70	.20
Y	.20	.85
Z	.30	.70

eigenvalue for this factor. The r^2 values for Variables V, W, X, Y, and Z are .0025, .01, .04, .72, and .49, and the sum of these values is 1.26; this is the eigenvalue of the second factor.

Principal Components Versus Factor Analysis

Causal Flow

We have indicated that the component is analogous to the dependent variable in a multiple regression analysis. This is because principal components are latent or composites descriptive of the information contained in the measured variables (the variables in the analysis). In some sense, the

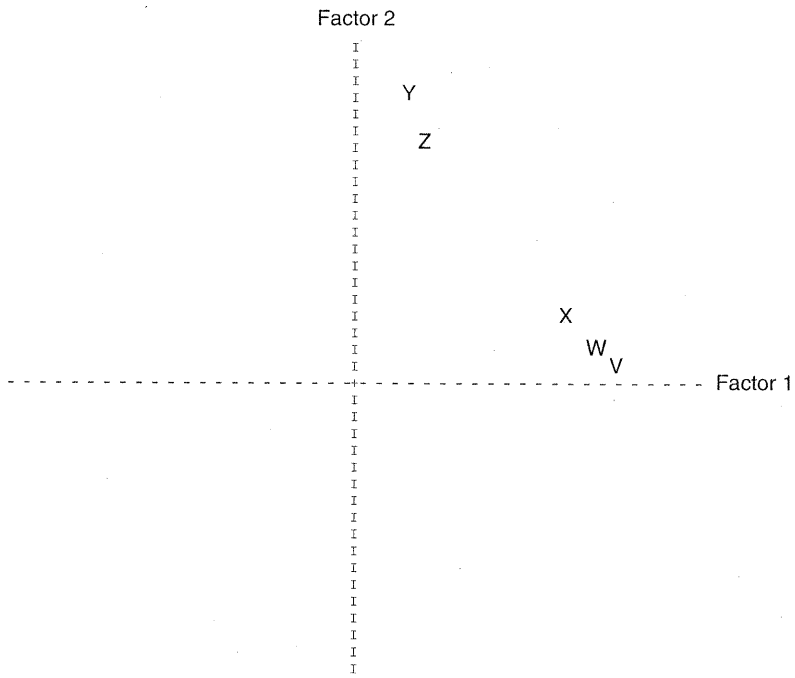

Figure 12a.8 Factor Loadings of Five Variables on Two Factors Shown in Graphic Form

components "arise from" the measured variables. From a causal modeling perspective, the causal flow is from the measured variables to the latent components. Because we think of independent variables as causes of dependent variables, the measured variables are analogous to independent variables and the components are analogous to dependent variables. Factor analysis shifts this conception around. The measured variables are taken as "indicators" of the factors. Here, the causal flow is from the factor (still a latent variable) to the measured indicator variables. Thus, the factors are analogous to the independent variables, and the measured variables are analogous to the dependent variables. This whole conception is discussed in much greater detail in Chapter 13A.

Variance Explained

Another difference between principal components and factor analysis is how each identifies the target variance that is to be explained. Principal components analysis, as we have seen, deals with the total amount of variance in the data set. We thus measured the magnitudes of the first few components

by how much of the total variance they can account for. The total variance on which we focus in principal components analysis is partitioned in factor analysis.

In factor analysis, the total variance of each variable is partitioned into two segments. One segment is common or shared (overlapping) variance—variance that is shared with the other variables in the analysis. The other segment is variance unique to the variable itself. Only the common variance is used in factor analysis. This approach is adopted because it is argued that if the point of factor analysis is to describe the structure of the set of variables in the analysis, then we should really be looking only at the shared or common variance of the variables. That is, the factor structure should be descriptive of what the variables have in common rather than including what is idiosyncratic (unique) about them. Factor analysis therefore deals only with the shared variance—the covariance—of the variables in the analysis. The magnitudes of the factors are thus measured by how much of the common variance (the covariance) they can account for.

These two perspectives partly play themselves out in how the correlation matrix is prepared for the component or factor extraction process. As we have already seen, the extraction process begins with the correlation matrix, examples of which are shown in the tables near the beginning of this chapter. The main issue at hand is the value that is to be used on the diagonal of the matrix, the place representing the correlation of each variable to itself. In those tables, we did not actually place values on the diagonal. This was done to enable the tables to be more easily read by having fewer numbers in them but also to avoid raising the issue of what to place on the diagonal before we were ready to discuss it more fully.

Students are taught in their early statistics course that a variable correlated with itself should yield a value of 1.00. This value of 1.00, however, represents all the variance (the total variance) of the variable. It is therefore an acceptable value to be used in principal components analysis because this procedure analyzes the total variance.

But using 1s on the diagonal of the correlation matrix is not appropriate in factor analysis. Rather, factor analysis will use some estimate of a variable's *communality* on the diagonal. Communality estimates are intended to represent how much a variable has in common with the remaining variables in the analysis. For example, if Variable J shares 83% of its variance with the other variables (with the remaining variance being unique to that variable), then the value that would be used for the diagonal element for Variable J in many variations of factor analyses would be .83 instead of 1.00.

Each variable in the analysis is associated with its own communality. It is the squared multiple correlation resulting from the multiple regression prediction model in which the variable is being predicted by the other variables.

It is also the sum of the squared correlations between the variable and the other variables (we square the correlations in the factor matrix and add across the rows). Communalities thus give an indication of the degree to which a variable was represented in the factor solution.

We can briefly illustrate the idea of communality, symbolized in texts as h^2, by contrasting principal components analysis with principal axis factor analysis. In the full principal components solution, each variable's communality will be equal to 1 because all its variance will have been accounted for in the solution. For a less than full solution, say a three-component solution, a variable's communality will be less than that; it will be the sum of the squared correlations in the three-component factor matrix.

In the full principal axis solution, each variable's initial communality will be less than 1 because only the common variance and not the unique variance will be considered. It is these values that are originally placed on the diagonal of the correlation matrix; it will be estimated in an iterative fashion in the extraction process until a stabilized (and somewhat lower) value is reached. Finally, in a less than full solution, a variable's communality will be computed in the same manner as was done for principal components (sum the squared correlations across the row of the factor matrix), but again because only the common variance is analyzed, these communalities will be less than those corresponding communalities computed for a principal components analysis.

Different Extraction Methods

Partitioning the variance of the variables into common and unique variance and then deciding on how to perform the extraction process have given rise to several alternative procedures developed under the aegis of factor analysis. We summarize some of the more popular ones here and mention principal components analysis as well.

- *Principal components analysis* uses 1s on the diagonal of the correlation matrix in the process of extraction and is the method described above.
- *Principal axis factoring* (also called *principal factors*) differs from principal components analysis mainly by using estimates of communalities (a measure of shared variance) on the diagonal in the extraction process. SPSS does this by starting with squared multiple correlations (a measure of the relationship between a particular variable and the set of remaining variables in the analysis—hence the term *communality*), but alternative methods (e.g., choosing the largest correlation in the row of the correlation matrix) can be used instead. Many programs such as SPSS then reestimate these

communality values via an iterative process to achieve a "better" estimate of this shared variance. Once stable estimates have been achieved, a factor analysis is performed in the manner already described above for principal components analysis.

- *Generalized least squares* factoring and *unweighted least squares* factoring both attempt to minimize the off-diagonal differences between the observed and reproduced correlation matrices (because factors are meant to represent the relationships between the variables, one should be able to perform a "back-to-the-future" operation of "reproducing" [guessing] what the original correlation matrix must have been like to begin with). Communalities are derived from the solution rather than estimated from it. The two procedures differ in that generalized least squares factoring gives greater weight to variables more strongly related to other variables in the set (those with higher communality values) whereas unweighted least squares weights all variables equally.
- *Maximum likelihood* factor extraction is related to the generalized least squares method in that it, too, uses weights proportional to the relatedness of the variables. However, maximum likelihood calculates weights for the variables on the factors that maximize the probability of having sampled the correlation matrix from a multivariate normally distributed population. Thompson (2004) contrasts the ordinary least squares approach with maximum likelihood theory by noting that rather than trying to reproduce the sample data, the maximum likelihood procedure attempts to use the sample data to directly estimate the population covariance matrix. He does go on to say that "estimating population parameters with a statistical theory that optimizes the estimation . . . is very appealing. Better estimation of population parameters should yield more replicable results" (p. 127). Maximum likelihood, as we will see in the following chapters, is extensively used in confirmatory factor analysis, path analysis, and structural equation modeling.
- *Alpha* factoring also starts with communality estimates on the diagonal. Its extraction strategy is designed to maximize the reliability (defined by Cronbach's coefficient alpha) of the factors.
- In *image factoring,* the common (shared) factor variance is estimated via multiple regression to predict the value of the variable. This predicted value is known as the "image" of the variable. The "anti-image" of the variable is the difference between the actual and predicted values. Factor extraction is performed on the image covariance matrix containing the squared images on the diagonal.

Recommendations Concerning Extraction

Based on the preceding discussion, it may seem that principal components analysis and the various forms of factor analysis might each yield vastly different results in this extraction phase. One of the pleasant surprises experienced by students when they actually perform these various procedures on the same data set is that the extraction results are quite often remarkably similar given a relatively large sample size and reasonable reliability of the measurements. The primary difference between the output of these two general approaches is that principal components analysis will account for more variance than the factor analyses. This discrepancy occurs because factor analysis targets only the common variance, which is always less than the total variance dealt with by principal components.

So what extraction procedure should you use in your own research? Some researchers (e.g., Kazelskis, 1978) have argued that principal components analysis, by focusing on the total variance, can provide a somewhat inflated solution (e.g., it will account for more variance and will show higher correlations between the variables and the factors) than the data really warrant, and the argument can be made to either correct the component correlations or to use one of the factor analysis methods. At the same time, a look through the research journals suggests that principal components analysis is an increasingly popular extraction technique (Bryant & Yarnold, 1995). Traditional wisdom, however, suggests that there is probably not going to be much difference in the interpretation based on either principal components or factor analysis, especially when you are analyzing more than two dozen or so variables that each are reasonably related to the dimensions tapped into by the components or factors (Gorsuch, 1983; Stevens, 2002).

Theoretically, principal components analysis is conceptually simpler than factor analysis. Principal components analysis attempts to summarize or aggregate sets of correlated variables and in that sense is relatively empirical (inductive). From the same standpoint, factor analysis is more complex in that factors are thought of more as causes that underlie or drive the correlations between the variables. Although both procedures are considered "exploratory," factor analysis is more related to theory development (Tabachnick & Fidell, 2001b). We will discuss the tie between factor analysis and theory more explicitly in our treatment of confirmatory factor analysis (see Chapter 13A).

Here is our general recommendation. For many applications that you are likely to encounter, principal components analysis should probably be sufficient to meet your needs. If you opt for factor analysis, our tastes run toward either generalized least squares, maximum likelihood, or principal axis solutions.

In discussing the next phase of the analysis, the rotation phase, there is little value for us to distinguish between principal components and factors. We will therefore talk in terms of factor analysis to keep our language simple, but please note that what we say is applicable to principal components analysis as well.

The Second Phase: Factor Rotation

By virtue of the mathematical procedure underlying the extraction process, the first extracted factor accounts for the most variance, the second extracted factor accounts for the next largest portion of the variance, and so on. Researchers and statisticians have argued that the mathematically based placement of these factors does not optimize the interpretability of the solution. Hence, once the number of factors to be interpreted is selected, the factors are then rotated. We mean by *rotation* the pivoting of the first n number of extracted factors around their point of intersection. The outcome of this rotation phase is what researchers ordinarily interpret.

Factor rotation is designed to achieve simple structure (we will discuss this in a moment). From the histories presented by Harmon (1960) and Harris (2001), we learn that Thurstone (1947) was the first person to explicate the concept of rotation to achieve his standards for simple structure. However, Thurstone's stringent criteria for realizing this state were unable to be satisfied by most data sets, prompting several statisticians in the early 1950s (Carroll, 1953; Ferguson, 1954; Neuhaus & Wrigley, 1954; Tucker, 1955) to develop an alternative rotational procedure known as *quartimax*.

Although quartimax was an improvement over Thurstone's method, it did not generally meet the needs of most researchers. Quartimax tended to identify one general factor that correlated with most of the variables in the analysis. This outcome was problematic to many investigators in that some of the variables captured by this general factor were not necessarily related to each other. Thus, the quartimax rotation strategy did not really distinguish subsets of related variables (i.e., separate factors) that would be of use to researchers. This problem was finally solved in 1958 when Kaiser published the results of his dissertation on another alternative strategy to factor rotation. Kaiser's publication introduced to the field the rotation strategy known as *varimax* that most researchers currently prefer.

The Goal of Simple Structure: The Least Squares Rule in Operation

The least squares rule, as used in the factor extraction process, has an awful lot to be said in favor of its use, but at least one of its properties

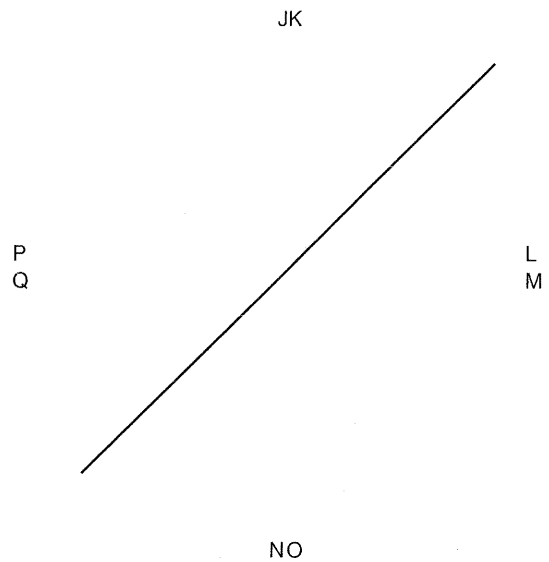

Figure 12a.9 The First Factor Fit to the Factor Space

becomes something of a liability here. This rule really represents moderation, compromise, and a lack of commitment. It is the "one solution fits all" approach that does not work in many situations. To minimize the squared distance from all the variables taken together, the line (factor) cannot afford to come too close to some variables because it would have to then be terribly far from others. In the world of squared distances, large distances become magnified in importance. Positioning of the factor thus becomes a very delicate matter.

The result of this strategy of compromise is shown in Figure 12a.9. It is a very simplified picture of the first extracted factor fitted to a set of eight variables, J through Q, before any rotation process has been accomplished. As you can see, the line of best fit is not all that far from the eight variables, but neither does it come all that near any of them.

In Figure 12a.10, we have added the second extracted factor to the mix (we are still in the extraction phase here). This second factor is also subject to the least squares rule, and it is positioned as conservatively as the first. This situation is a bit at odds with the goal of simple structure.

Achieving Simple Structure

Rotation of the extracted factor structure does not change the amount of variance accounted for but simply redistributes the variance across the

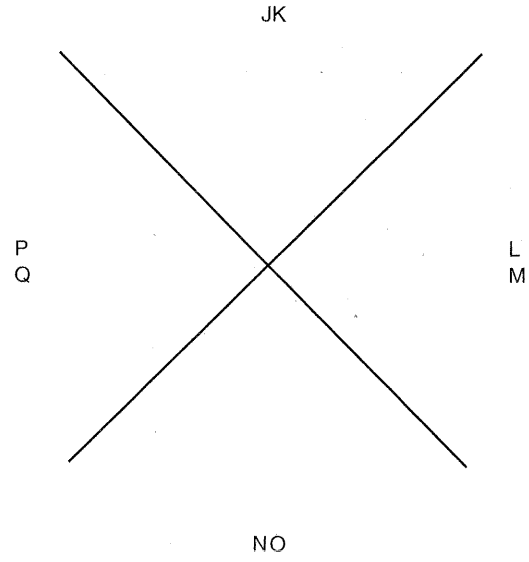

Figure 12a.10 The First Two Factors Fit to the Factor Space

factors to facilitate interpretation by trying to achieve "simple structure." Simple structure is obtained when the correlations of the variables with the factors would be either very high (values near 1) or very low (values near 0). Rotation is an iterative process of "twisting" or "pivoting" the factor structure in the multidimensional space until a satisfactory "fit" is established.

To be moved closer to some variables in this rotation process (to achieve correlations closer to 1), each factor must actually move farther away from other variables (and thus reduce those correlations closer to the target of 0). Because the factors are being rotated together, we care less about newly computed eigenvalues but more about how clear the simple structure becomes. Most researchers do not even compute the new eigenvalues of the factors once rotation has been done. In fact, we do not ordinarily keep track of which rotated factor was originally which extracted factor.

Figure 12a.10 showed the outcome after extracting the second factor from the data. For this example, assume that we have made the decision to stop the extraction process at this point and to rotate these first two factors. We show the results of this rotation in Figure 12a.11.

In the rotation process that we are using (a varimax rotation—we will describe this shortly), the factors are kept at a 90° angle of intersection with each other and so will remain orthogonal to each other after the rotation is completed (we will explain this in the following section of this chapter).

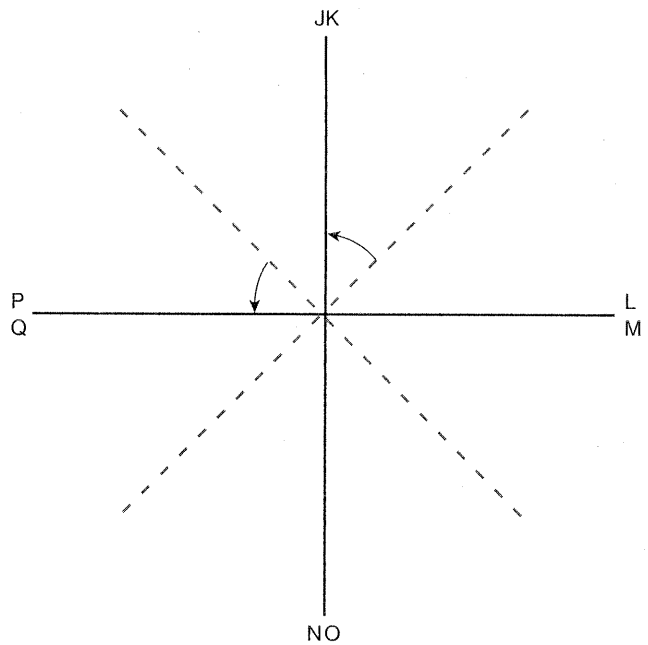

Figure 12a.11 Rotating the Two Factors to Achieve Simple Structure

In your mind's eye when you imagine this rotation process, also imagine putting some super glue on the intersection point of the factors so that when they are rotated or pivoted, they will move as a single structure.

We have used lighter dashed lines in Figure 12a.11 to show the original position of the factors and have used a solid black line to show the position of the factors following this rotation process. Each factor has been repositioned closer to four of the variables at the expense of moving away from the other four. We have not changed the total amount of variance accounted for, but that total variance has been redistributed in the process of striving to achieve simple structure. If we were interpreting the two-factor solution, it is this rotated solution with which we would be working.

We can illustrate this process numerically as well by drawing on one of our previous examples. We presented toward the beginning of this chapter (in Table 12a.3) the rotated factor matrix for the four-variable example of Items A through D, which demonstrated a two-factor structure in a numerical manner. Here, in Table 12a.9, we reproduce that information again, but this time we also show you the factor matrix prior to rotation (the factor matrix resulting from the extraction process).

Table 12a.9 Extracted (Pre-Rotated) and Rotated Factor Matrix for Variables A Through D

		Factor 1	*Factor 2*
Extraction	B	.73530	–.62913
(Pre-Rotated)	D	.73292	.62701
Factor Matrix	A	.73073	–.63204
	C	.69914	.66497
	B	.96563	.06356
	A	.96439	.05828
Rotated Factor Matrix			
	C	.03566	.96422
	D	.08635	.96065

As we can see from the top portion of Table 12a.9, the least squares solution shown for the pre-rotation result (the factor matrix resulting from the extraction phase) was truly a compromise. Variable B, for example, correlated about .74 with Factor 1 and about –.63 with Factor 2. And Variable C, as another example, correlated almost equally with both factors.

The extraction process placed the two factors such that they were moderately close to all the variables. As a consequence, although the factors were not all that far from the variables, neither were they especially close to at least some of them. Thus, all four variables are correlated with both factors, with correlations ranging from the low .6s to the low .7s.

It is now possible to appreciate why we prefer to work with the results of the rotation process. Recall the correlation matrix, the place where the factor analysis all started. In our discussion of that matrix, it was very clear that there were two viable factors represented by the four variables—one factor composed primarily of A and B and another composed primarily of C and D. But that structure is not apparent from the factor matrix following the extraction phase. Even though Variables A and B are not related to Variables C and D, all four are substantially correlated to each of the factors. The extraction process, driven by the least squares rule, has succeeded in having all the variables as related as possible to all the factors but has not differentiated sets of variables for us.

The rotation phase picks up from where the extraction process leaves off. Once we identify the number of factors with which we want to work, we abandon the least squares rule. Instead, we allow the rotation process to be driven by the goal of simple structure so that the relationships between the variables can reemerge from the blandness of ordinary least squares.

You can see in the bottom portion of Table 12a.9 the dramatic structural change as a result of the rotation process. The factors were given an opportunity to make a commitment—to achieve simple structure. Factor 1 was thus placed quite close to (it was allowed to correlate quite strongly with) A and B, and Factor 2 was thus placed quite close to C and D. Of course, in the process of committing to A and B, Factor 1 wound up much farther away from C and D. That turned out to be acceptable, however, because the position of Factor 2 worked out well to complement that of Factor 1. Thus, the rotation process accomplished the goal of achieving simple structure and our effort to interpret the factor analysis results was thus facilitated.

Although the factors were repositioned in the factor space, the total amount of variance accounted for in this two-factor solution remained unchanged. This can be easily verified by simply computing the sum of all the squared pre-rotated correlations and comparing them with the sum of all the squared rotated correlations. If you do the math, both sets of squared correlations sum to the value of 3.73669.

Orthogonal Factor Rotation

The term *orthogonal* is a statistical concept indicating that two or more entities are independent of (not correlated with) one another. An orthogonal rotation strategy keeps the factors independent of each other during the rotation process. Geometrically, factors are orthogonal if they cross each other at 90°. In Figure 12a.11, we retained the 90° angle between the factors when we rotated them, and in all the previous examples of rotated factor matrices involving Variables A through H, the results were based on the orthogonal rotation strategy known as varimax.

Three methods of orthogonal rotation are available in most statistical packages:

- *Varimax* rotation, introduced by Kaiser (1958) as an alternative to the quartimax strategy, simplifies the correlations within each factor (the columns of the factor matrix) by striving toward 1s and 0s. It simplifies the factors (the columns of the factor matrix). Varimax rotation works toward having some factors correlate quite strongly with some variables (correlations nearer to 1) but more weakly with the other variables (correlations nearer to 0). Because it is focused on the factors (as are the researchers most of the time), this method is the most frequently used orthogonal rotation strategy.
- *Quartimax* rotation, resulting from the work done in the early 1950s, simplifies the variables (the rows of the factor matrix). It does this by having a variable correlate more strongly (nearer to 1)

to one factor and more weakly (nearer to 0) to all other factors. As we have already described, this strategy tends to drive the rotated solution toward a single general factor. Quartimax rotation is infrequently used today not only because it is variable- rather than factor-oriented but also because researchers do not often prefer one dominant factor.

- *Equimax* rotation attempts to effect a compromise between varimax and quartimax rotational strategies; it, too, has failed to electrify the interests of most researchers.

Oblique Factor Rotation

The key difference between orthogonal and oblique rotation is that oblique rotation does not require the rotation process to keep the factors uncorrelated. In geometric terms, the mandate that the factors remain at 90° is no longer in effect. Instead of super gluing the factors together as we had you imagine for orthogonal rotation, here we suggest that you form an image of the factors connected to a hinge at their point of intersection. This would allow the factors to be pivoted if need be to form an oblique angle (hence the term *oblique rotation*) rather than being forced to remain perpendicular to each other.

There are two oblique rotation methods available in SPSS: *Direct Oblimin* and *Promax*. The former is the more processor-intensive procedure and is the one that you should probably use in your own research if you are doing an oblique solution; it is the one we focus on in this text. Promax is a simpler and faster method to obtain an oblique solution and should probably be considered for use if you need to save processing time when you have a very large data set.

In the Direct Oblimin rotation, the amount of correlation permitted between factors is under the control of the researcher, although most computer programs such as SPSS have default values that are usually pretty acceptable (SPSS uses the term "delta value") for how much correlation is permitted if the researchers do not specify a value (SPSS's default is a delta value of zero).

To permit a particular degree of correlation to exist between factors is not to require that the factors achieve that much correlation. Rather, the correlation value is a ceiling or maximum that is permitted should the factors exhibit that much relationship. The idea here is to permit the factors some leeway so that if they do correlate somewhat, the factor structure would be allowed to incorporate that in the solution.

So why set a limit at all? Because the situation is similar to the hands of an analogue clock where the minute hand can come closer and closer to the

hour han
placed ex
or 10°, or
each othe
the factor
We ha
and 12a.1
the factor
pattern. T
achieve th
As sh
these two
factors int
(or 45° if
those of y
X is the ra
is the hyp
the cosine
crossing a
pendent,
because th

Figure 12a.

hour hand and at some point overlap it. If two factors are allowed to be placed extremely close to each other, perhaps intersecting each other at 5° or 10°, one can question whether the two factors are all that different from each other. It is therefore useful to set some sort of limit, then, to prevent the factors from being located too close to each other.

We have provided an illustration of an oblique rotation in Figures 12a.12 and 12a.13. Figure 12a.12 shows variables labeled as A through H arrayed in the factor space, which we have divided into quadrants to help you see the pattern. Take the two lines as factors that are being obliquely rotated to achieve the best simple structure they can.

As shown in Figure 12a.13, in this simplified example, the best fit of these two factors to the eight variables in the analysis is obtained when the factors intersect at about 135° if you look at the oblique angle of intersection (or 45° if you look at the acute angle of intersection) rather than 90°. For those of you who remember geometry, in a right triangle, the cosine of angle X is the ratio of A to B where A is the side that is adjacent to angle X and B is the hypotenuse. It turns out that the correlation of the factors is equal to the cosine of the angle formed by the two factors. In the case of the factors crossing at 90°, we have learned that these factors are orthogonal (independent, uncorrelated). Well, the cosine of 90° is zero (which it has to be because the factors are uncorrelated in this configuration, i.e., $r = 0$).

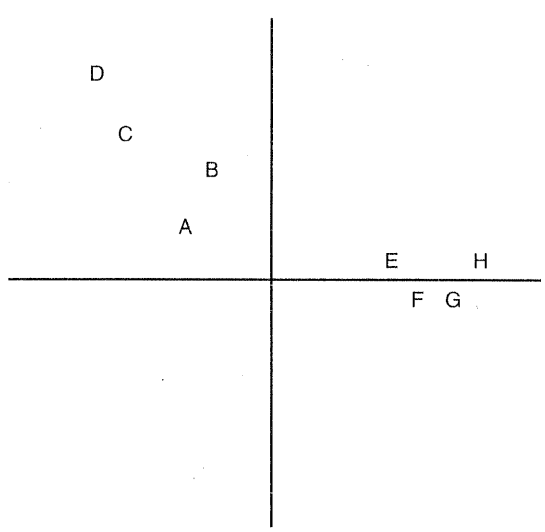

Figure 12a.12 An Array of Variables Shown in Two Dimensions

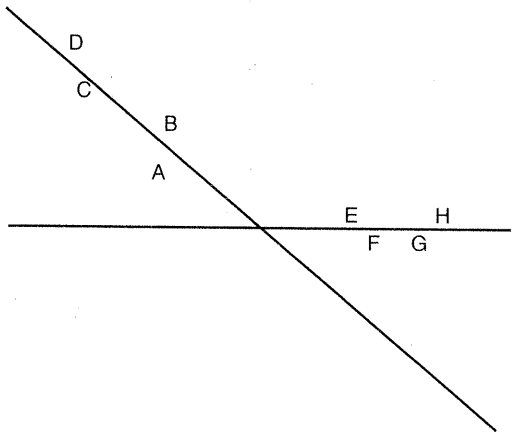

Figure 12a.13 Oblique Rotation of Factors Can Sometimes Achieve Simple Structure Better Than Orthogonal Rotation

Choosing Between Orthogonal and Oblique Strategies

In most studies, researchers present the results of just one of the factor rotation strategies. Very often, the decision as to which one is more appropriate is made on the basis of the theory or model from which the study was generated. That is, ideally, the model has indicated whether the factors should be correlated or not, and that direction may well be sufficient for researchers to commit to one or the other strategy.

In the absence of any solid expectations, it is not uncommon for researchers to initially perform an oblique rotation solution. This allows them to examine the factor correlations that are provided as part of the oblique rotation output. Based on these computed factor correlations, researchers determine whether an oblique or orthogonal rotation strategy is more appropriate for their analysis.

We can supply you with a very general rule of thumb to start your decision-making process if you need to decide between oblique and orthogonal rotations based on the statistical analysis: With factor correlations generally in the range of the high .3s or better, most researchers would probably opt to work with an oblique rotation; with factor correlations in the teens or less, many researchers would probably opt to work with an orthogonal (most likely varimax) solution; with factor correlations between these two ranges, the decision is less clear cut and may vary substantially among researchers.

Reading Features of the Factor Analysis Printout

Although we will cover this in a more hands-on way in our worked example in Chapter 12B, there are certain parts of the printout that have immediate use helping researchers accept a particular solution to the factor analysis. We will briefly cover these features here by returning to the context of our simplified and hypothetical eight-variable example with the Variables A through H.

Variance Accounted For

As we have seen, eigenvalues indicate the amount of variance each component has accounted for. For the extraction phase, the first component will have the largest eigenvalue, the second component the next largest eigenvalue, and so on. This is one of the first portions of the output that researchers will examine. You will find a table in any printout that you examine that looks similar to what we have presented in Table 12a.10. The example used in this table shows the factor analysis solution of Variables A through H from an earlier example in this chapter. The correlation matrix for these variables, which you may wish to review, was presented in Table 12a.6.

Because there are eight variables in the analysis, the total amount of variance is equal to 8. After the first factor is in place, we can compute the correlation between each variable and the factor. We then square these correlations and sum them to achieve the eigenvalue of the factor. Eigenvalues are directly interpretable in terms of variance accounted for.

In Table 12a.10 we see that the first factor has an eigenvalue of 3.64341; that is, the first factor has accounted for about 3.64 units of variance. Because the total amount of variance is 8, dividing 3.64 by 8 tells us the proportion of the total variance extracted by the first factor. That turns out to be 45.5%, and this is shown in the next column of Table 12a.10. By the same reasoning, the second factor accounts for 35.2% of the total variance. Together, as shown in the last column of the table, the first two factors have accounted for almost 81% of the variance.

As shown in Table 12a.10, the extraction process has gone to completion in the sense that all eight factors have been extracted. Together, these eight factors have cumulatively accounted for all the variance. Because the usual goal in such an analysis is to account for the most variance with the fewest number of factors, we certainly want to work with fewer than all the eight factors extracted here.

Although this solution is a bit exaggerated, it does demonstrate some valid and important points.

Table 12a.10 Eigenvalues and Variance Accounted for in the Eight-Variable Factor Analysis of Variables A Through H

Factor	*Eigenvalue*	*% Variance Accounted For*	*% Cumulative Variance*
1	3.64341	45.5	45.5
2	2.81393	35.2	80.7
3	.58015	7.3	88.0
4	.53570	6.7	94.7
5	.19279	2.4	97.1
6	.10758	1.3	98.4
7	.07462	.9	99.4
8	.05181	.7	100.0

- The first factor always accounts for the most variance, the second factor for the next most variance, and so forth.
- Our focus is almost always on the cumulative percentage of variance accounted for. Here, the first two factors accounted for an incredible 80.7% of the variance.
- It is not an issue here, but factors that cannot muster enough strength to achieve an eigenvalue of 1 are ordinarily not going to be part of the solution. That is, researchers are not usually going to go that far down the list of possible factors to select.
- Researchers have the desire to work with factor solutions that appear to capture enough of the variance to provide confidence that they did not lose too much information in the data crunching process. A very rough rule of thumb proposed by some authors (e.g., Tabachnick & Fidell, 2001b) is that the solution should account for at least 50% of the variance.
- The factor solution is cumulative starting with the first factor. When we select a factor solution, what we are really saying is that we select the first n number of factors.
- In the example, we would select the two-factor solution as most viable. We saw this in the correlation matrix shown earlier in this chapter. As you can see from Table 12a.10, we gain a chunk of variance by reaching down to the second factor. The third factor is not contributing an awful lot of explained variance and its eigenvalue is below the value of 1.

Eigenvalue Plot

Another part of the printout that we would want to examine is the plot of eigenvalues. It is a visual depiction of the first two columns of the table

illustrated in Table 12a.10. In real-world data sets where the solution is not oversimplified by textbook writers, researchers can sometimes see the pattern more clearly than in the table. The idea behind this graphical way to determine the number of factors in a solution was introduced by Raymond B. Cattell in 1966.

In an eigenvalue plot, eigenvalues are represented on the *Y* axis and the factors are represented on the *X* axis. An idealized curve might look like a negatively decelerating curve (similar to Ebbinghaus's forgetting curve). This plot of eigenvalues is known as a *scree distribution* or *scree plot.* Generally, as one part of the decision process aimed at determining the number of factors there should be in the solution, one looks for where the scree plot is still reasonably dropping (factors are still contributing to variance accounted for in potentially meaningful ways) and where it starts to straighten out (the point of diminishing returns where no gain is made by choosing additional factors).

The scree plot for our eight-variable example is presented in Figure 12a.14. As can be seen in Figure 12a.14, by the time we reach the third factor the function is pretty much attenuated, visually suggesting that the two-factor solution is probably the most viable.

An example of a somewhat more complex scree plot is shown in Figure 12a.15. We clearly make gains from the first to the second factor and from the second to the third factor. There may some gain by incorporating the fourth factor and a bit more with the fifth factor, but it is not clear from the scree plot if these gains are both worthwhile. The sixth and later factors are probably not strong enough to warrant inclusion in the final solution.

In this example, we would thus be inclined to nominate the two-factor through five-factor solutions as candidates for solution, with the most likely number of factors in the final decision probably being either three or four based on this scree plot. We would therefore want to examine the rotated factor matrix for each alternative solution, determine which variables were associated with which factors, and generate a label for each factor in each analysis. Based on several considerations discussed later, we would single out one of these sets as our solution.

The Rotated Factor Matrix

Varimax Rotated Factor Matrix

What is finally interpreted by researchers is the rotated factor matrix. Using our simplified eight-variable example, this matrix, originally shown in Table 12a.7, is once again reproduced in Table 12a.11. These results represent the varimax rotation of the two factors.

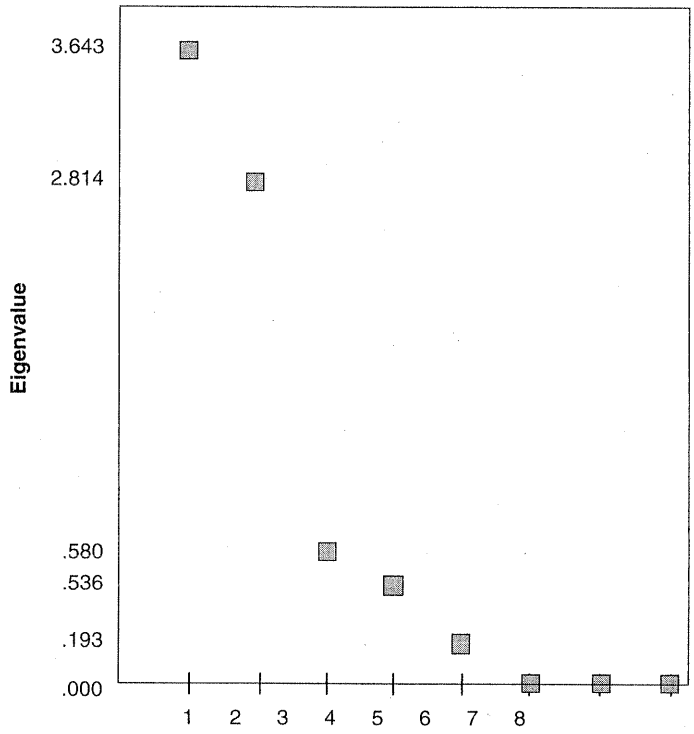

Figure 12a.14 A Scree Plot of Eigenvalues

Looking down the columns, we can see that varimax rotation has done a good job in achieving simple structure (of course, we constructed this example to exaggerate this point). With a couple of exceptions as we note below, the correlations of the variables to the factors are either relatively high or relatively low.

The Magnitude of the Variable Weights

In the context of a rotated factor matrix, the attention of researchers is focused on the magnitude of the structure and/or pattern coefficients (these are the correlations in an orthogonal rotation and are at the same time the structure and pattern coefficients; as we will see, in an oblique rotation these coefficients are different). Many researchers call these values by the colloquial and generic label of *loadings*. The key question here is how large a coefficient or "loading" should be before we accept it as sufficiently relating to the factor to commit ourselves to that interpretation. Comrey and Lee

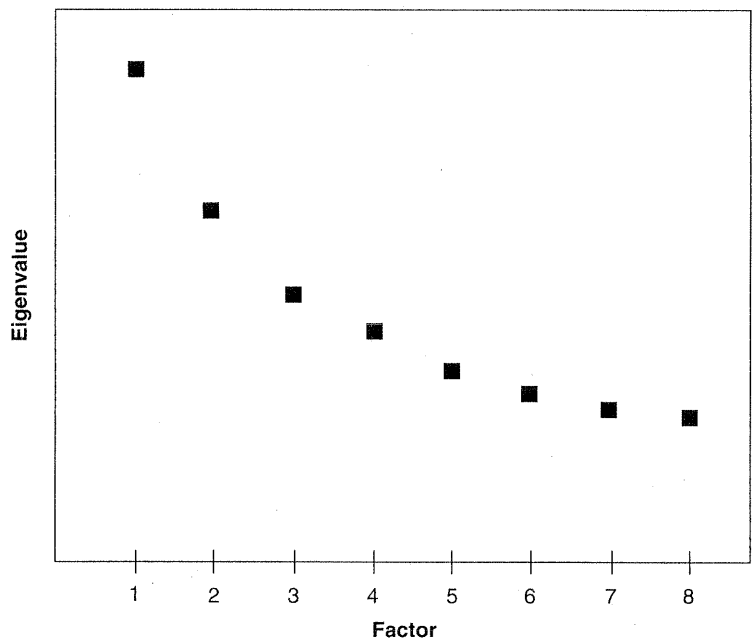

Figure 12a.15 Another Somewhat More Complex Scree Plot

(1992) have characterized coefficients of .7 as excellent, .63 as very good, .55 as good, .45 as fair, and .32 as close to minimal. Few if any authors recommend going below about .3. Tabachnick and Fidell (2001b) have argued for .32 as rock bottom, Gorsuch (1983) has suggested that .3 may be too small, especially if there are many variables in the analysis, and Stevens (2002) has recommended .40 in order to achieve practical worth.

Table 12a.11 Varimax Rotated Factor Matrix for the Two-Factor Solution of Variables A Through H

	Factor 1	*Factor 2*
A	.95558	.04616
E	.94828	.04119
B	.93463	.08318
G	.73099	.06213
F	.03521	.95125
D	.08330	.94641
C	.05697	.92905
H	.05845	.73754

Our recommendation is to base your decision in part on your sample size. With less than about 200 cases, think in terms of values in the .40 to .45 range as minimal. With larger samples, we suggest using about .40 as your rough guide, but variables with coefficients or loadings in the high .3s might be captured as well with a clear factor structure.

These recommendations address the interpretation of a rotated factor solution but may be too lenient if you wish to build a composite subscale based only on the subset of variables that you define as loading on the factor. At least under certain circumstances, variables with coefficients much below .5 can actually reduce the reliability of scales, especially those built from a combination of close to a dozen variables.

Interpreting Factors

For the purposes of this example, we will accept a variable as related to a factor if it has produced a structure or pattern coefficient of .4 or better. The interpretation process starts by examining the variables in Table 12a.11 row by row. For each row where we get a relatively high coefficient, we mark the result. Thus, we would say that A and B both "load" on Factor 1, C and D load on Factor 2, E loads on Factor 1, and F loads on Factor 2. Variables G and H are not associated with such extraordinary high coefficients but certainly correlate strongly enough on their respective factors to be included in the interpretation of the solution.

Interpretation of the solution is now possible to make by examining the columns. Recall two points: (a) A factor or component is what the variables correlated with it have in common; (b) variables more highly correlated with the factor or component share more in common with it. Thus, Factor 1 can be identified by emphasizing the content of A, E, and B, and Factor 2 can be identified by emphasizing the content of F, D, and C.

To illustrate this idea, assume that A, E, and B were the following items on an inventory:

I try to avoid being in a crowd.

I tend to stay away from parties.

I don't enjoy social events.

These three variables all correlate relatively strongly with the factor. What is the factor? One way to approach this is to think of the game show where you are given specific instances and must come up with the general category. In this case, the instances are as follows: avoid crowds, stay away from parties, and dislike social events. So what is the category? It is something like social avoidance, and thus "social avoidance" is the interpretation of the factor.

Negative Factor Coefficients

This illustration, by the way, is based on Scale 0, Social Introversion, of the MMPI-2. Although we have given the essence of the items rather than reproducing the items verbatim, it is worth noting that the second and third items are reverse worded on the inventory. The items would thus be something like the following:

I try to avoid being in a crowd.

I like parties.

I enjoy social events.

If the factor is truly social avoidance, then the last two items would be negatively correlated with the factor. That is, those who were more socially avoidant would try harder to avoid crowds (positive correlation) but would be less enthused with parties (negative correlation) and would enjoy social events to a lesser extent (negative correlation). The variable weights on the factor would look like the following:

Variable	*Factor 1*
A	.95558
E	−.94828
B	−.93463

The negative values would simply indicate that the relationship between the variable and the factor is an inverse one. Remember that a correlation of .94 represents the same strength of relationship as −.94 because r^2 is always positive. Positive and negative structure or pattern coefficients simply reflect the way the variable is measured (e.g., the way an item is worded); stronger relationships, whether direct or inverse, still reflect more association on the factor.

Obliquely Rotated Factor Matrix

In an orthogonal rotation, the values in the rotated factor matrix represent both the correlations (structure coefficients) of the variables with the factors and the weights of the variables (pattern coefficients) in the linear function defining the variate (factor). The structure coefficients are equal to the weights in an orthogonal rotation precisely because the factors are independent of each other.

In an oblique rotation strategy, the factors are permitted to be correlated (and virtually always are, at least a little). Because the factors are no longer independent of (orthogonal to) each other, the weights of the variables in the linear function (variate, factor) no longer must be equal to the structure coefficients. Therefore, instead of the single rotated factor matrix produced by an orthogonal rotation strategy, an oblique rotation strategy generates two somewhat different rotated factor matrices, one presenting the weights of the variables in the linear function (pattern matrix) and one presenting the correlations of the variables with the factors (structure matrix).

- The *pattern matrix* presents what are called pattern coefficients. Pattern coefficients are the weights that have been assigned to variables in the linear functions (variates, factors). This is analogous to what is done in regression, and pattern coefficients are essentially multiple regression beta weights. To the extent that the factors are relatively strongly correlated these coefficients can (and occasionally do) go beyond the range of ±1.
- The *structure matrix* presents the structure coefficients. Structure coefficients depict the correlations between the variables and the factors (variates). They are true correlations and are therefore bounded by the range of ±1. Generally, these values tend to be just a bit higher than those of the pattern coefficients (except when the factors are very highly correlated), but using these coefficients in interpreting the factor solution should often result in the same characterization of the factors.

Recommendations

It may not surprise you to learn that different authors have offered alternative viewpoints on which obliquely rotated factor matrix researchers should use for their interpretation and hence which they should report to readers. Hair et al. (1998), Mertler and Vannatta (2001), and Tabachnick and Fidell (2001b) have proposed that the pattern matrix be used. On the other hand, Thompson (1984, 1991) has stressed the importance of the structure coefficients in interpreting classical parametric analyses, and Courville and Thompson (2001) have argued that the structure coefficients are important in factor analysis as well as in regression. In addition, Diekhoff (1992), Gorsuch (1983), and Pedhazur and Schmelkin (1991) urge that both structure and pattern coefficients be considered in interpreting factors.

Although it is possible that the pattern and structure matrices would provide somewhat different impressions about the nature of the underlying factor structure, it is frequently the case that the two pictures are relatively

consonant with one another. That is, although the pattern coefficients may be somewhat larger if the factors are highly correlated and the structure coefficients may be somewhat higher if the factors are not highly correlated, they will very often result in a very similar interpretation. Thus, variables easily meeting your criterion for being accepted as relatively strongly representing a factor (e.g., variables loading in the range of .70 where your loading criterion is .40) are likely to emerge on the same factor in both the pattern and structure matrices. We do urge you to examine both matrices to determine if this sort of an outcome holds for your particular data analysis.

For the purposes of this discussion, we will assume that both the pattern and structure matrices yield similar interpretations. Because both ways of viewing the results are viable and useful, the choice as to which rotated factor matrix to present probably comes down to the language used in articulating the framework from which the research was derived. That is, if the theoretical framework conceptualized the factor loadings in terms of weights or in terms of correlations to the variate, then that is very likely the language you would want to maintain and that, in turn, tells you which of the two matrices you would present in communicating your results.

If that framework is neutral on which viewpoint is more reasonable to take, then your choice of which matrix to present may be driven by how comfortable you are in talking about weights versus structure coefficients. It seems to us that somewhat more researchers report the pattern matrix, and that alternative is a perfectly reasonable reporting strategy to follow. But the structure matrix presenting the variable-factor correlations may sometimes turn out to be a little clearer and a little easier to discuss and might therefore be considered for reporting on those occasions.

Selecting the Factor Solution

As you can deduce from our discussions in this chapter, there is no agreement on a single factor extraction method to use. And unless the factors are substantially correlated, in which case you should really use an oblique rotation, it is often a close call as to whether to report a varimax or an oblique rotation. Thus, you cannot follow a prescribed formula to achieve your factor analytic solution. In one of the editions of his classic testing textbook, Lee Cronbach (1970) said it very plainly: "There is no one 'right' way to do a factor analysis any more than there is a 'right' way to photograph Waikiki Beach" (p. 315). Cronbach's words are still true today.

The solution finally accepted by researchers is an informed but ultimately a subjective choice. Tabachnick and Fidell (2001b) suggested that at least for analyses containing up to about 40 variables, the number of factors whose

eigenvalue is above 1.00 can be estimated in advance by dividing the number of variables by 3 to 5. In the majority of analyses, however, researchers are unlikely to accept that many components or factors in their final decision.

The choice of the factor solution is based on multiple features of the results, including the following:

- *The theoretical and empirical milieu.* Research is carried out in a context that allows researchers to speculate in advance of data collection what some of the themes underlying the data might be. Thus, although the guidelines here may appear quite data driven, they must always be taken with respect to what the researchers already know about the content domain under study.
- *The plot of the eigenvalues.* You do not want to get into the region of diminishing returns and will most likely select a solution before you reach that point.
- *The amount of variance accounted for by different solutions.* Accepting more factors will allow you to account for more of the variance, but at some point you may be identifying more factors than is reasonable for the construct. You can begin to see this when you find that the content of and the labels for a couple of the factors are difficult to tell apart. At that point, the increased amount of accounted-for variance is probably doing more harm than good.
- *The number of variables used to represent factors.* It is appropriate to have a decent number of variables taken as illustrating the factor. This is true for at least two reasons: (a) You want to be clear what the factor is, and more variables loading on the factor give you more information on which to base your interpretation, and (b) if you plan to create subscales based on the factor structure, you need enough variables that are highly correlated to the factor to achieve an acceptable level of reliability. Although it is difficult to provide an exact number of variables that will always meet these criteria, generally we can say that four or five items per factor is usually as small a count as you would ordinarily want to have.
- *The strength of the coefficients.* Even if you use a lenient loading criterion of .40, variables correlated with the factor near this lower limit are clearly not strongly related to the factor. Ideally, you should have enough variables in the .7 range or higher to not worry about bringing in variables that are in the .4s.
- *The reasonableness of the interpretation.* If you have to articulate a bottom line for characterizing what researchers finally select, this feature is it. As characterized by Johnson and Wichern (1998),

ultimately, the selection of a factor structure is not judged on any single quantitative basis but rather on what they call a "Wow" criterion: "If, while scrutinizing the factor analysis, the investigator can shout 'Wow, I understand these factors,' the application is deemed successful" (p. 565). As it turns out, reasonableness emerges from the other considerations already mentioned. That is, solutions accounting for relatively small percentages of the variance (e.g., less than 40%) will tend not to have many variables with strong loadings showing up on the factors and are unlikely to be conducive to clear and forthright interpretation.

One strategy that we have found to be helpful in selecting the factor structure that is most reasonably interpreted is trying out a small range of factor solutions based on the scree plot and the amount of variance accounted for. Our intent in this process is to try to "surround" the solution that is most likely to be accepted (e.g., if we guess that either the three- or four-factor solution is going to be eventually selected, we would run the two-, three-, four-, and five-factor solutions). This process can be thought of as laying out the evolution of the factor structure, and it very often helps researchers with factor interpretation by making them aware of which variables have formed stable groupings over the multiple solution set. The bottom line, however, always rests squarely on the sensibility of the interpretation of the factor structure as formulated by the researchers.

Recommended Readings

Conway, J. M., & Huffcutt, A. I. (2004). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods, 6,* 147–168.

Eysenck, H. J. (1952). The uses and abuses of factor analysis. *Applied Statistics, 1,* 45–49.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4,* 272–299.

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

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods, 7,* 191–205.

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

Pett, M. A., Lackey, N. R., & Sullivan, J. L. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research.* Thousand Oaks, CA: Sage.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology, 15,* 201–293.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications.* Washington, DC: American Psychological Association.

Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychological Methods, 1,* 254–365.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99,* 432–442.