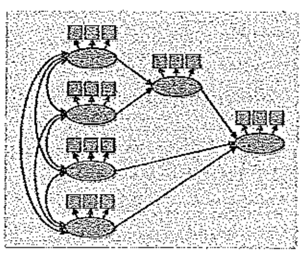
A Step-by-Step Approach to Using the SAS® System for Factor Analysis and Structural Equation Modeling



1994

CARY, NC: SAS INSTITUTE

Larry Hatcher, Ph.D.

Introduction: When Is Exploratory Factor Analysis Appropriate?

Exploratory factor analysis may be appropriate when you have obtained measures on a number of variables, and want to identify the number and nature of the underlying factors that are responsible for covariation in the data. In other words, exploratory factor analysis is appropriate when you want to identify the factor structure underlying a set of data.

For example, imagine that you are a political scientist who has developed a 50-item questionnaire to assess political attitudes. You administer the questionnaire to 500 people, and perform a factor analysis on their responses. The results of the analysis suggest that, although the questionnaire contained 50 items, it was really just measuring two underlying factors, or constructs. You decided to label the first construct the **social conservatism** factor. Individuals who scored high on this construct tended to agree with statements such as "People should be married before living together" and "Children should respect their elders." You chose to label the second construct the **economic conservatism** factor. Individuals who scored high on this factor tended to agree with statments such as "The size of the federal government should be reduced" and "Our taxes should be lowered."

In short, by performing a factor analysis on responses to this questionnaire, you were able to determine the number of constructs measured by this questionnaire (two) as well as the nature of those constructs. The results of the analysis showed you which questionnaire items were measuring the social conservatism factor, and which were measuring the economic conservatism factor.

The use of factor analysis assumes that each of the observed variables being analyzed are measured on an interval or ratio scale. Some additional assumptions underlying the use of factor analysis are listed in an appendix at the end of this chapter.

Note: You will see a lot of similarity between the issues discussed in this chapter and those discussed in the preceding chapter on principal component analysis. This is because although there are conceptual differences between principal component analysis and exploratory factor analysis, there are also many similarities in terms of how the procedures are conducted. Some of these differences and similarities are discussed in a later section, "Exploratory Factor Analysis versus Principal Component Analysis."

It is likely that some users will read this chapter without first reviewing the chapter on principal component analysis; this made it necessary to present in this chapter much of the material that was already covered in the principal component chapter. Readers who have already covered the principal component chapter should be able to skim over this material fairly quickly.

Introduction to the Common Factor Model

Example: Investment Model Questionnaire

Exploratory factor analysis will be demonstrated by performing a factor analysis on fictitious data from a questionnaire designed to measure investment model constructs (Rusbult, 1980). The investment model was introduced in the preceding chapter; remember that this model describes certain constructs that affect an individual's **commitment** to a romantic relationship, that is, his or her intention to maintain the relationship. Two of the constructs that are believed to influence commitment are alternative value and investment size. **Alternative value** refers to the attractiveness of a person's alternatives to his or her current romantic partner. For example, a woman would score high on alternative value if it would be appealing for her to leave her current partner for a different partner, or simply leave her current partner and be unattached to anyone. **Investment size** refers to the time or personal resources that a person has put into his or her relationship with the current partner. For example, a woman would score high on investment size if she has invested a lot of time and effort in developing her current relationship, or if she and her partner have many mutual friendships that would be lost if the relationship were to end.

Imagine that you have developed a short questionnaire to assess alternative value and investment size. The questionnaire is to be completed by persons who are currently involved in romantic associations. With this questionnaire, items 1-3 were designed to assess investment size, and items 4-6 were designed to assess alternative value. Part of the questionnaire is reproduced here:

```
Please rate each of the following items to indicate the extent to
which you agree or disagree with each statement.  Use a response
scale in which 1=Strongly Disagree and 7=Strongly Agree.

_____   1. I have invested a lot of time and effort in developing
           my relationship with my current partner.
_____   2. My current partner and I have developed interest in a
           lot of fun activities that I would lose if our
           relationship were to end.
_____   3. My current partner and I have developed a lot of mutual
           friendships that I would lose if our relationship were
           to end.
_____   4. It would be more attractive for me to be involved in a
           relationship with someone else rather than continue in
           a relationship with my current partner.
_____   5. It would be more attractive for me to be by myself than
           to continue my relationship with my current partner.
_____   6. In general, my alternatives to remaining in this
           relationship are quite attractive.
```

Assume that this questionnaire was administered to 200 subjects, and their responses were keyed so that responses to question 1 were coded as variable V1, responses to question 2 were coded as variable V2, and so forth. The correlations between the six variables are presented in Table 2.1.

Table 2.1.

Correlations between Questions Assessing Investment Size and Alternative Value

	Intercorrelations					
Question	V1	V2	V3	V4	V5	V6
V1	1.00					
V2	.81	1.00				
V3	.79	.92	1.00			
V4	-.03	-.07	-.01	1.00		
V5	-.06	-.01	-.11	.78	1.00	
V6	-.10	-.08	-.04	.79	.85	1.00

Note. N = 200.

The preceding matrix of correlations consists of six rows (running horizontally) and six columns (running vertically). Where the row for one variable intersects with the column for a second variable, you can find the correlation for that pair of variables. For example, where the row for V2 intersects with the column for V1, you can see that the correlation between these variables is .81.

Notice the pattern of intercorrelations in Table 2.1. Questions 1, 2, and 3 are strongly correlated with one another, but these variables are essentially uncorrelated with questions 4, 5, and 6. Similarly, questions 4, 5, and 6 are strongly correlated with one another, but are essentially uncorrelated with questions 1, 2, and 3. Reviewing the complete matrix reveals that there are two sets of variables that seem to hang together: Variables 1, 2, and 3 form one group, and variables 4, 5, and 6 form the second. But why are the variables grouping together in this way?

The Common Factor Model: Basic Concepts

One possible explanation for this pattern of intercorrelations may be found in the path model of Figure 2.1. In this figure, responses to questions 1 through 6 are represented as the six squares labeled V1 through V6. This path model suggests that variables V1, V2, and V3 are correlated with one another because they are all influenced by the same underlying factor. A **factor** is an unobserved variable (or latent variable). Being latent means that you cannot measure a factor directly, as you would measure an observed variable such as height or weight. A factor is a hypothetical construct: you believe that it exists, and you believe that it influences certain manifest variables (or observed variables) that you can measure directly. In the present study, the manifest variables, or observed variables, are subject responses to items 1 through 6.

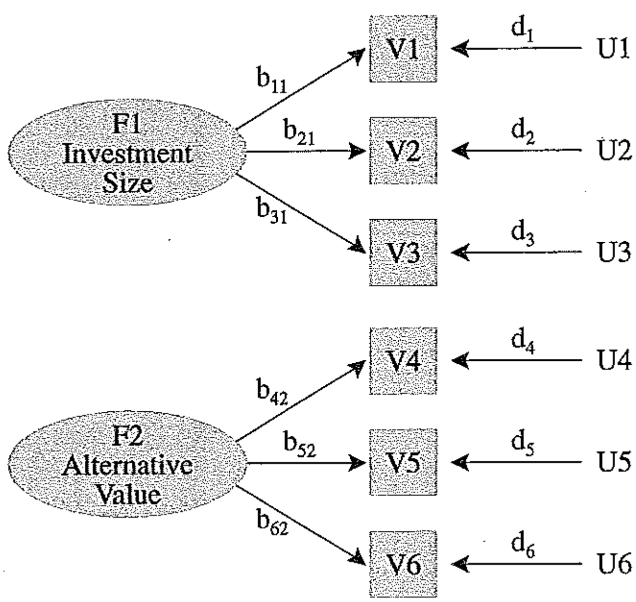


Figure 2.1: Path Model for a 6-Variable, 2-Factor Model, Orthogonal Factors, Factorial Complexity = 1

When representing factor models in figures, it is conventional to represent observed variables as squares or rectangles, and to represent latent factors as circles or ovals. You can see how the two factors appear in Figure 2.1. The first is labeled "F1: Investment Size," and the second is labeled "F2: Alternative Value."

Now return to the original question: Why do variables V1, V2, and V3 correlate so strongly with one another? According to the path model presented in Figure 2.1, these variables are intercorrelated because they are all caused by the same latent factor: the subjects' standing on the underlying investment size construct. This model proposes that, within each subject's belief system, there is a construct that you might call "investment size." Furthermore, this construct has a powerful influence on the way that subjects respond to questions 1, 2, and 3 (notice the arrows going from the oval factor to the squares). Therefore, even though you cannot directly measure someone's standing on the factor (it is a hypothetical construct, after all), you can infer that it really does exist by

- noting that questions 1, 2, and 3 correlate highly with one another
- reviewing the content of questions 1, 2, and 3 (that is, noting what these questions actually say)
- noting that all three questions seem to be measuring the same basic construct, a construct that could reasonably be named "investment size."

Note: The preceding is not a description of how one actually performs a factor analysis; it is just an attempt to help convey the conceptual meaning of the path model presented in the figure.

Common factors. The investment size factor (F1) presented in Figure 2.1 is known as a common factor. A **common factor** is a factor that influences more than one observed variable. In this case, you can see that variables V1, V2, and V3 are all influenced by the investment size factor. It is called a *common* factor because more than one variable has it in common. Because of this terminology, the type of analysis discussed in this chapter is often referred to as **common factor analysis.**

In the lower half of Figure 2.1, you can see that there is a second common factor (F2) representing the alternative value hypothetical construct. This factor affects responses to items 4, 5, and 6 (notice the directional arrows). In short, variables V4, V5, and V6 are intercorrelated because they have this alternative value factor in common. Variables V4, V5, and V6 are not influenced by the investment size factor (notice that there are no causal arrows going from F1 to these variables), and, similarly, V1, V2, and V3 are not influenced by the alternative value factor, F2. This should help clarify why variables V1, V2, and V3 tended to be uncorrelated with variables V4, V5, and V6.

Orthogonal versus oblique models. A few more points must be made in order to more completely understand the factor model presented in Figure 2.1. Notice that there is no arrow connecting F1 and F2. If it were hypothesized that the factors were correlated with one another, there would be a curved, double-headed arrow (a bidirectional arrow) connecting the two ovals. A double-headed arrow indicates that the researcher believes that the two constructs are correlated, but the arrow is not specifying any cause-and-effect relationship. The lack of such an arrow in Figure 2.1 means that the researcher expects these factors to be uncorrelated, or **orthogonal**. If a double-headed arrow did connect them, you would say that the factors are correlated, or **oblique**. Oblique factor models are discussed later in this chapter.

In some factor models a single-headed arrow connects two latent factors, indicating that one factor is expected to have a causal effect on the other. However, such models are normally not investigated with exploratory factor analysis, and will not be discussed in this chapter. For information on models that predict causal relationships between latent factors, see Chapter 5, "Developing Measurement Models with Confirmatory Factor Analysis," and Chapter 6, "Path Analysis with Latent Variables."

Unique factors. Notice that the two common factors are not the only factors that influence the observed variables. For example, you can see that there are actually two factors that influence variable V1: the common factor, F1, and a second factor labeled "U1." Here, U1 is a **unique factor**: a factor that influences only one observed variable. A unique factor represents all of the independent factors that are unique to that single variable (including the error component that is unique to that variable). In the figure, the unique factor U1 affects only V1, U2 affects only V2, and so forth.

Factor loadings. In Figure 2.1, each of the arrows going from a common factor to a variable is identified with a specific coefficient such as b_{11}, b_{21}, or b_{42}. The convention used in labeling these coefficients is quite simple: the first number in the subscript represents the number of the variable that the arrow points toward, and the second number in the subscript represents the number of the factor where the arrow originates. In this way, the coefficient b_{21} represents the arrow that goes to variable 2 from factor 1, the coefficient b_{52} represents the arrow that goes to variable 5 from factor 2, and so forth.

These coefficients represent **factor loadings**. But what exactly is a factor loading? Technically, it is a coefficient that appears in either a factor pattern matrix or a factor structure matrix (these matrices are included in the output of an oblique factor analysis). When you conduct an oblique factor analysis, the loadings in the pattern matrix will have a definition that is different from the definition given to loadings in the structure matrix (these definitions are discussed later in the chapter). To keep things simple, however, skip the oblique analysis for the moment, and focus on what the loadings represent when you perform an analysis in which the factors are orthogonal (uncorrelated). Factor loadings have a more simple interpretation in an orthogonal solution.

When investigating orthogonal factors, the b coefficients may be thought of in a number of different ways. For example, they may be viewed as:

- **Standardized regression coefficients**. The factor loadings obtained in an analysis with orthogonal factors may be thought of as standardized regression weights. If all of the variables (including the factors) are standardized to have unit variance (variance = 1.00), the b coefficients are analogous to the standardized regression coefficients (or regression weights) obtained in regression analysis. In other words, the b weights may be thought of as optimal linear weights by which the F factors are multiplied in calculating subject scores on the V variables (i.e., the weights used in predicting the variables from the factors).

- **Correlation coefficients.** Factor loadings also represent the product-moment correlation between an observed variable and an underlying factor. For example, if $b_{52} = .85$, this would indicate that the correlation between V5 and F2 is .85. This may surprise you if you are familiar with multiple regression, because most textbooks on multiple regression point out that standardized multiple regression coefficients and correlation coefficients are two different things. However, standardized regression coefficients are equivalent to correlation coefficients when the predictor variables are completely uncorrelated with each other. And that is the case in factor analysis with orthogonal factors: the factors serve as *predictor* variables in predicting the observed variables. And because the factors are uncorrelated with each other, the factor loadings may be interpreted as both standardized regression weights and as correlation coefficients.

- **Path coefficients.** Finally, the b coefficients are also analogous to the path coefficients obtained in path analysis. That is, they may be seen as standardized linear weights that represent the size of the effect that an underlying factor has in causing variability in the observed variable (path analysis is covered in Chapter 4, "Path Analysis with Manifest Variables").

Factor loadings are important because they help you interpret the factors that are responsible for the covariation in the data. This means that, after the factors are rotated, you can review the nature of the variables that have significant loadings for a given factor (i.e., the variables that are most strongly related to the factor). The nature of these variables will help you understand the nature of that factor.

Factorial complexity. **Factorial complexity** is a characteristic of an observed variable. The factorial complexity of a variable refers to the number of common factors that have a significant loading for that variable. For example, in Figure 2.1, you can see that the factorial complexity of V1 is one: V1 displays a significant loading for F1, but not for F2. The factorial complexity of V4 is also one: it displays a significant loading for F2, but not for F1.

Although the factor model of Figure 2.1 is fairly simple, a more complex model is presented in Figure 2.2. As with the previous model, two common factors are again responsible for the covariation in the data set. However, you can see that each of the common factors in Figure 2.2 has significant loadings on all six observed variables. In the same way, you can see that each variable is influenced by both common factors. Because each variable in the figure has significant loadings for two common factors, it may be said that each variable has a factorial complexity of two.

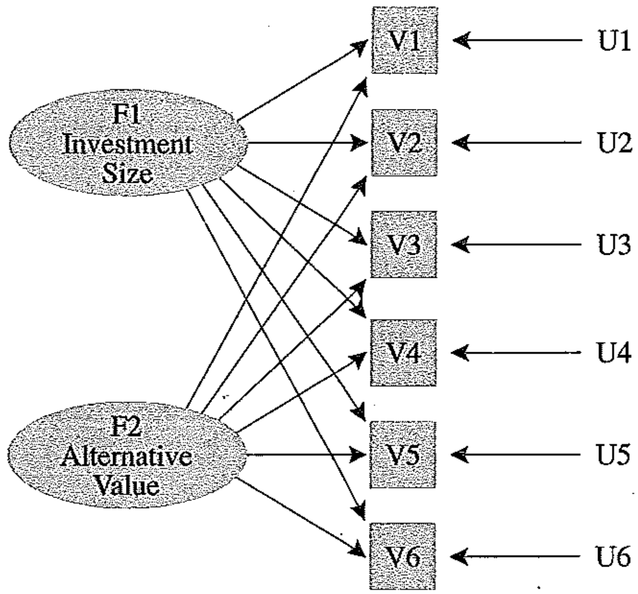


Figure 2.2: Path Model for a 6-Variable, 2-Factor Model, Orthogonal Factors, Factorial Complexity = 2

Observed variables as linear combinations of underlying factors. It is possible to think of a given observed variable, such as V1, as being a weighted sum of the underlying factors included in the factor model. For example, notice that in Figure 2.2, there are three factors that affect V1: two common factors (F1 and F2), and one unique factor (U1). By multiplying these factors by the appropriate weights, it is possible to calculate any subject's score on V1. In algebraic form, this would be done with the following equation:

$$V1 = b_{11}(F1) + b_{12}(F2) + d_{1}(U1)$$

In this equation, b_{11} is the regression weight for F1 (the amount of weight given to F1 in the prediction of V1), b_{12} is the regression weight for F2, and d_1 is the regression weight for the unique factor associated with V1. You can see that a given person's score on V1 is determined by multiplying the underlying factors by the appropriate regression weights, and summing the resulting products. This is why, in factor analysis, the observed variables are viewed as linear combinations of the underlying factors.

The preceding equation is therefore similar to the multiple regression equation as it is described in most statistics texts; in factor analysis, the observed variable (i.e., V1) serves as counterpart to the criterion variable (Y) in multiple regression, and the latent factors (i.e., F1, F2, and U1) serve as counterparts to the predictor variables (the X variables) in multiple regression. You would generally expect to obtain a different set of factor weights, and thus a different predictive equation, for each observed variable in a factor analysis.

Where do you find the regression weights for the common factors in factor analysis? In the **factor pattern matrix**. An example of a pattern matrix is presented here:

Factor Pattern

Variable	Factor 1	Factor 2
V1	.87	.26
V2	.80	.48
V3	.77	.34
V4	-.56	.49
V5	-.58	.52
V6	-.50	.59

You can see that the rows (running horizontally) in the factor pattern represent the different observed variables such as V1 and V2. The columns in the factor pattern represent the different factors such as F1 and F2. Where a row and column intersect, you will find a factor loading (or standardized regression coefficient). For example, in determining values of variable V1, F1 is given a weight of .87, and F2 is given a weight of .26; in determining values of V2, F1 is given a weight of .80, and F2 is given a weight of .48.

Communality versus the unique component. A **communality** is a characteristic of an observed variable. It refers to the variance in an observed variable that is accounted for by the common factors. If a variable demonstrates a large communality, it means that this variable is strongly influenced by at least one of the common factors. The symbol for communality is h^2. The communality for a given variable is computed by squaring that variable's factor loadings for all retained common factors, and summing these squares. For example, using the factor loadings from the previous factor pattern, you may compute the communality for V1 in the following way:

$$h_1^2 = b_{11}^2 + b_{12}^2$$

$$= (.87)^2 + (.26)^2$$

$$= .756 + .068$$

$$= .82$$

So the communality for V1 is approximately 82. This means that 82% of the variance in V1 is accounted for by the two common factors. You can now compute the communality for each variable, and add these values to the table that contains the pattern matrix:

Factor Pattern

Variable	Factor 1	Factor 2	h^2
V1	.87	.26	.82
V2	.80	.48	.87
V3	.77	.34	.71
V4	-.56	.49	.55
V5	-.58	.52	.61
V6	-.50	.59	.60

In contrast to the communality, the **unique component** refers to the proportion of variance in a given observed variable that is not accounted for by the common factors. Once communalities are computed, it is easy to calculate the unique component: simply subtract the communality from one. The unique component for V1 can be calculated in this way:

$$d_1^2 = 1 - h_1^2$$

$$= 1 - .82$$

$$= .18$$

This shows that 18% of the variance in V1 is not accounted for by the common factors; alternatively, you could say that 18% of the variance in V1 is accounted for by the unique factor, U1.

If you then proceed to take the square root of the unique component, you can compute the coefficient, d. This should look familiar, because earlier d was defined as the weight given to a unique factor in determining values on the observed variable. For variable V1, the unique component was calculated as .18. The square root of .18 is approximately .42. Therefore, the unique factor U1 would be given a weight of .42 in determining values of V1 (that is, $d_1 = .42$).

Exploratory Factor Analysis versus Principal Component Analysis

Some readers are likely to notice the many similarities between exploratory factor analysis and principal component analysis. In fact, these similarities have even led some researchers to incorrectly report that they have conducted factor analysis when, in fact, they have conducted principal component analysis. Because of this common misunderstanding, this section reviews some of the similarities and differences between the two procedures.

How Factor Analysis Differs from Principal Component Analysis

Purpose. Only factor analysis may be used to identify the factor structure underlying a set of variables. In other words, if you want to identify the number and nature of the latent factors that are responsible for covariation in the data set, then factor analysis, and not principal component analysis, should be used.

Principal components versus common factors. A principal component is an artificial variable; it is a linear combination of the (optimally weighted) observed variables. It is possible to calculate where a given subject stands on a principal component by simply summing that subject's (optimally weighted) scores on the observed variable being analyzed. For example, you could determine each subject's score on principal component 1 by using the following formula:

$$C_1 = b_{11}(X_1) + b_{12}(X_2) + \ldots b_{1p}(X_p)$$

where

C_1 = the subject's score on principal component 1 (the first component extracted)
b_{1p} = the regression coefficient (or weight) for observed variable p, as used in creating principal component 1
X_p = the subject's score on observed variable p.

In contrast, a common factor is a hypothetical latent variable that is assumed to be responsible for the covariation between two or more observed variables. Because factors are unmeasured latent variables, you may never know exactly where a given subject stands on an underlying factor (although it is possible to arrive at estimates, as you will see later).

In common factor analysis, the factors are not assumed to be linear combinations of the observed variables (as is the case in principal component analysis). Factor analysis assumes just the opposite, that the observed variables are linear combinations of the underlying factors. This is illustrated in the following equation:

$$X_1 = b_1(F_1) + b_2(F_2) + \ldots b_q(F_q) + d_1(U_1)$$

where

X_1 = the subject's score on observed variable 1
b_q = the regression coefficient (or weight) for underlying common factor q, as used in determining the subject's score on X_1
F_q = the subject's score on underlying factor q

d_1 = the regression weight for the unique factor associated with X_1
U_1 = the unique factor associated with X_1.

Because similar steps are followed in extracting principal components and common factors, it is easy to incorrectly assume that they are conceptually identical. However, the preceding equations show that they differ in an important way: in principal components analysis, the principal components are linear combinations of the observed variables, whereas the factors of factor analysis are not viewed in this way. In factor analysis the observed variables are viewed as linear combinations of the underlying factors.

You might be confused by this point if you know that it is possible to compute factor scores in exploratory factor analysis, and that these factor scores are essentially linear composites of observed variables. However, in reality these factor scores are merely *estimates* of where the subjects stand on the underlying factors. These so-called factor scores generally do not correlate perfectly with scores on the actual underlying factor (for this reason, they will be referred to as **estimated factor scores** in this text).

On the other hand, the principal component scores obtained in principal component analysis are not estimates; they are perfect representations of the extracted components. Remember that a principal component is simply a mathematical transformation (a linear combination) of the observed variables. So a given subject's component score represents with perfect accuracy where that subject stands on the principal component. It is therefore proper to discuss *actual* component scores rather than estimated component scores.

Variance accounted for. Factor analysis and principal component analysis also differ with respect to the type of variance accounted for. The factors of factor analysis account for common variance in a data set, while the components of principal component analysis account for total variance in the data set. This difference may be understood with reference to Figure 2.3.

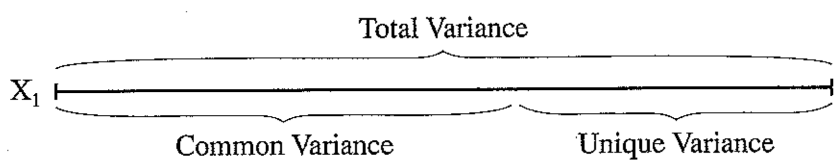


Figure 2.3: Total Variation in Variable X_1, as Divided into Common and Unique Components

Assume that the length of the line in Figure 2.3 represents the total variance in observed variable X_1, and that variables X_1 through X_6 are subjected to factor analysis. The figure shows that the total variance in X_1 may be divided into two components: the common variance and the unique variance. The **common variance** corresponds to the communality of X_1: the proportion of total variance in the variable that is accounted for by the common factors. The remaining variance is the unique component: that variance (whether systematic or random) which is specific to variable X_1.

In factor analysis, factors are extracted to account for only the common variance, and the remaining unique variance remains unanalyzed. This is accomplished by analyzing an **adjusted correlation matrix**: a correlation matrix with communality estimates on the diagonal. You cannot know a variable's actual communality prior to the factor analysis, and so it must be estimated using one of a number of alternative procedures. This chapter recommends that squared multiple correlations be used as prior communality estimates. A variable's squared multiple correlation is obtained by using multiple regression to regress it on the remaining observed variables (later, you will find that these values can be obtained easily by using the PRIORS= option with PROC FACTOR). The adjusted correlation matrix that is analyzed in factor analysis has correlations between the observed variables off the diagonal, and communality estimates on the diagonal.

In principal component analysis, however, components are extracted to account for the **total variance** in the data set, not just the common variance. This is accomplished by analyzing an **unadjusted correlation matrix**: a correlation matrix with ones (1.00) on the diagonal. Why ones? Since all variables are standardized in the analysis, each has a variance of one. Because the correlation matrix contains ones (rather than communalities) on the diagonal, 100% of each variable's variance will be accounted for by the combined components, not just the variance that the variable shares in common with other variables.

It is this difference that explains why only factor analysis, and not principal component analysis, can be used to identify the number and nature of the factors that are responsible for covariation in the data set. Because principal component analysis makes no attempt to separate the common component from the unique component of each variable's variance, this procedure can provide a misleading picture of the factor structure underlying the data. Either procedure may be used to reduce a number of variables to a more manageable number; however, if you want to identify the factor structure of a data set (such as that portrayed in the path model of Figure 2.1), only factor analysis will do.

How Factor Analysis Is Similar to Principal Component Analysis

Purpose (in some cases). Both factor analysis and principal component analysis may be used as **variable reduction procedures**; that is, both may be used to reduce a number of variables to a smaller, more manageable number. This is why both procedures are so widely used in analyzing data from multiple-item questionnaires in the social sciences. Both procedures can be used to reduce a large number of survey questions into a smaller number of scales.

Extraction methods (in some cases). In this chapter, you first learn how to use the principal axis method for extracting factors. This is the same mathematical procedure used to extract principal components in Chapter 1 (a later example will also show how to use the maximum likelihood method: an extraction method that is typically used only with factor analysis).

Results (in some cases). Principal component analysis and factor analysis often lead to similar conclusions regarding the appropriate number of factors (or components) to retain, as well as similar conclusions regarding how the factors (or components) should be interpreted. This is especially true when the variable communalities are high (near 1.00). The reason for this should be obvious: when the principal axis extraction method is used, the only real difference between the two procedures involves the values that appear on the diagonal of the correlation matrix. If the communalities are very high (near 1.00), there is little difference between the matrix that is analyzed in principal component analysis and the matrix that is analyzed in factor analysis; hence, the similar solutions.

Preparing and Administering the Investment Model Questionnaire

As was discussed at the beginning of this chapter, assume that you are interested in measuring two constructs that constitute important components of the investment model (Rusbult, 1980). One construct is investment size: the amount of time or personal resources that the person has put into his or her relationship with the current partner. The other construct is alternative value: the attractiveness of a person's alternatives to his or her current romantic partner.

Writing the Questionnaire Items

The questionnaire used earlier in this chapter is again reproduced here. Note that items 1–3 were designed to assess investment size, and items 4–6 were designed to assess alternative value.

```
Please rate each of the following items to indicate the
extent to which you agree or disagree with each statement.
Use a response scale in which 1=Strongly Disagree and
7=Strongly Agree.

_____  1. I have invested a lot of time and effort in developing
          my relationship with my current partner.
_____  2. My current partner and I have developed interest in a
          lot of fun activities that I would lose if our
          relationship were to end.
_____  3. My current partner and I have developed a lot of mutual
          friendships that I would lose if our relationship were
          to end.
_____  4. It would be more attractive for me to be involved in a
          relationship with someone else rather than continue in
          a relationship with my current partner.
_____  5. It would be more attractive for me to be by myself than
          to continue my relationship with my current partner.
_____  6. In general, my alternatives to remaining in this
          relationship are quite attractive.
```

Number of Items per Factor

As was mentioned in Chapter 1, it is highly desirable to have at least three (and preferably more) variables loading on each factor when the analysis is complete. Because some of the items may be dropped during the course of the analysis, it is generally good practice to write at least five items for each construct that you wish to measure; in this way, you increase the chances that at least three items per factor will survive the analysis (you can see that the preceding questionnaire violates this recommendation by including only three items for each factor at the outset).

Note: Remember that the recommendation of three items per scale actually constitutes a *lower bound*. In practice, test and attitude scale developers normally desire that their scales contain many more than just three items to measure a given construct. It is not unusual to see individual scales that include 10, 20, or even more items to assess a single construct. Other things held constant, the more items in the scale, the more reliable it will be. The recommendation of three items per scale should therefore be viewed as a rock-bottom lower bound, appropriate only if practical concerns (such as total questionnaire length) prevent you from including more items. For more information on scale construction, see Spector (1992).

Minimally Adequate Sample Size

Factor analysis is a large-sample procedure, so it is important to use guidelines to choose the sample size which will be minimally adequate for an analysis. *The minimal number of subjects in the sample should be the larger of 100 subjects, or 5 times the number of variables being analyzed.* If responses to a questionnaire are being analyzed, then the number of variables is equal to the number of items on the questionnaire. To illustrate, assume that you want to perform an analysis on responses to a 50-item questionnaire. Five times the number of items on the questionnaire equals 250. Therefore, it would be best if your final sample provides usable (complete) data from at least 250 subjects. It should be remembered, however, that any subject who fails to answer just one item will not provide usable data for the factor analysis, and will therefore be dropped from the final sample. A certain number of subjects can always be expected to leave at least one question blank; therefore, to ensure that the final sample includes at least 250 usable responses, you should administer the questionnaire to perhaps 300 subjects.

These rules regarding the number of subjects per variable again constitute a lower bound, and some have argued that they should apply only under two optimal conditions for factor analysis: (a) when many variables are expected to load on each factor, and (b) when variable communalities are high. Under less optimal conditions, larger samples may be required.

SAS Program and Analysis Results

This section provides instructions on writing the SAS program, along with an overview of the SAS output. A subsequent section will provide a more detailed treatment of the steps followed in the analysis, and the decisions to be made at each step.