Stevens, J. P. (2009). Applied multivariate statistics for the social sciences (5th Ed.). New York: Routledge.

11

Exploratory and Confirmatory Factor Analysis

11.1 Introduction

Consider the following two common classes of research situations:

1. Exploratory regression analysis: An experimenter has gathered a moderate to large number of predictors (say 15 to 40) to predict some dependent variable.
2. Scale development: An investigator has assembled a set of items (say 20 to 50) designed to measure some construct (e.g., attitude toward education, anxiety, sociability). Here we think of the items as the variables.

In both of these situations the number of simple correlations among the variables is very large, and it is quite difficult to summarize by inspection precisely what the pattern of correlations represents. For example, with 30 variables, there are 435 simple correlations. Some means is needed for determining if there is a small number of underlying constructs that might account for the main sources of variation in such a complex set of correlations.

Furthermore, if there are 30 variables (whether predictors or items), we are undoubtedly not measuring 30 different constructs; hence, it makes sense to find some variable reduction scheme that will indicate how the variables cluster or hang together. Now, if sample size is not large enough (how large N needs to be is discussed in Section 11.7), then we need to resort to a logical clustering (grouping) based on theoretical or substantive grounds. On the other hand, with adequate sample size an empirical approach is preferable. Two basic empirical approaches are (a) principal components analysis and (b) factor analysis. In both approaches linear combinations of the original variables (the factors) are derived, and often a small number of these account for most of the variation or the pattern of correlations. In factor analysis a mathematical model is set up, and the factors can only be estimated, whereas in components analysis we are simply transforming the original variables into the new set of linear combinations (the principal components).

Both methods often yield similar results. We prefer to discuss principal components for several reasons:

1. It is a psychometrically sound procedure.
2. It is simpler mathematically, relatively speaking, than factor analysis. And a main theme in this text is to keep the mathematics as simple as possible.
3. The factor indeterminacy issue associated with common factor analysis (Steiger, 1979) is a troublesome feature.
4. A thorough discussion of factor analysis would require hundreds of pages, and there are other good sources on the subject (Gorsuch, 1983).

Recall that for discriminant analysis uncorrelated linear combinations of the original variables were used to additively partition the association between the classification variable and the set of dependent variables. Here we are again using uncorrelated linear combinations of the original variables (the principal components), but this time to additively partition the variance for a set of variables.

In this chapter we consider in some detail two fundamentally different approaches to factor analysis. The first approach, just discussed, is called exploratory factor analysis. Here the researcher is attempting to determine how many factors are present and whether the factors are correlated, and wishes to name the factors. The other approach, called confirmatory factor analysis, rests on a solid theoretical or empirical base. Here, the researcher "knows" how many factors there are and whether the factors should be correlated. Also, the researcher generally forces items to load only on a specific factor and wishes to "confirm" a hypothesized factor structure with data. There is an overall statistical test for doing so. First, however, we turn to the exploratory mode.

11.2 Exploratory Factor Analysis

11.2.1 The Nature of Principal Components

If we have a single group of subjects measured on a set of variables, then principal components partition the total variance (i.e., the sum of the variances for the original variables) by first finding the linear combination of the variables that accounts for the maximum account of variance:

$$y_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1p}x_p$$

y_1 is called the first principal component, and if the coefficients are scaled such that $\mathbf{a}_1'\mathbf{a}_1 = 1$ [where $\mathbf{a}_1' = (a_{11}, a_{12}, \ldots, a_{1p})$] then the variance of y_1 is equal to the *largest* eigenvalue of the sample covariance matrix (Morrison, 1967, p. 224). The coefficients of the principal component are the elements of the eigenvector corresponding to the largest eigenvalue.

Then the procedure finds a second linear combination, *uncorrelated* with the first component, such that it accounts for the next largest amount of variance (after the variance attributable to the first component has been removed) in the system. This second component y_2 is

$$y_2 = a_{21}x_1 + a_{22}x_2 + \cdots a_{2p}x_p$$

and the coefficients are scaled so that $\mathbf{a}_2'\mathbf{a}_2 = 1$, as for the first component. The fact that the two components are constructed to be uncorrelated means that the Pearson correlation between y_1 and y_2 is 0. The coefficients of the second component are simply the elements of the eigenvector associated with the second largest eigenvalue of the covariance matrix, and the sample variance of y_2 is equal to the second largest eigenvalue.

The third principal component is constructed to be uncorrelated with the first two, and accounts for the third largest amount of variance in the system, and so on. Principal components analysis is therefore still another example of a mathematical maximation procedure, where each successive component accounts for the maximum amount of the variance that is left.

Thus, through the use of principal components, a set of correlated variables is transformed into a set of uncorrelated variables (the components). The hope is that a much smaller number of these components will account for most of the variance in the original set of variables, and of course that we can meaningfully interpret the components. By most of the variance we mean about 75% or more, and often this can be accomplished with five or fewer components.

The components are interpreted by using the component-variable correlations (called *factor loadings*) that are largest in absolute magnitude. For example, if the first component loaded high and positive on variables 1, 3, 5, and 6, then we would interpret that component by attempting to determine what those four variables have in common. *The component procedure has empirically clustered the four variables, and the job of the psychologist is to give a name to the construct that underlies variability and thus identify the component substantively.*

In the preceding example we assumed that the loadings were all in the same direction (all positive). Of course, it is possible to have a mixture of high positive and negative loadings on a particular component. In this case we have what is called a *bipolar* factor. For example, in components analyses of IQ tests, the second component may be a bipolar factor contrasting verbal abilities against spatial-perceptual abilities.

Social science researchers would be used to extracting components from a correlation matrix. The reason for this standardization is that scales for tests used in educational, sociological, and psychological research are usually arbitrary. If, however, the scales are reasonably commensurable, performing a components analysis on the *covariance* matrix is preferable for statistical reasons (Morrison, 1967, p. 222). The components obtained from the correlation and covariance matrices are, in general, *not* the same. The option of doing the components analysis on either the correlation or covariance matrix is available on SAS and SPSS.

A precaution that researchers contemplating a components analysis with a small sample size (certainly any n around 100) should take, especially if most of the elements in the sample correlation matrix are small, is to apply Bartlett's sphericity test (Cooley & Lohnes, 1971, p. 103). This procedure tests the null hypothesis that the variables in the *population* correlation matrix are uncorrelated. If one fails to reject with this test, then there is no reason to do the components analysis because the variables are already uncorrelated. The sphericity test is available on both the SAS and SPSS packages.

11.3 Three Uses for Components as a Variable Reducing Scheme

We now consider three cases in which the use of components as a variable reducing scheme can be very valuable.

1. The first use has already been mentioned, and that is to determine empirically how many dimensions (underlying constructs) account for most of the variance on an instrument (scale). The original variables in this case are the items on the scale.

2. In a multiple regression context, if the number of predictors is large relative to the number of subjects, then we may wish to use principal components on the predictors to reduce markedly the number of predictors. If so, then the N/variable ratio increases considerably and the possibility of the regression equation's holding up under cross-validation is much better (see Herzberg, 1969). We show later in the chapter (Example 11.3) how to do this on SAS and SPSS.

   The use of principal components on the predictors is also one way of attacking the multicollinearity problem (correlated predictors). Furthermore, because the new predictors (i.e., the components) are uncorrelated, the order in which they enter the regression equation makes no difference in terms of how much variance in the dependent variable they will account for.

3. In the chapter on k-group MANOVA we indicated several reasons (reliability consideration, robustness, etc.) that generally mitigate against the use of a large number of criterion variables. Therefore, if there is initially a large number of potential criterion variables, it probably would be wise to perform a principal components analysis on them in an attempt to work with a smaller set of new criterion variables. We show later in the chapter (in Example 11.4) how to do this for SAS and SPSS. It must be recognized, however, that the components are *artificial* variables and are not necessarily going to be interpretable. Nevertheless, there are techniques for improving their interpretability, and we discuss these later.

11.4 Criteria for Deciding on How Many Components to Retain

Four methods can be used in deciding how many components to retain:

1. Probably the most widely used criterion is that of Kaiser (1960): Retain only those components whose eigenvalues are greater than 1. Unless something else is specified, this is the rule that is used by SPSS, but not by SAS. Although using this rule generally will result in retention of only the most important factors, blind use could lead to retaining factors that may have no practical significance (in terms of percent of variance accounted for).

   Studies by Cattell and Jaspers (1967), Browne (1968), and Linn (1968) evaluated the accuracy of the eigenvalue > 1 criterion. In all three studies, the authors determined how often the criterion would identify the correct number of factors from matrices with a known number of factors. The number of variables in the studies ranged from 10 to 40. Generally, the criterion was accurate to fairly accurate, with gross overestimation occurring only with a large number of variables (40) *and* low communalities (around .40). The criterion is more accurate when the number of variables is small (10 to 15) or moderate (20 to 30) and the communalities are high (>.70). The communality of a variable is the amount of variance on a variable accounted for by the set of factors. We see how it is computed later in this chapter.

2. A graphical method called the *scree test* has been proposed by Cattell (1966). In this method the magnitude of the eigenvalues (vertical axis) is plotted against their ordinal numbers (whether it was the first eigenvalue, the second, etc.). Generally what happens is that the magnitude of successive eigenvalues drops

off sharply (steep descent) and then tends to level off. The recommendation is to retain all eigenvalues (and hence components) in the sharp descent *before* the first one on the line where they start to level off. In one of our examples we illustrate this test. This method will generally retain components that account for large or fairly large and distinct amounts of variances (e.g., 31%, 20%, 13%, and 9%). Here, however, blind use might lead to not retaining factors which, although they account for a smaller amount of variance, might be practically significant. For example, if the first eigenvalue at the break point accounted for 8.3% of variance and then the next three eigenvalues accounted for 7.1%, 6%, and 5.2%, then 5% or more might well be considered significant in some contexts, and retaining the first and dropping the next three seems somewhat arbitrary. The scree plot is available on SPSS (in FACTOR program) and in the SAS package. Several studies have investigated the accuracy of the scree test. Tucker, Koopman, and Linn (1969) found it gave the correct number of factors in 12 of 18 cases. Linn (1968) found it to yield the correct number of factors in seven of 10 cases, whereas Cattell and Jaspers (1967) found it to be correct in six of eight cases.

A later, more extensive study on the number of factors problem (Hakstian, Rogers, & Cattell, 1982) adds some additional information. They note that for $N > 250$ and a mean communality $\geq .60$, either the Kaiser or Scree rules will yield an accurate estimate for the number of true factors. They add that such an estimate will be just that much more credible if the Q/P ratio is $< .30$ (P is the number of variables and Q is the number of factors). With mean communality .30 or $Q/P > .3$, the Kaiser rule is less accurate and the Scree rule much less accurate.

3. There is a statistical significance test for the number of factors to retain that was developed by Lawley (1940). However, as with all statistical tests, it is influenced by sample size, and large sample size may lead to the retention of too many factors.
4. Retain as many factors as will account for a specified amount of total variance. Generally, one would want to account for at least 70% of the total variance, although in some cases the investigator may not be satisfied unless 80 to 85% of the variance is accounted for. This method could lead to the retention of factors that are essentially variable specific, that is, load highly on only a single variable.

So what criterion should be used in deciding how many factors to retain? *Since the Kaiser criterion has been shown to be quite accurate when the number of variables is <30 and the communalities are >.70, or when $N > 250$ and the mean communality is $\geq .60$, we would use it under these circumstances.* For other situations, use of the scree test with an $N > 200$ will probably not lead us too far astray, provided that most of the communalities are reasonably large.

In all of the above we have assumed that we will retain only so many components, which will hopefully account for a sizable amount of the total variance and simply discard the rest of the information, that is, not worry about the 20 or 30% of the variance that is not accounted for. However, it seems to us that in some cases the following suggestion of Morrison (1967, p. 228) has merit:

> Frequently, it is better to summarize the complex in terms of the first components with large and markedly distinct variances and include as highly specific and unique variates those responses which are generally independent in the system. Such unique responses could probably be represented by high loadings in the later components but only in the presence of considerable noise from the other unrelated variates.

In other words, if we did a components analysis on, say, 20 variables and only the first four components accounted for large and distinct amounts of variance, then we should summarize the complex of 20 variables in terms of the four components *and* those particular variables that had high correlations (loadings) with the latter components. In this way more of the total information in the complex is retained, although some parsimony is sacrificed.

11.5 Increasing Interpretability of Factors by Rotation

Although the principal components are fine for summarizing most of the variance in a large set of variables with a small number of components, often the components are not easily interpretable. The components are artificial variates designed to maximize variance accounted for, not designed for interpretability. Two major classes of rotations are available:

1. Orthogonal (rigid) rotations—here the new factors are still uncorrelated, as were the original components.
2. Oblique rotations—here the new factors will be correlated.

11.5.1 Orthogonal Rotations

We discuss two such rotations:

1. Quartimax—Here the idea is to clean up the variables. That is, the rotation is done so that each variable loads mainly on one factor. Then that variable can be considered to be a relatively pure measure of the factor. The problem with this approach is that most of the variables tend to load on a single factor (producing the so called "g" factor in analyses of IQ tests), making interpretation of the factor difficult.
2. Varimax—Kaiser (1960) took a different tack. He designed a rotation to clean up the factors. That is, with his rotation, each factor tends to load high on a smaller number of variables and low or very low on the other variables. This will generally make interpretation of the resulting factors easier. The varimax rotation is the default option in SPSS.

It should be mentioned that when the varimax rotation is done, the *maximum variance property* of the original components is destroyed. The rotation essentially reallocates the loadings. Thus, the first rotated factor will no longer *necessarily* account for the maximum amount of variance. The amount of variance accounted for by each rotated factor has to be recalculated. You will see this on the printout from SAS and SPSS. Even though this is true, and somewhat unfortunate, it is more important to be able to interpret the factors.

11.5.2 Oblique Rotations

Numerous oblique rotations have been proposed: for example, oblimax, quartimin, maxplane, orthoblique (Harris–Kaiser), promax, and oblimin. Promax and orthoblique are available on SAS, and oblimin is available on SPSS.

Many have argued that correlated factors are much more reasonable to assume in most cases (Cliff, 1987; Pedhazur & Schmelkin, 1991; *SAS STAT User's Guide,* Vol. 1, p. 776, 1990), and therefore oblique rotations are quite reasonable. The following from Pedhazur and Schmelkin (1991) is interesting:

> From the perspective of construct validation, the decision whether to rotate factors orthogonally or obliquely reflects one's conception regarding the structure of the construct under consideration. It boils down to the question: Are aspects of a postulated multidimensional construct intercorrelated? The answer to this question is relegated to the status of an assumption when an orthogonal rotation is employed.... The preferred course of action is, in our opinion, to rotate both orthogonally and obliquely. When, on the basis of the latter, it is concluded that the correlations among the factors are negligible, the interpretation of the simpler orthogonal solution becomes tenable. (p. 615)

It has also been argued that there is no such thing as a "best" oblique rotation. The following from the *SAS STAT User's Guide* (Vol. 1, 1990) strongly expresses this view:

> You cannot say that any rotation is better than any other rotation from a statistical point of view; all rotations are equally good statistically. Therefore, the choice among different rotations must be based on nonstatistical grounds. ... If two rotations give rise to different interpretations, those two interpretations must not be regarded as conflicting. Rather, they are two different ways of looking at the same thing, two different points of view in the common factor space. (p. 776)

In the two computer examples we simply did the components analysis and a varimax rotation, that is, an orthogonal rotation. The solutions obtained may or may not be the most reasonable ones. We also did an oblique rotation (promax) on the Personality Research Form using SAS. Interestingly, the correlations among the factors were very small (all $<.10$ in absolute value), suggesting that the original orthogonal solution is quite reasonable. We leave it to the reader to run an oblique rotation (oblimin) on the California Psychological Inventory using SPSS, and to compare the orthogonal and oblique solutions.

The reader needs to be aware that when an oblique solution is more reasonable, interpretation of the factors becomes more complicated. Two matrices need to be examined:

1. Factor pattern matrix—The elements here are analogous to standardized regression coefficients from a multiple regression analysis. That is, a given element indicates the importance of that variable to the factor with the influence of the other variables partialled out.
2. Factor structure matrix—The elements here are the simple correlations of the variables with the factors; that is, they are the factor loadings.

For orthogonal factors these two matrices are the same.

11.6 What Loadings Should Be Used for Interpretation?

Recall that a loading is simply the Pearson correlation between the variable and the factor (linear combination of the variables). Now, certainly any loading that is going to be used to interpret a factor should be statistically significant at a minimum. The formula for the standard error of a correlation coefficient is given in elementary statistics books as

$1/\sqrt{N-1}$ and one might think it could be used to determine which loadings are significant. But, in components analysis (where we are maximizing again), and in rotating, there is considerable opportunity for capitalization on chance. This is especially true for small or moderate sample sizes, or even for fairly large sample size (200 or 300) if the number of variables being factored is large (say 40 or 50). Because of this capitalization on chance, the formula for the standard error of correlation can *seriously underestimate* the actual amount of error in the factor loadings.

A study by Cliff and Hamburger (1967) showed that the standard errors of factor loadings for orthogonally rotated solutions in all cases were considerably greater (150 to 200% in most cases) than the standard error for an ordinary correlation. Thus, a rough check as to whether a loading is statistically significant can be obtained by *doubling* the standard error, that is, doubling the critical value required for significance for an ordinary correlation. This kind of statistical check is most crucial when sample size is small, or small relative to the number of variables being factor analyzed. When sample size is quite large (say 1,000), or large relative to the number of variables (N = 500 for 20 variables), then significance is ensured. It may be that doubling the standard error in general is too conservative, because for the case where a statistical check is more crucial (N = 100), the errors were generally less than 1½ times greater. However, because Cliff and Hamburger (1967, p. 438) suggested that the sampling error might be greater in situations that aren't as clean as the one they analyzed, it probably is advisable to be conservative until more evidence becomes available.

Given the Cliff and Hamburger results, we feel it is time that investigators stopped blindly using the rule of interpreting factors with loadings greater than $|.30|$, and take sample size into account. Also, because in checking to determine which loadings are significant, many statistical tests will be done, it is advisable to set the α level more stringently for each test. This is done to control on overall α, that is, the probability of at least one false rejection. We would recommend testing each loading for significance at α = .01 (two-tailed test). To aid the reader in this task we present in Table 11.1 the critical values for a simple correlation at α = .01 for sample size ranging from 50 to 1,000. *Remember that the critical values in Table 11.1 should be doubled, and it is the doubled value that is used as the critical value for testing the significance of a loading.* To illustrate the use of Table 11.1, suppose a factor analysis had been run with 140 subjects. Then, only loadings >2(.217) = .434 in absolute value would be declared statistically significant. If sample size in this example had been 160, then interpolation between 140 and 180 would give a very good approximation to the critical value.

Once one is confident that the loadings being used for interpretation are significant (because of a significance test or because of large sample size), then the question becomes which loadings are large enough to be practically significant. For example, a loading of .20 could well be significant with large sample size, but this indicates only 4% shared variance between the variable and the factor. It would seem that one would want in general a variable to share *at least* 15% of its variance with the construct (factor) it is going to be used to

TABLE 11.1

Critical Values for a Correlation Coefficient at α = .01 for a Two-Tailed Test

n	CV	n	CV	n	CV
50	.361	180	.192	400	.129
80	.286	200	.182	600	.105
100	.256	250	.163	800	.091
140	.217	300	.149	1000	.081

help name. This m
purposes. To interp
the factor (construc

11.7 Sample Siz

Various rules have
Many of the popul
number of variable
20 subjects per vari
jects per variable a
and Velicer (1988) i
are component sat
size. Also, number
dations for the appl

1. Components
regardless of
2. Components
size is greater
3. Components
ple size is at le

An additional reaso
at least three loadin
These results are
the-pants" judgmen
as with any study, t
across a component
a reliable componen
a strict empirical ba
The third recomm
low loadings be int
that a factor defined
are as close as we ca
Velicer also indic
average of the three
munication, August

11.8 Four Compu

We now consider fo
max rotation in pra
Psychological Inven
input a correlation m

help name. This means using only loadings that are about .40 or greater for interpretation purposes. To interpret what the variables with high loadings have in common, i.e., to name the factor (construct), a substantive specialist is needed.

11.7 Sample Size and Reliable Factors

Various rules have been suggested in terms of the sample size required for reliable factors. Many of the popular rules suggest that sample size be determined as a function of the number of variables being analyzed, ranging anywhere from two subjects per variable to 20 subjects per variable. And indeed, in a previous edition of this text, I suggested five subjects per variable as the minimum needed. However, a Monte Carlo study by Guadagnoli and Velicer (1988) indicated, contrary to the popular rules, that the most important factors are component saturation (the absolute magnitude of the loadings) and absolute sample size. Also, number of variables per component is somewhat important. Their recommendations for the applied researcher were as follows:

1. Components with four or more loadings above .60 in absolute value are reliable, regardless of sample size.
2. Components with about 10 or more low (.40) loadings are reliable as long as sample size is greater than about 150.
3. Components with only a few low loadings should not be interpreted unless sample size is at least 300.

An additional reasonable conclusion to draw from their study is that any component with at least three loadings above .80 will be reliable.

These results are nice in establishing at least some empirical basis, rather than "seat-of-the-pants" judgment, for assessing what components we can have confidence in. However, as with any study, they cover only a certain set of situations. For example, what if we run across a component that has two loadings above .60 and six loadings of at least .40; is this a reliable component? My guess is that it probably would be, but at this time we don't have a strict empirical basis for saying so.

The third recommendation of Guadagnoli and Velicer, that components with only a few low loadings be interpreted tenuously, doesn't seem that important to me. The reason is that a factor defined by only a few loadings is not much of a factor; as a matter of fact, we are as close as we can get to the factor's being variable specific.

Velicer also indicated that when the *average* of the four largest loadings is >.60 or the *average* of the three largest loadings is >.80, then the factors will be reliable (personal communication, August, 1992). This broadens considerably when the factors will be reliable.

11.8 Four Computer Examples

We now consider four examples to illustrate the use of components analysis and the varimax rotation in practice. The first two involve popular personality scales: the California Psychological Inventory and the Personality Research Form. Example 11.1 shows how to input a correlation matrix using the SPSS FACTOR program, and Example 11.2 illustrates correlation matrix input for the SAS FACTOR program. Example 11.3 shows how to do a components analysis on a set of predictors and then pass the new predictors (the factor scores) to a regression program for both SAS and SPSS. Example 11.4 illustrates a components analysis and varimax rotation on a set of dependent variables and then passing the factor scores to a MANOVA program for both SAS and SPSS.

Example 11.1: California Psychological Inventory on SPSS

The first example is a components analysis of the California Psychological Inventory followed by a varimax rotation. The data was collected on 180 college freshmen (90 males and 90 females) by Smith (1975). He was interested in gathering evidence to support the uniqueness of death anxiety as a construct. Thus, he wanted to determine to what extent death anxiety could be predicted from general anxiety, other personality variables (hence the use of the CPI), and situational variables related to death (recent loss of a love one, recent experiences with a deathly situation, etc.). In this use of multiple regression Smith was hoping for a *small* R^2; that is, he wanted only a small amount of the variance in death anxiety scores to be accounted for by the other variables.

Table 11.2 presents the SPSS control lines for the factor analysis, along with annotation explaining what several of the commands mean. Table 11.3 presents part of the printout from SPSS. The printout indicates that the first component (factor) accounted for 37.1% of the total variance. This is arrived at by dividing the eigenvalue for the first component (6.679), which tells how much variance that component accounts for, by the total variance (which for a correlation matrix is just the sum of the diagonal elements, or 18 here). The second component accounts for $2.935/18 \times 100 = 16.3\%$ of the variance, and so on.

As to how many components to retain, Kaiser's rule of using only those components whose eigenvalues are greater than 1 would indicate that we should retain only the first four components (which is what has been done on the printout; remember Kaiser's rule is the default option for SPSS). Thus, as the printout indicates, we account for 71.4% of the total variance. Cattell's screen test (see Table 11.3) would not agree with the Kaiser rule, because there are only three eigenvalues (associated with the first three factors) before the breaking point, the point where the steep descent stops and the eigenvalues start to level off. The results of a study by Zwick and Velicer (1986) would lead us to use only three factors here. These three factors, as Table 11.3 shows, account for 65.2% of the total variance.

Table 11.4 gives the unrotated loadings and the varimax rotated loadings. From Table 11.1, the critical value for a significant loading is 2(.192) = .384. Thus, this is an absolute minimum value for us to be confident that we are dealing with nonchance loadings. The original components are somewhat difficult to interpret, especially the first component, because 14 of the loadings are "significant." Therefore, we focus our interpretation on the rotated factors. The variables that we use in interpretation are boxed in on Table 11.4. The first rotated factor still has significant loadings on 11 variables, although because one of these (.410 for CS) is just barely significant, and is also substantially less than the other significant loadings (the next smallest is .535), we disregard it for interpretation purposes. Among the adjectives that characterize high scores on the other 10 variables, from the CPI manual, are: calm, patient, thorough, nonaggressive, conscientious, cooperative, modest, diligent, and organized. Thus, this first rotated factor appears to be a "conforming, mature, inward tendencies" dimension. That is, it reveals a low-profile individual, who is conforming, industrious, thorough, and nonaggressive.

The loadings that are significant on the second rotated factor are also strong loadings (the smallest is .666): .774 for dominance, .666 for capacity for status, .855 for sociability, .780 for social presence, and .879 for self-acceptance. Adjectives from the CPI manual used to characterize high scores on these variables are: aggressive, ambitious, spontaneous, outspoken, self-centered, quick, and enterprising. Thus, this factor appears to describe an "aggressive, outward tendencies" dimension. High scores on this dimension reveal a high-profile individual who is aggressive, dynamic, and outspoken.

TABLE 11.2

SPSS Factor Control Lines for Principal Components on California Psychological Inventory

```
TITLE 'PRINCIPAL COMPONENTS ON CPI'.
MATRIX DATA VARIABLES=DOM CAPSTAT SOCIAL SOCPRES SELFACP WELLBEG RESPON SOCLIZ SELFCTRL TOLER
GOODIMP COMMUNAL ACHCONF ACHINDEP INTELEFF PSYMIND FLEX FEMIN/CONTENTS=N_SCALAR CORR/.
BEGIN DATA.
 180
1.000
 .467  1.000
 .681   .600  1.000
 .447   .585   .643  1.000
 .610   .466   .673   .612  1.000
 .236   .324   .339   .357   .077  1.000
 .401   .346   .344   .081   .056   .518  1.000
 .214   .179   .242   .003  -.029   .517   .632  1.000
-.062   .105  -.001  -.130  -.352   .619   .476   .544  1.000
 .227   .465   .295   .330   .004   .698   .502   .517   .575  1.000
 .238   .392   .367   .178   .023   .542   .381   .367   .697   .501  1.000
 .189   .146   .227   .159   .117   .336   .380   .384   .084   .192  -.001  1.000
 .401   .374   .479   .296   .154   .676   .567   .589   .633   .588   .610   .307  1.000
 .075   .400   .140   .289  -.027   .513   .369   .280   .464   .720   .359   .175   .465  1.000
 .314   .590   .451   .457   .192   .671   .500   .442   .456   .716   .460   .333   .616   .688  1.000
 .167   .337   .239   .336   .011   .463   .217   .182   .410   .502   .397  -.060   .393   .519   .466  1.000
 .148   .203  -.028   .236   .037   .051  -.155  -.300  -.043   .218   .079  -.149  -.120   .444   .199   .276  1.000
 .099   .061  -.069  -.158  -.097  -.038   .275   .159   .215   .032   .091   .139   .071   .033  -.031  -.145  -.344  1.000
END DATA.
FACTOR MATRIX IN(COR=*)/ ①
  CRITERIA=FACTORS(3)/ ②
  PRINT=CORRELATION DEFAULT/
  PLOT=EIGEN/
  FORMAT=BLANK(.384)/. ③
```

① To read in matrices in FACTOR the matrix subcommand is used. The keyword IN specifies the file from which the matrix is read. The COR=* means we are reading the correlation matrix from the active file.

② This subcommand means we are requesting three factors.

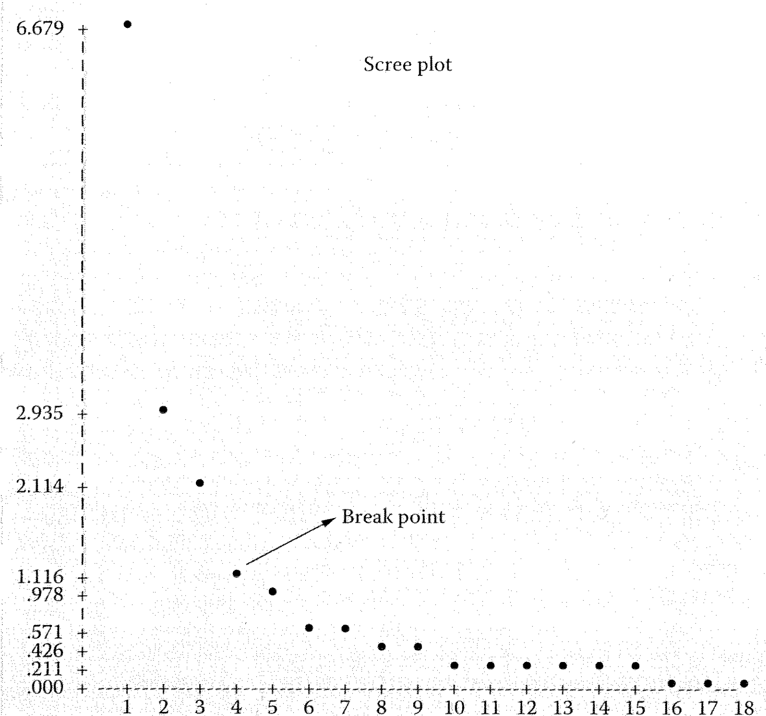
③ The BLANK .384 is very useful for zeroing in the most important loadings. It means that all loadings less than .384 in absolute value will not be printed.

TABLE 11.3

Eigenvalues, Communalities, and Scree Plot for CPI from SPSS Factor Analysis Program

FINAL STATISTICS:

VARIABLE	COMMUNALITY	*	FACTOR	EIGENVALUE	① PCT of VAR	CUM PCT
		*				
DOM	.64619	*	1	6.67904	37.1	37.1
CAPSTAT	.61477	*	2	2.93494	16.3	53.4
SOCIAL	.79929	*	3	2.11381	11.7	② 65.2
SOCPRES	.72447	*				
SELFACP	.79781	*				
WELLBEG	.69046	*				
RESPON	.65899	*				
SOCLIZ	.68243	*				
SELFCTRL	.83300	*				
TOLER	.75739	*				
GOODIMP	.50292	*				
COMMUNAL	.31968	*				
ACHCONF	.72748	*				
ACHINDEP	.69383	*				
INTELEFF	.73794	*				
PSYMIND	.55269	*				
FLEX	.66568	*				
FEMIN	.32275	*				



① The eigenvalue indicates the amount of variance accounted for by each factor.
② The three factors account for 65.2% of total variance.

Exploratory and (

TABLE 11.4

Unrotated Compon

FACTOR MATRIX:

INTELEFF
ACHCONF
TOLER
WELLBEG
ACHINDEP
RESPON
GOODIMP
CAPSTAT
SOCLIZ
SOCIAL
PSYMIND
SELFACP
SELFCTRL
SOCPRES
DOM
FLEX
FEMIN
COMMUNAL
VARIMAX CONVER
ROTATED FACTOR

TOLER
SELFCTRL
ACHINDEP
WELLBEG
INTELEFF
ACHCONF
GOODIMP
PSYMIND
SELFACP
SOCIAL
SOCPRES
DOM
CAPSTAT
FLEX
SOCLIZ
RESPON
FEMIN
COMMUNAL

Note: Only three factc
in SPSS—which
usually within 1
to 0 (see Table 1
tors indicated by
the following su

TABLE 11.4

Unrotated Components Loadings and Varimax Rotated Loadings for California Psychological Inventory

FACTOR MATRIX:

	FACTOR 1	FACTOR 2	FACTOR 3
INTELEFF	.84602		
ACHCONF	.81978		
TOLER	.81618		
WELLBEG	.80596		
ACHINDEP	.67844		−.45209
RESPON	.67775		.38887
GOODIMP	.67347		
CAPSTAT	.64991	.43580	
SOCLIZ	.61110		.41036
SOCIAL	.60980	.60145	
PSYMIND	.57314		−.47158
SELFACP		.82106	
SELFCTRL	.60942	−.67659	
SOCPRES	.51248	.66551	
DOM	.50137	.55616	
FLEX			−.76714
FEMIN			.49437
COMMUNAL			.43941

VARIMAX CONVERGED IN 5 ITERATIONS.

ROTATED FACTOR MATRIX:

	FACTOR 1	FACTOR 2	FACTOR 3
TOLER	.85516		
SELFCTRL	.80528		
ACHINDEP	.80019		
WELLBEG	.78605		
INTELEFF	.77170		
ACHCONF	.70442		
GOODIMP	.68552		
PSYMIND	.66676		
SELFACP		.87923	
SOCIAL		.85542	
SOCPRES		.77968	
DOM		.77396	
CAPSTAT	.40969	.66550	
FLEX			−.76248
SOCLIZ	.53450		.62776
RESPON	.53971		.56861
FEMIN			.56029
COMMUNAL			.47917

Note: Only three factors are displayed in this table, because there is evidence that the Kaiser criterion (the default in SPSS—which yields four factors) can yield too many factors (Zwick & Velicer, 1986), while the scree test is usually within 1 or 2 of true number of factors. Note also that all loadings less than |.384| have been set equal to 0 (see Table 11.2). Both of these are changes from the third edition of this text. To obtain just the three factors indicated by the scree test, you need to insert in the control lines in Table 11.2 after the Print subcommand the following subcommand: CRITERIA MINEIGEN(2)/CRITERIA = FACTORS(3)/

Factor 3 is somewhat dominated by the flexibility variable (loading = −.76248), although the loadings for socialization, responsibility, femininity, and communality are also fairly substantial (ranging from .628 to .479). Low scores on flexibility from the CPI manual characterize an individual as cautious, guarded, mannerly, and overly deferential to authority. High scores on femininity reflect an individual who is patient, gentle, and respectful and accepting of others. Factor 3 thus seems to be measuring a "demure inflexibility in intellectual and social matters."

Before proceeding to another example, we wish to make a few additional points. Nunnally (1978, pp. 433–436) indicated, in an excellent discussion, several ways in which one can be fooled by factor analysis. One point he made that we wish to elaborate on is that of ignoring the simple correlations among the variables after the factors have been derived; that is, not checking the correlations among the variables that have been used to define a factor, to see if there is communality among them in the simple sense. As Nunnally noted, in some cases, variables used to define a factor may have simple correlations near 0.

For our example this is not the case. Examination of the simple correlations in Table 11.2 for the 10 variables used to define Factor 1 shows that most of the correlations are in the moderate to fairly strong range. The correlations among the five variables used to define Factor 2 are also in the moderate to fairly strong range.

An additional point concerning Factor 2 is of interest. The empirical clustering of the variables coincides almost exactly with the logical clustering of the variables given in the CPI manual. The only difference is that Wellbeg is in the logical cluster but not in the empirical cluster (i.e., not on the factor).

Example 11.2: Personality Research Form on SAS

We now consider the interpretation of a principal components analysis and varimax rotation on the Personality Research Form for 231 undergraduate males from a study by Golding and Seidman (1974). The control lines for running the analysis on the SAS FACTOR program and the correlation matrix are presented in Table 11.5. It is important to note here that SAS is different from the other major package (SPSS) in that (a) a varimax rotation is *not* a default option—the default is no rotation, and (b) the Kaiser criterion (retaining only those factors whose eigenvalues are >1) is not a default option. In Table 11.5 we have requested the Kaiser criterion be used by specifying MINEIGEN = 1.0, and have requested the varimax rotation by specifying ROTATE = VARIMAX.

To indicate to SAS that we are inputting a correlation matrix, the TYPE = CORR in parentheses after the name for the data set is necessary. The TYPE = 'CORR' on the next line is also required. Note that the name for each variable precedes the correlations for it with all the other variables. Also, note that there are 14 periods for the ABASE variable, 13 periods for the ACH variable, 12 periods for AGGRESS, and so on. These periods need to be inserted. Finally, the correlations for each row of the matrix must be on a separate record. Thus, although we may need two lines for the correlations of ORDER with all other variables, once we put the last correlation there (which is a 1) we must start the correlations for the next variable (PLAY) on a *new* line. The same is true for the SPSS FACTOR program.

The CORR in this statement yields the correlation matrix for the variables. The FUZZ = .34 prints correlations and factor loadings with absolute value less than .34 as missing values. Our purpose in using FUZZ is to think of values <|.34| as chance values, and to treat them as 0. The SCREE is inserted to obtain Cattell's scree test, useful in determining the number of factors to retain.

The first part of the printout appears in Table 11.6, and the output at the top indicates that according to the Kaiser criterion only four factors will be retained because there are only four eigenvalues >1. Will the Kaiser criterion accurately identify the true number of factors in this case? To answer this question it is helpful to refer back to the Hakstian et al. (1982) study cited earlier. They noted that for $N > 250$ and a mean communality >.60, the Kaiser criterion is accurate. Because the total of the communality estimates in Table 11.6 is given as 9.338987, the mean communality here is 9.338987/15 = .622. Although N is not >250, it is close (N = 231), and we feel the Kaiser rule will be accurate.

TABLE 11.5
SAS Factor Control

```
DATA PRF(TYPE = C
TYPE = 'CORR';
INPUT NAME $ ABA
ENDUR EXHIB HAR
CARDS;
ABASE      1.0
ACH         .01
AGGRESS    −.32
AUTON       .13
CHANGE      .15
COGSTR     −.23
DEF        −.42
DOMIN      −.22
ENDUR       .01
EXHIB      −.09
HARAVOD    −.22
IMPLUS      .14
NUTUR       .33
ORDER      −.11
PLAY        .05
PROC FACTOR COR
```

TABLE 11.6
Eigenvalues and Sc

	1
Eigenvalue	3.16
Difference	0.68
Proportion	0.21
Cumulative	0.21
	9
Eigenvalue	0.541
Difference	0.10
Proportion	0.03
Cumulative	0.85

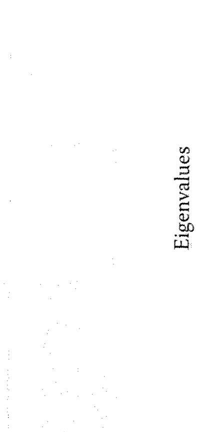


TABLE 11.5
SAS Factor Control Lines for Components Analysis and Varimax Rotation on the Personality Research Form

```
DATA PRF(TYPE = CORR);
TYPE = 'CORR';
INPUT NAME $ ABASE ACH AGGRESS AUTON CHANGE COGSTR DEF DOMIN
ENDUR EXHIB HARAVOD IMPLUS NUTUR ORDER PLAY;
CARDS;
ABASE     1.0  ...
ACH        .01  1.0  ...
AGGRESS   -.32 -.08  1.0  ...
AUTON      .13  .03   .04  1.0  ...
CHANGE     .15  .09   .06   .28  1.0  ...
COGSTR    -.23  .22   .02  -.17 -.27  1.0  ...
DEF       -.42  .06   .57   .04 -.01   .14  1.0  ...
DOMIN     -.22  .37   .25   .08  .17  -.05   .32  1.0  ...
ENDUR      .01  .65  -.11   .09  .03   .20   .02  .39  1.0  ...
EXHIB     -.09  .13   .28  -.07  .15  -.24   .10  .52   .08  1.0  ...
HARAVOD   -.22 -.02  -.01  -.28 -.33   .45   .08 -.21  -.08 -.22  1.0  ...
IMPLUS     .14 -.16   .30   .16  .33  -.46   .14  .07  -.23  .34  -.31  1.0  ...
NUTUR      .33  .30  -.23  -.24  .03  -.05  -.19  .16   .20  .22  -.04   .04  1.0  ...
ORDER     -.11  .29   .01  -.13 -.17   .53   .09  .08   .27 -.11   .22  -.35  0.0  1.0  ...
PLAY       .05 -.25   .27  -.02  .12  -.31  -.02  .11  -.27  .43  -.26   .48 -.10 -.25  1.0
PROC FACTOR CORR FUZZ = .34 MINEIGEN = 1.0 REORDER ROTATE = VARIMAX SCREE;
```

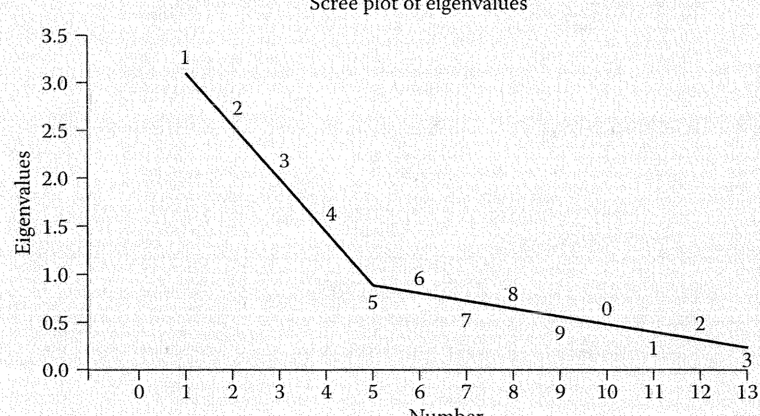
TABLE 11.6
Eigenvalues and Scree Plot from the SAS Factor Program for Personality Research Form

Eigenvalues of the Correlation Matrix: Total = 15 Average = 1

	1	2	3	4	5	6	7	8
Eigenvalue	3.1684	2.4821	2.2464	1.4422	0.8591	0.8326	0.6859	0.6047
Difference	0.6862	0.2358	0.8042	0.5830	0.0266	0.1466	0.0812	0.0636
Proportion	0.2112	0.1655	0.1498	0.0961	0.0573	0.0555	0.0457	0.0403
Cumulative	0.2112	0.3767	0.5265	0.6226	0.6799	0.7354	0.7811	0.8214

	9	10	11	12	13	14	15
Eigenvalue	0.5411	0.4382	0.4060	0.3826	0.3283	0.3108	0.2717
Difference	0.1029	0.0322	0.0234	0.0543	0.0175	0.0391	
Proportion	0.0361	0.0292	0.0271	0.0255	0.0219	0.0207	0.0181
Cumulative	0.8575	0.8867	0.8867	0.9393	0.9612	0.9819	1.0000

Scree plot of eigenvalues



The scree plot in Table 11.6 also supports using four factors, because the break point occurs at the fifth eigenvalue. That is, the eigenvalues level off from the fifth eigenvalue on. To further support the claim of four true factors, note that the Q/P ratio is $4/15 = .267 < .30$, and Hakstian et al. (1982) indicated that when this is the case the estimate of the number of factors will be just that much more credible.

To interpret the four factors, the sorted, rotated loadings in Table 11.7 are very useful. Referring back to Table 11.1, we see that the critical value for a significant loading at the .01 level is $2(.17) = .34$. So, we certainly would not want to pay any attention to loadings less than .34 in absolute value. That is why we have had SAS print those loadings as a period. This helps to sharpen our focus on the salient loadings. The loadings that most strongly characterize the first three factors (and are of the same order of magnitude) are boxed in on Table 11.7. In terms of interpretation, Factor 1 represents an "unstructured, free spirit tendency," with the loadings on Factor 2 suggesting a "structured, hard driving tendency" construct. Factor 3 appears to represent a "non-demeaning aggressive tendency," while the loadings on Factor 4, which are dominated by the very high loading on autonomy, imply a "somewhat fearless tendency to act on one's own."

As mentioned in the first edition of this text, it would help if there were a statistical test, even a rough one, for determining when one loading on a factor is significantly greater than another loading on the same factor. This would then provide a more solid basis for including one variable in the interpretation of a factor and excluding another, assuming we can be confident that both are nonchance loadings. I remain unaware of such a test.

Example 11.3: Regression Analysis on Factor Scores—SAS and SPSS

We mentioned earlier in this chapter that one of the uses of components analysis is to reduce the number of predictors in regression analysis. This makes good statistical and conceptual sense for several reasons. First, if there is a fairly large number of initial predictors (say 15), we are undoubtedly not measuring 15 different constructs, and hence it makes sense to determine what the main constructs are that we are measuring. Second, this is desirable from the viewpoint of scientific parsimony. Third, we reduce from 15 initial predictors to, say, four new predictors (the components or rotated factors), our N/k ratio increases dramatically and this helps cross-validation prospects considerably. Fourth, our new predictors are uncorrelated, which means we have eliminated multicollinearity, which is a major factor in causing unstable regression equations. Fifth, because the new predictors are uncorrelated, we can talk about the unique contribution of each predictor in accounting for variance on y; that is, there is an unambiguous interpretation of the importance of each predictor.

We illustrate the process of doing the components analysis on the predictors and then passing the factor scores (as the new predictors) for a regression analysis for both SAS and SPSS using the National Academy of Science data introduced in Chapter 3 on multiple regression. Although there is not a compelling need for a factor analysis here because there are just six predictors, this example is simply meant to show the process. The new predictors, that is, the retained factors, will then be used to predict quality of the graduate psychology program. The control lines for doing both the factor analysis and the regression analysis for both packages are given in Table 11.8.

Note in the SAS control lines that the output data set from the principal components procedure contains the original variables *and* the factor scores for the first two components. It is this data set that we are accessing in the PROC REG procedure. Similarly, for SPSS the factor scores for the first two components are saved and *added* to the active file (as they call it), and it is this file that the regression procedure is dealing with.

So that the results are comparable for the SAS and SPSS runs, a couple of things *must* be done. First, as mentioned in Table 11.8, one must insert STANDARD into the control lines for SAS, so that the components have a variance of 1, as they have by default for SPSS. Second, because SPSS does a varimax rotation by default and SAS does not, we must insert the subcommand ROTATION=NOROTATE into the SPSS control lines so that is the principal components scores that are being used by the regression procedure in each case. If one does not insert the NOROTATE subcommand, then the regression analysis will use the *rotated* factors as the predictors.

TABLE 11.7

Factor Loading and

TABLE 11.7

Factor Loading and Rotated, Sorted Loadings for Personality Research Form

Factor Pattern

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
IMPLUS	0.76960	•	•	•
PLAY	0.66312	•	•	•
CHANGE	0.46746	•	•	−0.36271
HARMAVOD	−0.58060	•	−0.35665	•
ORDER	−0.60035	•	•	•
COGSTR	−0.73891	•	•	•
DOMIN	•	0.80853	•	•
ACH	•	0.61394	0.48781	•
ENDUR	•	0.57943	0.49114	•
EXHIB	0.48854	0.53279	•	0.44574
ABASE	•	−0.37413	0.62691	•
NUTUR	•	•	0.60007	0.52851
DEF	•	0.54265	−0.56778	•
AGGRESS	•	0.45762	−0.61053	•
AUTON	•	•	•	−0.77911

NOTE: Values less than 0.34 have been printed as (•).

Variance explained by each

FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
3.168359	2.482114	2.246351	1.442163

Final Community Estimates: Total = 9.338987

ABASE	ACH	AGGRESS	AUTON	CHANGE	COGSTR	DEF	DOMIN
0.567546	0.715861	0.670982	0.701144	0.448672	0.624114	0.644643	0.701961
ENDUR	EXHIB	HARMAVOD	IMPLUS	NUTUR	ORDER	PLAY	
0.713278	0.724334	0.537959	0.502875	0.659155	0.452917	0.573546	

Rotated Factor Pattern

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
PLAY	0.73149	•	•	•
IMPLUS	0.73013	•	•	•
EXHIB	0.66060	0.47003	•	•
ORDER	−0.53072	•	•	•
COGSTR	−0.66102	•	•	•
ASH	•	0.78676	•	•
ENDUR	•	0.75731	•	•
DOMIN	•	0.71173	0.35986	•
NUTUR	•	0.51149	−0.50100	•
DEF	•	•	0.79311	•
AGGRESS	•	•	0.76624	•
ABASE	•	•	−0.71271	•
AUTON	•	•	•	0.83214
CHANGE	•	•	•	0.57560
HARMAVOD	−0.44237	•	•	−0.53376

Variance explained by each

FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
2.891095	2.405032	2.297653	1.745206

TABLE 11.8

SAS and SPSS Control Lines for Components Analysis on National Academy of Science Data and Then Passing Factor Scores for a Regression Analysis

SAS

```
   DATA REGRESS;
   INPUT QUALITY NFACUL NGRADS PCTSUPP PCTGRT NARTIC PCTPUB @@;
   CARDS;
   DATA IN BACK OF TEXT
①  PROC PRINCOMP N = 2 STANDARD OUT = FSCORES;
②  VAR = NFACUL NGRADS PCTSUPP PCTGRT NARTIC PCTPUB; PROC REG DATA = FSCORES;
③  MODEL QUALITY = PRIN1 PRIN2;
   SELECTION = STEPWISE;
   PROC PRINT DATA = FSCORES;
```

SPSS

```
   DATA LIST FREE/QUALITY NFACUL NGRADS PCTSUPP PCTGRT NARTIC
      PCTPUB.
   BEGIN DATA.
      DATA IN BACK OF TEXT
   END DATA.
④  FACTOR VARIABLES = NFACUL TO PCTPUB/
⑤     ROTATION = NOROTATE/
⑥     SAVE REG (ALL FSCORE)/.
   LIST.
   REGRESSION DESCRIPTIVES = DEFAULT/
⑦     VARIABLES = QUALITY FSCOREl FSCORE2/
      DEPENDENT = QUALITY/
      METHOD = STEPWISE/.
```

① The N = 2 specifies the number of components to be computed; here we just want two. STANDARD is necessary for the components to have variance of 1; otherwise the variance will equal the eigenvalue for the component (see *SAS STAT User's Guide,* Vol. 2, p. 1247). The OUT data set (here called FSCORES) contains the original variables and the component scores.
② In this VAR statement we "pick off" just those variables we wish to do the components analysis on, that is, the predictors.
③ The principal component variables are denoted by default as PRIN1, PRIN2, etc.
④ Recall that TO enables one to refer to a consecutive string of variables more concisely.
⑤ By default in SPSS the VARIMAX rotation would be done, and the factor scores obtained would be those for the rotated factors. Therefore, we specify NOROTATE so that no rotation is done.
⑥ There are three different methods for computing factor scores, but for components analysis they all yield the same scores. Thus, we have used the default method REG (regression method).
⑦ In saving the factor scores we have used the rootname FSCORE; the maximum number of characters for this name is 7. This rootname is then used along with a number to refer to consecutive factor scores. Thus, FSCORE1 for the factor scores on component 1, FSCORE2 for the factor scores on component 2, etc.

Example 11.4: MANOVA on Factor Scores—SAS and SPSS

In Table 11.9 we illustrate a components analysis on a hypothetical set of seven variables, and then pass the first two components to do a two-group MANOVA on these "new" variables. Because the components are uncorrelated, one might argue for performing just three univariate tests, for in this case an exact estimate of overall α is available from $1 - (1 - .05)^3 = .145$. Although an exact estimate is available, the multivariate approach covers a possibility that the univariate approach would miss, that is, the case where there are small nonsignificant differences on each of the variables, but cumulatively (with the multivariate test) there is a significant difference.

TABLE 11.9

SAS and SPSS Control Lines for Components Analysis on Set of Dependent Variables and Then Passing Factor Scores for Two-Group MANOVA

SAS

```
DATA MANOVA;
INPUT GP X1 X2 X3 X4 X5 X6 X7;
CARDS;
1 23  4 45 43 34  8 89 1 34 46 54 46 27  6  93
1 31 34 45 43 56  5 78 1 36  8 65 57 56  3 104
1 43 56 67 54 67 78 92 1 23 43 54 76 54  2 112
2 21 32 65 47 65 56 69 2 34 54 32 45 67 65  74
2 31 23 43 45 76 86 61 2 17 23 43 25 46 65  66
PROC PRINCOMP N = 2 STANDARD OUT = FSCORES;
VAR X1 X2 X3 X4 X5 X6 X7;
PROC GLM DATA = FSCORES;
MODEL PRIN1 PRIN2 = GP;
MANOVA H = GP;
PROC PRINT DATA = FSCORES;
```

SPSS

```
DATA LIST FREE/GP X1 X2 X3 X4 X5 X6 X7.
BEGIN DATA.
1 23  4 45 43 34  8 89 1 34 46 54 46 27  6  93
1 31 34 45 43 56  5 78 1 36  8 65 57 56  3 104
1 43 56 67 54 67 78 92 1 23 43 54 76 54  2 112
2 21 32 65 47 65 56 69 2 34 54 32 45 67 65  74
2 31 23 43 45 76 86 61 2 17 23 43 25 46 65  66
END DATA.
FACTOR VARIABLES = X1 TO X7/
   ROTATION = NOROTATE/
   SAVE REG (ALL FSCORE)/.
LIST.
MANOVA FSCORE1 FSCORE2 BY GP(1,2)/.
```

Also, if we had done an oblique rotation, and hence were passing correlated factors, then the case for a multivariate analysis is even more compelling because an exact estimate of overall α is not available. Another case where some of the variables would be correlated is if we did a factor analysis and retained three factors and two of the original variables (which were relatively independent of the factors). Then there would be correlations between the original variables retained and between those variables and the factors.

11.9 The Communality Issue

In principal components analysis we simply transform the original variables into linear combinations of these variables, and often three or four of these combinations (i.e., the components) account for most of the total variance. Also, we used 1's in the diagonal of the correlation matrix. Factor analysis per se differs from components analysis in two ways: (a) The hypothetical factors that are derived can only be *estimated* from the original variables, whereas in components analysis, because the components are specific linear combinations, no estimate is involved, and (b) numbers less than 1, called communalities, are put in the main diagonal of the correlation matrix in factor analysis. A relevant question is, "Will different factors emerge if 1's are put in the main diagonal (as in components analysis) than will emerge if communalities (the squared multiple correlation of each variable with all the others is one of the most popular) are placed in the main diagonal?"

The following quotes from five different sources give a pretty good sense of what might be expected in practice. Cliff (1987) noted that, "the choice of common factors or components methods often makes virtually no difference to the conclusions of a study" (p. 349). Guadagnoli and Velicer (1988) cited several studies by Velicer et al. that "have demonstrated that principal components solutions differ little from the solutions generated from factor analysis methods" (p. 266).

Harman (1967) stated, "As a saving grace, there is much evidence in the literature that for all but very small sets of variables, the resulting factorial solutions are little affected by the particular choice of communalities in the principal diagonal of the correlation matrix" (p. 83). Nunnally (1978) noted, "It is very safe to say that if there are as many as 20 variables in the analysis, as there are in nearly all exploratory factor analyses, then it does not matter what one puts in the diagonal spaces" (p. 418). Gorsuch (1983) took a somewhat more conservative position: "If communalities are reasonably high (e.g., .7 and up), even unities are probably adequate communality estimates in a problem with more than 35 variables" (p. 108). A general, somewhat conservative conclusion from these is that when the number of variables is moderately large (say >30), and the analysis contains virtually no variables expected to have low communalities (e.g., .4), then practically any of the factor procedures will lead to the same interpretations. Differences can occur when the number of variables is fairly small (<20), and some communalities are low.

11.10 A Few Concluding Comments

We have focused on an internal criterion in evaluating the factor solution, i.e., how interpretable the factors are. However, an important external criterion is the reliability of the solution. If the sample size is large, then one should randomly split the sample to check the consistency (reliability) of the factor solution on both random samples. In checking to determine whether the same factors have appeared in both cases it is not sufficient to just examine the factor loadings. One needs to obtain the correlations between the factor scores for corresponding pairs of factors. If these correlations are high, then one may have confidence of factor stability.

Finally, there is the issue of "factor indeterminancy" when estimating factors as in the common factor model. This refers to the fact that the factors are not uniquely determined. The importance of this for the common factor model has been the subject of much hot debate in the literature. We tend to side with Steiger (1979), who stated, "My opinion is that indeterminacy and related problems of the factor model counterbalance the model's theoretical advantages, and that the elevated status of the common factor model (relative to, say, components analysis) is largely undeserved" (p. 157).

11.11 Exploratory and Confirmatory Factor Analysis

The principal component analyses presented previously in this chapter are a form of what are commonly termed *exploratory factor analyses* (EFAs). The purpose of exploratory analysis is to identify the factor structure or model for a set of variables. This often involves determining how many factors exist, as well as the pattern of the factor loadings. Although most EFA programs allow for the number of factors to be specified in advance, it is not possible in these programs to force variables to load only on certain factors. EFA is generally considered to be more of a theory-generating than a theory-testing procedure. In contrast, *confirmatory factor analysis* (CFA) is generally based on a strong theoretical or empirical foundation that allows the researcher to specify an exact factor model in advance. This model usually specifies which variables will load on which factors, as well as such things as which factors are correlated. It is more of a theory-testing procedure than is EFA. Although, in practice, studies may contain aspects of both exploratory and confimatory analyses, it is useful to distinguish between the two techniques in terms of the situations in which they are commonly used. The following table displays some of the general differences between the two approaches.

Exploratory—Theory Generating	Confirmatory—Theory Testing
Heuristic—weak literature base	Strong theory or strong empirical base
Determine the number of factors	Number of factors fixed *a priori*
Determine whether the factors are correlated or uncorrelated	Factors fixed *a priori* as correlated or uncorrelated
Variables free to load on all factors	Variables fixed to load on a specific factor or factors

Let us consider an example of an EFA. Suppose a researcher is developing a scale to measure self-concept. The researcher does not conceptualize specific self-concept factors in advance, and simply writes a variety of items designed to tap into various aspects of self-concept. An EFA or components analysis of these items may yield three factors that the researcher then identifies as physical (PSC), social (SSC), and academic (ASC) self-concept. The researcher notes that items with large loadings on one of the three factors tend to have very small loadings on the other two, and interprets this as further support for the presence of three distinct factors or dimensions underlying self-concept.

A less common variation on this EFA example would be one in which the researcher had hypothesized the three factors *a priori* and intentionally written items to tap each dimension. In this case, the EFA would be carried out in the same way, except that the researcher might specify in advance that three factors should be extracted. Note, however, that in both of these EFA situations, the researcher would *not* be able to force items to load on certain factors, even though in the second example the pattern of loadings was hypothesized in advance. Also, there is no overall statistical test to help the researcher determine whether the observed pattern of loadings confirms the three factor structure. Both of these are limitations of EFA.

Before we turn to how a CFA would be done for this example, it is important to consider examples of the types of situations in which CFA would be appropriate; that is, situations in which a strong theory or empirical base exists.

11.11.1 Strong Theory

The four-factor model of self-concept (Shavelson, Hubner, and Stanton, 1976), which includes general self-concept, academic self-concept, English self-concept, and math self-concept, has a strong underlying theory. This model was presented and tested by Byrne (1994).

11.11.2 Strong Empirical Base

The "big five" factors of personality—extraversion, agreeableness, conscientiousness, neuroticism, and intellect—is an example. Goldberg (1990), among others, provided some strong empirical evidence for the five-factor trait model of personality. The five-factor model is not without its critics; see, for example, Block (1995). Using English trait adjectives obtained from three studies, Goldberg employed five different EFA methods, each one rotated orthogonally and obliquely, and found essentially the same five uncorrelated factors or personality in each analysis. Another confirmatory analysis of these five personality factors by Church and Burke (1994) again found evidence for the five factors, although these authors concluded that some of the factors may be correlated.

The Maslach Burnout Inventory was examined by Byrne (1994), who indicated that considerable empirical evidence exists to suggest the existence of three factors for this instrument. She conducted a confirmatory factor analysis to test this theory.

In this chapter we consider what are called by many people "measurement models." As Jöreskog and Sörbom put it (1993, p. 15), "The purpose of a measurement model is to describe how well the observed indicators serve as a measurement instrument for the latent variables."

Karl Jöreskog (1967, 1969; Jöreskog & Lawley, 1968) is generally credited with overcoming the limitations of exploratory factor analysis through his development of confirmatory factor analysis. In CFA, researchers can specify the structure of their factor models *a priori*, according to their theories about how the variables ought to be related to the factors. For example, in the second EFA situation just presented, the researcher could constrain the ASC items to load on the ASC factor, and to have loadings of zero on the other two factors; the other loadings could be similarly constrained.

Figure 11.1 gives a pictorial representation of the hypothesized three-factor structure. This type of representation, usually referred to as a *path model*, is a common way of showing the hypothesized or actual relationships among observed variables and the factors they were designed to measure.

The path model shown in Figure 11.1 indicates that three factors are hypothesized, as represented by the three circles. The curved arrows connecting the circles indicate that all three factors are hypothesized to be correlated. The items are represented by squares and are connected to the factors by straight arrows, which indicate causal relationships.

In CFA, each observed variable has an error term associated with it. These error terms are similar to the residuals in a regression analysis in that they are the part of each observed variable that is not explained by the factors. In CFA, however, the error terms also contain measurement error due to the lack of reliability of the observed variables.

The error terms are represented by the symbol δ in Figure 11.1 and are referred to in this chapter as *measurement errors*. The straight arrows from the δ's to the observed variables indicate that the observed variables are influenced by measurement error in addition to being influenced by the factors.

We could write equations to specify the relationships of the observed variables to the factors and measurement errors. These equations would be written as:

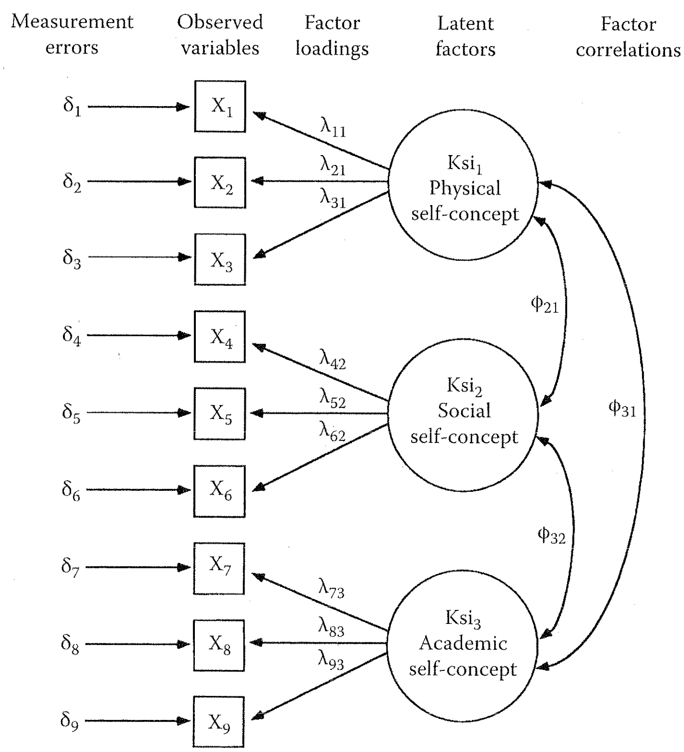


FIGURE 11.1
Three-factor self-concept model with three indicators per factor.

$$X = \lambda\xi + \delta$$

where the symbol λ stands for a factor loading and the symbol ξ represents the factor itself. This is similar to the regression equation

$$Y = \beta X + e$$

where β corresponds to λ and e corresponds to δ. One difference between the two equations is that in the regression equation, X and Y are both observed variables, whereas in the CFA equation, X is an observed variable but ξ is a latent factor. One implication of this is that we cannot obtain solutions for the values of λ and δ through typical regression methods. Instead, the correlation or covariance matrix of the observed variables is used to find solutions for elements of the matrices. This matrix is usually symbolized by **S** for a sample matrix and Σ for a population matrix. The relationships between the elements of **S** or Σ and the elements of λ, ξ, and δ can be obtained by expressing each side of the equation

$$X = \lambda\xi + \delta$$

as a covariance matrix. The algebra is not presented here (cf. Bollen, 1989, p. 35), but results in the following equality:

$$\Sigma = \lambda\phi\lambda' + \theta\delta$$

where ϕ is a matrix of correlations or covariances among the factors (ξs) and $\theta\delta$ is a matrix of correlations or covariances among the measurement error terms. Typically, $\theta\delta$ is a diagonal matrix, containing only the variances of the measurement errors. This matrix equation shows that the covariances among the X variables (Σ) can be broken down into the CFA matrices λ, ϕ, and $\theta\delta$. It is this equation that is solved to find values for the elements of λ, ϕ, and $\theta\delta$.

As the first step in any CFA, the researcher must therefore fully specify the structure or form of the matrices λ, ϕ, and $\theta\delta$ in terms of which elements are to be included. In our example, the λ matrix would be specified to include only the loadings of the three items designated to measure each factor, represented in Figure 11.1 by the straight arrows from the factors to the variables. The ϕ matrix would include all of the factor correlations, represented by the curved arrows between each pair of factors in Figure 11.1. Finally, one measurement error variance for each item would be estimated.

These specifications are based on the researcher's theory about the relationships among the observed variables, latent factors, and measurement errors. This theory may be based on previous empirical research, the current thinking in a particular field, the researcher's own hypotheses about the variables, or any combination of these. It is essential that the researcher be able to base a model on theory, however, because, as we show later, it is not always possible to distinguish between different models on statistical grounds alone. In many cases, theoretical considerations are the only way in which one model can be distinguished from another.

In the following sections, two examples using the LISREL program's (Jöreskog & Sörbom, 1986, 1988, 1993) new simplified language, known as SIMPLIS, are presented and discussed in order to demonstrate the steps involved in carrying out a CFA. Because CFAs always involve the analysis of a covariance or correlation matrix, we begin in Section 11.12 with a brief introduction to the PRELIS program that has been designed to create matrices that LISREL can easily use.

11.12 PRELIS

The PRELIS program is sometimes referred to as a "preprocessor" for LISREL. The PRELIS program is usually used by researchers to prepare covariance or correlation matrices that can be analyzed by LISREL. Although correlation and covariance matrices can be output from statistics packages such as SPSS or SAS, the PRELIS program has been especially designed to prepare data in a way that is compatible with the LISREL program, and has several useful features.

PRELIS 1 was introduced in 1986, and was updated in 1993 with the introduction of PRELIS 2. PRELIS 2 offers several features that were unavailable in PRELIS 1, including facilities for transforming and combining variables, recoding, and more options for handling missing data. Among the missing data options is an imputation procedure in which values obtained from a case with a similar response pattern on a set of matching variables are substituted for missing values on another case (see Jöreskog & Sörbom, 1996, p. 77 for more information). PRELIS 2 also offers tests of univariate and multivariate normality. As Jöreskog and Sörbom noted (1996, p. 168), "For each continuous variable, PRELIS 2 gives tests of zero skewness and zero kurtosis. For all continuous variables, PRELIS 2 gives tests of zero multivariate skewness and zero multivariate kurtosis." Other useful features of the

TABLE 11.10

PRELIS Command Lines for Health Beliefs Model Example

```
Title: Amlung Dissertation: Health Belief Model; Correlated Factors
da ni=27 no=527                                                    ①
ra fi=a:\amlung.dta fo                                             ②
(27f1.0)
la                                                                 ③
sus1 sus2 sus3 sus4 sus5 ser1 ser2 ser3 ser4 ser5 ser6 ser7 ser8 ben1
ben4 ben7 ben10 ben11 ben12 ben13 bar1 bar2 bar3 bar4 bar5 bar6 bar7
mi all (0)                                                         ④
co all                                                             ⑤
ou cm=amlung.cov                                                   ⑥
```

① The da (data) line specifies that there are 27 variables (ni=27) and 527 observations. The number of observations need not be included here—PRELIS will count them if this information is not given.
② The raw data (ra) are read from a file called "amlung.dta" that is on a disk on the a drive. The command "fo" indicates that a FORTRAN type format will be given on the next line. This must be enclosed in parentheses. The format 27f1.0 means that there are 27 variables, each taking up one column, with no decimals.
③ Names of up to 8 characters can be given to the variables.
④ Missing values for all variables are coded as 0.
⑤ The variables are all declared to be continuous (co).
⑥ A covariance matrix (cm) is requested, which will be written to a file called "amlung.cov."

PRELIS 2 program include facilities for conducting bootstrapping procedures and Monte Carlo or simulation studies. These procedures are described in the PRELIS 2 manual (Jöreskog & Sörbom, 1996, Appendix C, pp. 185–206). Another improvement implemented in PRELIS 2 has to do with the computation of the weight matrix needed for weighted least squares (WLS) estimation. The weight matrix computed in PRELIS 1 was based on a simplifying assumption that was later found to yield inaccurate results. This has been corrected in PRELIS 2.

Although LISREL can read in raw data, it has no facilities for data screening or for handling missing values. For this reason, most researchers prefer to use programs such as PRELIS to create their covariance matrix, which can then be easily read into LISREL. The PRELIS program can read in raw data and compute various covariance matrices as well as various types of correlation matrices (Pearson, polychoric, polyserial, tetrachoric, etc.). At the same time, PRELIS will compute descriptive statistics, handle missing data, perform data transformations such as recoding or transforming variables, and provide tests of normality assumptions.

Table 11.10 shows the PRELIS command lines used to create the covariance matrix used by Amlung (1996) in testing two competing CFA models of the Health Belief Model (HBM). In this study, Amlung reanalyzed data from Champion and Miller's 1996 study in which 527 women responded to items designed to measure the four theoretically derived HBM dimensions of seriousness, susceptibility, benefits, and barriers. Through preliminary reliability analyses and EFAs, Amlung selected 27 of the HBM items with which to test two CFA models.

The PRELIS language is not case sensitive; either upper- or lowercase letters can be used. Note that unless the raw data are in free format, with at least one space between each variable, a FORTRAN format, enclosed in parentheses, must be given in the line directly after the "ra" line. This is indicated by the keyword "fo" on the "ra" line. Those readers who are unfamiliar with this type of format are encouraged to refer to the examples given in the PRELIS manual.

In addition to the covariance matrix, which is written to an external file, an output file containing descriptive statistics and other useful information is created when the PRELIS program is run. Selected output for the HBM example is shown in Table 11.11.

As can be seen in Table 11.11, some of the HBM items have fairly high levels of non-normality. PRELIS provides statistical tests of whether the distributions of the individual variables are significantly skewed and kurtotic. For example, in looking at the first part of the table, we can see that the variable SER1 has a skewness value of −2.043 and a kurtosis value of 7.157. In the next section of the table we see that these skewness and kurtosis values resulted in highly significant *z* values of −4.603 and 9.202, respectively. These values indicate that the distribution of the item SER1 deviates significantly from normality with regard to both skewness and kurtosis. This is confirmed by the highly significant

TABLE 11.11

PRELIS 2 Output for Health Belief Model

TOTAL SAMPLE SIZE = 527
UNIVARIATE SUMMARY STATISTICS FOR CONTINUOUS VARIABLES

VARIABLE	MEAN	S. DEV.	SKEW	KURT	MIN	FREQ	MAX	FREQ
SUS1	2.528	0.893	0.448	0.131	1.000	52	5.000	13
SUS2	2.512	0.843	0.315	0.204	1.000	51	5.000	9
SUS3	2.615	0.882	0.216	−0.419	1.000	43	5.000	6
SUS4	2.510	0.953	0.638	−0.124	1.000	51	5.000	15
SUS5	2.493	1.032	0.685	−0.240	1.000	65	5.000	22
SER1	4.539	0.657	−2.043	7.157	1.000	5	5.000	314
SER2	4.220	0.837	−1.331	2.310	1.000	7	5.000	216
SER3	3.421	1.054	−0.261	−0.712	1.000	16	5.000	82
SER4	2.979	1.124	0.089	−1.090	1.000	36	5.000	42
SER5	3.789	0.891	−0.707	0.155	1.000	4	5.000	99
SER6	2.643	1.126	0.374	−0.695	1.000	78	5.000	33
SER7	3.268	1.085	−0.180	−1.057	1.000	17	5.000	58
SER8	2.421	0.952	0.811	0.439	1.000	63	5.000	20
BEN1	3.824	0.671	−0.765	1.780	1.000	3	5.000	57
BEN4	3.715	0.729	−0.865	0.617	2.000	46	5.000	40
BEN7	3.486	0.804	−0.417	0.263	1.000	7	5.000	38
BEN10	4.021	0.679	−1.121	3.180	1.000	3	5.000	100
BEN11	3.888	0.804	−1.114	1.779	1.000	6	5.000	90
BEN12	3.759	0.898	−0.897	0.586	1.000	8	5.000	84
BEN13	4.066	0.627	−1.258	5.096	1.000	4	5.000	100
BAR1	2.408	0.996	0.587	−0.303	1.000	82	5.000	12
BAR2	2.125	0.818	0.896	0.812	1.000	95	5.000	2
BAR3	1.943	0.763	0.947	1.478	1.000	138	5.000	2
BAR4	1.913	0.644	0.811	2.328	1.000	118	5.000	1
BAR5	1.937	0.731	0.977	2.072	1.000	131	5.000	3
BAR6	3.224	1.116	−0.368	−0.968	1.000	34	5.000	44
BAR7	1.808	0.616	1.220	5.601	1.000	142	5.000	4

TABLE 11.11 (continued)

PRELIS 2 Output for Health Belief Model

TEST OF UNIVARIATE NORMALITY FOR CONTINUOUS VARIABLES

	SKEWNESS		KURTOSIS		SKEWNESS AND KURTOSIS	
	Z-SCORE	P-VALUE	Z-SCORE	P-VALUE	CHI-SQUARE	P-VALUE
SUS1	2.813	0.002	0.742	0.229	8.466	0.015
SUS2	2.401	0.008	1.041	0.149	6.848	0.033
SUS3	1.971	0.024	−2.361	0.009	9.458	0.009
SUS4	3.235	0.001	−0.472	0.318	10.685	0.005
SUS5	3.320	0.000	−1.138	0.128	12.316	0.002
SER1	−4.630	0.000	9.202	0.000	106.113	0.000
SER2	−4.115	0.000	5.694	0.000	49.351	0.000
SER3	−2.184	0.014	−5.186	0.000	31.659	0.000
SER4	1.082	0.140	−12.904	0.000	167.682	0.000
SER5	−3.357	0.000	0.841	0.200	11.979	0.003
SER6	2.601	0.005	−4.983	0.000	31.597	0.000
SER7	−1.768	0.039	−11.794	0.000	142.221	0.000
SER8	3.521	0.000	1.889	0.029	15.965	0.000
BEN1	−3.450	0.000	4.933	0.000	36.237	0.000
BEN4	−3.598	0.000	2.445	0.007	18.920	0.000
BEN7	−2.730	0.003	1.269	0.102	9.061	0.011
BEN10	−3.909	0.000	6.672	0.000	59.792	0.000
BEN11	−3.902	0.000	4.931	0.000	39.536	0.000
BEN12	−3.642	0.000	2.352	0.009	18.797	0.000
BEN13	−4.047	0.000	8.151	0.000	82.823	0.000
BAR1	3.134	0.001	−1.540	0.062	12.192	0.002
BAR2	3.641	0.000	2.978	0.001	22.125	0.000
BAR3	3.706	0.000	4.420	0.000	33.270	0.000
BAR4	3.521	0.000	5.717	0.000	45.085	0.000
BAR5	3.744	0.000	5.372	0.000	42.876	0.000
BAR6	−2.583	0.005	−9.443	0.000	95.833	0.000
BAR7	4.011	0.000	8.446	0.000	87.425	0.000

TEST OF MULTIVARIATE NORMALITY FOR CONTINUOUS VARIABLES

SKEWNESS		KURTOSIS		SKEWNESS AND KURTOSIS	
Z-SCORE	P-VALUE	Z-SCORE	P-VALUE	CHI-SQUARE	P-VALUE
83.599	0.000	36.469	0.000	8318.704	0.000

chi-square value of 106.113, which is a combined test of both skewness and kurtosis. Finally, tests of multivariate skewness and kurtosis, both individually and in combination, are given. For the HBM data, these tests indicate significant departures from multivariate normality that may bias the tests of fit for this model (see, e.g., West, Finch, & Curran, 1995; Muthén & Kaplan, 1992).

In section 11.13 a LISREL 8 example using the HBM data is presented in order to demonstrate the steps involved in carrying out a CFA. The next sections explain each step in more detail.

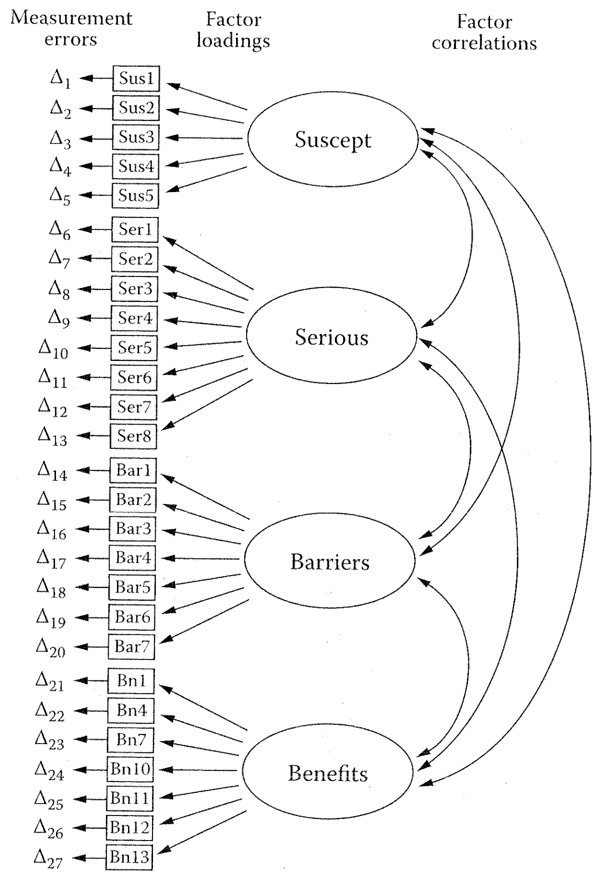


FIGURE 11.2
Model 1: Correlated factors for the health belief model.

11.13 A LISREL Example Comparing Two *a priori* Models

In this section, the new SIMPLIS language of the LISREL program is used to analyze data from the common situation in which one wishes to test a hypothesis about the underlying factor structure of a set of observed variables. The researcher usually has several hypotheses about the nature of the matrices λ (factor loadings), φ (factor correlations), and $\theta\delta$ (measurement error variances and covariances). Common hypotheses are that the items load on the appropriate factors, the factors are correlated in a certain way, or are uncorrelated, that the measurement errors are uncorrelated, or, in some cases, that some of these are correlated. These hypotheses can all be tested simultaneously using CFA methods.

In our example, the observed variables are the 27 HBM items. Amlung (1996) wanted to compare the fit of two models. In the first model, based on studies by Kegeles (1963) and Maiman et al. (1977), all the factors were free to correlate. The second model, based more on theoretical considerations (Rosenstock, 1974), allowed for correlations among only two pairs of factors. These two models are shown in Figure 11.2 and Figure 11.3 respectively. For simplicity, the measurement errors are omitted in Figure 11.3.

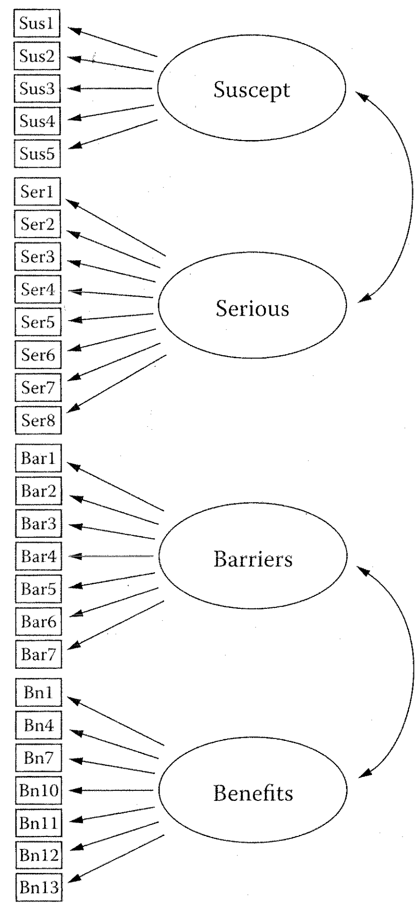


FIGURE 11.3
Model 2: Health Belief Model with two pairs of correlated factors.

The LISREL 8 SIMPLIS language program for Model 2 is shown in Table 11.12.

In both models, items were allowed to load only on the factor on which they were written to measure. This is accomplished in LISREL 8 by the first four lines under the keyword "relationships" shown in Table 11.12. As can be seen from the figures, all factors were hypothesized to correlate in the first model, whereas in the second only the two pairs of factors Seriousness and Susceptibility and Benefits and Barriers were allowed to correlate. Because *in LISREL 8 factors are all correlated by default*, this was accomplished by including the four lines under "relationships" that set the other correlations to zero. To run the first model, in which all factors were allowed to correlate, one would need to delete only those four lines from the LISREL 8 program. Finally, the measurement error variances are always included by default in LISREL 8.

Table 11.13 shows the estimates of the factor loadings and measurement error variances for Model 2. The standard error of each parameter estimate and a so-called *t* value obtained by dividing the estimate by its standard error are shown below each one. Table 11.14 shows the factor correlations for Model 2, along with their standard errors and *t* values.

Values of *t* greater than about |2.0| are commonly taken to be significant. Of course, these values are greatly influenced by the sample size, which is quite large in this example.

TABLE 11.12

SIMPLIS Command Lines for HBM with Two Pairs of Correlated Factors

```
title: Amlung Dissertation: Model with 2 pairs of correlated factors
observed variables:
sus1 sus2 sus3 sus4 sus5 ser1 ser2 ser3 ser4 ser5 ser6 ser7 ser8 ben1
ben4 ben7 ben10 ben11 ben12 ben13 bar1 bar2 bar3 bar4 bar5 bar6 bar7        ①
covariance matrix from file: AMLUNG.COV                                     ②
sample size 527
latent variables: suscept serious benefits barriers                         ③
relationships:                                                              ④
sus1 sus2 sus3 sus4 sus5 = suscept
ser1 ser2 ser3 ser4 ser5 ser6 ser7 ser8 = serious
ben1 ben4 ben7 ben10 ben11 ben12 ben13 = benefits
bar1 bar2 bar3 bar4 bar5 bar6 bar7 = barriers
set the correlation of benefits and serious to 0                            ⑤
set the correlation of benefits and suscept to 0
set the correlation of suscept and barriers to 0
set the correlation of barriers and serious to 0
end of problem
```

① Up to 8 characters can be used to name each observed variable.
② Here, the matrix created by PRELIS 2 is used by the LISREL 8 program.
③ Names (8 characters or less) are given to the latent variables (factors).
④ Here, under relationships, we link the observed variables to the factors.
⑤ In these four lines, the correlations among certain pairs of factors are set to zero.

Although all of the *t* values for the parameters in Model 2 are statistically significant, it is evident that the items on the Benefits scale have loadings that are much lower than those of the other scales. Several other items, such as Ser1, also have very low loadings. We saw in our PRELIS output that the distribution of Ser1 was quite nonnormal. This probably resulted in a lack of variance for this item, which in turn has caused its low loading.

The factor correlations are of particular interest in this study. Amlung (1996) hypothesized that only the two factor pairs Seriousness/Susceptibility and Benefits/Barriers would be significantly correlated. The results shown in Table 11.14 support the hypothesis that these two pairs of factors are significantly correlated. To see whether these were the only pairs with significant correlations, we must look at the factor correlations obtained from Model 1, in which all of the factors were allowed to correlate. These factor correlations, along with their standard errors and *t* values, are shown in Table 11.15.

Although the highest factor correlations are found between the factors Barriers/Benefits, and Seriousness/Susceptibility, all other factor pairs, with the exception of Seriousness/Benefits, are significantly correlated. None of the factor correlations are particularly large in magnitude, however, and the statistical significance may be due primarily to the large sample size. Based on our inspection of the parameter values and *t* statistics, support for Model 2 over Model 1 appears to be somewhat equivocal. However, note that these statistics are tests of *individual* model parameters. There are also statistics that test all model parameters simultaneously. Many such statistics, commonly called *overall fit* statistics, have been developed. These are discussed in more detail in Section 11.15. For now, we consider only the chi-square test and the goodness-of-fit index (GFI).

The chi-square statistic in CFA tests the hypothesis that the model fits, or is consistent with, the pattern of covariation of the observed variables. If this hypothesis were rejected, it would mean that the hypothesized model is not reasonable, or does not fit with our data. Therefore, contrary to the usual hypothesis testing procedures, we do *not* want to reject

TABLE 11.13

Factor Loadings a
for Health Belief

```
①
sus1 = 0.79*suscept,
  (0.031) ③
  25.28 ④
sus2 = 0.77*suscept,
  (0.029)
  26.79
sus3 = 0.74*suscept,
  (0.032)
  23.38
sus4 = 0.77*suscept,
  (0.035)
  22.15
sus5 = 0.81*suscept,
  (0.038)
  20.98
ser1 = 0.18*serious, l
  (0.031)
  5.76
ser2 = 0.47*serious, l
  (0.037)
  12.48
ser3 = 0.67*serious, l
  (0.046)
  14.61
ser4 = 0.70*serious, l
  (0.049)
  14.19
ser5 = 0.56*serious, E
  (0.039)
  14.60
ser6 = 0.63*serious, E
  (0.050)
  12.66
ser7 = 0.75*serious, E
  (0.046)
  16.27
ser8 = 0.45*serious, E
  (0.043)
  10.41
```

① Factor loading.
② Measurement error v
③ Standard error.
④ *t* Value.

TABLE 11.13

Factor Loadings and Measurement Error Variances with Standard Errors and *t* Values for Health Belief Model 2

LISREL ESTIMATES (MAXIMUM LIKELIHOOD)

①	②		
sus1 = 0.79*suscept,	Errorvar. = 0.18, R* = 0.78	ben1 = 0.29*benefits,	Errorvar. = 0.37, R* = 0.18
(0.031) ③	(0.015) ③	(0.031)	(0.024)
25.28 ④	12.09 ④	9.36	15.48
sus2 = 0.77*suscept,	Errorvar. = 0.12, R* = 0.83	ben4 = 0.35*benefits,	Errorvar. = 0.41, R* = 0.23
(0.029)	(0.012)	(0.033)	(0.027)
26.79	10.38	10.78	15.20
sus3 = 0.74*suscept,	Errorvar. = 0.23, R* = 0.70	ben7 = 0.20*benefits,	Errorvar. = 0.61, R* = 0.059
(0.032)	(0.017)	(0.038)	(0.038)
23.38	13.42	5.17	16.01
sus4 = 0.77*suscept,	Errorvar. = 0.31, R* = 0.66	ben10 = 0.49*benefits,	Errorvar. = 0.22, R* = 0.53
(0.035)	(0.022)	(0.028)	(0.018)
22.15	13.99	17.73	12.37
sus5 = 0.81*suscept,	Errorvar. = 0.42, R* = 0.61	ben11 = 0.62*benefits,	Errorvar. = 0.27, R* = 0.59
(0.038)	(0.029)	(0.032)	(0.024)
20.98	14.40	19.00	11.35
ser1 = 0.18*serious,	Errorvar. = 0.40, R* = 0.075	ben12 = 0.62*benefits,	Errorvar. = 0.42, R* = 0.48
(0.031)	(0.025)	(0.037)	(0.032)
5.76	15.90	16.70	13.04
ser2 = 0.47*serious,	Errorvar. = 0.48, R* = 0.31	ben13 = 0.41*benefits,	Errorvar. = 0.22, R* = 0.43
(0.037)	(0.033)	(0.026)	(0.016)
12.48	14.48	15.61	13.62
ser3 = 0.67*serious,	Errorvar. = 0.67, R* = 0.40	bar1 = 0.75*barriers,	Errorvar. = 0.43, R* = 0.56
(0.046)	(0.049)	(0.038)	(0.031)
14.61	13.60	19.50	13.94
ser4 = 0.70*serious,	Errorvar. = 0.78, R* = 0.38	bar2 = 0.53*barriers,	Errorvar. = 0.39, R* = 0.42
(0.049)	(0.057)	(0.033)	(0.026)
14.19	13.79	15.91	14.97
ser5 = 0.56*serious,	Errorvar. = 0.48, R* = 0.40	bar3 = 0.64*barriers,	Errorvar. = 0.17, R* = 0.71
(0.039)	(0.035)	(0.028)	(0.014)
14.60	13.60	23.10	11.86
ser6 = 0.63*serious,	Errorvar. = 0.87, R* = 0.32	bar4 = 0.48*barriers,	Errorvar. = 0.19, R* = 0.55
(0.050)	(0.060)	(0.025)	(0.013)
12.66	14.41	19.17	14.06
ser7 = 0.75*serious,	Errorvar. = 0.62, R* = 0.48	bar5 = 0.62*barriers,	Errorvar. = 0.15, R* = 0.73
(0.046)	(0.049)	(0.026)	(0.013)
16.27	12.65	23.60	11.42
ser8 = 0.45*serious,	Errorvar. = 0.70, R* = 0.23	bar6 = 0.33*barriers,	Errorvar. = 1.14, R* = 0.089
(0.043)	(0.047)	(0.050)	(0.071)
10.41	15.09	6.67	16.05
		bar7 = 0.42*barriers,	Errorvar. = 0.20, R* = 0.47
		(0.025)	(0.014)
		17.19	14.67

① Factor loading.
② Measurement error variance.
③ Standard error.
④ *t* Value.

TABLE 11.14

Factor Correlations, Standard Errors, and t Values for Health Belief Model 2

CORRELATION MATRIX OF INDEPENDENT VARIABLES				
	suscept	serious	benefits	barriers
suscept	1.00 ①			
serious	0.24 ②	1.00		
	(0.05) ③			
	4.92 ④			
benefits	— ⑤	—	1.00	
barriers	—	—	−0.27	1.00
			(0.05)	
			−5.66	

① Factor variances were set equal to 1.0 in order to give a metric to the factors.
② Factor correlation.
③ Standard error.
④ t Value.
⑤ Indicates that this correlation was not estimated.

TABLE 11.15

Factor Correlations, Standard Errors, and t Values for HBM Model 1

CORRELATION MATRIX OF INDEPENDENT VARIABLES				
	suscept	serious	benefits	barriers
suscept	1.00			
serious	0.24 ①	1.00		
	(0.05) ②			
	4.93 ③			
benefits	−0.16	−0.02	1.00	
	(0.05)	(0.05)		
	−3.37	−0.43		
barriers	0.15	0.20	−0.27	1.00
	(0.05)	(0.05)	(0.05)	
	3.33	4.14	−5.66	

① Factor correlation.
② Standard error.
③ t Value.

the null hypothesis. Unfortunately, the chi-square statistic used in CFA is very sensitive to sample size, such that, with a large enough sample size, almost any hypothesis will be rejected. This dilemma, which is discussed in more detail in Section 11.15, has led to the development of many other statistics designed to assess overall model fit in some way. One of these is the goodness-of-fit index (GFI) produced by the LISREL program. This index is roughly analogous to the multiple R^2 value in multiple regression in that it represents the overall amount of the covariation among the observed variables that can be accounted for by the hypothesized model.

TABLE 11.16

Goodness-of-Fit Statistics for Model 1 (All Factors Correlated)

CHI-SQUARE WITH 318 DEGREES OF FREEDOM = 1147.45 (P = 0.0)
ROOT MEAN SQUARE ERROR OF APPROXIMATION (RMSEA) = 0.070
P-VALUE FOR TEST OF CLOSE FIT (RMSEA < 0.05) = 0.00000037
EXPECTED CROSS-VALIDATION INDEX (ECVI) = 2.41
ECVI FOR SATURATED MODEL = 1.44
INDEPENDENCE AIC = 6590.16
MODEL AIC = 1267.45
ROOT MEAN SQUARE RESIDUAL (RMR) = 0.047
STANDARDIZED RMR = 0.063
GOODNESS OF FIT INDEX (GFI) = 0.86
ADJUSTED GOODNESS OF FIT INDEX (AGFI) = 0.83
PARSIMONY GOODNESS OF FIT INDEX (PGFI) = 0.72
NORMED FIT INDEX (NFI) = 0.82
NON-NORMED FIT INDEX (NNFI) = 0.85
PARSIMONY NORMED FIT INDEX (PNFI) = 0.75

TABLE 11.17

Goodness-of-Fit Statistics for Model 2 (Two Pairs of Correlated Factors)

CHI-SQUARE WITH 322 DEGREES OF FREEDOM = 1177.93 (P = 0.0)
ROOT MEAN SQUARE ERROR OF APPROXIMATION (RMSEA) = 0.071
P-VALUE FOR TEST OF CLOSE FIT (RMSEA < 0.05) = 0.00000038
EXPECTED CROSS-VALIDATION INDEX (ECVI) = 2.45
ECVI FOR SATURATED MODEL = 1.44
INDEPENDENCE AIC = 6590.16
MODEL AIC = 1289.93
ROOT MEAN SQUARE RESIDUAL (RMR) = 0.062
STANDARDIZED RMR = 0.081
GOODNESS OF FIT INDEX (GFI) = 0.85
ADJUSTED GOODNESS OF FIT INDEX (AGFI) = 0.83
PARSIMONY GOODNESS OF FIT INDEX (PGFI) = 0.73
NORMED FIT INDEX (NFI) = 0.82
NON-NORMED FIT INDEX (NNFI) = 0.85
PARSIMONY NORMED FIT INDEX (PNFI) = 0.75

Values of the chi-square statistic and GFI obtained for Models 1 and 2, as well as many other overall fit indices produced by the LISREL 8 program, are presented in Table 11.16 and Table 11.17, respectively.

The chi-square values for Models 1 and 2 are 1147.45 and 1177.93, respectively, with 318 and 322 degrees of freedom. Both chi-square values are highly significant, indicating that neither model adequately accounts for the observed covariation among the HBM items. The GFI values for the two models are almost identical at .86 and .85 for Models 1 and 2, respectively. In many cases, models that provide a good fit to the data have GFI values above .9, so again the two models tested here do not seem to fit well. The large chi-square values may be due, at least in part, to the large sample size, rather than to any substantial misspecification of the model. However, it is also possible that the model is misspecified in some fundamental way. For example, one or more of the items may actually load on more than one of the factors, instead of loading on only one, as specified in our model. Before making any decisions about the two models, we must examine such possibilities. We learn more about how to do this in the following sections, in which model identification, estimation, assessment, and modification are discussed more thoroughly.

11.14 Identification

The topic of identification is complex, and a thorough treatment is beyond the scope of this chapter. The interested reader is encouraged to consult Bollen (1989).

Identification of a CFA model is a prerequisite for obtaining correct estimates of the parameter values. A simple algebraic example can be used to illustrate this concept. Given the equations $X + Y = 5$, we cannot obtain *unique* solutions for X and Y, because an infinite number of values for X and Y will produce the same solution (5 and 0, 100 and −95, 2.5 and 2.5, etc.). However, if we impose another constraint on our solution by specifying that $2X = 4$, we can obtain one and only one solution: $X = 2$ and $Y = 3$. After imposing the additional constraint, we have two unknowns, X and Y, and two pieces of information, $X + Y = 5$, and $2X = 4$. Note that in the first situation with two unknowns and only one piece of information, the problem was not that we could not find a solution, but that we could find too many solutions. When this is the case, there is no way of determining which solution is "best" without imposing further constraints. Identification refers, therefore, to whether the parameters of a model can be *uniquely* determined.

Models that have more unknown parameters than pieces of information are called unidentified or *underidentified* models, and cannot be solved uniquely. Models with just as many unknowns as pieces of information are referred to as *just-identified* models, and can be solved, but cannot be tested statistically. Models with more information than unknowns are called *overidentified* models, or sometimes simply identified models, and can be solved uniquely. In addition, as we show in Section 11.15, overidentified models can be tested statistically.

As we have seen, one condition for identification is that the number of unknown parameters must be less than or equal to the number of pieces of information. In CFA, the unknown parameters are the factor loadings, factor correlations, and measurement error variances (and possibly covariances) that are to be estimated, and the information available to solve for these is the elements of the covariance matrix for the observed variables. In the HBM example, the number of parameters to be estimated for Model 1 would be the 27 factor loadings, plus the six factor correlations, plus the 27 measurement error variances, for a total of 60 parameters. In Model 2, we estimated only two factor correlations, giving us a total of 56 parameters for that model. The number of unique values in a covariance matrix is equal to $p(p + 1)/2$, where p is the number of observed variables. This number represents the number of covariance elements below the diagonal plus the number of variance elements. Above-diagonal elements are not counted because they must be the same as the below-diagonal elements. For the 27 items in our HBM example, the number of elements in the covariance matrix would be $(27 \times 28)/2$, or 378. Because the number of pieces of information is much greater than the number of parameters to be estimated, we should have enough information to identify these two models.

Bollen (1989) gave several rules that enable researchers to determine the identification status of their models. In general, CFA models should be identified if they have at least

three items for each
case, and applied re
factor loadings or c
larger in magnitude
than 1.0 (for further
One more piece of i
factor must have a un
scale. Instead, they ar
this is to set the varia
gram, this is done aut
contain the factor cc
Once the identifica
factor correlations, a
is the subject of the

11.15 Estimation

Recall that in CFA it
can be explained by
relationships are rep
and measurement er
and the researchers'
Estimation is concern
the matrix Σ. This is a
are sought that will r
In reality, we do no
is this sample matrix
parameters in λ, φ, an
In practice, our mo
ally do is to find para
tion that measures ho
symbolized as $F(S;\Sigma)$.
probably the most co

where tr stands for the
p is the number of vari
The criterion for find
values of the fit functio
minology, we are tryin
the differences between
some type of model mis
maximizing a quantity
squares criterion in mu

three items for each factor. However, there are some situations in which this will not be the case, and applied researchers should be alert for signs of underidentification. These include factor loadings or correlations that seem to have the wrong sign or are much smaller or larger in magnitude than what was expected, negative variances, and correlations greater than 1.0 (for further discussion see Wothke, 1993).

One more piece of information is necessary in order to assure identification of CFA models: each factor must have a unit of measurement. Because the factors are unobservable, they have no inherent scale. Instead, they are usually assigned scales in a convenient metric. One common way of doing this is to set the variances of the factors equal to one (Bentler, 1992a, p. 22). In the LISREL 8 program, this is done automatically. Note that one consequence of this is that the matrix φ will contain the factor correlations rather than the factor covariances.

Once the identification of a model has been established, estimation of the factor loadings, factor correlations, and measurement error variances can proceed. The estimation process is the subject of the next section.

11.15 Estimation

Recall that in CFA it is hypothesized that the relationships among the observed variables can be explained by the factors. The researchers' hypotheses about the form of these relationships are represented by the structure of the factor loadings, factor correlations, and measurement error variances. Thus, the relationship between the observed variables and the researchers' hypotheses or model is represented by the equation $\Sigma = \lambda\varphi\lambda' + \theta\delta$. Estimation is concerned with finding the values for λ, φ, and $\theta\delta$ that will best reproduce the matrix Σ. This is analogous to the situation in multiple regression in which values of β are sought that will reproduce the original Y values as closely as possible.

In reality, we do not have the population matrix Σ, but rather the sample matrix **S**. It is this sample matrix that is compared to the matrix reproduced by the estimates of the parameters in λ, φ, and $\theta\delta$, referred to as $\Sigma(\theta)$.

In practice, our model will probably not reproduce S perfectly. The best we can usually do is to find parameter estimates that result in matrix $\bar{\Sigma}$ that is close to S. A function that measures how close $\bar{\Sigma}$ is to S is called a *discrepancy* or *fit function*, and is usually symbolized as $F(S;\bar{\Sigma})$. Many different fit functions are available in CFA programs, but probably the most commonly used is the maximum likelihood function, defined as:

$$F(\mathbf{S};\Sigma(\theta)) = \mathrm{tr}(\mathbf{S}\Sigma(\theta)^{-1}) + \left[\ln|\Sigma(\theta)| - \ln|\mathbf{S}|\right] - p$$

where tr stands for the trace of a matrix, defined as the sum of its diagonal elements, and p is the number of variables.

The criterion for finding estimates of the parameters in λ, φ, and $\theta\delta$ is that they result in values of the fit function $F(\mathbf{S};\Sigma(\theta))$ that are as small as possible. In maximum likelihood terminology, we are trying to find parameter estimates that will *maximize* the likelihood that the differences between **S** and $\Sigma(\theta)$ are due to random sampling fluctuations, rather than to some type of model misspecification. Although the maximum likelihood criterion involves maximizing a quantity rather than minimizing one, it is similar in purpose to the least squares criterion in multiple regression, in which the quantity $\Sigma(Y - Y')^2$ is minimized.

Unlike the least squares criterion, however, the criterion used in maximum likelihood estimation of CFA parameters cannot usually be solved algebraically. Instead, computer programs have been developed that use an iterative process for finding the parameter estimates. In an iterative solution, a set of initial values for the parameters of λ, φ, and $\theta\delta$ are used as starting points. The matrix $\Sigma(\theta) = \lambda\varphi\lambda' + \theta\delta$ is then calculated based on these values, and is compared with **S**. If $\Sigma(\theta)$ does not match **S**, one or more of the initial values is changed in such a way as to improve the fit, and $\Sigma(\theta)$ is recomputed. This process continues until $\Sigma(\theta)$ is sufficiently close to **S**, or until no further improvements can be made that are within the constraints of the hypothesized model. The final values obtained are taken as the estimates of the elements of λ, φ, and $\theta\delta$.

The value of the fit function $F(\mathbf{S};\Sigma(\theta))$ that is based on these final values can be used to determine how well the hypothesized model fits the observed matrix **S**. The topic of assessing model fit is taken up in the next section. Before concluding our discussion about estimation, however, a few words concerning the appropriate sample size are essential. Several studies have focused on the question of how large a sample should be in order to obtain accurate results in maximum likelihood CFA. Most have recommended minimum sample sizes of 200, although smaller sample sizes may be adequate for models with a relatively small number of parameters to be estimated (see, e.g., Anderson & Gerbing, 1984; Boomsma, 1982). For models with many paths to be estimated, larger sample sizes are needed.

11.16 Assessment of Model Fit

The appropriate way to assess the fit of CFA models has been a subject of debate since the 1970s. A plethora of fit statistics has been developed and discussed in the literature. In this chapter, I focus only on the most commonly used fit statistics and present some general guidelines for model assessment. For more detailed information, the reader is directed to the excellent presentations in Bollen (1989), Bollen and Long (1993), Hayduk (1987), and Loehlin (1992).

It is useful to divide statistics for assessing the fit of a model, commonly called *fit statistics*, into two categories: those that measure the overall fit of the model, and those that are concerned with individual model parameters, such as factor loadings or correlations.

Probably the most well-known measure of overall model fit is the chi-square (χ^2) statistic, which was presented briefly in Section 11.13. This statistic is calculated as

$$(n-1)F(\mathbf{S};\Sigma(\theta))$$

and is distributed as a chi-square with degrees of freedom equal to the number of elements in **S**, $p(p+1)/2$ minus the number of parameters estimated, if certain conditions are met. These conditions include having a large enough sample size and variables that follow a multivariate normal distribution. Notice that, for a just-identified model, the degrees of freedom are zero, because the number of parameters estimated are equal to the number of elements in **S**. This means that just-identified models cannot be tested. However, recall that just-identified models will *always* exactly reproduce **S** perfectly; therefore a test of such a model would be pointless, as we already know the answer.

The chi-square statistic can be used to test the hypothesis that $\Sigma = \Sigma(\theta)$, or that the original population matrix is equal to the matrix reproduced from one's model. Remember that, contrary to the general rule in hypothesis testing, the researcher would *not* want to reject the null hypothesis, as finding $\Sigma \neq \Sigma(\theta)$ would mean that the hypothesized model parameters were unable to reproduce **S**. Thus, smaller rather than larger chi-square values are indicative of a good fit.

From the chi-square formula we can see that, as n increases, the value of chi-square will increase to the point at which, for a large enough value of n, even trivial differences between Σ and $\Sigma(\theta)$ will be found significant. Largely because of this, as early as 1969 Jöreskog recommended that the chi-square statistic be used more as a descriptive index of fit rather than as a statistical test. Accordingly, Jöreskog and Sörbom (1993) included other fit indices in the LISREL output. The GFI was introduced in Section 11.12. This index was defined by Jöreskog and Sörbom as:

$$\text{GFI} = 1 - \frac{F(\mathbf{S};\Sigma(\theta))}{F(\mathbf{S};\Sigma(0))}$$

where $F(\mathbf{S}; \Sigma(0))$ is the value of the fit function for a null model in which all parameters except the variances of the variables have values of zero. In other words, the null model is one that posits no relationships among the variables. The GFI can be thought of as the amount of the overall variance and covariance in **S** that can be accounted for by $\Sigma(\theta)$ and is roughly analogous to the multiple R^2 in multiple regression. The adjusted GFI (AGFI) is given as

$$\text{AGFI} = 1 - \frac{p(p+1)}{2df}(1 - \text{GFI})$$

(Jöreskog & Sörbom, 1993), where p represents the number of variables in the model and df stands for degrees of freedom. The AGFI adjusts the GFI for degrees of freedom, resulting in lower values for models with more parameters. The rationale behind this adjustment is that models can always be made to reproduce **S** more closely by adding more parameters to the model. The ultimate example of this is the just-identified model, which always reproduces **S** exactly because it includes all possible parameters. In our HBM examples, Model 1 resulted in values of .86 and .83 for the GFI and AGFI, and the corresponding values for Model 2 were .85 and .83. The AGFI was not substantially lower than the GFI for these models because the number of parameters estimated was not overly large, given the number of pieces of information (covariance elements) that were available to estimate them.

Another measure of overall fit is the difference between the matrices **S** and $\Sigma(\theta)$. These differences are called *residuals* and can be obtained as output from CFA computer programs. *Standardized* residuals are residuals that have been standardized to have a mean of zero and a standard deviation of one, making them easier to interpret. Standardized residuals larger than $|2.0|$ are usually considered to be suggestive of a lack of fit.

Bentler and Bonett (1980) introduced a class of fit indexes commonly called *comparative fit indexes*. These indexes compare the fit of the hypothesized model to a baseline or null model, in order to determine the amount by which the fit is improved by using the hypothesized model instead of the a model. The most commonly used null model is that described earlier in which the variables are completely uncorrelated.

The normed fit index (NFI; Bentler & Bonett, 1980) can be computed as

$$\chi_0^2 - \chi_1^2/\chi_0^2$$

where χ_0^2 and χ_1^2 are the χ^2 values for the null and hypothesized models, respectively. The NFI represents the increment in fit obtained by using the hypothesized model relative to the fit of the null model. Values range from zero to one, with higher values indicative of a greater improvement in fit.

Bentler and Bonett's nonnormed fit index (NNFI) can be calculated as

$$\text{NNFI} = \frac{(\chi_0^2/df_0 - \chi_1^2/df_1)}{(\chi_0^2/df_0 - 1)}$$

where χ_0^2 and χ_1^2 are as before and df_0 and df_1 are the degrees of freedom for the null and hypothesized models, respectively. This index is referred to as nonnormed because it is not constrained to have values between zero and one, as is common for comparative fit indexes. The NNFI can be interpreted as the increment in fit per degree of freedom obtained by using the hypothesized model, relative to the best possible fit that could be obtained by using the hypothesized model. As with the NFI, higher values are suggestive of more improvement in fit. Although NFI and NNFI values greater than .9 have typically been considered indicative of a good fit, this rule of thumb has recently been called into question (see, e.g., Hu & Bentler, 1995). Values of the NFI and NNFI were .82 and .85, respectively, for both HBM models, indicating that these two models resulted in identical improvements in fit over a null model.

Because a better fit can always be obtained by adding more parameters to the model, James, Mulaik, and Brett (1982) suggested a modification of the NFI to adjust for the loss of degrees of freedom associated with such improvements in fit. This parsimony adjustment is obtained by multiplying the NFI by the ratio of degrees of freedom of the hypothesized model to those of the null model. A similar adjustment to the GFI was suggested by Mulaik et al. (1989). These two parsimony-adjusted indices are implemented in LISREL 8 as the parsimony goodness-of-fit index (PGFI) and the parsimony normed fit index (PNFI). For the two HBM models, the values of the PGFI and PNFI were .72 and .75, respectively, for Model 1, and .73 and .75 for Model 2. Because the two models differed by only four degrees of freedom, the parsimony adjustments had almost identical effects on them.

Several researchers (see, e.g., Cudeck & Henly, 1991) suggested that it may be unrealistic to suppose that the null hypothesis $\Sigma = \Sigma(\theta)$ will hold exactly, even in the population, because this would mean that the model can correctly specify all of the relationships among the variables. The lack of fit of the hypothesized model to the population is known as the *error of approximation*. The root mean square error of approximation (Steiger, 1990) is a standardized measure of error of approximation

$$\text{RMSEA} = \sqrt{\max\left\{\left(\frac{f(\theta)}{df} - \frac{1}{n}\right), 0\right\}}$$

where $F(\theta)$ is the maximum likelihood fit function discussed earlier, and df and n are as before.

MacCallum (1995, pp. 29–30), in arguing for RMSEA, discussed the disconfirmability of a model:

> A model is disconfirmable to the degree that it is possible for the model to be inconsistent with observed data ... if a model is not disconfirmable to any reasonable degree, then a finding of good fit is essentially useless and meaningless. Therefore, in the model specification process, researchers are very strongly encouraged to keep in mind the

> principle of disconfirmability and to construct models that are not highly parametrized .... Researchers are thus strongly urged to consider an index such as the root mean square error of approximation (RMSEA), which is essentially a measure of lack of fit per degree of freedom.

Based on their experience, Browne and Cudeck (1993) suggested that RMSEA values of .05 or less indicate a close approximation and that values of up to .08 suggest a reasonable fit of the model in the population. For our two HBM models, the RMSEA values were .07 and .071 for Models 1 and 2, respectively.

Finally, Browne and Cudeck (1989) proposed a single-sample cross-validation index developed to assess the degree to which a set of parameter estimates estimated in one sample would fit if used in another similar sample. This index is roughly analogous to the adjusted or "shrunken" R^2 value obtained in multiple regression. It is given as the ECVI, or expected cross-validation index, in the LISREL program. Because the ECVI is based on the chi-square statistic, smaller values are desired, which would indicate a greater likelihood that the model would cross-validate in another sample. A similar index is reported as part of the output from the LISREL 8 as well as the EQS (Bentler, 1989, 1992a) program. This is the Akaike (1987) Information Criterion (AIC), calculated as $\chi^2 - 2df$. As with the ECVI, smaller values of the AIC represent a greater likelihood of cross-validation. In a recent study by Bandalos (1993), values of the ECVI and AIC were compared with the values obtained by carrying out an actual two-sample cross-validation procedure in CFA. It was found that, although both indices provided very accurate estimates of the actual two-sample cross-validation values, the ECVI was slightly more accurate, especially with smaller sample sizes.

Thus far, the overall fit indices for the two HBM models have not provided us with a compelling statistical basis for preferring one model over the other. Values of the GFI, AGFI, NFI, NNFI, the parsimony-adjusted indices, and the RMSEA are almost identical for these two models. However, these two models are *nested* models, meaning that one can be obtained from the other by eliminating one or more paths. More specifically, Model 2 is nested within Model 1 because we can obtain the former from the latter by eliminating four of the factor correlations. The difference between the chi-square values of two nested models is itself distributed as a chi-square statistic, with degrees of freedom equal to the difference between the degrees of freedom for the two models. For Model 1, the chi-square value and degrees of freedom were 1147.45 and 318, while the corresponding values for Model 2 were 1177.93 and 322. The chi-square difference test is thus 30.38 with four degrees of freedom. The chi-square critical value at the .05 level of significance is 9.488. We would therefore find the chi-square difference statistically significant, which indicates that Model 2 (with a significantly higher chi-square value) fit significantly *worse* than Model 1.

In addition to the overall fit indices, individual parameter values should be scrutinized closely. Computer programs such as LISREL and EQS provide tests of each parameter estimate, computed by dividing the parameter estimate by its standard error. (These are referred to as t tests in LISREL.) These values can be used to test the hypothesis that the parameter value is significantly different from zero. The actual values of the parameter estimates should also be examined to determine whether any appear to be out of range. Out-of-range parameter values may take the form of negative variances in φ or $\theta\delta$, factor correlations greater than one, parameter estimates that seem much too high or too low, or parameter estimates that have the opposite sign from what was expected. Models resulting in any of these problems should be studied carefully to determine the cause of the error. One possible reason for problems of this type is that the model is not identified. However, it may also be that the researcher has inadvertently set the model up incorrectly.

It should be clear from this discussion that the assessment of model fit is not a simple process, nor is there a definitive answer to the question of how well a model fits the data. However, *several criteria with which most experts are in agreement have been developed over the years*. These have been discussed by Bollen and Long (1993) and are summarized here.

1. Hypothesize *at least* one model *a priori*, based on the best theory available. Often, theoretical knowledge in an area may be ambiguous or contradictory, and more than one model may be tenable. The relative fit of the different models can be compared using such indexes as the NFI, NNFI, PNFI, ECVI, and AIC.
2. Do not rely on the chi-square statistic as the only basis for assessing fit. The use of several indexes is encouraged.
3. Examine the values of individual parameter estimates in addition to assessing the overall fit.
4. Assessment of model fit should be made in the context of prior studies in the area. In fields in which little research has been done, less stringent standards may be acceptable than in areas in which well-developed theory is available.
5. As in any statistical analysis, data should be screened for outliers and for violations of distributional assumptions. Multivariate normality is one assumption underlying the use of maximum likelihood estimation in CFA.

The following quote from MacCallum (1995) concerning model fit touches on several issues that researchers must bear in mind during the process of model specification and evaluation, and thus makes a fitting conclusion to this section:

> A critical principle in model specification and evaluation is the fact that all of the models that we would be interested in specifying and evaluating are wrong to some degree. Models at their best can be expected to provide only a close approximation to observed data, rather than an exact fit. In the case of SEM, the real-world phenomena that give rise to our observed correlational data are far more complex than we can hope to represent using a linear structural equation model and associated assumptions. Thus we must define as an optimal outcome a finding that a particular model fits our observed data closely and yields a highly interpretable solution. Furthermore, one must understand that even when such an outcome is obtained, one can conclude only that the particular model is a plausible one. There will virtually always be other models that fit the data to exactly the same degree, or very nearly so, thereby representing models with different substantive interpretation but equivalent fit to the observed data. The number of such models may be extremely large, and they can be distinguished only in terms of their substantive meaning. (p. 17)

11.17 Model Modification

It is not uncommon in practice to find large discrepancies between **S** and $\Sigma(\theta)$, indicating that the hypothesized model was unable to accurately reproduce the original covariance matrix. Assuming that the hypothesized model was based on the best available theory, changes based on theoretical considerations may not be feasible. Given this state of affairs, the researcher may opt to modify the model in a post hoc fashion by adding or deleting parameters suggested by the fit statistics obtained. Statistics are available from both the LISREL and EQS programs that suggest possible changes to the model that will improve fit.

Two caveats are in order before we begin our discussion of these statistics. First, as in any post hoc statistical analysis, modifications made on the basis of information derived from a given sample cannot properly be tested on that same sample. This is because the results obtained from any sample data will have been fitted to the idiosyncrasies of that data, and may not generalize to other samples. For this reason, *post hoc model modifications must be regarded as tentative until they have been replicated on a different sample.* The second point that must be kept in mind is that the modifications suggested by programs such as LISREL and EQS can only tell us what additions or deletions of parameters will result in a better *statistical* fit. These modifications may or may not be defensible from a theoretical point of view. Changes that cannot be justified theoretically should be made.

Bollen (1989), in discussing modification of models, wrote:

> Researchers with inadequate models have many ways—in fact, too many ways—in which to modify their specification. An incredible number of major or minor alterations are possible, and the analyst needs some procedure to narrow the choices. The empirical means can be helpful, but they can also lead to nonsensical respecifications. Furthermore, empirical means work best in detecting simple alterations and are less helpful when major changes in structure are needed. The potentially richest source of ideas for respecification is the theoretical or substantive knowledge of the researcher. (pp. 296–297)

With these caveats, we can turn our attention to the indices that may be useful in suggesting possible model modifications. One obvious possibility is to delete parameters that are nonsignificant. For example, a factor loading may be found for which the reported t value in LISREL is less than $|2.0|$, indicating that the value of that loading is not significantly different from zero. Deleting a parameter from the model will not result in a better fit, but will gain a degree of freedom, resulting in a lower critical value. However, if the same data are used to both obtain and modify the model, this increase in degrees of freedom is not justified. This is because the degree of freedom has already been used to obtain the estimate in the original model. In subsequent analyses on other data sets, however, the researcher could omit the parameter, thus gaining a degree of freedom and obtaining a simpler model. Simpler models are generally preferred over more complex models for reasons of parsimony.

Another type of model modification that might be considered is to add parameters to the model. For example, a variable that had been constrained to load on only one factor might be allowed to have loadings on two factors. In the LISREL program, modification indexes (MIs) are provided. These are estimates of the decrease in the chi-square value that would result if a given parameter, such as a factor loading, were to be added to the model. MIs are available for all parameters that were constrained to be zero in the original model. They are accompanied by the expected parameter change (EPC) statistics. These represent the value a given parameter would have if it were added to the model. As is the case with the deletion of parameters, parameters should be added one at a time, with the model being reestimated after each addition. In the EQS program, the Lagrange Multiplier (LM) statistics serve the same function as the MIs in LISREL. EQS also provides multivariate LM statistics that take into account the correlations among the parameters.

The modification indexes for the factor loading and measurement error variance matrices from Model 1 of the HBM data are shown in Table 11.18. Because all of the factor correlations were included in that model, no modification indexes were computed for these.

TABLE 11.18

Modification Indexes for Health Belief Model 1

THE MODIFICATION INDICES SUGGEST TO ADD THE			
PATH TO	FROM	DECREASE IN CHI-SQUARE	NEW ESTIMATE
ser1	benefits	8.9	0.09
ser1	barriers	21.5	−0.14
ser8	suscept	13.9	0.15
ser8	benefits	14.1	−0.16
ser8	barriers	12.6	0.15
bar2	serious	10.9	0.11
bar2	benefits	11.4	−0.11
bar3	benefits	9.1	0.07
THE MODIFICATION INDICES SUGGEST TO ADD AN ERROR COVARIANCE			
BETWEEN	AND	DECREASE IN CHI-SQUARE	NEW ESTIMATE
sus2	sus1	26.8	0.07
sus3	sus2	16.1	0.05
sus4	sus2	44.2	−0.09
sus5	sus2	11.5	−0.05
sus5	sus4	93.1	0.18
ser2	ser1	56.6	0.16
ser3	ser2	65.4	0.24
ser4	sus1	12.4	−0.07
ser4	ser3	77.2	0.35
ser5	ser3	24.9	−0.15
ser5	ser4	21.7	−0.15
ser6	ser1	18.0	−0.12
ser6	ser2	24.5	−0.16
ser6	ser3	13.3	−0.15
ser7	ser2	19.7	−0.13
ser7	ser3	29.3	−0.20
ser7	ser4	17.9	−0.17
ser7	ser5	33.9	0.18
ser7	ser6	42.3	0.26
ser8	ser2	19.5	−0.12
ser8	ser7	8.2	0.10
ben4	ben1	70.8	0.15
ben7	ben1	9.6	0.07
ben10	ben1	9.2	−0.04
ben11	ser4	8.2	−0.07
ben12	ser5	9.7	−0.07
ben12	ben11	23.1	0.11
ben13	ben4	10.9	−0.05
ben13	ben10	41.1	0.09
bar1	ben1	8.1	−0.05
bar3	ser1	13.7	−0.05
bar3	bar1	44.2	0.11
bar4	bar1	26.2	−0.08
bar4	bar3	21.1	−0.05
bar5	bar4	17.2	0.04
bar6	bar4	15.8	0.09
bar7	bar4	26.3	0.05

STIMATE
0.09
0.14
0.15
0.16
0.15
0.11
0.11
0.07
ARIANCE
STIMATE
0.07
0.05
0.09
0.05
.18
.16
.24
.07
.35
.15
.15
.12
.16
.15
.13
.20
.17
.18
.26
.12
.10
.15
.07
.04
.07
.07
.11
.05
.09
.05
.05
.11
.08
.05
.04
.09
.05

The MIs suggest that the largest drop in chi-square (93.1) would be obtained if we were to add a measurement error covariance for items 4 and 5 on the Susceptibility scale. Several other large MIs have been obtained for pairs of measurement error covariances. However, most researchers share the view of Hoyle and Panter (1995), who stated that correlated errors of measurement are among the most problematic types of post hoc modifications because they are rarely theoretically justified and are unlikely to replicate. The need for correlated measurement errors is an indication that the factor model has been unable to account for all of the covariation among the variables. This may occur if, for example, more factors are needed or if method variance is present. These possibilities should be evaluated before any decision is made with regard to freeing these measurement error covariances.

If changes are made on the basis of MIs, the model must be reestimated following each change, as it is likely that the other parameter estimates and their MI values, as well as the chi-square value, would also change. This is the reason for the common recommendation that only one modification be made to a model at a time. Finally, no model modifications should be made unless they are theoretically defensible.

Section 11.20 provides a discussion of some concerns that have been voiced about current practices in CFA studies. One of these pertains to the use of MIs in making post hoc model modifications; the other has to do with the issue of equivalent or alternative models. Before discussing these issues, however, we consider two more examples: one that has been analyzed using the LISREL 8 program and one using the EQS (Bentler, 1989) program.

11.18 LISREL 8 Example

In this example, the observed variables are items from a measure of test anxiety known as the Reactions to Tests (RTT) scale. The RTT was developed by Sarason (1984) to measure the four hypothesized dimensions of worry, tension, test-irrelevant thinking, and bodily symptoms. The data are drawn from a study of the scale by Benson and Bandalos (1992) in which the items were found to be approximately normally distributed. For simplicity, only three items from each scale are used.

The factor structure tested is shown in Figure 11.4. As can be seen from the figure, each of the three items for each scale is hypothesized to load only on the scale it was written to measure, and the factors are hypothesized to correlate with each other. The 12 diagonal elements of $\theta\delta$, or the measurement error variances, are included in the model. The absence of curved arrows connecting the δ's in Figure 11.4 means that the measurement errors were not hypothesized to be correlated.

The SIMPLIS command lines are shown in Table 11.19. For those readers who do not have access to LISREL 8 and the new SIMPLIS command language, the LISREL 7 commands for this problem are shown in the Appendix to this chapter.

Table 11.20 shows the estimates of the factor loadings with their standard errors and *t* values in matrix format rather than the equation format used for the Health Belief examples. This format is used in older versions of the LISREL program and is preferred by some researchers. Table 11.21 presents similar information for the factor correlations. To conserve space, estimates of the measurement error variances are not shown.

An inspection of the *t* values for these parameter estimates reveals that all are significant with the exception of the correlation between the tension and test-irrelevant thinking factors, which has a *t* value of 1.76. There are no unreasonable parameter estimates such as negative variances or correlations greater than 1. All of the parameter estimates appear to be in the expected range of values and to have the expected signs. This is important because unreasonable values usually indicate a problem with the model, such as a lack of identification.

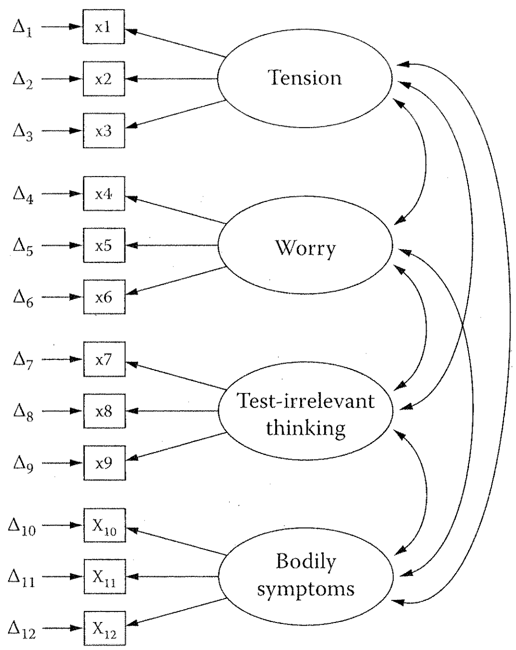


FIGURE 11.4
Four-factor test anxiety model with three indicators per factor.

The significance of the factor loadings is of special interest as these indicate that the items did have significant loadings on the factors they were intended to measure. The lack of a significant correlation between tension and test-irrelevant thinking is not surprising; other studies have also found the test-irrelevant thinking factor to be the most distinct of the four factors. The magnitudes and statistical significance of the remaining factor correlations support the hypothesis that the four factors are distinct, yet related, dimensions of test anxiety.

Our inspection of the parameter values and t statistics indicate support for the hypothesized four-factor structure. Some selected tests of overall fit are shown in Table 11.22.

The chi-square value of 88.396 with 48 degrees of freedom is significant with a probability of .0003, indicating that the model does not adequately account for the observed covariation among the variables. However, many of the other fit indexes suggest that the fit of the model is fairly good. It may be that the significant chi-square is, at least in part, due to the fairly large sample size, rather than to any serious misspecification of the model.

Setting the loadings equal to 0 does not make them 0. Thus, previous empirical work should be done to ensure that the items are relatively pure measures of the constructs they are designed to measure. In Table 11.23 we have allowed TEN1, WOR1, IRTHK1, and BODY1 to load on all four factors, to see if they are relatively pure measures of TEN, WOR, IRTHK, and BODY. We have done the same for TEN2 and such, and for TEN3 and such. The loadings on the other factors are in almost all cases close to 0, which is reassuring.

11.19 EQS Example

Having presented an example using the LISREL 8 program, I now discuss one using Bentler's (1989) EQS program. This chapter is not intended as a comprehensive guide to the program, however. The interested reader should consult the program manual or the excellent reference on this program by Byrne (1994).

TABLE 11.19

SIMPLIS Commar

TITLE: FOUR FACT
OBSERVED VARIAE
IRTHK2 IRTHK3 BC
COVARIANCE MAT
.7821
.5602 .9299
.5695 .6281 .9
.1969 .2599 .2
.2289 .2835 .3
.2609 .3670 .3
.0556 .0740 .0
.0025 .0279 .0
.0180 .0753 .0
.1617 .1919 .2
.2628 .3047 .4
.2966 .3040 .3
SAMPLE SIZE: 318
LATENT VARIABLES
RELATIONSHIPS:
TEN1 TEN2 TEN3 =
WOR1 WOR2 WOF
IRTHK1 IRTHK2 IRT
BODY1 BODY2 BO
END OF PROBLEM

① Up to 8 character
② Only the lower ha
③ Names (8 charact
④ Here, under relati

In this example,
EQS program. Altl
the inclusion of the
matrix, which, you
skills measured on
example, the data
in order to demor
earlier and the CF
factor structure re
and exploratory ar
set of data would
of contrasting the
The EQS comm
Table 11.25, along
included in a sepa
and because our p
An inspection o
They also appear
loadings differ sor
substantial. Recall
but loadings less tl

TABLE 11.19

SIMPLIS Command Lines for Test Anxiety Example

```
TITLE: FOUR FACTOR STRUCTURE FOR ANXIETY
OBSERVED VARIABLES: TEN1 TEN2 TEN3 WOR1 WOR2 WOR3 IRTHK1
IRTHK2 IRTHK3 BODY1 BODY2 BODY3                                      ①
COVARIANCE MATRIX:
.7821
.5602  .9299                                                         ②
.5695  .6281  .9751
.1969  .2599  .2362  .6352
.2289  .2835  .3079  .4575  .7943
.2609  .3670  .3575  .4327  .4151  .6783
.0556  .0740  .0981  .2094  .2306  .2503  .6855
.0025  .0279  .0798  .2047  .2270  .2257  .4224  .6952
.0180  .0753  .0744  .1892  .2352  .2008  .4343  .4514  .6065
.1617  .1919  .2892  .1376  .1744  .1845  .0645  .0731  .0921  .4068
.2628  .3047  .4043  .1742  .2066  .2547  .1356  .1336  .1283  .1958  .7015
.2966  .3040  .3919  .1942  .1864  .2402  .1073  .0988  .0599  .2233  .3033  .5786
SAMPLE SIZE: 318
LATENT VARIABLES: TENSION WORRY TIRT BODY                            ③
RELATIONSHIPS:
TEN1 TEN2 TEN3 = TENSION                                             ④
WOR1 WOR2 WOR3 = WORRY
IRTHK1 IRTHK2 IRTHK3 = TIRT
BODY1 BODY2 BODY3 = BODY
END OF PROBLEM
```

① Up to 8 characters can be used to name each observed variable.
② Only the lower half of the covariance matrix need be inserted.
③ Names (8 characters or less) are given to the latent variables (factor).
④ Here, under relationships, we link the observed variables to the factors.

In this example, the data given later in Exercise 7 of this chapter have been reanalyzed using the EQS program. Although the data presented in Exercise 7 are in the form of a correlation matrix, the inclusion of the standard deviations makes it possible for the program to calculate a covariance matrix, which, you will recall, is preferred for use in CFA. The data consist of 10 communication skills measured on 159 deaf rehabilitation candidates. A two-factor solution was obtained. In this example, the data has been reanalyzed using this two-factor solution as the hypothesized model in order to demonstrate the similarities of, and differences between, the EFA solution presented earlier and the CFA procedures. I emphasize that the reanalysis cannot be used as a test of the factor structure reported earlier because the same data are being used in both the confirmatory and exploratory analyses. If our objective were to test the structure obtained from the EFA, a new set of data would have to be obtained. This analysis is therefore introduced only for the purpose of contrasting the exploratory and confirmatory procedures.

The EQS command lines are presented in Table 11.24. The factor loadings are shown in Table 11.25, along with their standard errors and t values. The measurement error variances are included in a separate matrix labeled "variances of independent variables." To conserve space, and because our primary interest is in the factor loadings, this matrix is not reproduced here.

An inspection of the t values for the factor loadings reveals that all are statistically significant. They also appear to be reasonable and of the expected magnitude and direction. The factor loadings differ somewhat from those from the EFA, but these differences do not appear to be substantial. Recall that in the original EFA, each variable actually had loadings on both factors, but loadings less than .30 were not reported. In the current analysis, loadings less than .30 were

TABLE 11.20

Factor Loadings, Standard Errors, and t Values for Test Anxiety Example

LISREL ESTIMATES (MAXIMUM LIKELIHOOD)

LAMBDA-X (factor loadings)

	tension	worry	tirt	body
TEN1	.69 ①	— ④	—	—
	(.04) ②			
	15.59 ③			
TEN2	.76	—	—	—
	(.05)			
	16.01			
TEN3	.84	—	—	—
	(.05)			
	17.70			
WOR1	—	.64	—	—
		(.04)		
		16.18		
WOR2	—	.66	—	—
		(.05)		
		14.51		
WOR3	—	.67	—	—
		(.04)		
		16.30		
IRTHK1	—	—	.64	—
			(.04)	
			15.47	
IRTHK2	—	—	.67	—
			(.04)	
			16.09	
IRTHK3	—	—	.67	—
			(.04)	
			17.69	
BODY1	—	—	—	.38
				(.04)
				10.51
BODY2	—	—	—	.54
				(.05)
				11.52
BODY3	—	—	—	.56
				(.04)
				13.29

① Factor loading.
② Standard error.
③ t Value.
④ — Indicates a factor loading that was constrained to equal zero by the program.

TABLE 11.21

Factor Correlations, Standard Errors, and *t* Values for Test Anxiety Example

	PHI (factor correlations)			
	tension	worry	tirt	body
tension	1.00 ①			
worry	.55 ②	1.00		
	(.05) ③			
	11.01 ④			
tirt	.11	.49	1.00	
	(.06)	(.05)		
	1.76	9.28		
body	.78	.59	.29	1.00
	(.04)	(.05)	(.07)	
	18.73	10.89	4.25	

① Factor variances were set to equal 1.0 in order to give a metric to the factors.
② Factor correlation.
③ Standard error.
④ *t* Value.

TABLE 11.22

Goodness-of-Fit Statistics for Test Anxiety Example

CHI-SQUARE WITH 48 DEGREES OF FREEDOM = 88.396 (P = 0.000345)
GOODNESS OF FIT INDEX (GFI) = 0.957
ADJUSTED GOODNESS OF FIT INDEX (AGFI) = 0.929
EXPECTED CROSS-VALIDATION INDEX (ECVI) = 0.477
90 PERCENT CONFIDENCE INTERVAL FOR ECVI = (0.397; 0.564)
ECVI FOR SATURATED MODEL = 0.492
INDEPENDENCE AIC = 1634.054
MODEL AIC = −7.604
NORMED FIT INDEX (NFI) = 0.95
NON-NORMED FIT INDEX (NNFI) = 0.967
PARSIMONY NORMED FIT INDEX = 0.691

constrained to be zero. This probably accounts for most of the discrepancies between the two sets of factor loadings.

To ascertain whether the two-factor model represents a good fit to the data, the fit statistics must be considered. Some of these are presented in Table 11.26.

Overall, the fit statistics do not suggest a good fit to the data. The chi-square value is highly significant, indicating that the model has not adequately reproduced the original covariance matrix. Both the NFI and the NNFI are well below .9. The AIC value of 262.6 for our model is considerably lower than the independence model AIC value of 1602.04, but this indicates only that the hypothesized model represents a considerable improvement over a model in which the variables are hypothesized to be uncorrelated.

The results of the LM and Wald tests are often useful in identifying the sources of model misfit. The LM test is equivalent to the modification index (MI) in LISREL and represents the amount by which the overall chi-square value should decrease if a parameter were to be added to the model. In contrast, values of the Wald test represent the amount by which the overall chi-square value will increase if a parameter were to be dropped from the model. The LM and Wald tests are thus

TABLE 11.23

Loadings with TEN1, WOR1, IRTHK1, and BODY1 Free to Load on All Factors; TEN2, WOR2, IRTHK2, and BODY2 Free to Load on All Factors; and TEN3, WOR3, IRTHK3, and BODY3 Free to Load on All Factors

	TENSION	WORRY	TIRT	BODY		TENSION	WORRY	TIRT	BODY		TENSION	WORRY	TIRT	BODY
TEN1	0.98	−0.01	−0.01	−0.31	TEN1	0.68	—	—	—	TEN1	0.70	—	—	—
	(0.17)	(0.08)	(0.06)	(0.17)		(0.04)					(0.04)			
	5.92	−0.11	−0.12	−1.87		15.44					15.85			
TEN2	0.75	—	—	—	TEN2	0.98	0.15	0.00	−0.35	TEN2	0.80	—	—	—
	(0.05)					(0.14)	(0.08)	(0.06)	(0.15)		(0.05)			
	15.81					6.90	1.96	−0.03	−2.39		16.70			
TEN3	0.83	—	—	—	TEN3	0.83	—	—	—	TEN3	0.58	−0.08	−0.02	0.38
	(0.05)					(0.05)					(0.08)	(0.06)	(0.06)	(0.08)
	17.50					17.41					7.12	−1.26	−0.34	4.50
WOR1	−0.18	0.94	−0.14	−0.10	WOR1	—	0.64	—	—	WOR1	—	0.68	—	—
	(0.12)	(0.11)	(0.07)	(0.13)			(0.04)					(0.04)		
	−1.50	8.20	−1.97	−0.80			16.13					16.82		
WOR2	—	0.65	—	—	WOR2	0.01	0.65	0.04	−0.03	WOR2	—	0.67	—	—
		(0.05)				(0.09)	(0.07)	(0.05)	(0.10)			(0.05)		
		14.07				0.11	8.80	0.84	−0.28			14.57		
WOR3	—	0.65	—	—	WOR3	—	0.67	—	—	WOR3	0.11	0.50	0.05	0.11
		(0.04)					(0.04)				(0.07)	(0.06)	(0.05)	(0.07)
		15.50					16.37				1.63	8.82	1.02	1.51

IRTHK1	0.09	0.05	0.63	−0.09	IRTHK1	—	—	0.64	—	IRTHK1	—	—	0.64	—
	(0.10)	(0.06)	(0.06)	(0.11)				(0.04)					(0.04)	
	0.87	0.75	11.46	−0.81				15.42					15.20	
IRTHK2	—	—	0.67	—	IRTHK2	−0.14	0.01	0.64	0.13	IRTHK2	—	—	0.66	—
			(0.04)			(0.08)	(0.06)	(0.05)	(0.09)				(0.04)	
			16.09			−1.79	0.20	12.67	1.55				15.83	

IRTHK1	0.09	0.05	0.63	−0.09	IRTHK1	—	—	0.64	—	IRTHK1	—	—	0.64	—
	(0.10)	(0.06)	(0.06)	(0.11)				(0.04)					(0.04)	
	0.87	0.75	11.46	−0.81				15.42					15.20	
IRTHK2	—	—	0.67	—	IRTHK2	−0.14	0.01	0.64	0.13	IRTHK2	—	—	0.66	—
			(0.04)			(0.08)	(0.06)	(0.05)	(0.09)				(0.04)	
			16.09			−1.79	0.20	12.67	1.55				15.83	
IRTHK3	—	—	0.67	—	IRTHK3	—	—	0.68	—	IRTHK3	0.09	−0.06	0.74	−0.09
			(0.04)					(0.04)			(0.07)	(0.06)	(0.06)	(0.07)
			17.69					17.76			1.35	−1.17	12.84	−1.27
BODY1	−0.17	0.02	−0.03	0.55	BODY1	—	—	—	0.39	BODY1	—	—	—	0.39
	(0.14)	(0.07)	(0.06)	(0.15)					(0.04)					(0.04)
	−1.26	0.34	−0.47	3.54					10.56					10.50
BODY2	—	—	—	0.53	BODY2	0.12	−0.03	0.10	0.43	BODY2	—	—	—	0.55
				(0.05)		(0.10)	(0.07)	(0.05)	(0.12)					(0.05)
				11.24		1.14	−0.48	1.77	3.67					11.36
BODY3	—	—	—	0.53	BODY3	—	—	—	0.57	BODY3	0.00	0.02	−0.08	0.57
				(0.04)					(0.04)		(0.10)	(0.07)	(0.06)	(0.11)
				12.47					13.09		0.01	0.28	−1.34	5.39

TABLE 11.24

Command Lines for EQS CFA of Bolton Data

```
EXAMPLE 2: BOLTON DATA;
 /SPECS                                                        ①
CAS = 159;VARS = 10;ME = ML;MA = COV;
 /EQUATIONS
  V1 =            1.000*F2 + E1;
  V2 =            1.000*F2 + E2;
  V3 = 1.000*F1 + 1.000*F2 + E3;
  V4 = 1.000*F1 + 1.000*F2 + E4;
  V5 = 1.000*F1            + E5;                               ②
  V6 = 1.000*F1            + E6;
  V7 =            1.000*F2 + E7;
  V8 = 1.000*F1 + 1.000*F2 + E8;
  V9 = 1.000*F1            + E9;
 V10 = 1.000*F1            + E10;
 /VARIANCES
E1 TO E10 = .500*;                                             ③
F1 TO F2 = 1.000;
  /MATRIX
    1.0
    .59   1.0
    .30   .34   1.0
    .16   .24   .62   1.0
  -.02  -.13   .28   .37   1.0
    .00  -.05   .42   .51   .90   1.0
    .39   .61   .70   .59   .05   .20   1.0
    .17   .29   .57   .88   .30   .46   .60   1.0              ④
  -.04  -.14   .28   .33   .93   .86   .04   .28   1.0
  -.04  -.08   .42   .50   .87   .94   .17   .45   .90   1.0
  /STANDARD DEVIATIONS
    .45  1.06  1.17  1.11  1.50  .144  1.31  1.04
    /LMTEST; /WTEST;
    /END
```

① ME = ML means that the method of estimation (ME) is maximum likelihood covariance matrix (COV).
② These lines give the structure for the factor loadings. The asterisks design
③ The measurement error variances (E1) must be given starting values; here fixed at 1.00.
④ Only the lower half of the correlation matrix is required, along with the start

tests of whether parameters should be added to or deleted from the model, respectively. The results of the Wald and LM tests are presented in Table 11.27.

The results of the Wald test indicate that there are no model parameters that could be dropped without significantly worsening overall model fit. This is not surprising, as all parameter estimates were found to be highly significant.

The EQS program computes both univariate and multivariate forms of the LM test. The multivariate LM is generally preferred because it takes into account the correlations among the parameters. It may be that two parameters have high values for the univariate LM tests, but that these two parameters are highly correlated with one another. In such a case, adding both parameters will not decrease the overall chi-square value much more than would adding only one. The multivariate LM tests take the intercorrelations among the parameters into account in computing the estimated

①

②

③

④

:e matrix (COV).

).

, respectively. The

could be dropped
arameter estimates

test. The multivari-
ng the parameters.
but that these two
parameters will not
e. The multivariate
uting the estimated

TABLE 11.25

Factor Loadings, Standard Errors, and *t* Values for CFA of Bolton Data

MEASUREMENT EQUATIONS WITH STANDARD ERRORS AND TEST STATISTICS

V1 = V1 =	.302 ① *F2	+	1.000 ④ E1		
	.081 ②				
	3.711 ③				
V2 = V2 =	.461*F2	+	1.000 E2		
	.079				
	5.863				
V3 = V3 =	.349*F1	+	.585*F2	+	1.000 E3
	.068		.069		
	5.150		8.527		
V4 = V4 =	.428*F1	+	.796*F2	+	1.000 E4
	.056		.057		
	7.613		14.022		
V5 = V5 =	.934*F1	+	1.000 E5		
	.060				
	15.560				
V6 = V6 =	.960*F1	+	1.000 E6		
	.059				
	16.382				
V7 = V7 =	.728*F2	+	1.000 E7		
	.071				
	10.276				
V8 = V8 =	.371*F1	+	.815*F2	+	1.000 E8
	.057		.059		
	6.530		13.883		
V9 = V9 =	.930*F1	+	1.000 E9		
	.060				
	15.464				
V10 = V10 =	.965*F1	+	1.000 E10		
	.058				
	16.518				

① Factor loading.
② Factor correlation.
③ *t* Value.
④ Values of 1.000 here serve only to indicate that the measurement error was included in the model, and should not be interpreted as parameter estimates.

TABLE 11.26

Goodness-of-Fit Statistics from CFA of Bolton Data

GOODNESS OF FIT SUMMARY
INDEPENDENCE AIC = 1602.04359
MODEL AIC = 262.60282
CHI-SQUARE = 326.603 BASED ON 32 DEGREES OF FREEDOM
PROBABILITY VALUE FOR THE CHI-SQUARE STATISTIC IS LESS THAN 0.001
BENTLER-BONETT NORMED FIT INDEX = .807
BENTLER-BONETT NONNORMED FIT INDEX = .748

TABLE 11.27

Results of Wald and LM Tests From CFA of Bolton Data

WALD TEST (FOR DROPPING PARAMETERS)
NONE OF THE FREE PARAMETERS IS DROPPED IN THIS PROCESS.

LAGRANGE MULTIPLIER TEST (FOR ADDING PARAMETERS)
ORDERED UNIVARIATE TEST STATISTICS:

NO	PARAMETER	CHI-SQUARE	PROBABILITY	PARAMETER CHANGE
1	V9,F2 ①	22.643 ②	.000 ③	−.159 ④
2	V6,F2	15.386	.000	.110
3	V5,F2	13.986	.000	−.123
4	V10,F2	10.107	.001	.087
5	V7,F1	5.646	.017	.177
6	V2,F1	3.651	.056	−.145
7	F2,F1	.882	.348	.098
8	V1,F1	.434	.510	−.052
9	F2,F2	.000	1.000	.000
10	F1,F1	.000	1.000	.000

MULTIVARIATE LAGRANGE MULTIPLIER TEST BY SIMULTANEOUS PROCESS IN STAGE 1

CUMULATIVE MULTIVARIATE STATISTICS					UNIVARIATE INCREMENT	
STEP	PARAMETER	CHI-SQUARE	D.F.	PROB	CHI-SQUARE	PROB
1	V9,F2	22.643	1	.000	22.643	.000
2	V5,F2	45.826	2	.000	23.183	.000
3	V7,F1	53.946 ⑤	3	.000	8.119 ⑥	.004

① V9,F2 represents the loading of variable 9 on factor 2.
② Amount by which the overall chi-square value would decrease if the parameter V9,F2 were added to the model.
③ Probability of the associated chi-square value with 1 degree of freedom.
④ Value that the associated parameter would have if added to the model.
⑤ Total amount by which the overall chi-square value would decrease if the three parameters V9,F2, V5,F2, and V7,F1 were all added to the model.
⑥ Amount of the decrease in chi-square that would be accounted for by adding the parameter V7,F1 to the model.

decrease in the overall chi-square. This is why some parameters that are included in the univariate LM tests are not included in the multivariate test. For example, the parameter V6,F2 has a univariate value of 15.386, indicating that if variable 6 were allowed to load on factor 2, the overall chi-square value would decrease by 15.386. However, note that this decrease in chi-square would result only if that were the only parameter added to the model. The fact that the parameter V6,F2 is not included in the multivariate LM test probably indicates that it is so highly correlated with one or more of the other parameters that it would not result in a large decrease if other parameters were also added. The results of the multivariate LM test indicate that the greatest decrease in the overall chi-square value would occur if variables 9 and 5 were allowed to load on factor 2 and variable 7 were allowed to load on factor 1. Of course, these changes should be made only if they can be supported by theory.

Overall, then, the structure obtained from the EFA for these 10 items cannot be shown to fit the data optimally. Even the addition of the three-factor loadings described in the preceding paragraph would not result in a nonsignificant chi-square value. It may be that more than two factors are needed, or that the factors should be allowed to correlate. At this point, however, the researcher should carefully consider whether any proposed changes in the model can be justified theoretically.

11.20 Some Caveats Regarding Structural Equation Modeling

Covariance structure modeling (CSM) or structural equation modeling (SEM) techniques, which include CFA, have been used extensively since the 1980s. They have been touted as one of the most important advances in quantitative methodology in many years. One of the advantages of these techniques is that they allow for measurement error to be taken into account, which traditional procedures do not. Although these techniques are very sophisticated mathematically, they can now be implemented easily with the latest releases of programs such as LISREL and EQS. The availability of Windows versions of these programs has made their implementation still easier. However, Cliff (1983, 1989), among others, has cautioned researchers that the sophistication of these techniques and the facility with which they can now be applied should not blind researchers to some basic research principles.

One of these principles concerns the issue of capitalization on chance, which has been a major theme of this book. MacCallum, Roznowski, and Necowitz (1992) reported a compelling study on this issue. As noted earlier, it is not uncommon in practice for researchers to modify their models in a post hoc fashion, based on indices such as the LISREL MIs, or the LM and Wald tests given by the EQS program. This process of post hoc model modification is often called a *specification search*. What most researchers appear to be unaware of is that this is a *data-driven* process that is very susceptible to capitalization on chance. Because of this, modifications made in this way are likely to be very unstable, and are unlikely to cross-validate. This is particularly true when sample size is small, the number of modifications is large, and the modifications are not theoretically defensible (MacCallum, 1986). As MacCallum et al. (1992), noted:

> *Model modification in practice is usually done with no substantive justification and no cross-validation, often involves a substantial number of modifications, and is often based on samples that may be too small for such analyses...* . We consider this to be an unfortunate state of affairs, representing a dangerous and misleading methodological trend (p. 494, italics added).

The MacCallum et al. (1992) study found that no searches were based on a sequence of four modifications that resulted in the same modified models when sample size was 250 or less. Only when the sample size reached 400 was there *some* consistency. Unfortunately, most studies reported in the literature have sample sizes between 100 and 350.

A recent Monte Carlo study by Hutchinson (1994) is right on target in having investigated the stability of post hoc model modifications for some CFA models. Two population models were created involving two- and four-factor oblique factor structures. In each model, all factors had four primary indicators (with population loadings varying from .60 to .80) and two secondary population loadings of .40. Four levels of misspecification were imposed on the two models by incorrectly setting certain loadings to zero. I discuss only the first two levels of misspecification here. For Model 1, the first level had two secondary loadings incorrectly set to zero, and the second level had one primary and one secondary loading incorrectly set to zero. For the four-factor model, Level 1 misspecification involved incorrectly setting four secondary loadings to zero, and Level 2 had two primary and two secondary loadings incorrectly set to zero.

Sample sizes of 200, 400, 800, and 1,200 were chosen. One hundred samples were generated for each model and sample size combination, for a total of 800 samples. Hutchinson found that:

TABLE 11.28

Number of Times (Out of 100) Population Models Recovered

	Level of misspecification			
	Two-factor model		Four-factor model	
n	1	2	1	2
200	23	26	19	30
400	64	78	64	71
800	94	93	96	99
1,200	94	93	100	100

> Conditions with marked declines in values of MIs were also those that exhibited greater modification consistency .... When values of MIs seem to gradually decrease, even if still statistically significant, it suggests that there may be a number of specification errors present, but none of substantial size. Errors of this type are more likely to reflect chance characteristics of the data. Consequently, in practice one should probably try to limit modifications to correction of noticeably large specification errors which would be more likely to replicate in other samples (p. 25).

Table 11.28 shows that if the specification errors are relatively minor, a sample size of 400 gives one a good chance (from 64% to 78%) of recovering a known population model. More severe misspecification requires at least 800 subjects to obtain similar results.

Another problem encountered in SEM analyses is that researchers too often seem to interpret the finding that their model fits the data as meaning it is the *only* model that can do so. Various individuals (Bollen, 1989, p. 71; Cliff, 1983; Jöreskog, 1993, p. 298) noted that there are always other models that can fit the data as well, if not better, than the one originally hypothesized. These alternative models represent competing hypotheses that must be ruled out if the originally hypothesized model is to be supported.

In a 1993 paper, MacCallum, Wegener, Uchino, and Fabrigar discussed this issue in the context of mathematically equivalent models. These are models that cannot be distinguished from the originally hypothesized model on the basis of their goodness of fit. For example, in CFA, one model with items that load on more than one factor and another model with items that load on only one factor but have correlated measurement errors may fit equally well in terms of their chi-square values, even though they represent fundamentally different hypotheses. In cases like this, there is no statistical basis for choosing one model over another, and such decisions must be made on the basis of theoretical considerations. In their 1993 study, MacCallum et al. catalogued all applications of CSM in three prominent journals (*Journal of Educational Psychology*, *Journal of Applied Psychology*, and *Journal of Personality and Social Psychology*) for the years 1988 through 1991. For these articles the median number of equivalent models was quite large, as shown here:

Journal	Number of articles	Percent with equivalent models	Median number of models
Educ. Psych	14	86	16.5
Appl. Psych	19	74	12.0
Personality & Social Psych	20	100	21.0

They selected one article from each of these three journals and presented an analysis of three of the plausible equivalent models for each case. As MacCallum et al. (1993) noted:

> Importantly, the presented equivalent models have theoretical implications that differ substantially from the models preferred by the authors of the published applications. We know of no compelling evidence that would suggest that these equivalent models are theoretically less plausible than the original models.... The gravity of this issue for empirical research is increased by the fact that the phenomenon of equivalent models has been virtually ignored in practice. Of the 72 published applications examined by Becker (1990) and the additional 53 studies considered in this article, only *one* study contained an explicit acknowledgment of the existence of even a single equivalent model.... *Without adequate consideration of alternative equivalent models, support for one model from a class of equivalent models is suspect at best and potentially groundless and misleading.* (p. 196)

When taken in conjunction with their 1992 paper, the picture painted by MacCallum and his colleagues is a bleak one with regard to the amount of confidence one is justified in placing in the results of many CSM studies. What then should be done in CSM studies to enhance meaningfulness and generalizability? First, if post hoc model modifications are to be made, sample size must be adequate (probably 400 subjects for most studies, although this will depend on the size of the model). Also, no modifications should be made without a clear theoretical justification. Any model obtained as a result of such modifications should be treated *very tentatively* until the model has been validated on an independent sample of data. This is the issue of cross-validation, which has been stressed in this text, and which several prominent CSM researchers (Breckler, 1990; Browne & Cudeck, 1989; Cudeck & Browne, 1983; Jöreskog, 1993; MacCallum et al., 1992) have indicated is crucial in CSM research.

The problem of equivalent or alternative models must also be seriously considered in CSM studies. Although Jöreskog (1993, p. 295) indicated that the consideration of several *a priori* models is rare in practice, Bollen and Long (1993, p. 7), in the same volume, stated that one point of consensus among CSM researchers is that "it is better to consider several alternative models than to examine only a single model." While not all alternative models will be plausible, those that are should be estimated along with the originally hypothesized model. The values of such indexes as the AIC, ECVI, PGFI, and PNFI can then be used as a basis for comparing the fit of the various models.

In concluding this section, the following from Cliff (1987) is important. Most of all, one wonders at the personal arrogance and disrespect for the scientific process that is shown by some of these authors. Do they really think causal relations are established by a simple statistical analysis of a few, often adventitiously available, variables?

11.21 Summary

1. There are two types of factor analysis: exploratory and confirmatory. As pointed out, exploratory is more theory generating, whereas confirmatory is more theory testing.
2. The components are uncorrelated by how they are derived. For 30 variables the number of correlations is high (435), and it is very difficult to summarize by inspection precisely what this pattern of correlations represents. Principal components analysis is a means of "boiling down" the main sources of variation in such a complex set of correlations, and often a small number of components will account for most of the variance.
3. Three uses for components as a variable reducing scheme are (a) determining the number of dimensions underlying a test (b) reducing the number of predictors, prior to a regression analysis, and (c) reducing the number of dependent variables, prior to a MANOVA.
4. The absolute magnitude and number of loadings are crucial in determining reliable components. Components with at least four loadings $> |.60|$, or with at least three loadings $> |.80|$ are reliable. Also, components with at least 10 loadings $> |.40|$ are reliable for $N > 150$. Finally, Velicer has indicated (personal communication) that when the *average* of the 4 largest loadings (in absolute value) is > .60 or when the *average* of the 3 largest loadings (in absolute value) is > .80, then the components will be reliable.
5. I suggest *doubling* the critical value for an ordinary correlation and using that, at the .01 level, to determine whether a loading is significant.
6. For increasing interpretability of factors, there are two types of rotations: (a) orthogonal — the rotated factors are still uncorrelated, and (b) oblique — the rotated factors are correlated.
7. Only consider CFA (confirmatory factor analysis) if there is strong theory or a solid empirical basis (one or more previous studies that used exploratory factor analysis).
8. We have used two software packages (LISREL and EQS) to illustrate CFA. In particular, we have illustrated SIMPLIS (simple LISREL), which makes it very easy to run analyses.
9. Anyone using SEM (structural equation modeling) would be well advised to pay attention to the following warnings from McCallum regarding model modification and equivalent models: Model modification in practice is usually done with no substantive justification and no cross validation, often involves a substantial number of modifications, and is often based on samples that may be too small. Without adequate consideration of equivalent models, support for one model from a class of equivalent models is suspect at best and potentially groundless and misleading.
10. The McCallum warning on equivalent models occurred many years ago (1993), and one may think that things have improved a lot since then. Apparently that is NOT the case, at least that is how Hershberger (2006, *Structural Equation Modeling: A Second Course*: Hancock and Mueller (eds)) sees it. He notes that none of the major software packages that do SEM generate equivalent models. For example, AMOS, LISREL AND EQS do not generate equivalent models. To him and this writer, that is unacceptable.

11.22 Exercises

1. The notion o
ear combinat
analysis but
analysis and
ances for the
relation) mat
fundamental
make up the
set of p varial

The varian

where $\mathbf{a}' = (a_{11}$

(a) Write out
three vari

(b) The covar

and the

What is

2. Golding and
undergraduat
and obtained
lic speaking,
practice, milit
cal, nature, ag
service, religic

11.22 Exercises

1. The notion of a linear combination of variables and how much variance that linear combination accounts for is fundamental not only in principal components analysis but also in other forms of multivariate analysis such as discriminant analysis and canonical correlation. We indicated in this chapter that the variances for the successive components are equal to eigenvalues of covariance (correlation) matrix. However, the variance for a linear combination is defined more fundamentally in terms of the variances and covariances of the variables which make up the composite. We denote the matrix of variances and covariances for a set of p variables as:

$$\mathbf{S} = \begin{bmatrix} s_1^2 & s_{12} & \cdots & s_{1p} \\ s_{21} & s_2^2 & \cdots & s_{2p} \\ \vdots & \vdots & & \vdots \\ s_{p1} & s_{p2} & & s_p^2 \end{bmatrix}$$

The variance of a linear combination is defined as:

$$\text{var}(a_{11}x_1 + a_{12}x_2 + \cdots + a_{1p}x_p) = \text{var}(\mathbf{a}'\mathbf{x}) = \mathbf{a}'\mathbf{S}\mathbf{a},$$

where $\mathbf{a}' = (a_{11}, a_{12}, \ldots, a_{1p})$.

(a) Write out what the formula for the variance of a linear combination of two and three variables will be.

(b) The covariance matrix $\mathbf{S}$ for a set of three variables was:

$$\mathbf{S} = \begin{bmatrix} 451.4 & 271.2 & 168.7 \\ & 171.7 & 103.3 \\ \text{symm} & & 66.7 \end{bmatrix}$$

and the first principal component of $\mathbf{S}$ was

$$y_1 = .81x_1 + .50x_2 + .31x_3$$

What is the variance of y_1?

2. Golding and Seidman (1974) measured 231 undergraduate males enrolled in an undergraduate psychology course on the *Strong Vocational Interest Blank for Men*, and obtained the following correlation matrix on the 22 basic interest scales: public speaking, law/politics, business management, sales, merchandising, office practice, military activities, technical supervision, mathematics, science, mechanical, nature, agriculture, adventure, recreational leadership, medical service, social service, religious activities, teaching, music, art, and writing.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	1.0																					
2	.77	1.0																				
3	53	50	1.0																			
4	54	44	74	1.0																		
5	54	48	91	82	1.0																	
6	30	28	72	63	75	1.0																
7	16	20	28	19	26	31	1.0															
8	36	34	79	56	70	63	38	1.0														
9	−11	−05	08	02	05	20	03	14	1.0													
10	−10	−09	−03	−07	−08	02	15	05	50	1.0												
11	−02	−07	22	23	21	27	29	37	44	62	1.0											
12	14	−02	04	05	07	−03	23	11	−04	37	31	1.0										
13	09	−01	06	10	09	−03	24	11	−10	08	21	73	1.0									
14	21	18	15	15	14	−01	16	13	13	11	28	12	31	1.0								
15	16	21	22	22	22	23	29	18	03	−07	09	10	32	41	1.0							
16	23	24	09	12	12	05	19	08	08	41	24	33	05	12	10	1.0						
17	38	36	13	21	14	10	07	00	−19	−04	−07	23	09	−01	18	29	1.0					
18	32	17	18	22	17	27	17	13	−01	12	14	33	19	00	19	20	47	1.0				
19	37	23	29	35	28	30	15	20	−03	18	16	36	12	−02	12	22	51	41	1.0			
20	22	04	−01	05	06	−05	−22	−06	01	22	11	31	00	−05	−28	26	27	37	42	1.0		
21	19	−01	−06	04	05	−13	−15	−10	02	22	12	49	17	02	−22	23	26	25	34	73	1.0	
22	49	26	04	16	10	−08	−10	−06	−23	−04	−12	28	09	08	−02	15	42	31	42	57	62	1.0

Run a components analysis on this matrix. Also, do a varimax rotation, and compare the interpretations.

3. In which, if either, of the following cases would it be advisable to apply Bartlett's sphericity test before proceeding with a components analysis?

Case 1

1	.31	.45	.18	.56	.41	.50	
	1	.27	.36	.04	.30	.21	
		1	.63	.16	.41	.25	
			1	.28	.15	.32	125 subjects
				1	.46	.53	
					1	.39	
						1	

Case 2

1	.29	.18	.04	.11	.15	
	1	.07	.40	.12	.03	
		1	.23	.06	.13	111 subjects
			1	−.08	−.14	
				1	.12	
					1	

The actual sphericity test statistic is:

$$\chi^2 = -\left(N-1-\frac{2p+5}{6}\right)\ln|\mathbf{R}|, \text{ with } 1/2p(p-1)df$$

However, Lawley has shown that a good approximation to this statistic is:

$$\chi^2 = \left(N - 1 - \frac{2p+5}{6}\right)\sum\sum r_{ij}^2,$$

where the sum extends only over the correlations (r_{ij}) above the main diagonal.

Use the Lawley approximation for the two cases given here to determine whether you would reject the null hypothesis of uncorrelated variables in the population.

4. Consider the following correlation matrix:

$$\mathbf{R} = \begin{bmatrix} 1 & .6579 & .0034 \\ & 1 & -.0738 \\ & & 1 \end{bmatrix}$$

A principal components analysis on this matrix produced the following factor structure, that is, component–variable correlations:

	Principal Components		
Variables	1	2	3
1	.906	.112	.408
2	.912	−.005	−.411
3	−.097	.994	−.048

We denote the column of component–variable correlations for the first component by $\mathbf{h}_1$, for the second component by $\mathbf{h}_2$, and for the third component by $\mathbf{h}_3$. Show that the original correlation matrix $\mathbf{R}$ will be reproduced, within rounding error, by $\mathbf{h}_1\mathbf{h}_1' + \mathbf{h}_2\mathbf{h}_2' + \mathbf{h}_3\mathbf{h}_3'$. As you are doing this, observe what part of $\mathbf{R}$ the matrix $\mathbf{h}_1\mathbf{h}_1'$ reproduces, etc.

5. Consider the following principal components solution on five variables and the corresponding varimax rotated solution. Only the first two components are given, because the eigenvalues corresponding to the remaining components were very small (<.3).

			Varimax solution	
Variables	Comp 1	Comp 2	Factor 1	Factor 2
1	.581	.806	.016	.994
2	.767	−.545	.941	−.009
3	.672	.726	.137	.980
4	.932	−.104	.825	.447
5	.791	−.558	.968	−.006

(a) Find the percent of variance accounted for by each principal component.

(b) Find the percent of variance accounted for by each varimax rotated factor.

(c) Compare the variance accounted for by Component 1 (2) with variance accounted for by each corresponding rotated factor.

(d) Compare the total percent of variance accounted for by the two components with the total percent of variance accounted for by the two rotated factors.

6. Consider the following correlation matrix for the 12 variables on the General Aptitude Test Battery (GATB):

NAMES	1.000											
ARITh	.697	1.000										
DIM	.360	.366	1.00									
VOCAB	.637	.580	.528	1.00								
TOOLS	.586	.471	.554	.425	1.00							
MATH	.552	.760	.468	.616	.369	1.000						
SHAPES	.496	.411	.580	.444	.531	.400	1.00					
MARK	.561	.501	.249	.465	.444	.407	.387	1.000				
PLACE	.338	.297	.276	.211	.292	.300	.323	.494	1.00			
TURN	.349	.247	.279	.209	.336	.234	.401	.540	.773	1.00		
ASMBL	.390	.319	.358	.267	.361	.208	.444	.439	.468	.476	1.00	
DASMBL	.354	.325	.234	.283	.267	.311	.428	.422	.453	.482	.676	1.00

(a) Run a components analysis and varimax rotation on the SAS factor program.
(b) Interpret the components and the varimax rotated factors.
(c) Use the oblique rotation PROMAX, and interpret the oblique factors.
(d) What are the correlations among the oblique factors?
(e) Which factors seem more reasonable to use here?

7. Bolton (1971) measured 159 deaf rehabilitation candidates on 10 communication skills, of which six were reception skills in unaided hearing, aided hearing, speech reading, reading, manual signs, and fingerspellings. The other four communication skills were expression skills: oral speech, writing, manual signs, and fingerspelling. Bolton did what is called a principal axis analysis, which is identical to a components analysis, except that the factors are extracted from a correlation matrix with communality estimates on the main diagonal rather than 1's, as in components analysis. He obtained the following correlation matrix and varimax factor solution:

Correlation Matrix of Communication Variables for 159 Deaf Persons

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	M	S
C_1	*39*										1.10	0.45
C_2	59	*55*									1.49	1.06
C_3	30	34	*61*								2.56	1.17
C_4	16	24	62	*81*							2.63	1.11
C_5	−02	−13	28	37	*92*						3.30	1.50
C_6	00	−05	42	51	90	*94*					2.90	1.44
C_7	39	61	70	59	05	20	*71*				2.14	1.31
C_8	17	29	57	88	30	46	60	*78*			2.42	1.04
C_9	−04	−14	28	33	93	86	04	28	*92*		3.25	1.49
C_{10}	−04	−08	42	50	87	94	17	45	90	*94*	2.89	1.41

Note: The italicized diagonal values are squared multiple correlations.

(a) Interpret t
(b) Does the v
are correla
cipal axis
the origina
8. (a) As sugges
California
(b) Do the obli
max factor
(c) What are t
(d) Which fact
9. (a) Consider a
rotated fac
ity of thes
Velicer (198
(b) Now consi
Research F
of the four
which fact
10. Look at the tal
Tellegen's three
relations were u
varimax rotatio
that is, a princip
that should loa
(a) Concerning
1? How abo
(b) Concerning
to be a goo
are from Cl
Associatior

Varimax Factor Solution for 10 Communication Variables for 159 Deaf Persons

		I	II
C_1	Hearing (unaided)		49
C_2	Hearing (aided)		66
C_3	Speech reading	32	70
C_4	Reading	45	71
C_5	Manual signs	94	
C_6	Fingerspelling	94	
C_7	Speech		86
C_8	Writing	38	72
C_9	Manual signs	94	
C_{10}	Fingerspelling	96	
Percent of common variance		53.8	39.3

Note: Factor loadings less than .30 are omitted.

(a) Interpret the varimax factors. What does each of them represent?

(b) Does the way the variables that defined Factor 1 correspond to the way they are correlated? That is, is the empirical clustering of the variables by the principal axis technique consistent with the way those variables "go together" in the original correlation matrix?

8. (a) As suggested in the chapter, do the SPSS oblique rotation OBLIMIN on the California Psychological Inventory.

   (b) Do the oblique factors seem to be easier to interpret than the uncorrelated, varimax factors?

   (c) What are the correlations among the oblique factors?

   (d) Which factors, correlated or uncorrelated, would you prefer here?

9. (a) Consider again the factor analysis of the CPI, and in particular, the first two rotated factors presented in Table 11.4. Can we have confidence in the reliability of these factors according to the Monte Carlo results of Guadagnoli and Velicer (1988)?

   (b) Now consider the rotated factor loadings for the SAS run on the Personality Research Form given in Table 11.7. Can we have confidence in the reliability of the four rotated factors according to the Guadagnoli and Velicer study? For which factor(s) is the evidence strong, but not totally conclusive?

10. Look at the tables that follow. The first involves an exploratory factor analysis on Tellegen's three-factor model, that is, a principal axis analysis (squared multiple correlations were used as the communality estimates in the main diagonal) followed by a varimax rotation. The second is an exploratory factor analysis on the "big five" model, that is, a principal axis analysis followed by a varimax rotation. Note that those scales that should load on each factor according to the models have been boxed in.

    (a) Concerning Tellegen's model in the top table, do we have a good fit for Factor 1? How about for Factors 2 and 3?

    (b) Concerning the big five model (NEO scales), for which factors does there seem to be a good fit? For which factor(s) is the fit not so good? (Note: These tables are from Church and Burke, copyright © 1994, by the American Psychological Association. Reprinted with permission.)

Varimax-Rotated Factor Matrix for Multidimensional Personality Questionnaire (MPQ) Scales

	Factors			
MPQ scale	1	2	3	h^2
Well being	**−.55**	−.04	**.37**	.43
Social potency	−.03	.23	**.50**	.30
Achievement	−.07	.22	**.67**	.50
Social closeness	**−.49**	.05	−.02	.24
Stress reaction	**.44**	.01	−.03	.19
Alienation	**.66**	−.16	.12	.48
Aggression	**.42**	**−.38**	.19	.36
Control	−.04	**.76**	.12	.59
Harmavoidance	−.09	**.54**	−.23	.35
Traditionalism	−.03	**.41**	−.05	.17
Absorption	.12	−.19	**.31**	.15

Note: Factor loadings greater than |.30| are shown in boldface.

Varimax-Rotated Factor Matrix for NEO Scales

	Factors					
NEO scale	1	2	3	4	5	h^2
Neuroticism facets						
Anxiety	**.79**	−.01	.07	−.02	−.04	63
Hostility	**.43**	**−.56**	.03	.14	.06	.53
Depression	**.76**	−.17	.05	−.16	−.05	.64
Self-consciousness	**.73**	.02	−.16	−.14	−.16	.60
Impulsiveness	**.45**	.00	.23	−.05	**.39**	.40
Vulnerability	**.66**	−.18	−.14	−.28	.07	.57
Extraversion facets						
Warmth	−.06	**.65**	.15	.25	**.37**	.65
Gregariousness	−.10	.20	−.02	.02	**.60**	.41
Assertiveness	**−.36**	−.18	.14	**.61**	.21	.60
Activity	−.08	.07	−.02	**.53**	.08	.30
Excitement-seeking	−.01	05	.14	.04	**.56**	.33
Positive emotions	−.12	**.59**	.22	.27	**.35**	.61
Openness facets						
Fantasy	.15	.05	**.54**	−.11	.20	.37
Aesthetics	.13	.20	**.63**	.04	−.01	.45
Feelings	.17	**.45**	**.50**	.23	.13	.55
Actions	−.24	−.02	**.48**	−.15	.16	.33
Ideas	−.18	−.06	**.63**	.16	−.13	.48
Values	−.04	.18	**.43**	.03	.11	.23
Agreeableness	−.08	**.80**	.19	.01	.05	.69
Conscientiousness	−.13	**.32**	−.05	**.65**	−.25	.60

Note: Factor loadings greater than |.30| are shown in boldface. Primary loadings hypothesized in the NEO Big Five model are shown in boxes.

11. Consider th
(a) Draw th
(b) Are the
they we
(c) Does th
(d) Does th
(e) The mo
square :
PLACE
doing th

TITLE : G.
OBSERVE
SHAPES N
CORREL/
1.00
.697 1.00
.360 .366
.637 .580
.586 .471
.552 .760
.496 .411
.561 .501
.338 .297
.349 .247
.390 .319
.354 .325
SAMPLE S
LATENT V
RELATION
ARITH M/
TURN PL/
DIM SHAI
END OF P

GATB – T
CORREL

NAMES
ARITH
DIM
VOCAB
TOOLS
MATH
SHAPES
PLACE
TURN
ASMBL
DASMBL

11. Consider the following confirmatory factor analysis output from LISREL 8.
    (a) Draw the path diagram, labeling all "variables" clearly.
    (b) Are the variables loading significantly (test each at the .01 level) on the factor they were supposed to measure?
    (c) Does the chi square test indicate a good fit at the .05 level?
    (d) Does the value of RMSEA indicate a good fit?
    (e) The modification indices near the end indicate that we could reduce the chi square statistic considerably by adding an error covariance for TURN and PLACE and an error covariance for DASMBL and ASMBL. Should we consider doing this?

SIMPLIS INPUT FILE

```
TITLE : GATB – THREE CORRELATED FACTORS
OBSERVED VARIABLES: NAMES ARITH DIM VOCAB TOOLS MATH
SHAPES MARK PLACE TURN ASMBL DASMBL
CORRELATION MATRIX:
1.00
.697  1.00
.360  .366  1.00
.637  .580  .528  1.00
.586  .471  .554  .425  1.00
.552  .760  .468  .616  .369  1.00
.496  .411  .580  .444  .531  .400  1.00
.561  .501  .249  .465  .444  .407  .387  1.00
.338  .297  .276  .211  .292  .300  .323  .494  1.00
.349  .247  .279  .209  .336  .234  .401  .540  .773  1.00
.390  .319  .358  .267  .361  .208  .444  .439  .468  .476  1.00
.354  .325  .234  .283  .267  .311  .428  .422  .453  .482  .676  1.00
SAMPLE SIZE=200
LATENT VARIABLES=FACTOR1 FACTOR2 FACTOR3
RELATIONSHIPS:
ARITH MATH NAMES VOCAB=FACTOR1
TURN PLACE DASMBL ASMBL=FACTOR2
DIM SHAPES TOOLS=FACTOR3
END OF PROBLEM
```

OUTPUT

GATB – THREE CORRELATED
CORRELATION MATRIX TO BE ANALYZED

	NAMES	ARITH	DIM	VOCAB	TOOLS	MATH
NAMES	1.00					
ARITH	0.70	1.00				
DIM	0.36	0.37	1.00			
VOCAB	0.64	0.58	0.53	1.00		
TOOLS	0.59	0.47	0.55	0.42	1.00	
MATH	0.55	0.76	0.47	0.62	0.37	1.00
SHAPES	0.50	0.41	0.58	0.44	0.53	0.40
PLACE	0.34	0.30	0.28	0.21	0.29	0.30
TURN	0.35	0.25	0.28	0.21	0.34	0.23
ASMBL	0.39	0.32	0.36	0.27	0.36	0.21
DASMBL	0.35	0.32	0.23	0.28	0.27	0.31

CORRELATION MATRIX TO BE ANALYZED

	SHAPES	PLACE	TURN	ASMBL	DASMBL
SHAPES	1.00				
PLACE	0.32	1.00			
TURN	0.40	0.77	1.00		
ASMBL	0.44	0.47	0.48	1.00	
DASMBL	0.43	0.45	0.48	0.68	1.00

GATB – THREE CORRELATED
Number of Iterations = 11

LISREL ESTIMATES (MAXIMUM LIKELIHOOD)

NAMES = 0.79*FACTOR1, Errorvar. = 0.37, R^2 = 0.63
(0.061) (0.047)
12.88 7.93

ARITH = 0.86*FACTOR1, Errorvar. = 0.25, R^2 = 0.75
(0.059) (0.040)
14.71 6.36

DIM = 0.74*FACTOR3, Errorvar. = 0.46, R^2 = 0.54
(0.067) (0.061)
11.07 7.46

VOCAB = 0.74*FACTOR1, Errorvar. = 0.45, R^2 = 0.55
(0.063) (0.053)
11.74 8.50

TOOLS = 0.74*FACTOR3, Errorvar. = 0.46, R^2 = 0.54
(0.067) (0.061)
11.08 7.45

MATH = 0.81*FACTOR1, Errorvar. = 0.34, R^2 = 0.66
(0.061) (0.045)
13.34 7.62

SHAPES = 0.76*FACTOR3, Errorvar. = 0.42, R^2 = 0.58
(0.066) (0.060)
11.54 7.05

PLACE = 0.84*FACTOR2, Errorvar. = 0.29, R^2 = 0.71
(0.062) (0.049)
13.66 6.00

TURN = 0.86*FACTOR2, Errorvar. = 0.26, R^2 = 0.74
(0.061) (0.048)
14.12 5.39

ASMBL = 0.63*FACTOR2, Errorvar = 0.60, R^2 = 0.40
(0.068) (0.067)
9.31 9.00

DASMBL = 0.62*FACTOR2, Errorvar. = 0.62, R^2 = 0.38
(0.068) (0.068)
9.13 9.05

GOODNESS (
CHI-SQUARE
ESTIMATED
MINIMUM FI
POPULATIO
ROOT MEAN
P-VALUE FO
EXPECTED C
ECVI FOR SA
ECVI FOR IN
CHI-SQUARE
INDEPENDE
MODEL AIC =
SATURATED
INDEPENDE
MODEL CAIC
SATURATED
ROOT MEAN
STANDARDIZ
GOODNESS C
ADJUSTED G
PARSIMONY
NORMED FIT
NON-NORME
PARSIMONY
COMPARATIV
INCREMENTA
RELATIVE FIT
CRITICAL N (
CONFIDENCE
FOR CHI-SQU
GATB – THREE
SUMMARY ST
SMALLEST FIT
MEDIAN FITT
LARGEST FITT
STEMLEAF PL
- 0 | 9937777776
- 0 | 3322220000
0 | 1111123444
0 | 5555666889
1 | 0034
1 | 6778
2 |
2 | 9

MBL

0


```
GOODNESS OF FIT STATISTICS

CHI-SQUARE WITH 41 DEGREES OF FREEDOM = 225.61 (P = 0.0)
ESTIMATED NON-CENTRALITY PARAMETER (NCP) = 184.61

MINIMUM FIT FUNCTION VALUE = 1.13
POPULATION DISCREPANCY FUNCTION VALUE (F0) = 0.93
ROOT MEAN SQUARE ERROR OF APPROXIMATION (RMSEA) = 0.15
P-VALUE FOR TEST OF CLOSE FIT (RMSEA < 0.05) = 0.00000037

EXPECTED CROSS-VALIDATION INDEX (ECVI) = 1.38
ECVI FOR SATURATED MODEL = 0.66
ECVI FOR INDEPENDENCE MODEL = 6.43

CHI-SQUARE FOR INDEPENDENCE MODEL WITH 55 DEGREES OF FREEDOM = 1257.96
INDEPENDENCE AIC = 1279.96
MODEL AIC = 275.61
SATURATED AIC = 132.00
INDEPENDENCE CAIC = 1327.24
MODEL CAIC = 383.07
SATURATED CAIC = 415.69

ROOT MEAN SQUARE RESIDUAL (RMR) = 0.078
STANDARDIZED RMR = 0.078
GOODNESS OF FIT INDEX (GFI) = 0.84
ADJUSTED GOODNESS OF FIT INDEX (AGFI) = 0.74
PARSIMONY GOODNESS OF FIT INDEX (PGFI) = 0.52

NORMED FIT INDEX (NFI) = 0.82
NON-NORMED FIT INDEX (NNFI) = 0.79
PARSIMONY NORMED FIT INDEX (PNFI) = 0.61
COMPARATIVE FIT INDEX (CFI) = 0.85
INCREMENTAL FIT INDEX (IFI) = 0.85
RELATIVE FIT INDEX (RFI) = 0.76

CRITICAL N (CN) = 58.29


CONFIDENCE LIMITS COULD NOT BE COMPUTED DUE TO TOO SMALL P-VALUE
FOR CHI-SQUARE

GATB – THREE CORRELATED

SUMMARY STATISTICS FOR FITTED RESIDUALS
SMALLEST FITTED RESIDUAL = -0.09
MEDIAN FITTED RESIDUAL = 0.00
LARGEST FITTED RESIDUAL = 0.29
STEMLEAF PLOT
- 0 | 9937777777666666555
- 0 | 3322220000000000000
  0 | 1111123444
  0 | 55556668899
  1 | 0034
  1 | 6778
  2 |
  2 | 9
```

```
SUMMARY STATISTICS FOR STANDARDIZED RESIDUALS
SMALLEST STANDARDIZED RESIDUAL = -5.36
MEDIAN STANDARDIZED RESIDUAL = 0.00
LARGEST STANDARDIZED RESIDUAL = 8.83

STEMLEAF PLOT
- 4 | 410
- 2 | 5200420
- 0 | 9887664339644300000000000
  0 | 346780001111356778
  2 | 00362336
  4 | 54
  6 | 5
  8 | 8
LARGEST NEGATIVE STANDARDIZED RESIDUALS
RESIDUAL FOR    DIM AND      ARITH    -3.01
RESIDUAL FOR    VOCAB AND    ARITH    -4.13
RESIDUAL FOR    MATH AND     NAMES    -5.36
RESIDUAL FOR    ASMBL AND    PLACE    -3.22
RESIDUAL FOR    ASMBL AND    TURN     -3.98
RESIDUAL FOR    DASMBL AND   PLACE    -3.50
RESIDUAL FOR    DASMBL AND   TURN     -3.03
LARGEST POSITIVE STANDARDIZED RESIDUALS
RESIDUAL FOR    VOCAB AND    DIM       3.33
RESIDUAL FOR    TOOLS AND    NAMES     4.48
RESIDUAL FOR    MATH AND     ARITH     5.44
RESIDUAL FOR    TURN AND     PLACE     8.83
RESIDUAL FOR    ASMBL AND    NAMES     3.21
RESIDUAL FOR    ASMBL AND    SHAPES    3.55
RESIDUAL FOR    DASMBL AND   NAMES     2.58
RESIDUAL FOR    DASMBL AND   SHAPES    3.28
RESIDUAL FOR    DASMBL AND   ASMBL     7.48
```

THE MODIFICATION INDICES SUGGEST TO ADD THE

PATH TO	FROM	DECREASE IN CHI-SQUARE	NEW ESTIMATE
ARITH	FACTOR3	9.3	−0.28
ASMBL	FACTOR3	12.0	0.29

THE MODIFICATION INDICES SUGGEST TO ADD AN ERROR COVARIANCE

BETWEEN	AND	DECREASE IN CHI-SQUARE	NEW ESTIMATE
DIM	NAMES	14.5	−0.14
DIM	ARITH	8.0	−0.09
VOCAB	ARITH	17.1	−0.16
VOCAB	DIM	16.4	0.16
TOOLS	NAMES	18.7	0.16
MATH	NAMES	28.8	−0.21
MATH	ARITH	29.5	0.22
MATH	DIM	8.8	0.11
MATH	TOOLS	10.0	−0.11
TURN	PLACE	78.0	0.60
ASMBL	PLACE	10.4	−0.15
ASMBL	TURN	15.8	−0.19
DASMBL	PLACE	12.3	−0.17
DASMBL	TURN	9.2	−0.14
DASMBL	ASMBL	56.0	0.37

12. Consider the following confirmatory factor analysis output from LISREL 8.
    (a) Draw the path diagram, labeling all "variables" clearly.
    (b) Are VISUAL and VERBAL significantly correlated at the .01 level?
    (c) Are the indicators for VERBAL significantly linked to it at the .01 level?
    (d) Does the chi square test indicate a good fit at the .05 level?
    (e) Do some of the other indices (e.g., AGFI and NNFI) also indicate a good fit?

SIMPLIS INPUT FILE

```
TITLE : THREE FACTOR FROM JORESKOG
OBSERVED VARIABLES: VISPERC CUBES LOZENGES PARCOMP SENCOMP WORD ADD
COUNT SCCAPS
CORRELATION MATRIX:
1.00
.318  1.00
.436  .419  1.00
.335  .234  .323  1.00
.304  .157  .283  .722  1.00
.326  .195  .350  .714  .685  1.00
.116  .057  .056  .203  .246  .170  1.00
.314  .145  .229  .095  .181  .113  .585  1.00
.489  .239  .361  .309  .345  .280  .408  .512  1.00
SAMPLE SIZE: 145
LATENT VARIABLES: VISUAL VERBAL SPEED
RELATIONSHIPS:
VISPERC CUBES LOZENGES SCCAPS=VISUAL
PARCOMP SENCOMP WORD=VERBAL
ADD COUNT SCCAPS=SPEED
END OF PROBLEM
```

OUTPUT

THREE FACTOR FROM

CORRELATION MATRIX TO BE ANALYZED

	VISPERC	CUBES	LOZENGES	PARCOMP	SENCOMP	WORD
VISPERC	1.00					
CUBES	0.32	1.00				
LOZENGES	0.44	0.42	1.00			
PARCOMP	0.34	0.23	0.32	1.00		
SENCOMP	0.30	0.16	0.28	0.72	1.00	
WORD	0.33	0.20	0.35	0.71	0.68	1.00
ADD	0.12	0.06	0.06	0.20	0.25	0.17
COUNT	0.31	0.14	0.23	0.10	0.18	0.11
SCCAPS	0.49	0.24	0.36	0.31	0.34	0.28

CORRELATION MATRIX TO BE ANALYZED

	ADD	COUNT	SCCAPS
ADD	1.00		
COUNT	0.58	1.00	
SCCAPS	0.41	0.51	1.00

THREE FACTOR FROM
Number of Iterations = 8

LISREL ESTIMATES (MAXIMUM LIKELIHOOD)

VISPERC = 0.71*VISUAL, Errorvar.= 0.50, R^2 = 0.50
(0.087) (0.090)
8.16 5.53

CUBES = 0.48*VISUAL, Errorvar.= 0.77, R^2 = 0.23
(0.091) (0.10)
5.33 7.62

LOZENGES = 0.65*VISUAL, Errorvar.= 0.58, R^2 = 0.42
(0.087) (0.091)
7.43 6.34

PARCOMP = 0.87*VERBAL, Errorvar.= 0.25, R^2 = 0.75
(0.070) (0.051)
12.37 4.81

SENCOMP = 0.83*VERBAL, Errorvar.= 0.31, R^2 = 0.69
(0.071) (0.054)
11.61 5.80

WORD = 0.83*VERBAL, Errorvar.= 0.32, R^2 = 0.68
(0.072) (0.054)
11.51 5.91

ADD = 0.68*SPEED, Errorvar.= 0.54, R^2 = 0.46
(0.089) (0.093)
7.68 5.76

COUNT = 0.86*SPEED, Errorvar.= 0.26, R^2 = 0.74
(0.092) (0.11)
9.37 2.31

SCCAPS = 0.46*VISUAL + 0.42*SPEED, Errorvar.= 0.47, R^2 = 0.53
(0.089) (0.088) (0.073)
5.15 4.73 6.42

CORRELATION MATRIX OF INDEPENDENT VARIABLES

	VISUAL	VERBAL	SPEED
VISUAL	1.00		
VERBAL	0.56		1.00
	(0.08)		
	6.87		
SPEED	0.39	0.22	1.00
	(0.10)	(0.10)	
	3.73	2.32	

GOODNES

CHI-SQUA
ESTIMATE
90 PERCEN
MINIMUM
POPULATI
90 PERCEN
ROOT MEA
90 PERCEN
P-VALUE F

EXPECTED
90 PERCEN
ECVI FOR S
ECVI FOR I

CHI-SQUAR
INDEPEND
MODEL AIC
SATURATE
INDEPEND
MODEL CA
SATURATE

ROOT MEA
STANDARD
GOODNESS
ADJUSTED
PARSIMONY

NORMED FI
NON-NORM
PARSIMONY
COMPARATI
INCREMENT
RELATIVE FI

CRITICAL N

THREE FACT

SUMMARY S
SMALLEST FI
MEDIAN FIT
LARGEST FIT

STEMLEAF PI
- 10 | 6
- 8 |
- 6 | 21171
- 4 | 52
- 2 | 7444
- 0 | 878721000
0 | 22791
2 | 2

```
GOODNESS OF FIT STATISTICS

CHI-SQUARE WITH 23 DEGREES OF FREEDOM = 29.01 (P = 0.18)
ESTIMATED NON-CENTRALITY PARAMETER (NCP) = 6.01
90 PERCENT CONFIDENCE INTERVAL FOR NCP = (0.0 ; 23.96)
MINIMUM FIT FUNCTION VALUE = 0.20
POPULATION DISCREPANCY FUNCTION VALUE (F0) = 0.042
90 PERCENT CONFIDENCE INTERVAL FOR F0 = (0.0 ; 0.17)
ROOT MEAN SQUARE ERROR OF APPROXIMATION (RMSEA) = 0.043
90 PERCENT CONFIDENCE INTERVAL FOR RMSEA = (0.0 ; 0.085)
P-VALUE FOR TEST OF CLOSE FIT (RMSEA < 0.05) = 0.57

EXPECTED CROSS-VALIDATION INDEX (ECVI) = 0.51
90 PERCENT CONFIDENCE INTERVAL FOR ECVI = (0.47 ; 0.63)
ECVI FOR SATURATED MODEL = 0.62
ECVI FOR INDEPENDENCE MODEL = 3.57

CHI-SQUARE FOR INDEPENDENCE MODEL WITH 36 DEGREES OF FREEDOM = 496.67
INDEPENDENCE AIC = 514.67
MODEL AIC = 73.01
SATURATED AIC = 90.00
INDEPENDENCE CAIC = 550.46
MODEL CAIC= 160.50
SATURATED CAIC = 268.95

ROOT MEAN SQUARE RESIDUAL (RMR) = 0.045
STANDARDIZED RMR = 0.045
GOODNESS OF FIT INDEX (GFI) = 0.96
ADJUSTED GOODNESS OF FIT INDEX (AGFI) = 0.92
PARSIMONY GOODNESS OF FIT INDEX (PGFI) = 0.49

NORMED FIT INDEX (NFI) = 0.94
NON-NORMED FIT INDEX (NNFI) = 0.98
PARSIMONY NORMED FIT INDEX (PNFI) = 0.60
COMPARATIVE FIT INDEX (CFI) = 0.99
INCREMENTAL FIT INDEX (IFI) = 0.99
RELATIVE FIT INDEX (RFI) = 0.91

CRITICAL N (CN) = 207.70


THREE FACTOR FROM

SUMMARY STATISTICS FOR FITTED RESIDUALS
SMALLEST FITTED RESIDUAL = -0.12
MEDIAN FITTED RESIDUAL = 0.00
LARGEST FITTED RESIDUAL = 0.12

STEMLEAF PLOT
- 10|6
 - 8|
 - 6|21171
 - 4|52
 - 2|7444
 - 0|8787210000000000000
   0|22791
   2|2
```

```
 4 | 5016
 6 | 17
 8 |
10 | 5
12 | 0

SUMMARY STATISTICS FOR STANDARDIZED RESIDUALS
SMALLEST STANDARDIZED RESIDUAL = -1.96
MEDIAN STANDARDIZED RESIDUAL = 0.00
```

13. For exercise 2, use ONLY the first 15 variables. Obtain three factors for the following runs:
    (a) Run a components analysis and varimax rotation.
    (b) Run a components analysis and oblique rotation.
    (c) Which of the above solutions would you prefer?
14. Consider the RMSEA's in Tables 11.16 and 11.17. Do they offer us a clear choice as to which model is to be preferred?
15. Consider the following part of the quote from Pedhazur and Schmelkin (1991), "... It boils down to the question: Are aspects of a postulated multidimensional construct intercorrelated? The answer to this question is relegated to the status of an assumption when an orthogonal rotation is employed."
    (a) What did they mean by the last part of this statement?

12

Canonical

12.1 Introducti

In Chapter 3, we e
using multivariat
is in prediction, a
ables (both predic
Canonical correlati
variables, and *is a*
mutually independe
the use of pairs of
Because the com
ing of the total bet
ponents analysis (
uncorrelated linea
combinations (the
variables, whereas
so-called *canonical*
interpreting the pr
ables and the com
variables and the c
One could consi
stated, investigator
dependent variable
Let us now consi
An investigator wi
(say, as measured b
and a battery of ac
pair of canonical va
linear combination
most highly with th
achievement (as rev
ond pair of canonic
ated with a differen
As a second exam
the *same* set of varia
of the personality p
years. Canonical co