

characteristic of subjects that related to the treatment (which might well be the baseline condition) and not consider it in advance). You can then question. You then incorporate the two-way factorial design. Because the sample sizes will probably be unequal, the ANOVA will be

matching, there will be an interpretive effect. The treatment variable interacts strongly. The treatment effect depends on whatever if subjects were matched on a baseline. The interaction might force you to qualify its effects of the drugs. You might find that behavior only among subjects that were not those subjects who were initially matched from the matching variable \times treatment. At the point that all interactions force main effects.

ary along three or more dimensions. There are independent group manipulated measures or blocked variables. In some cases, cells themselves can be missing. In the same, but the problems increase designs.

riates are removed before treatment analysis. Various examples. First-order interactions of main effect codes. Second-order interactions involving first-order interaction codes by

table efficiency and generality as you have more than one independent variable. Designs with more than three factors are difficult to interpret, especially when cell sizes are unequal.

In some detail, I will now turn to actually the observed variables and the observed variables that are continuous than the original variables. This occupies the entire next chapter, is. Many forms of factor analysis involve analysis of mathematical (atheoretical)

6 Exploratory Factor Analysis

Chapter Overview

Factor analysis may be viewed as a set of models for transforming a group of variables into a simpler and more useful form. In essence, linear combinations are formed from variables, and the resulting linear combinations are used to "predict" the original variables. The reason for using this apparently circular process is that a small number of linear combinations, which define a new set of variables, may be able to describe all or nearly all of the meaning of the larger set of original variables.

There are many forms of factor analysis. One basic distinction, discussed in Chapter 1, is that between *exploratory* and *confirmatory*¹ factor analysis. In exploratory factor analysis, the linear combinations are formed in such a way as to satisfy some *mathematical (nonsubstantive)* criteria, such as explaining the highest percentage of variance inherent in the original set of variables. The researcher's major task is to interpret the meaning of the factors so obtained. In confirmatory factor analysis, the linear combinations are formed to satisfy some *substantive* criteria implicitly or explicitly suggested by a theory. Here, the goal of the inquiry is to see if the defined factors fit the data well. The present chapter is focused on exploratory factor analysis; confirmatory factor analysis will be covered in Chapter 7. This chapter is divided into eight parts.

I. THE BASIC FACTOR ANALYTIC MODEL—All factor analytic models contain a *factor equation* that relates an observable *raw-score matrix* to the product of two new matrices that are not directly observable: (a) the *pattern matrix*, which contains beta weights to predict variables from factors, and (b) the *factor score matrix*, which contains scores for each subject on each factor. A *classic* or *common factor model* contains two additional matrices, namely, the pattern and factor score matrices for *unique factors*. These unique

¹ My definition of confirmatory factor analysis follows from Nunnally's (1978) point of view. Those who work within the LISREL tradition discussed in the next chapter adopt a more restricted definition.

factors, in essence, represent random or measurement error. On the other hand, a *component model* does not contain an explicit concept of measurement error.

Using simple algebra, one can derive a *covariance equation* from the factor equation. The covariance equation defines the observed *correlation* or *variance-covariance matrix* as the product of the aforementioned pattern matrix and the *factor correlation matrix*, which describes the correlations among the factors. The factor correlation matrix is an *identity matrix* in an *orthogonal solution*. When the factor correlation matrix is not an identity matrix, the result is known as an *oblique solution*. The covariance equation contains one more term in the common factor model, the *uniqueness matrix*, which contains the error variance for each variable. An infinite number of pairs of pattern and factor score matrices provide mathematically equivalent solutions to the factor equation. Similarly, an infinite number of pairs of pattern and factor correlation matrices provide mathematically equivalent solutions to the covariance equation. Consequently, *no factor analytic solution is unique*. This is termed *the first form of factor indeterminacy*.

II. COMMON USES OF FACTOR ANALYSIS—There is no such thing as “a” factor analysis of a set of data. One reason there are alternative strategies is that factor analyses may be used for different purposes. Some of the major functions of factor analysis are (1) *orthogonalization* of a set of variables, (2) *reduction* in the size of a set of variables, (3) *dimensional analysis* of latent dimensions, (4) generation of *factor scores*, and (5) *statistical control*.

III. AN OVERVIEW OF THE EXPLORATORY FACTORING PROCESS—In this brief section, four steps common to nearly all factor analyses are described: (1) *entering the data*, (2) forming the *correlation matrix*, (3) obtaining the *unrotated (principal component)* solution, and (4) *rotating factors* to a *secondary solution*, which is normally the final product of interest.

IV. PRINCIPAL COMPONENTS—Even though a principal component (PC) solution is usually an interim result, it is very important to understand it. A simulated problem is presented in which the solution is shown to depend on an eigenanalysis (Chapter 2) of the correlation or variance-covariance matrix. Specifically, the principal components are demonstrated to be eigenvectors, and the variance they account for are their associated eigenvalues. Among the most important properties of a PC solution is the fact that the first, second, third, ..., PC account for the most available variance, both individually and collectively.

One reason a principal component solution is important is that its results commonly determine the number of factors retained in the rotation. Some of the major criteria for deciding on the number of factors are discussed. Although the most common criterion for determining the number of factors is the number of eigenvalues greater than 1.0, it is not the only criterion, and others may be more suitable in different situations. In particular, what is known as the *Scree criterion* may be a better alternative in item-level factoring.

Various matrices derivable from the basic matrices contained in the factor

and covariance equations are intro-
of these is the matrix containing cor-
the *structure matrix*.

V. FACTOR DEFINITION A
among different kinds of factors are
all variables in the set load (contai-
factor on which some, but not all, va-
only one variable loads (other varia-
unipolar factor (which may be a g-
variables either load in the same dir-
or do not load at all; and (5) a *bipol-*
a group factor), which some variab-
tively. *Factor definition* refers to the
to a group of variables that load on

The reason these distinctions are i-
of factors which provide totally equi-
form of factor indeterminacy). Put
are *mathematically* equivalent. How-
logical utility. Different schools of th-
popular view is Thurstone's (1947) s-
is undesirable. Holzinger (Holzinger)
emphasized the importance of a gen-

Alternate, equivalent solutions can
of *rotation*. Principles of rotation
orthogonal rotation. *Graphic* repre-
more so before the invention of c-
analytically, i.e., to satisfy one of se-
logic of the two most popular meth-
especially, *varimax*, is covered. This
oblique rotations.

VI. THE COMMON FACTOR
the common factor model to the c-
noted earlier is that the common fac-
measurement error; the component
a given number of common factors
matrix better than the same numbe-
standpoint, the common factor mode-
numbers less than unity in the diag-
component models operate upon a
diagonal. Unfortunately, several met-
munalities, but there is no general a;
general approaches to the commun-
using the reliability coefficient to defi-
which is used by LISREL and relate-
by-product of other operations. For

measurement error. On the other hand, there is an explicit concept of measurement error.

The variance equation from the factor analysis model is the observed correlation or variance matrix. This matrix describes the correlations among the variables. The identity matrix in an orthogonal factor solution is not an identity matrix, the covariance equation contains one unique solution. The covariance equation contains one infinite number of pairs of pattern matrices. There are two types of equivalent solutions to the covariance equation: (1) pattern matrices that are mathematically equivalent and (2) pattern matrices that are analytically equivalent. This is true for any analytic solution.

ANALYSIS—There is no such thing as a "best" factor analysis. One reason there are alternative factor analyses is that they serve different purposes. Some of the purposes are (1) *orthogonalization* of a set of variables, (2) *dimensional analysis*, (3) *correlation scores*, and (4) *statistical control*.
EXPLORATORY FACTORING PROCEDURES—In nearly all factor analyses, the first step is to obtain the correlation matrix, (2) to find the eigenvalues and eigenvectors, and (3) to rotate the factors to a final product of interest.

Although a principal component analysis is very important to understand the solution, it is shown to depend on the correlation or covariance matrix. The eigenvalues and eigenvectors are demonstrated to be eigenvalues and eigenvectors of the correlation matrix. The PC solution is the fact that the most available variance, both

and covariance equations are introduced here. Perhaps the most important of these is the matrix containing correlations between variables and factors—the *structure matrix*.

V. FACTOR DEFINITION AND ROTATION—Some distinctions among different kinds of factors are introduced: (1) a *general factor* on which all variables in the set load (contain a substantial beta weight); (2) a *group factor* on which some, but not all, variables load; (3) a *singlet factor* on which only one variable loads (other variables may have small loadings on it); (4) a *unipolar factor* (which may be a general or a group factor) on which all variables either load in the same direction, i.e., are all positive or all negative, or do not load at all; and (5) a *bipolar factor* (which may also be a general or a group factor), which some variables load positively and others load negatively. *Factor definition* refers to the identification of the properties common to a group of variables that load on a particular factor.

The reason these distinctions are important is that there exist an infinite set of factors which provide totally equivalent representation of the data (the first form of factor indeterminacy). Put in another way, various factor solutions are *mathematically* equivalent. However, not all solutions have equal *psychological* utility. Different schools of thought on this issue will be discussed. The popular view is Thurstone's (1947) *simple structure*, in which a general factor is undesirable. Holzinger (Holzinger & Swineford, 1937), on the other hand, emphasized the importance of a general factor in his writings.

Alternate, equivalent solutions can relate to each other through the concept of *rotation*. Principles of rotation are introduced with emphasis given to *orthogonal rotation*. Graphic representations of rotations are quite useful and more so before the invention of computers. Most rotation now is done *analytically*, i.e., to satisfy one of several possible *mathematical* criteria. The logic of the two most popular methods of analytic rotation—*quartimax* and, especially, *varimax*, is covered. This section concludes with a discussion of *oblique* rotations.

VI. THE COMMON FACTOR MODEL—Many investigators prefer the common factor model to the component model. One reason that was noted earlier is that the common factor model contains an explicit concept of measurement error; the component model does not. A second reason is that a given number of common factors will reproduce the original correlation matrix better than the same number of components. From a computational standpoint, the common factor model operates upon a correlation matrix with numbers less than unity in the diagonal (termed *communalities*), whereas the component models operate upon a correlation matrix with unities in the diagonal. Unfortunately, several methods have been suggested to define communalities, but there is no general agreement upon the "best" method. Three general approaches to the communality problem are (1) *formal definition*, e.g., using the reliability coefficient to define the communality; (2) *direct estimation*, which is used by LISREL and related programs; and (3) communalities as a *by-product* of other operations. For example, one method, known as *Alpha*

(Kaiser & Caffey, 1965) determines the most reliable factors in a sense to be discussed in Chapter 12. The communalities emerge once the set of factors is defined.

VII. AN EXAMPLE OF THE COMMON FACTOR MODEL—The simulated example used for the component problem is modified slightly here to demonstrate a common factor solution.

Among the complexities of the common factor model noted in this section is that common factors are not completely specified by the variables, i.e., the multiple correlation between variables and a given factor is usually less than one—a fact that gives rise to what will be termed *the second form of factor indeterminacy*.

VIII. FACTOR SCORES—In most factor analytic work, it is not necessary to compute the scores individual subjects obtained on each factor (factor scores). Hence, the *factor score matrix*, which transforms variables into factors, is of minor interest. However, situations do arise in which factor scores are needed. The component model allows *exact* factor scores to be obtained that completely fit the data in the original sample; the common factor model does not because of the second form of factor indeterminacy. However, this comparative shortcoming of the common factor model is minor because when a component model obtained from one sample is applied to a new sample, its factor scores will not fit the model exactly because of sampling error. In this section, I cover (1) procedures to obtain component scores, (2) procedures to obtain common factor scores, and (3) criteria to evaluate the alternative procedures. It is noted that in a great many applications, *approximating* the factor score weight matrix from *salient variables* (those variables loading most highly upon a factor) will prove satisfactory although their properties should be evaluated.

The more abstract topic of factor analysis will now be considered. The previous two chapters dealt with how one set of variables could be used to predict a criterion. Predictors and criteria are all rather tangible entities and distinct from each other conceptually; you can easily visualize a set of drugs, on the one hand, and an activity measure, on the other hand. Factor analysis is more like regression analysis than it is different—one set of variables is used to predict another set of variables. However, in factor analysis, the “*predictors*” are linear combinations of the “*criteria*” and vice versa. This may sound like it goes around in circles. In order to show you that this is not the case, I will devote some space to defining the various goals of factor analysis. First, though, we will present the basic model common to all forms and applications of factor analysis.

In reading both this chapter and the next one, keep in mind that one use of factor analysis (and, for that matter, any time you derive linear combinations, which includes virtually the entire field of multivariate analysis) is to pool what different measures have in common to obtain a “purer” measure of a construct. In that sense, you are performing what Garner, Hake, and

Eriksen (1957) termed *converging evidence* defining a factor “converge” or merge.

The Basic Factor Analytic Model

The Factor Equation

The basic model for factor analysis is to show you the similarities between the variables in the data matrix. Equation (6-1b) defines factor equation:

$$\begin{aligned} x_{ij} &= f_{ii} b_{ji} + f_{ii} b_{ji} \\ \mathbf{X} &= \mathbf{FB}' + \mathbf{F}_u \mathbf{B}'_u \end{aligned}$$

The Raw-Score Matrix

The symbol x_{ij} in Eq. (6-1a) denotes the raw score for subject i on variable j . The first subscript i denotes the number of variables, and the second subscript j denotes the number of subjects. Usually, N is much greater than K . Factor analysis reverse the role of subject and variable. The X matrix is simply a data matrix as the \mathbf{Z} matrix is identified in the basic relation, Eq. (6-1).

The factor analytic model does not take a particular form. By convention (and this is important), each column of \mathbf{X} is scaled to a mean of zero so that it is a \mathbf{Z} matrix in the sense of being uncorrelated with the mean of zero.

Factor Scores and the Factor Score Matrix

The individual elements identified as *factor scores* for subject i across the various variables identify the various factors. (Factors are like unobservable or *latent* variables, and the factor score matrix of factor scores is denoted as columns. Factor scores are f_{ii} and a standard deviation of 1.0, i.e., $\sqrt{f_{ii}}$.)

Pattern Elements and the Pattern Matrix

The individual elements identified as *pattern elements* for factors are called *pattern elements* for factor analysis.

most reliable factors in a sense to be seen once the set of factors is

MON FACTOR MODEL—The problem is modified slightly here

factor model noted in this section specified by the variables, i.e., the given factor is usually less than termed the second form of factor

ctor analytic work, it is not necessarily obtained on each factor (factor α transforms variables into factors, β arise in which factor scores are γ factor scores to be obtained that δ ; the common factor model does determinacy. However, this com- model is minor because when a ϵ is applied to a new sample, its because of sampling error. In this mponent scores, (2) procedures to teria to evaluate the alternative y applications, approximating the bles (those variables loading most although their properties should

ysis will now be considered. The set of variables could be used to represent all rather tangible entities and can easily visualize a set of drugs, on the other hand. Factor analysis different—one set of variables is used α , in factor analysis, the “predictor” and vice versa. This may sound now you that this is not the case,ous goals of factor analysis. First, mon to all forms and applications

at one, keep in mind that one use time you derive linear combinations of multivariate analysis) is to ion to obtain a “purer” measure orming what Garner, Hake, and

Eriksen (1957) termed *converging operations*. That is, the individual variables defining a factor “converge” or meet together to define a construct.

The Basic Factor Analytic Model

The Factor Equation

The basic model for factor analysis is given by Eq. (6-1a) in a form designed to show you the similarities between it and previous uses of the general linear model. Equation (6-1b) defines factor analysis in a more compact matrix form. Both equations define the *factor equation*:

$$x_{ij} = f_{i1}b_{j1} + f_{i2}b_{j2} + \cdots + f_{ip}b_{jp} \quad [+ f_{iu}b_{ju}] \quad (6-1a)$$

$$\mathbf{X} = \mathbf{FB}' + \mathbf{F}_u\mathbf{B}'_u \quad (6-1b)$$

The Raw-Score Matrix

The symbol x_{ij} in Eq. (6-1a) denotes an observed score (element of \mathbf{X}). Most commonly, the first subscript denotes subjects, varying from 1 to N , and the second denotes the number of variables that have been obtained, varying from 1 to K . Usually, N is much greater than K . However, some applications of factor analysis reverse the role of subjects and variables or even study a single subject for whom multiple measures are obtained on different occasions. Thus, \mathbf{X} is simply a data matrix as the term was introduced in Chapter 3 and is so identified in the basic relation, Eq. (6-1b).

The factor analytic model does not require raw scores to be scaled in any particular form. By convention (and in many computer programs, exclusively) each column of \mathbf{X} is scaled to a mean of .0 and a standard deviation of 1.0, so that it is a \mathbf{Z} matrix in the sense of Chapter 3. Assume this is the case unless told otherwise.

Factor Scores and the Factor Score Matrix

The individual elements identified in Eq. (6-1a) as f_{i1}, f_{i2}, \dots , are the *factor scores* for subject i across the various factors. (Roman numerals traditionally identify the various factors.) Factor scores are analogous to predictor scores in multiple regression. The main difference is that factor scores are defined on unobservable or *latent* variables, which is another name for factors. The matrix of factor scores is denoted \mathbf{F} in Eq. (6-1b) with subjects as rows and factors as columns. Factor scores are conventionally scaled to a mean of .0 and a standard deviation of 1.0, i.e., as z scores.

Pattern Elements and the Pattern Matrix

The individual elements identified in Eq. (6-1a) as b_{j1}, b_{j2}, \dots , for measure j are called *pattern elements* for factors I, II, ... Even though I will not use the

Greek symbol to describe these elements, they are in actuality beta weights. However, instead of relating one observed variable to another observed variable, pattern elements relate factors (latent variables) to observed variables. The collection of pattern elements forms the *pattern matrix* (**B**) whose transpose appears in Eq. (6-1b). The **B** matrix has variables as rows and factors as columns. Its transpose is used in the equation to conform to the rules of matrix algebra. The **B** matrix and matrices closely related to it are the most important products of most factor analyses because they are the primary means of telling one what a set of factors are measuring.

Error Scores and Error Loadings

The final term in Eq. (6-1a) is the product of two quantities that arise from a series of special factors called *unique factors*. There is one unique factor for each variable in the model. The loading of each variable on its "own" unique factor is given by b_{ju} , and all other variables are usually assumed to have loadings of .0 on that factor.

The subject's factor score on each unique factor, f_{iu} , has the same properties as do the factor scores for the other factors. Moreover, scores on a given unique factor are assumed to be uncorrelated with scores on any other factors, unique or not. The matrix of pattern elements, B_u in Eq. (6-1b), is therefore a diagonal matrix.

Brackets are placed around the b_{ju} and f_{iu} terms and their matrix counterparts because their inclusion is optional. Including the terms produces the *classical* or *common factor model*. The nonunique factors are known as *common factors*. Excluding the terms produces the *component model*.

Some people reserve the term "factor" for the common factor model. Others use it to cover both models, as will be done in this book. The unique factors are the factor analytic counterpart of error in the regression model. The common factor model has some definite benefits but also introduces certain complications. Lack of an explicit concept of error in the component model is often regarded as a very serious drawback, but it need not always be so.

In certain recent formulations of the common factor model (e.g., Jöreskog, 1974; Jöreskog & Sörbom, 1986), unique factors can correlate with other unique factors to represent correlated error but cannot correlate with the common factors themselves.

The Covariance Equation

There is a second form of the factor model given by Eq. (6-2); it is derivable from Eq. (6-1b) and is called the *covariance equation*:

$$\mathbf{R} = \mathbf{B}\Phi\mathbf{B}' + [\mathbf{U}] \quad (6-2)$$

The left-hand side consists of a matrix, **R**, whose precise properties depend on the original scaling of **X**. **R** may be a variance-covariance matrix (hence, the

name of the equation), in which the columns of **X** are z scores, **R** contains variables off the diagonal. In a covariance matrix, the diagonal entries contain numbers less than 1.0 called *common factor variance* of each variable. These are treated as systematic. The common variance is omitted because the model treats all variables as if they were unique.

The matrix **B** is the pattern matrix of correlations among the factors and the matrix **U** contains the variance of the unique factors. It is a diagonal matrix and the unique factors are omitted.

The First Form of Factor Index

Neither **B** nor **F** is unique to a given model. There are an infinite number of pairs of such matrices however, will be related to each other described in Chapter 3. The reason is that there is no unique answer to the question of how many factors a product is 8. What are the two numbers? As 1 and 8 or 4 and 2. You cannot enough information to define the product uniquely. Therefore, there are infinitely many solutions. This is known as the indeterminacy of the factor analysis. Alternative solutions will not be discussed.

The "=" sign as used in both covariance equations does not mean "equals" in most applications. In factor analysis results, you would find correlations perfectly, although in a different sense. The similarity of correlations can be used to evaluate different symbols for these two matrices. Different matrices used to a minimum.

² Equation (6-1b) leads fairly simply to the covariance equation. It also holds for the component model, but only for the first term of Eq. (6-1b) and divide by number of observations. The first term ($\mathbf{X}'\mathbf{X}$) contains sums of squared z scores off the diagonal under consideration. Hence, the left-hand side of the equation can be rewritten as $(1/N)(\mathbf{B}\mathbf{F}')(\mathbf{F}\mathbf{B}')$ because of the transposes in reverse order and the rows of \mathbf{F}' are z scores, $(1/N)(\mathbf{F}'\mathbf{F})$ is the raw scores in $\mathbf{R} = (1/N)(\mathbf{X}'\mathbf{X})$.

hey are in actuality beta weights.
d variable to another observed
tent variables) to observed vari-
rms the *pattern matrix* (**B**) whose
x has variables as rows and factors
uation to conform to the rules of
closely related to it are the most
es because they are the primary
e measuring.

f two quantities that arise from a
s. There is one unique factor for
ach variable on its "own" unique
e usually assumed to have loadings

actor, f_{iuj} , has the same properties.
Moreover, scores on a given unique
ores on any other factors, unique
Eq. (6-1b), is therefore a diagonal

, terms and their matrix counter-
cluding the terms produces the
nique factors are known as *com-*
he component model.

the common factor model. Others
in this book. The unique factors
r in the regression model. The
efits but also introduces certain
f error in the component model
, but it need not always be so.
non factor model (e.g., Jöreskog,
actors can correlate with other
r but cannot correlate with the

given by Eq. (6-2); it is derivable
equation:

$$U] \quad (6-2)$$

ose precise properties depend on
e-covariance matrix (hence, the

name of the equation), in which case **C** would replace **R**. Assuming that the columns of **X** are *z* scores, **R** contains the correlations among the observed variables off the diagonal. In a common factor model, the diagonal elements contain numbers less than 1.0 called *communalities*, which represent the *common factor variance* of each variable, i.e., the proportion of variance that is treated as systematic. The communalities in a component model are unities because the model treats *all* variance as if it was systematic.

The matrix **B** is the pattern matrix of Eq. (6-1b). The matrix **Φ** contains correlations among the factors and is called the *factor correlation matrix*. The matrix **U** contains the variance of the unique factors and is called the *uniqueness matrix*. It is a diagonal matrix by virtue of the assumed independence of the unique factors and is omitted in a component model.²

The First Form of Factor Indeterminacy

Neither **B** nor **F** is unique to a given set of data in the sense that an infinite number of pairs of such matrices will reproduce the data *exactly* as well. All, however, will be related to each other by a suitable transformation in the sense described in Chapter 3. The reason is the same as the reason you cannot give a unique answer to the question: "I am thinking of two numbers whose product is 8. What are the two numbers?" You can provide *an* answer, such as 1 and 8 or 4 and 2. You cannot provide *the* answer, since you do not have enough information to define the result uniquely. *Unless constraints are introduced, there are an infinite number of mathematically equivalent ways to factor a matrix. Therefore, there are infinite sets of factors that will reproduce R equally well.* This is known as the first form of factor indeterminacy. However, alternative solutions will not be equally meaningful by psychological criteria.

The "=" sign as used in both of the preceding equations does not literally mean "equals" in most applications. If you were to apply Eq. (6-2) to factor analytic results, you would find that it would not reproduce the individual correlations perfectly, although it would be "close" (typically, in the least squares sense). The similarity of the *reproduced* correlations to the *obtained* correlations can be used to evaluate the fit of the model. Many textbooks use different symbols for these two matrices, but I will try to keep the number of different matrices used to a minimum.

² Equation (6-1b) leads fairly simply to Eq. (6-2). The proof is most easily stated for a component model, but also holds for the common factor model. Premultiply both sides of Eq. (6-1b) and divide by number of subjects, N : $(1/N)\mathbf{X}'\mathbf{X} = (1/N)(\mathbf{F}\mathbf{B}')'(\mathbf{F}\mathbf{B}')$. The term $(\mathbf{X}'\mathbf{X})$ contains sums of squared *z* scores on its diagonal and sums of products of *z* scores off the diagonal under conventional scaling. Dividing by N produces the correlations. Hence, the left-hand side is a correlation matrix. The right-hand side can be rewritten as $(1/N)(\mathbf{F}\mathbf{B}')(\mathbf{F}\mathbf{B}')$ because the transpose of a product $(\mathbf{F}\mathbf{B}')'$ is the product of the transposes in reverse order and because $(\mathbf{B}')'$ is **B**. Since the columns of **F** (hence, the rows of **F'**) are *z* scores, $(1/N)(\mathbf{F}'\mathbf{F})$ is a correlation matrix, but of factor scores, not the raw scores in $\mathbf{R} = (1/N)(\mathbf{X}'\mathbf{X})$.

An Important Special Case

The definition of \mathbf{F} does not require factor scores to be independent of each other (uncorrelated), i.e., for $\Phi = (1/N)(\mathbf{F}'\mathbf{F})$ to be an identity matrix. When factor scores are uncorrelated, the result is termed an *orthogonal solution* and Eq. (6-2) can be reduced to a simple form, often termed the "fundamental equation of factor analysis." Solutions in which factors are not orthogonal and are hence correlated are known as *oblique*. In an oblique solution, Φ is not an identity matrix:

$$\mathbf{R} = \mathbf{B}\mathbf{B}' + [\mathbf{U}] \quad (\text{orthogonal factors}) \quad (6-3)$$

Using different notation, it was noted in Chapter 3 that when a matrix \mathbf{A} had the property that $\mathbf{A}'\mathbf{A} = \mathbf{R}$, \mathbf{A} was the (nonunique) square root of \mathbf{R} . If each element of a matrix of z scores (\mathbf{Z}) is divided by the square root of N , it can easily be shown that the resulting matrix is a square root of \mathbf{R} . Orthogonal factor analysis of a correlation matrix can be thought of as the search for other matrices that are also the square roots of \mathbf{R} . In that same section of Chapter 3, I also noted that if the square root matrix was triangular, the result was called a "Cholesky decomposition." Before computers, the "square root method," which involves a Cholesky decomposition of \mathbf{R} , was a popular method of factoring. The point to keep in mind is that an orthogonal pattern matrix is a "square root" of a raw SP (sum of products), corrected SP, covariance, or correlation matrix.

Up to now, a number of different matrices have been discussed:

- \mathbf{R} = the correlation matrix
- \mathbf{B} = the pattern matrix (beta weights to predict variables from factors)
- \mathbf{F} = the factor score matrix (scores of subjects on factors)
- \mathbf{B}_u = the pattern matrix for unique factors (common factor model)
- \mathbf{F}_u = the factor score matrix for unique factors (common factor model)
- Φ = the factor correlation matrix (an identity matrix, \mathbf{I} , for orthogonal solutions)

There will be a few more matrices to consider, although they will involve quantities derivable from the above. Before presenting any more, however, I will turn to a consideration of common uses of factor analysis.

Common Uses of Factor Analysis

There are several fundamental distinctions made regarding the types of models that are formulated accordingly to Eq. (6-1) or, equivalently, Eq. (6-2). I feel that the distinction between *exploratory* and *confirmatory* factor analysis is the most important one. In the former, factors are defined in such a way as to meet certain mathematical considerations, without regard to any theory. The resulting factors are then named based on the variables that correlate

most strongly with them and the tion. In the latter, factors are c substantive theory. Interest here fit the proposed theory. Most st methods, but they then used the ture of a set of data for which a this approach to be wrong. Recei & Sörbom, 1986) has become a p It is an important step forward, confirmatory method known as as problems because of its simplici

I will be exclusively concerne analysis in this chapter. These because there are certainly many idea of the structure of your da discussion will also allow me to i in confirmatory factor analysis. E use exploratory factor analysis if data, such as one which says wh Under such circumstances, confir

Gorsuch (1983, p. 374)³ descr analysis. Although he notes the f will be modified somewhat.

Orthogonalization

Orthogonalization involves repla set of mutually uncorrelated vai linear combinations of other varia in their own right. The new variat uncorrelated when they are appli tions will generally be low if the composition. The full component replaced by a like number of cor across samples.

³ Among the major reference works, Gorsuch's (1983) to be the best at conc in behavioral research. His work was many works are seminal to those int research. Cattell (1957, 1966) is the be Harman (1976) and Mulaik (1972) ar at showing calculational steps. Most ducted prior to 1950, were to stud attitudes about factoring developed f Thurstone (cf. Thurstone, 1947). The areas, particularly the study of person

scores to be independent of each other. It is often termed an *orthogonal solution* and is often termed the "fundamental solution." In an oblique solution, Φ is not

agonal factors) (6-3)

Chapter 3 that when a matrix A (nonunique) square root of R . If divided by the square root of N , it is a square root of R . Orthogonal thought of as the search for other factors. In that same section of Chapter 3, it was mentioned that if the matrix was triangular, the result for computers, the "square root decomposition of R , was a popular method. This is that an orthogonal pattern (sum of products), corrected SP,

s have been discussed:

dict variables from factors)
cts on factors)
common factor model)
ors (common factor model)
fy matrix, I , for orthogonal solu-

sider, although they will involve presenting any more, however, I feel of factor analysis.

ade regarding the types of models or, equivalently, Eq. (6-2). I feel that confirmatory factor analysis is a process defined in such a way as to be useful, without regard to any theory. It is based on the variables that correlate

most strongly with them and thereby contribute most heavily to their definition. In the latter, factors are defined according to the specification of a substantive theory. Interest here is directed toward seeing how well the data fit the proposed theory. Most students in the past learned only exploratory methods, but they then used these methods to examine the underlying structure of a set of data for which an underlying theory really existed. I believe this approach to be wrong. Recently, Jöreskog's LISREL program (Jöreskog & Sörbom, 1986) has become a popular tool for confirmatory factor analysis. It is an important step forward, although I personally favor a much older confirmatory method known as *oblique multiple groups* for certain types of problems because of its simplicity.

I will be exclusively concerned with traditional methods of exploratory analysis in this chapter. These methods are important in their own right because there are certainly many instances in which you will have little or no idea of the structure of your data before you collect them. Moreover, my discussion will also allow me to introduce some concepts that are also useful in confirmatory factor analysis. However, to perhaps overstate a point: *do not use exploratory factor analysis if you are testing a proposed organization of the data, such as one which says which variables "go together" to form factors. Under such circumstances, confirmatory methods offer a more direct approach.*

Gorsuch (1983, p. 374)³ describes four broad uses of exploratory factor analysis. Although he notes the four specific uses stated below, his discussion will be modified somewhat.

Orthogonalization

Orthogonalization involves replacing a set of correlated variables with a new set of mutually uncorrelated variables called factors. Although factors are linear combinations of other variables, they can also be thought of as variables in their own right. The new variables (factors) will not necessarily be mutually uncorrelated when they are applied to a new sample, but their intercorrelations will generally be low if the original and the new samples are similar in composition. The *full component model*, in which the entire set of variables is replaced by a like number of components, can be used to examine stability across samples.

³ Among the major reference works solely devoted to factor analysis, I consider Gorsuch's (1983) to be the best at concisely describing the overall role of factor analysis in behavioral research. His work was largely stimulated by Raymond Cattell, whose many works are seminal to those interested in factor analysis as a tool in personality research. Cattell (1957, 1966) is the best starting place. Cattell (1978) is another source. Harman (1976) and Mulaik (1972) are also excellent texts with the former quite good at showing calculational steps. Most factor analytic studies, particularly those conducted prior to 1950, were to studies of cognitive abilities (intelligence). Certain attitudes about factoring developed from those studies, particularly from the work of Thurstone (cf. Thurstone, 1947). These attitudes have been carried over into other areas, particularly the study of personality where the attitudes may be less applicable.

Reduction in the Number of Variables

One may choose to *reduce the size of a set of variables*. In general, one needs as many components as there are variables to reproduce all the information contained in the original set of variables. However, the information contained in the original set of variables can often be *approximated* by a much smaller number of components. Sometimes, this approximation is perfectly adequate. When that is the case, one can adopt the *truncated component* model, which uses only the most important components for estimation purposes. Some refer to truncated components as a "cheap" form of common factor analysis. However, the logic of truncated components is not unreasonable. The common factor model attempts to decide what is *shared* among variables versus what is *unique* to individual variables. A truncated component solution attempts to decide what is *important* versus what is *trivial*.

Bernstein, Schoenfeld, and Costello (1982) used truncated components of the MMPI's 14 major scales as predictors in multiple regression. These authors were concerned with predicting various performance measures from police cadets' MMPI profiles. This is an excellent example of data which produce a multicollinear correlation matrix because of the interrelations among the subscales. It was found that five components accounted for most of the predictive variance in performance and only excluded little (but, in one case, some) information important to prediction. Thus, a truncated component model was applied that greatly simplified the task of prediction.

Dimensional Analysis

A third use of exploratory factor analysis is *dimensional analysis*, which allows one to make statements such as " k factors contain *all* the variance that the p variables share with each other or can share with other variables, the remaining variance being error." A common factor model is more appropriate for dimensional analysis because the factor analysis is primarily being used as a theoretical tool to decide what is systematic and what is random. Dimensional analysis can also involve forming *higher-order factors*. If the derived factors are oblique, their correlation matrix, Φ , can be further factored using the same general methods as those used to factor the original correlation matrix, R itself, to obtain "factors of factors." The process produces a *hierarchy* of factors. The process can also be repeated, but the idea of a "hierarchy of a hierarchy of factors" has not proven fruitful thus far, perhaps understandably.

Determination of Factor Scores

Exploratory factor analysis may be used to obtain individual *factor scores* (F). Actually, for all the seemingly mystical aspects of factor analysis, prac-

tically every time you take a test, w Suppose your instructor scores o multiple choice or true-false test. you answer correctly; hence, it is Were all students' scores rescaled called the *first centroid factor*. forming successive equally weight sis.) If, in contrast, you take an e questions, the resulting z scores centroid, since the weighting is ur

The weights that are used to ol responses to individual items, fori *weight matrix*, which will be den weight has the same physical layo matrix, S , defined subsequently (v content is quite different from bo is used to estimate *factors* as lit *columnwise (factorwise)* and the w contrast, B is read *rowwise (variabi factors*. All three matrices are rela on whether one has chosen a com less of the model chosen, the elem as beta (regression) weights. How the criteria are factors, whereas in

Gorsuch (1983) also notes how be used for a fifth purpose, that c that one factor is age. (It is quite of a single observed variable.) Al definition, independent of the lir factoring is logically equivalent t multiple regression, a device also noted. (Similarly, oblique solutio multiple regression.)

Clearly, factor analysis can b Because the goals of data analysi methods are appropriate for differe tions: (1) there is no such thing as need to get a feel for the variety need to get a feel for what happens the defaults built into whichever pa of factor analysis will be discussed routines in the major packages per This is not necessary! You may u factors, a procedure known as *ad*

of variables. In general, one needs to reproduce all the information; however, the information contained *approximated* by a much smaller approximation is perfectly adequate. *Truncated component* model, which for estimation purposes. Some refer of common factor analysis. However, unreasonable. The common *red* among variables versus what 1 component solution attempts to *al.*

2) used truncated components of rs in multiple regression. These various performance measures from excellent example of data which fix because of the interrelations components accounted for most and only excluded little (but, in one exception. Thus, a truncated componented the task of prediction.

Dimensional analysis, which allows obtain *all* the variance that the *p* with other variables, the remainder model is more appropriate for lysis is primarily being used as a and what is random. Dimensional *ter factors*. If the derived factors be further factored using the same e original correlation matrix, **R** process produces a *hierarchy* of but the idea of a "hierarchy of a thus far, perhaps understandably.

to obtain individual *factor scores* aspects of factor analysis, prac-

tically every time you take a test, what you receive are, in essence, factor scores. Suppose your instructor scores one point for each item you get correct on a multiple choice or true-false test. Your score is the sum of the number of items you answer correctly; hence, it is a linear combination with equal weighting. Were all students' scores rescaled in the form of *z* scores, the result is what is called the *first centroid factor*. (Centroid factor analysis, which involves forming successive equally weighted sums, was an early form of factor analysis.) If, in contrast, you take an essay test with different weights for different questions, the resulting *z* scores would no longer be derived from the first centroid, since the weighting is unequal, but they would still be factor scores.

The weights that are used to obtain factor scores from raw scores, such as responses to individual items, form yet another matrix called the *factor score weight matrix*, which will be denoted by the symbol "**W**." The factor score weight has the same physical layout as the pattern matrix, **B**, and the structure matrix, **S**, defined subsequently (variables in rows; factors in columns), but its content is quite different from both matrices. The factor score weight matrix is used to estimate *factors* as linear combinations of *variables*. **W** is read *columnwise* (*factorwise*) and the weights are used to compute factor scores. In contrast, **B** is read *rowwise* (*variablewise*) and is used to estimate *variables* from *factors*. All three matrices are related, but in a complex manner that depends on whether one has chosen a component or a common factor model. Regardless of the model chosen, the elements of both **W** and **B** should be thought of as beta (regression) weights. However, in **W**, the predictors are variables and the criteria are factors, whereas in **B**, the converse is true.

Gorsuch (1983) also notes how the properties of orthogonal solutions can be used for a fifth purpose, that of *statistical control*. For example, suppose that one factor is age. (It is quite possible for a factor to be defined in terms of a single observed variable.) All factors orthogonal to that factor are, by definition, independent of the linear effects of age. In general, orthogonal factoring is logically equivalent to a hierarchical or successive solution in multiple regression, a device also used for statistical control, as previously noted. (Similarly, oblique solutions may be viewed as a type of simultaneous multiple regression.)

Clearly, factor analysis can be employed for many different purposes. Because the goals of data analysis are quite varied, different factor analytic methods are appropriate for different situations. This fact has three implications: (1) there is no such thing as "the" factor analysis of a data set, (2) you need to get a feel for the variety of strategies that are possible, and (3) you need to get a feel for what happens when you instruct the computer to override the defaults built into whichever package you used. Several specific techniques of factor analysis will be discussed here and in the Chapter 7. Factor analysis routines in the major packages perform the same type of analysis on all factors. This is not necessary! You may use different factoring methods for different factors, a procedure known as *ad lib factoring*.

No other factor can account for more variance than the first PC in the sample. Also note that the magnitude of the first eigenvalue (4.654) divided by N (6) yields the same estimate of variance accounted for. *The eigenvalues of a PC solution describe the total variance accounted for; when divided by the number of variables, they define the proportion of total variance accounted for.* Doing the same for PC_{II} leads to a value of .224.

Nunnally (1978) summarizes some important properties of a PC solution:

1. *Each factor maximizes the variance explained.* This has three important implications: (a) The sum of squared loadings going down a column is as large as possible, (b) the residual matrix for a given number of factors (explained subsequently) is as small as possible, and (c) the first PC explains more variance in \mathbf{X} than any other linear combination of the original variables.
2. *Any group of successive PCs explains at least as much variance as any like number of components obtained by any other method.*
3. *The sum of the squared loadings for a PC equals the eigenvalue of that PC.*
4. *All eigenvalues of a PC solution will be zero or greater when Pearson product-moment correlations have been used to define \mathbf{R} .*
5. *The number of nonzero eigenvalues* (two in the present case) *equals the number of factors underlying the data.* In most real applications, all eigenvalues will be greater than zero, however, unless a redundancy has been built into the data by including a variable that is linearly dependent on some other variables.
6. *The sum of the eigenvalues equals the sum of the diagonal elements of \mathbf{R}* (its trace) which, in turn, equals the number of variables.
7. *PCs are mutually orthogonal.*
8. *The factor scores produced by a PC solution are also uncorrelated.* This lack of correlation among factor scores holds true in all orthogonal component solutions, not just a PC solution.
9. *The sum of squared cross products of the factor pattern (the off-diagonal elements of $\mathbf{B}'\mathbf{B}$) are .0.*

When \mathbf{R} has what is known as *positive manifold* in the sense that all correlations among the original set of variables are positive, as is true here, the solution has another property. Loadings on the first factor will all be positive. On all remaining factors, half of the loadings will be positive and half will be negative, creating what is called a *bipolar* factor (see subsequent *Types of Factors* section.) Furthermore, because the sum of cross products on any pair of factors is .0, half of the loadings that were positive on any one PC will be negative on any other PC and vice versa (except for the first PC).

A Note on the Orthogonality of PCs

Although a formal proof of the mutual orthogonality of eigenvectors is complex, a graphical demonstration is useful. Reexamine Fig. 3-1, which

contains the envelope of a scatter in z-score form for convenience variables is fairly strong; hence, the right and bottom left quadrants are quadrants. Note, though, that the top-left quadrant (closest to the first PC) and the minor axis (second PC) divide the observation size. In other words, the observation used as coordinate axes instead of

How an Eigenanalysis Is Performed

An example of an eigenanalysis on 3. The eigenvalues and eigenvectors generally possible with larger matrices.

With large matrices, the eigenanalysis you will never have to do one by hand to perform the analysis.

1. In the simplest but not the most common case, choose an arbitrarily chosen nonzero vector v_{in} with many elements as there are variables. Normalize it, i.e., $v_{in}'v_{in} = 1.0$.
2. Form the product $v_{in}'\mathbf{R}$. Determine the result λ . Call the normalized v_{out} . If the two are equal within tolerance, they are sufficiently similar in magnitude.
3. Replace v_{in} with v_{out} and repeat.
4. The vector v_{in} or, equivalently, v_{out} , is an eigenvalue.
5. Form a residual variance-covariance matrix product of the eigenvalue. This process *partials (adjusts)* \mathbf{R} to produce \mathbf{PC}_1 . Its rank will be one less than the covariance matrix, it may be correlated, dividing each off-diagonal element by the diagonal entries.

After the first eigenvalue, its associated eigenvectors are obtained, the process is repeated. *Successive partialling is what makes defining a factor, partialing out that factor, and defining a new factor, etc., will produce orthogonal factors.*

It is useful to look at the structure of a matrix whose columns are the no diagonal matrix of eigenvalues. Equa-

e than the first PC in the sample. eigenvalue (4.654) divided by N (6) factored. The eigenvalues of a PC for; when divided by the number of variance accounted for. Doing

ant properties of a PC solution: xined. This has three important things going down a column is as for a given number of factors possible, and (c) the first PC explains a combination of the original

east as much variance as any like er method.

equals the eigenvalue of that PC. zero or greater when Pearson d to define \mathbf{R} .

in the present case) equals the most real applications, all eigenvalues, unless a redundancy has been found that is linearly dependent on some

of the diagonal elements of \mathbf{R} (its variables.

are also uncorrelated. This lack true in all orthogonal component

factor pattern (the off-diagonal

manifold in the sense that all loadings are positive, as is true here, is on the first factor will all be loadings will be positive and half of the factor (see subsequent Types the sum of cross products on any one PC will be positive on any one PC will except for the first PC).

rthogonality of eigenvectors is ful. Reexamine Fig. 3-1, which

contains the envelope of a scatterplot for two variables, z_x and z_y , which are in z-score form for convenience (see p. 85). The relation between the two variables is fairly strong; hence, observations are concentrated in the upper right and bottom left quadrants relative to the upper left and bottom right quadrants. Note, though, that the major axis of the ellipse (which corresponds to the first PC) and the minor axis of the ellipse (which corresponds to the second PC) divide the observations in the scatterplot into four areas of equal size. In other words, the observations are uncorrelated when the two PCs are used as coordinate axes instead of the original variables, z_x and z_y .

How an Eigenanalysis Is Performed

An example of an eigenanalysis on a matrix of rank 2 was presented in Chapter 3. The eigenvalues and eigenvectors could be obtained directly, but this is not generally possible with larger matrices.

With large matrices, the eigenanalysis requires iterative methods. Although you will never have to do one by hand, it is instructive to look at one way to perform the analysis.

1. In the simplest but not the most efficient method, one begins with an arbitrarily chosen nonzero vector, which will be called \mathbf{v}_{in} . It will have as many elements as there are variables to be factored. Assume that it is normalized, i.e., $\mathbf{v}'_{in}\mathbf{v}_{in} = 1.0$.
2. Form the product $\mathbf{v}_{in}\mathbf{R}$. Determine the ratio of the length of \mathbf{v}_{in} to $\mathbf{v}_{in}\mathbf{R}$ and call the result λ . Call the normalized vector product \mathbf{v}_{out} . Compare \mathbf{v}_{in} and \mathbf{v}_{out} . If the two are equal within tolerance, i.e., if the corresponding elements are sufficiently similar in magnitude, go to step (4).
3. Replace \mathbf{v}_{in} with \mathbf{v}_{out} and repeat step (2).
4. The vector \mathbf{v}_{in} or, equivalently, \mathbf{v}_{out} , is the eigenvector and λ is its associated eigenvalue.
5. Form a residual variance-covariance matrix by subtracting $\lambda\mathbf{v}_{in}\mathbf{v}'_{in}$ (the outer or matrix product of the eigenvector times the scalar eigenvalue) from \mathbf{R} . This process partials (adjusts) \mathbf{R} for the linear combination represented by PC_1 . Its rank will be one less than the rank of \mathbf{R} . As with any variance-covariance matrix, it may be converted to a residual correlation matrix by dividing each off-diagonal element by the square root of its corresponding diagonal entries.

After the first eigenvalue, its associated eigenvector, and the residual matrix are obtained, the process is repeated to obtain successive factors. The fact of successive partialling is what makes PC factors orthogonal. Any strategy that defines a factor, partials out that factor, forms a new factor, partials out that new factor, etc., will produce orthogonal factors.

It is useful to look at the structure of a PC solution symbolically. Let \mathbf{V} be a matrix whose columns are the normalized eigenvectors of \mathbf{R} , and let Λ be a diagonal matrix of eigenvalues. Equations (6-5) describe the relations:

$$\begin{aligned}
 R &= \mathbf{V}' \Lambda \mathbf{V} \\
 &= (\mathbf{V} \Lambda^{1/2})' (\mathbf{V} \Lambda^{1/2}) \quad (6-5)
 \end{aligned}$$

In other words, any correlation matrix can be expressed as the product of eigenvalues and associated eigenvectors. The eigenvalues, in turn, describe the variance explained by each factor.

Determining How Many Components to Retain

The next step in the analysis is to decide how many factors are needed in the rotated solution. This is true whether one remains within the component framework or adopts a common factor model. Several rules have been proposed:

1. *Choose only those eigenvectors whose associated eigenvalues are 1.0 or greater.* The Kaiser-Guttman (Guttman, 1954; Kaiser, 1960, 1970) “ $\lambda > 1$ ” criterion is historically the most widely used rule to determine the number of factors to keep. It says, in essence, “deem a component to be important if, and only if, it accounts for at least as much variance as an individual variable does.” Following this rule, the resulting number of factors does not depend on the number of subjects but it is influenced by the number of variables. For example, when you factor six variables, as in the above example, a factor must account for at least one-sixth the total variance to be “important.” (Note, the term “significant” is not being used because of its inferential connotations, which do not apply here.) If, however, you were factoring 100 variables, a factor would only have to account for 1% of the total variance to be considered “important.” Note that considerable “juggling” of the constants of Eq. (6-4) had to be done in order to obtain two components with eigenvalues greater than 1.0 even though the data actually did consist of two underlying factors.
2. *Infer the number of factors from the relations among successive eigenvalues.* This inference is usually made graphically by presenting eigenvalues along the Y axis and their serial positions along the X axis. It is known as a *scree plot*, after the geological term for the rubble at the bottom of a cliff (Cattell, 1966). The goal is to separate the overall curve into two functions with the early eigenvalues representing “important” factors and the later ones “unimportant” factors.
3. *Employ a significance test.* Bartlett (1950) has developed a chi-square tests to determine (a) if a correlation matrix with unity diagonals differs significantly from an identity matrix, and (b) if a residual matrix differs significantly from a null matrix following extraction of one or more PCs. Their respective equations are

$$\chi^2 = -[N - 1 - (2k + 5)/5] \ln |R| \quad (6-6)$$

$$df = k(k - 1)/2$$

$$\chi^2 = -[N - 1 - (2k + 5)/6 - 2p/3] \ln(c), \quad (6-7)$$

where

$$c = \frac{1}{\prod \lambda_i} \quad (k - p)$$

and

$$df = (k - p - N)$$

N is the number of subjects; k is the number of variables; p is the number of eigenvalues; and $\sum \lambda_i$ is the sum of the eigenvalues. (c is the ratio of the number of subjects and number of variables, a matrix containing trivial information, resulting in the extracted matrix, resulting in the extracted matrix.)

4. *Examine the residual correlations.* If these residuals are all small (less than about .10) and there is no apparent pattern, the *residual matrix* (its numbers are misleading in that they will also be small.)
5. *Use the variance explained as a percentage of the total variance,* to determine the number of factors.

There is nothing magical about these criteria. There are certain reasons to consider a particular number of factors, especially when you are comparing them to results from prior study. The number of factors is often determined by the subsequent step of rotation to a simple structure, which is decided upon three factors, perhaps. If your fourth eigenvalue is 1.000, you may conclude that your factor is important. You obtained the same measure with the same measures and with the same error (or perhaps even rounding errors in different packages). Look at solutions that you gain and what you lose in terms of bias (I have), consider a confirmatory factor analysis (see Chapter 7); if you have a theory which is well organized, test the factor structure against it.

Other Initial Component Selection Criteria
Although PCs are very important, I have already noted two others: eigenvalues and eigenanalyses. They are based on successive equally weighted eigenvectors (solid staircase,⁵ diagonal, pivoted).

⁵ So named because its pattern matrix has a solid diagonal equal to 1.00 and zeros elsewhere.

$$\lambda^{1/2} \quad (6-5)$$

can be expressed as the product of the eigenvalues, in turn, describe the

to Retain

how many factors are needed in the model remains within the component model. Several rules have been

associated eigenvalues are 1.0 or .1954; Kaiser, 1960, 1970) "λ > 1" used rule to determine the number of components. If a component to be important has as much variance as an individual variable, then the resulting number of factors does not depend on the number of variables. It is influenced by the number of variables, as in the above rule, one-sixth the total variance to account" is not being used because of (apply here.) If, however, you were only have to account for 1% of variance." Note that considerable work has to be done in order to obtain a factor greater than 1.0 even though the data is not orthogonal.

Scree plots among successive eigenvalues. This can be done by presenting eigenvalues along the X axis. It is known as a *scree plot*. It starts at the bottom of a cliff (Cattell, 1966) and then curves into two functions with "important" factors and the later ones.

has developed a chi-square test which compares the matrix with unity diagonals with a residual matrix. The test statistic is the difference between the number of significant eigenvalues and the number of PCs. Their

$$5)/5] \ln |R| \quad (6-6)$$

$$5)/6 - 2p/3] \ln(c), \quad (6-7)$$

where

$$c = \frac{|\mathbf{R}|}{\prod \lambda_i [(k - \sum \lambda_i)/(k - p)]^{k-p}}$$

and

$$df = (k - p - 1)(k - p - 2)/2$$

N is the number of subjects; k is the number of variables, \mathbf{R} is the correlation matrix; f is the number of factors; $\prod \lambda_i$ is the product of the eigenvalues, and $\sum \lambda_i$ is the sum of the eigenvalues. These formulas correct for both number of subjects and number of components. However, with large samples, a matrix containing trivial residual variance can still differ from a null matrix, resulting in the extraction of trivial factors.

4. *Examine the residual correlations.* Look at the residual correlation matrix. If these residuals are all small (as a rule, less than .3 in absolute value) and there is no apparent pattern, then stop. (Do not use the residual covariance matrix; its numbers are misleadingly small because the residual variances will also be small.)
5. *Use the variance explained as a criterion.* Choose some value, such as 90% of the total variance, to determine how many factors are in the data.

There is nothing magical about any of the preceding criteria. If you have certain reasons to consider a particular number of factors, use that number, especially when you are comparing your results to those obtained from a prior study. The number of factors extracted has a profound effect upon the subsequent step of rotation to a final solution. A prior study may have decided upon three factors, perhaps because the fourth eigenvalue was .9999. If your fourth eigenvalue is 1.00001 and you decide to keep four factors, you may conclude that your factor structure is different from those previously obtained with the same measures when the difference might just be sampling error (or perhaps even rounding error by the different algorithms used by various packages). Look at solutions with more and fewer factors to see what you gain and what you lose in the process. Better yet (and reflecting a real bias I have), consider a confirmatory factoring procedure as discussed in Chapter 7; if you have a theory which specifies how these variables should be organized, test the factor structure implied by that theory directly.

Other Initial Component Solutions

Although PCs are very important, they are not the only type of components. I have already noted two others that were quite popular in the days before computers made eigenanalysis rather simple. One is the *centroid* method, based on successive equally weighted sums, and the other is the *square root (solid staircase,⁵ diagonal, pivotal condensation)* method, which is simply a

⁵So named because its pattern matrix looks like a "staircase," with all entries above the diagonal equal to .0.

Cholesky decomposition of \mathbf{R} . The variances accounted for by successive centroid factors are often very similar to the variances accounted for by successive PCs, because there usually is little difference between the equal weighting of the centroid method and the differential (optimal) weighting of the PC solution (assuming variables are properly signed), another illustration that, often, “it don’t make no nevermind.” The square root method is also useful as a technique to obtain partial correlations (Nunnally, 1978).

Factor Definition and Rotation

Despite their many useful mathematical properties, PCs and related direct solutions usually do not provide the most useful way to describe data. The original solution is usually *rotated* to make it more interpretable.⁶

A suitable rotation can accomplish two objectives. One is that the resulting factor will fall “closer” to relevant variables in a geometric sense to be illustrated below. This allows the investigator to separate those variables that “belong” to the factor from those that do not. In contrast, correlations between individual variables and a given factor tend to be relatively homogeneous (except possibly for sign), so that the factor is not clearly defined by any particular group of variables. In addition, successive PCs explain maximal amounts of variance that remains, by definition. Consequently, they are of unequal importance. Rotated factors can be more nearly equal in the variance they explain and are thus of more nearly equal importance.

In order to appreciate the process of rotation more fully, familiarization with distinctions among various types of factors is useful.

1. A *general* factor is one on which all variables included in the data set have high loadings.
2. A *group* factor is one on which at least two but less than all variables have high loadings.
3. A *singlet* factor is one on which one and only variable has a high loading. It is somewhat like a *unique* factor in common factor analysis. However, a singlet factor may arise in a component solution and loadings of other variables are not constrained to be *exactly* zero, whereas unique factors are by-product of the common factor model and loadings of other variables are constrained to be .0.
4. A *unipolar* factor is a group or general factor on which all variables load in

⁶The MMPI is somewhat of an exception to be discussed in the Chapter 7. The first PC of the MMPI's 14 scales is very closely related to what is called “profile elevation” by users of that test—the tendency to endorse items that psychiatrically impaired individuals also endorse. Likewise, the second PC is closely related to the test taking attitudes of the respondent—whether they are trying to project a favorable or unfavorable self-image, and the third PC is closely related to what may be called optimism versus pessimism.

the *same* direction. Usually, if negative, all signs may be reversed. The first PC in Table 6-2 is a unipolar factor.

5. A *bipolar* factor is a factor on which all variables have positive loading and at least one has negative loading. The second PC in Table 6-2 is a bipolar factor.

“Loadings” of variables may be absolute (beta weights) or structure elements (eigenvalues). The structure elements are the same in an orthogonal solution as in a nonorthogonal solution, but the interpretation, regardless of the type of solution, is that it is less affected by sample size. The structure element is the same as was discussed in regression analysis. A point in favor of orthogonal factor analysis is that it preserves the relative weighting of each factor in predicting the variables.

A structure weight of .3 in absolute value indicates that 9% of the variance between variable and factor is due to define variables that are important. The choice of .3 as a cutoff is justified by the fact that it is a conservative decision to be based on the proportion of variance explained rather than use a higher cutoff when the variance explained is greater than .3.

Factor Definition

For all the mathematical elegance of factor analysis, the empirical oriented research has been limited by the lack of logical interpretability. As a matter of fact, exploratory factor analysis is to some extent a search for the best solution, given the considerations involved in defining factors. It is not a search for personal philosophy, preference for one factor solution over another, or sets of factors for the same set of variables. The choice of different factor solutions for the same set of variables can be factored to preserve a maximum of variance among several group factors. The choice of different factor solutions for the same set of variables is determined by the inherent ambiguity of the definitions of factors, whether to allow factors to be correlated or not.

Whether or not to retain a general factor was a question that Spearman (1904) reflected the British philosophy of factor analysis. He reasoned that the general factor reflects the commonalities among variables, intercorrelations among variables, and the particular structure—each test shot reflecting a general factor (intelligence or “g”) plus random error. Spearman's model is that if all systematic variance is removed, the remaining variance is due to random error.

es accounted for by successive he variances accounted for by le difference between the equal fferential (optimal) weighting of erly signed), another illustration The square root method is also tions (Nunnally, 1978).

roperties, PCs and related direct seful way to describe data. The more interpretable.⁶ ectives. One is that the resulting n a geometric sense to be illus- o separate those variables that n contrast, correlations between to be relatively homogeneous is not clearly defined by any ccessive PCs explain maximal tion. Consequently, they are of ore nearly equal in the variance al importance.

tion more fully, familiarization ors is useful.

es included in the data set have

but less than all variables have

nly variable has a high loading. non factor analysis. However, a solution and loadings of other zero, whereas unique factors are d loadings of other variables are

or on which all variables load in

iscussed in the Chapter 7. The first to what is called “profile elevation” tems that psychiatrically impaired is closely related to the test taking g to project a favorable or unfavor- d to what may be called optimism

the *same* direction. Usually, the loadings are all positive. (If all loadings are negative, all signs may be reversed, a process known as *reflection*.) The first PC in Table 6-2 is a unipolar factor.

5. A *bipolar* factor is a factor on which at least one variable has a substantial positive loading and at least one other variable has a substantial negative loading. The second PC in Table 6-2 is a bipolar factor.

“Loadings” of variables may be defined in terms of either pattern elements (beta weights) or structure elements (correlations), which will be numerically the same in an orthogonal solution. One reason to use the structure matrix for interpretation, regardless of whether the solution is orthogonal or oblique, is that it is less affected by sampling error than the pattern matrix. The issue is the same as was discussed in Chapters 4 and 5 relative to the instability of regression weights. A point in favor of using the pattern matrix is that the weighting of each factor in predicting the variable adjusts for the remaining factors.

A structure weight of .3 in absolute magnitude indicates a 10% overlap in variance between variable and factor. This .3 value is a commonly used cutoff to define variables that are important to the definition of a factor (*salients*). The choice of .3 as a cutoff is just a rule of thumb. It may also be used if a decision is to be based on the pattern matrix but it is a common practice to use a higher cutoff when the variance accounted for by that factor is high.

Factor Definition

For all the mathematical elegance of factor analysis, it is of little use to the empirically oriented researcher unless the emerging factors are psychologically interpretable. As a matter of fact, the most difficult problem in exploratory factor analysis is to define the resulting factors. There are several considerations involved in defining factors, some of which have much to do with personal philosophy, preference, and context in the sense that different factor solutions for the same set of data can each be meaningful. Many data sets can be factored to preserve a general factor or to distribute general factor variance among several group factors, simply by choice of different rotation methods. The choice of different rotation methods arises because of the inherent ambiguity of the definition of a factor structure. A related issue is whether to allow factors to be correlated or require them to be orthogonal.

Whether or not to retain a general factor is a historical issue. Spearman (1904) reflected the British philosophical tradition by believing in the unity of mental processes. He reasoned that if a single factor did underlie cognitive ability, intercorrelations among relevant ability tests should possess a particular structure—each test should measure one and only one trait (general intelligence or “*g*”) plus random error. One important implication of Spearman’s model is that *if all systematic variance is contained in a general factor*,

then a matrix of intercorrelations will be of rank one once measurement error is removed.

Thurstone's (cf. Thurstone, 1947) work, among others, makes it clear that more than one factor is needed to explain variation in cognitive ability. Once a general factor is removed from the data, group factors remain, reflecting systematic relationships among subsets of variables. However, the consequences of the need for group factors raised various issues. One issue was whether or not group factors should be orthogonal to each other. A second issue reflected a dispute between Karl Holzinger and L.L. Thurstone about the role of "g".

Holzinger and Swineford (1937) attempted to work within the Spearman tradition. He preferred solutions in which a "g" factor plus some group factors resulted. His approach to factoring provides what is called a *bifactor* pattern. On the other hand, Thurstone (1947) represented a contrasting philosophical tradition, which views the mind in terms of multiple "faculties" or loosely related cognitive abilities. This view led him to rotate in such a way as to minimize "g".

Three things are important to note about the preceding brief discussion of a very complex issue. One is that the Thurstonian tradition has dominated American psychology. A second is that a great deal of work with factor analysis has nothing to do with measures of cognitive ability. Thurstone's view may not be applicable in those areas. Therefore, the blind acceptance of default rotations may inadvertently bias individuals toward the Thurstone tradition. The third is that rather than being an argument against factor analysis, which is one point of view over the Spearman–Holzinger versus Thurstone dispute, one could argue that factor analysis demonstrates how different perspectives can each be valuable and that certain issues are not resolvable empirically.

Simple Structure

The dominant approach to exploratory factoring is the goal of achieving what Thurstone (1947) termed *simple structure*. Simple structure implies five general properties:

1. Each variable should have a loading that is nearly zero on at least one factor.
2. Each factor should contain more than one variable whose loadings are nearly .0.
3. Given any pair of factors, *A* and *B*, some variables should load on *A* but not on *B* and others should load on *B* but not on *A*.
4. For any pair of factors, *A* and *B*, many variables should not load on either *A* or *B*.
5. For any pair of factors, *A* and *B*, only a few variables should load on both *A* and *B*.

Normally, simple structure req
Many situations dictate that da
factors, making simple structure
dictate the retention of a general j

PC versus Simple Structure

Clearly, the PC solution for ou criteria for simple structure. Each of Thurstone's conditions for simp of many real-world situations, esp with positive manifold.

In any situation where *R* has p sures what all the tests have in cc maximizes the sum of squared fac tion of "g." The definition of tests however.

The PC_{II} is a bipolar factor, a positive manifold, which contrasts have in common versus what the L

One argument for simple struct measures and what that factor c cognitive abilities are not easiest t positive factor scores on PC_{II} m quantitative, and negative factor sc tive than verbal. Hence, PC_{II} does It describes the *difference* between however, such as personality assingful and unitary traits like intro

Another problem is that, by defi amounts of residual variance. Cor become progressively smaller (co PC_{II}), and it is progressively more sive factors, whether or not there quite possible for a component to tance" and be a singlet factor. Wh finding what specific property of a since it probably will also load on meet statistical criteria for "impot

The factor structure of the hypo by noting that it contains a verbal ; group factor. This interpretation perhaps test understood from a ge

rank one once measurement error

among others, makes it clear that variation in cognitive ability. Once , group factors remain, reflecting f variables. However, the conceded various issues. One issue was orthogonal to each other. A second Zinger and L.L. Thurstone about

ed to work within the Spearman "g" factor plus some group factors s what is called a *bifactor* pattern. ented a contrasting philosophical of multiple "faculties" or loosely m to rotate in such as way as to

the preceding brief discussion of stonian tradition has dominated great deal of work with factor cognitive ability. Thurstone's view ore, the blind acceptance of default s toward the Thurstone tradition. ent against factor analysis, which Izinger versus Thurstone dispute, strates how different perspectives are not resolvable empirically.

oring is the goal of achieving what nple structure implies five general

at is nearly zero on at least one one variable whose loadings are e variables should load on *A* but t not on *A*.

variables should not load on either ew variables should load on both

Normally, simple structure requires factors to be correlated.

Many situations dictate that data be described in terms of a series of group factors, making simple structure important. However, many other situations dictate the retention of a general factor.

PC versus Simple Structure

Clearly, the PC solution for our six hypothetical tests does not meet the criteria for simple structure. Each variable loads on both PCs; hence, only one of Thurstone's conditions for simple structure (5) holds. The example is typical of many real-world situations, especially those leading to correlation matrices with positive manifold.

In any situation where *R* has positive manifold like the present, PC_1 measures what all the tests have in common. Specifically, and by definition, PC_1 maximizes the sum of squared factor loadings. Hence, it provides one definition of "g." The definition of tests does depend on which tests were sampled, however.

The PC_{II} is a bipolar factor, as it will always be the case when there is positive manifold, which contrasts what the Tinkers, Evers, and Chance tests have in common versus what the Larry, Curly, and Moe tests have in common.

One argument for simple structure is that it is useful to know what a factor measures and what that factor does *not* measure. Also, bipolar factors of cognitive abilities are not easiest things to conceptualize. In the present case, positive factor scores on PC_{II} mean that the person is more verbal than quantitative, and negative factor scores mean that the person is more quantitative than verbal. Hence, PC_{II} does not describe a unitary type of intelligence. It describes the *difference* between two kinds of intelligence. In some settings, however, such as personality assessment, bipolar factors can describe meaningful and unitary traits like introversion-extroversion.

Another problem is that, by definition, successive PCs account for maximal amounts of residual variance. Consequently, the average factor loadings will become progressively smaller (compare the individual loadings of PC_1 and PC_{II}), and it is progressively more difficult to find salient variables on successive factors, whether or not there is positive manifold. Mathematically, it is quite possible for a component to meet a conventional criterion for "importance" and be a singlet factor. When that happens, one will have difficulty in finding what specific property of a given variable made that factor important, since it probably will also load on another factor. Even worse, the factor may meet statistical criteria for "importance" but contain only trivial loadings.

The factor structure of the hypothetical test data may be simply described by noting that it contains a verbal ability group factor and a numerical ability group factor. This interpretation simplifies discussion of the factors, but is perhaps best understood from a geometric standpoint.

Graphic Representation

Figure 6-1a contains the PC solution. All I have done is to use the loadings on PC_1 as the abscissa (X coordinate) and the loadings on PC_{II} as the ordinate (Y coordinate). Were the solution to demand three or more factors, it would become difficult to visualize the results in a single graph. The standard practice is to look at a series of graphs each containing a pair of factors.

Even though Fig. 6-1 contains no new information, it makes the "groupings" of X_1, X_2 , and X_3 and of X_4, X_5 , and X_6 more apparent to the eye. The graph also illustrates why the utility of an unrotated PC solution is limited. Note how the X and Y axes "float free" of the six variables in that none of the points are located near either axis. *A factor is most easily interpreted when points denoting variables fall near the axis that represents the factor.*

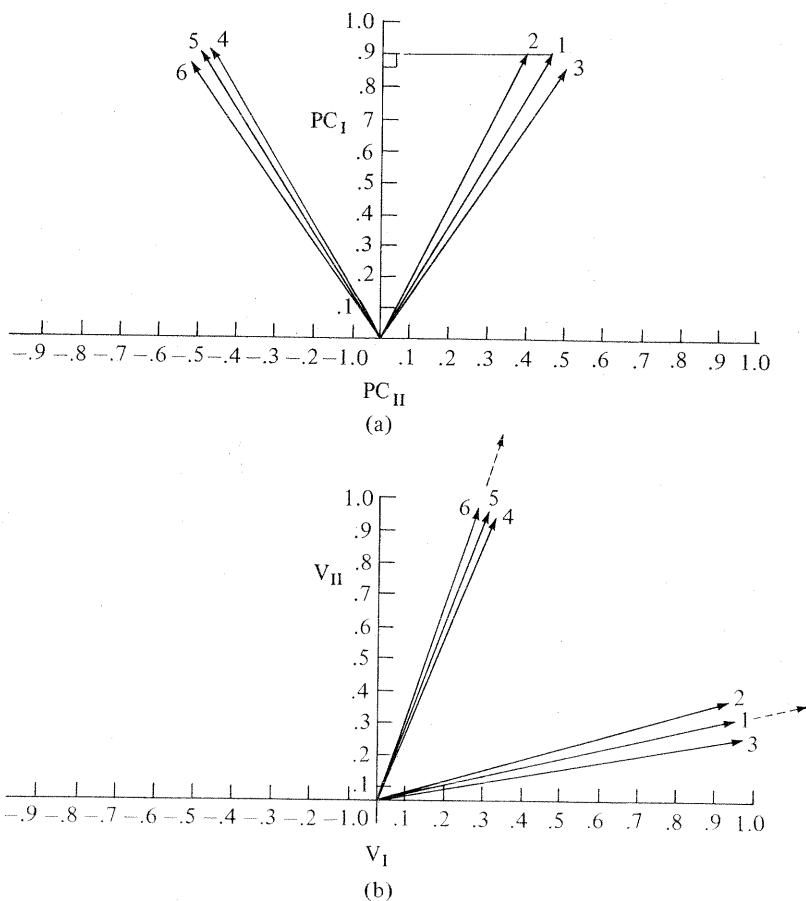


FIGURE 6-1. Factor structures corresponding to Table 6-2: (a) Principal component solution; (b) varimax rotation.

Figure 6-1b contains the poin These points were determined b I will discuss the varimax rotati ever, I could have obtained sim which I called V_1 , roughly 45 c tionally, the angle of rotation, sy from the original axis to the ne order to maintain orthogonality, the original Y axis (PC_{II}) by 45 panels should convince you that only the axes that provide a fra

Clearly, the points fall closer to axes (the PCs). Tests X_1, X_2 , ar tests X_4, X_5 , and X_6 fall very clc a relatively pure measure of nun relatively pure measure of verba even better solution is represen through the clumps of points. T Experienced factor analysts mig solution in this case. The proper later in this chapter.

Both graphs contain other im to each of the points. Note tha segments) are all of unit length. that treats all of the variance cc employed standardized variable the length of the vectors would ha of each variable's variance would

The more general outcome w each vector with respect to a give of the two squared factor loadin theorem, Eq. (4-3). The total len taking all factors into account, vector's length when factors are

Now, note the line drawn fro axis. The point of intersectio w connecting point and projection is of unit length, as it is here, the of the angle between that vector two angles is equivalent to the co (cf. Chapter 3), the projection rep Similarly, the projection of the pc of X_1 on PC_{II} .

As noted in Chapter 3, a rig mathematical by an *orthonorma*

have done is to use the loadings as loadings on PC_{II} as the ordinate and three or more factors, it would angle graph. The standard practice among a pair of factors.

information, it makes the "grouping" of X_6 more apparent to the eye. The unrotated PC solution is limited. The six variables in that none of the X_i is most easily interpreted when i represents the factor.

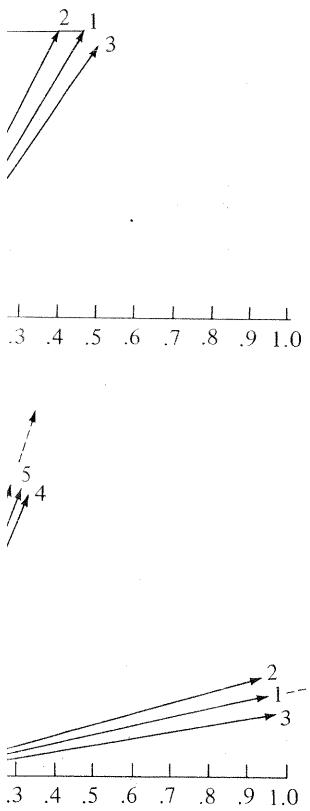


Figure 6-1b contains the points redrawn relative to a different pair of axes. These points were determined by a specific means of rotation called *varimax*. I will discuss the varimax rotation in more detail later in this chapter. However, I could have obtained similar results graphically. I located a new axis, which I called V_I , roughly 45 degrees from the old X axis, PC_I . (Conventionally, the angle of rotation, symbolized θ for the general case, is measured from the original axis to the new axis, in the *countrerclockwise* direction.) In order to maintain orthogonality, I also obtained a second axis (V_{II}) by rotating the original Y axis (PC_{II}) by 45 degrees. Physically rotating one of the two panels should convince you that the relations *among* the points are unchanged, only the *axes* that provide a frame of reference are changed.

Clearly, the points fall closer to the new axes (V_I and V_{II}) than to the original axes (the PCs). Tests X_1 , X_2 , and X_3 now fall very close to the V_{II} axis, and tests X_4 , X_5 , and X_6 fall very close to the V_I axis. V_I thus can be thought of as a relatively pure measure of numerical ability, and V_{II} can be thought of as a relatively pure measure of verbal ability. You might choose to argue that an even better solution is represented by the dashed lines that actually pass through the clumps of points. To do so is to argue for an oblique solution. Experienced factor analysts might prefer an oblique solution to an orthogonal solution in this case. The properties of an oblique rotation will be considered later in this chapter.

Both graphs contain other important features. Draw lines from the origin to each of the points. Note that the lines (*vectors*, since they describe line segments) are all of unit length. This is because I chose a *component* solution that treats all of the variance contained in each variable as systematic, and employed standardized variables. Had I chosen a common factor solution, the length of the vectors would have been less than one because only a portion of each variable's variance would have entered into the analysis.

The more general outcome with orthogonal factors is that the length of each vector with respect to a given pair of factors is the square root of the sum of the two squared factor loadings. The result follows from the Pythagorean theorem, Eq. (4-3). The total length of the vectors in a component solution, taking all factors into account, is 1.0. It is more complicated to describe a vector's length when factors are oblique.

Now, note the line drawn from the point corresponding to X_1 to the PC_I axis. The point of intersection with the axis is called its *projection*. The line connecting point and projection will also be parallel to the Y axis. If the vector is of unit length, as it is here, the length of the projection will equal the cosine of the angle between that vector and the PC_I axis. Because the cosine between two angles is equivalent to the correlation between two standardized variables (cf. Chapter 3), the projection represents the loading of that variable on PC_I . Similarly, the projection of the point onto the PC_{II} axis represents the loading of X_1 on PC_{II} .

As noted in Chapter 3, a rigid (orthogonal) rotation can be represented mathematically by an *orthonormal transformation matrix* of the form:

Table 6-2: (a) Principal component

$$\begin{bmatrix} \cos(x) & -\sin(x) \\ \sin(x) & \cos(x) \end{bmatrix}$$

If I refer to the matrix as \mathbf{T} , the matrix product \mathbf{BT} takes a set of loadings relative to one coordinate system (here, but not necessarily, a set of PCs) and expresses the relations among the variables relative to a new coordinate system (here, but not necessarily, a varimax rotation). The new pattern ($\mathbf{B}^* = \mathbf{BT}$) will preserve all the relations among the variables present in the original pattern matrix but can present the information in a more useful manner.

Analytic Orthogonal Rotation

In order to transform \mathbf{B} orthogonally into \mathbf{B}^* , all one needs is a suitable rotation angle, θ , for all pairs of axes. Choice of rotation angle was made visually before the computer era. Rotations are now usually performed to satisfy one of several possible mathematical criterion. Numerous criteria exist. However, *varimax* (Kaiser, 1958) is clearly the most popular one, primarily because it most closely approximates orthogonal simple structure. Varimax rotation has largely replaced an earlier analytic solution, *quartimax* (Ferguson, 1954) even though there are situations where thought would dictate that the latter is preferable. Generally speaking, varimax tends to represent the goals of the Thurstonian tradition, and quartimax tends to represent the goals of the Spearman–Holzinger tradition.

I will describe quartimax first because it is somewhat simpler. The goal of a quartimax rotation is to make the loadings on each *row* (that is, for each *variable*) of the pattern matrix as large (close to 1.0) or as small (close to .0) as possible in absolute value. The purpose is to maximize the variance of squared factor loadings in a rowwise manner. Since a variance involves squaring terms and the terms being squared are themselves squares (of factor loadings), one works with fourth powers, hence the prefix “quarti.”

The logic used by Ferguson (1954) and others working at about the same time (Carroll, 1953; Neuhaus & Wrigley, 1954) is most elegant. They noted that since the sum of squared factor loadings across rows is invariant:

$$h^2 = \sum b^2$$

This equality holds for all variables individually and, therefore, collectively. One may square both sides of the equation, leading to a sum of b^4 values on the left-hand side of the equation. The right-hand side consists of two types of expressions. One is the sum of fourth powers of b (b^4). The other consists of cross products of b^2 values for a given variable on pairs of factors. The result is nothing other than an expansion of $(X + Y)^2$ into $X^2 + Y^2 + 2XY$. The first two terms (X^2 and Y^2) represent the b^4 terms and $2XY$ corresponds to crossproducts of b^2 terms for different factors.

In quartimax, the b^4 terms are maximized, which is equivalent to minimizing the squared cross-product terms, as the two terms add to a constant.

Maximizing b^4 accomplishes the loadings across rows as large or a

For any pair of factors, the angle can be readily determined through available in all major statistical packages (pp. 285–286) for details.

When more than two factors are used. Factors I and II are rotated by the quartimax criterion. Rotations called I' and II' . The process is repeated until a *cycle* of rotation. Whenever the differences differ little, the process terminates eigenanalysis.

The objective of quartimax is to lead to a solution where each variable produces a general factor. The T_{II} as undesirable, but it need not be. *Varimax* was developed to minimize the distance of the loadings from zero, closer to Thurstone's idea. In its original form, quartimax gave loadings that were obtained from quartimax. *normalize* the loadings of pairs of variables. The rotation that is also difficult to understand (consult Harman, 1976, pp. 291–292). The original version of varimax and no and PC_{II} by forming a rotation angle on each of the two columns versus eigenvectors and rescaling the rotated loadings. The latter, in approximate simple structure as variances. In the present case, the quartimax was essentially identical to the varimax presented graphically in Fig. 6-1b.

One important thing to remember is that quartimax (and not be used when theoretical considerations are important; see Harman, 1966, 1983), since it tends to descriptively dominated by “g,” correlated and thus individually questionable following a varimax rotation. One uses of factor analysis is as follows: measures a single factor. Investigates the number of factors, uses a varimax to eliminate the single-factor structure.

$\begin{bmatrix} 1(x) \\ s(x) \end{bmatrix}$

product \mathbf{BT} takes a set of loadings (not necessarily, a set of PCs) and relative to a new coordinate system \mathbf{n}). The new pattern ($\mathbf{B}^* = \mathbf{BT}$) will be present in the original pattern in a more useful manner.

o \mathbf{B}^* , all one needs is a suitable choice of rotation angle was made. This is now usually performed to criterion. Numerous criteria exist. The most popular one, primarily orthogonal simple structure. Varimax analytic solution, *quartimax* (Ferguson's where thought would dictate king, varimax tends to represent quartimax tends to represent the

is somewhat simpler. The goal of loadings on each row (that is, for each i to 1.0) or as small (close to .0) as possible. Maximizing the variance of squared loadings involves squaring terms in squares (of factor loadings), one "quarti."

thers working at about the same time (1954) is most elegant. They noted that across rows is invariant:

ually and, therefore, collectively. leading to a sum of h^4 values on the left-hand side consists of two types of products of b (b^4). The other consists of variable on pairs of factors. The expansion of $(X + Y)^2$ into $X^2 + Y^2 + 2XY$. The b^4 terms and $2XY$ corresponds to factors. which is equivalent to minimizing the two terms add to a constant.

Maximizing b^4 accomplishes the objective of making the absolute values of loadings across rows as large or as small as possible.

For any pair of factors, the angle needed to achieve the quartimax criterion can be readily determined through calculus. Because it is complex and already available in all major statistical packages, I will refer you to Harman (1976, pp. 285–286) for details.

When more than two factors are rotated, the following procedure must be used. Factors I and II are rotated through an angle of θ degrees determined by the quartimax criterion. Rotation yields two new factors, which will be called I' and II'. The process is repeated for all possible pairs of factors, forming a cycle of rotation. Whenever the results obtained from two successive cycles differ little, the process terminates, just like any other iterative operation, e.g., eigenanalysis.

The objective of quartimax is to simplify rows. It can, and frequently does, lead to a solution where each variable loads maximally on the same factor, producing a general factor. The Thurstonian tradition regards a general factor as undesirable, but it need not be so if the remaining factors are well defined. Varimax was developed to minimize the role of general factors and is therefore closer to Thurstone's idea. In its original form, varimax simply consisted of applying the logic of quartimax to columns (factors) instead of rows (variables). This original form typically gave rise to results that were highly similar to those obtained from quartimax. Kaiser's (1958) ingenious "trick" was to normalize the loadings of pairs of factors before rotation. It leads to an angle of rotation that is also difficult to describe here (interested readers should consult Harman, 1976, pp. 291–292). To appreciate the difference between the original version of varimax and normalized varimax, consider (a) rotating \mathbf{PC}_I and \mathbf{PC}_{II} by forming a rotation angle to maximize the variance of the b^2 values on each of the two columns versus (b) forming the angle with the normalized eigenvectors and rescaling the rotating factors to unit length. The results will not be the same and the latter, in fact, leads to solutions that more closely approximate simple structure as well as more nearly "even out" factor variances. In the present case, the quartimax solution, which was not presented, was essentially identical to the varimax which is given in Table 6-2 and presented graphically in Fig. 6-1b.

One important thing to remember about varimax rotation is that it should not be used when theoretical considerations dictate a general factor (Gorsuch, 1966, 1983), since it tends to destroy that factor. Even so, if the data are overwhelmingly dominated by "g," implying the variables are highly intercorrelated and thus individually quite reliable, the general factor will remain following a varimax rotation. One of the most inappropriate, yet common, uses of factor analysis is as follows. Investigator A proposes that a scale measures a single factor. Investigator B uses a liberal rule to decide upon the number of factors, uses a varimax rotation and then "shows" that there are multiple factors. But the varimax rotation method is specifically designed to eliminate the single-factor structure proposed by Investigator A, which it

g items tend not to be very high on dictates a Holzinger bifactor thogonality, leave the first factor imax. The first factor will thus be s will have the requisite group ccomplish using standard pack-all factors.

to place axes as close to each of ificing their orthogonality. The described by their salient variables. may arise.

ger the same as the structure for o-order correlations in multiple tors are highly correlated.

need consider some properties of s two points, A and B, which are O_{II} .

ndicularly to O_I is the *structure* l the factor). I will designate the b_{ij} ." Now, consider the other line origin to the intersection is the b_{ij} to describe the pattern weight s the beta weight in predicting n exceed 1.0; structure weights eometry illustrated here to Figs.

B and the structure matrix S is x, Φ , as given in

$$(6-8)$$

analyzing items in Chapter 12. The of the correlations among the indi-elated because they are measuring orrelate poorly even though they s differ in shape. The former reflects lects *statistical* bases. If I asked you uld be measuring the same thing as yet the proportion of "yes" answers hich may lead the two questions to derived from the abilities literature) response distribution rather than re essentially artifacts.

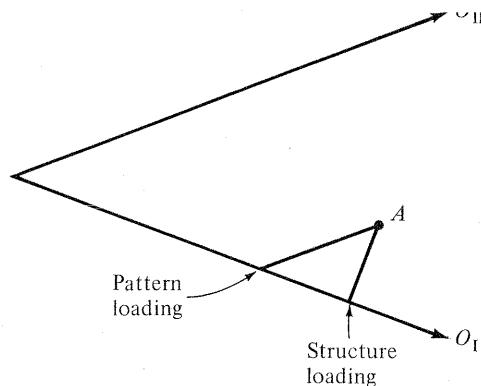


FIGURE 6-2. Graphical determination of a structure element and a pattern element.

Because the factors are correlated, h^2 (the multiple correlation between factors and a given variable) is no longer the sum of the squared factor loadings. (As noted previously, the term "loading" is ambiguous for oblique factors.) Instead, it is the sum of the products of the pattern elements and their corresponding structure elements over factors, i.e., $h_j^2 = \sum b_{jk}^* s_{jk}$, where j denotes the variable and k is an index for factors, paralleling Eq. (4-14). The matrix product $\mathbf{B}'\mathbf{S}$ contains the h^2 values of the individual variables in its diagonal and the estimated correlations among the variables off the diagonals. *Do not attempt to compute h^2 values by summing the squares of factor loadings across factors when the solution is oblique. Likewise, do not attempt to obtain the total variance by adding the individual factor variances.*

The rather different matrix product $\mathbf{B}'\mathbf{S}$ is also useful. It contains factor variances in its diagonal and covariances among the factors off the diagonals. The variance accounted for by an individual factor may be obtained just as in the orthogonal case.

It is very easy to be fooled by an oblique solution in exploratory factor analysis. Consequently, it should not be used until you first become very familiar with orthogonal solutions. Normally, people interpret the pattern elements rather than the structure elements. Pattern elements reflect only the *direct* contribution of a factor to a variable and not the *total* contribution, which is indexed by S. With very high factor correlations, it is quite possible for a variable to be "explained" by one factor, since it can have a high b weight for that factor and a low b weight for all other factors. However, its s weight can be large on all factors.

Reference Vectors

Although oblique rotations to simple structure could be done visually, Thurstone (1947) developed a procedure based on the concept of *reference vectors*.

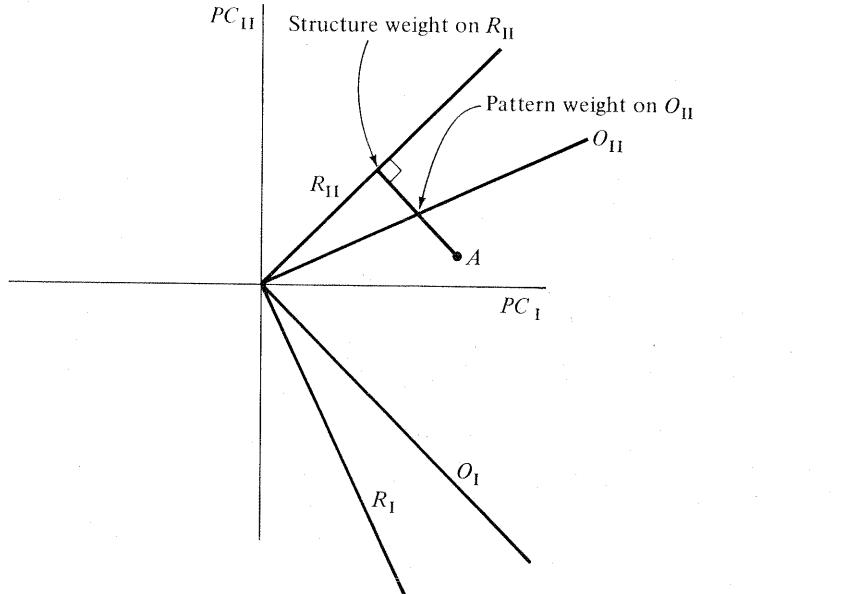


FIGURE 6-3. Principal components (PC_1 and PC_{II}), oblique rotations (O_1 and O_{II}), and reference vectors (R_1 and R_{II}).

Because hand rotation is seldom performed nowadays, reference vectors are rarely used for their original purpose, but the vectors have some interesting properties. Figure 6-3 contains a pair of such vectors.

Suppose O_1 and O_{II} are a pair of oblique factors. Now, choose an axis that is orthogonal to O_{II} (and any other axes except O_1). The axis is the reference vector for O_1 , R_1 . The reference vector for O_{II} , R_{II} , can be constructed in a like manner. (Do not confuse R_1 and R_{II} with the correlation matrix \mathbf{R} .) I have included a point (A) on the graph. Note that its *structure weights* on R_{II} are proportional to its pattern weights on O_{II} . By this construction, the reference structure weights contain correlations between a variable and a factor (the reference vector) *partialling out all original factors* (O_1 in the example), save one. Pattern weights on the original factor are proportional to reference structure weights. Consequently, pattern weights index partial correlations for the same reason that beta weights describe the change in criterion per unit change in the predictor, holding all other predictors constant.⁸ Reference vectors are part of the oblique rotation output in SAS and some other packages.

⁸In order to obtain the reference structure matrix, S_r , form the matrix product $\mathbf{B}'\mathbf{RB}$. Its diagonals contain the variances of the linear combinations of \mathbf{B} as applied to \mathbf{X} , another instance of Eq. (3-8). Form a diagonal matrix that contains the reciprocals of the variances and call the matrix \mathbf{K} . Then, $S_r = \mathbf{KB}$.

Analytic Oblique Rotation

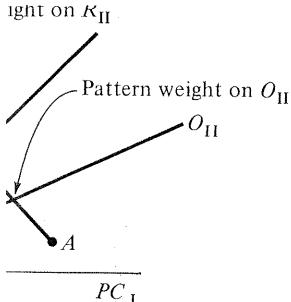
A bewildering number of analytic oblique rotation methods exist. As a result, I will not attempt to review them all here. The section on this topic is quite lucid in Harman (1960). In general, many such rotation methods attempt to control the degree of correlation between the resulting factors. One criterion controls the degree of correlation between the factors and a second criterion the degree of correlation between the factors and the original variables. The parameter determines the extent to which the factors are correlated with each other(s). If you wish to use an analytic oblique rotation, go on a bit of a "fishing expedition." Factor patterns with highly correlated factors will be highly redundant, and vice versa.

As was noted in the orthogonal rotation section, the first factor, PC_1 , is unrotated as a measure of " g " or the first common factor. The remaining factors, PC_2, PC_3, \dots , will be correlated among themselves. This is a *bifactor solution*. Unfortunately, bifactor solutions are not available in most existing packages.

The Common Factor Model

Considerable space has been devoted to the common factor model by investigators support, at least to some extent, the common factor model (e.g., Bernstein & Garbin, 1985a, 1985b, 1972). However, most researches support the common factor model, although it is not a particularly good model because most conception of the common factor model. One reason for this is that the common factor model is that the common factor is not a factor and it is clear that some of the common factor is not a factor. Random error is granted an explicit role in the common factor model, although the concept of *error variance*, which measures the amount of variance not explained by the common factor, is not explicitly mentioned. (See Cudeck, 1989, for a discussion of the reliability of the common factor model.) A second reason is that the common factor model is not a good model because it does not systematically, but what it is measuring. Other variables are measuring. Shared variance is shared with other variables in the common factor model. One goal of the common factor model is to explain the shared variance jointly and independently of the common factor.

⁹If you are familiar with the relationship between the common factor model and the structural equation modeling, you might recognize Scales 2 and 5 (M1 and M2) as being independent of the remaining major factors.



, oblique rotations (O_I and O_{II}), and

nowadays, reference vectors are
ie vectors have some interesting
vectors.

vectors. Now, choose an axis that
ept O_I). The axis is the reference
 R_{II} , can be constructed in a like
ie correlation matrix \mathbf{R} .) I have
t its *structure weights* on R_{II} are
y this construction, the reference
een a variable and a factor (the
factors (O_I in the example), save
r are proportional to reference
ghts index partial correlations for
the change in criterion per unit
predictors constant.⁸ Reference
utput in SAS and some other

, S_r , form the matrix product $\mathbf{B}'\mathbf{R}\mathbf{B}$.
combinations of \mathbf{B} as applied to \mathbf{X} ,
atrix that contains the reciprocals of
 \mathbf{B} .

Analytic Oblique Rotation

A bewildering number of analytic oblique rotations have been developed. As a result, I will not attempt to review them all here. As usual, Gorsuch's (1983) section on this topic is quite lucid. One important point to keep in mind is that many such rotation methods contain a parameter that allows you to control the degree of correlation among factors. The reason is that several procedures compromise between one criterion that produces highly correlated factors and a second criterion that produces nearly orthogonal factors. The parameter determines the extent to which one criterion is weighted relative to the other(s). If you wish to use an oblique exploratory rotation, you need to go on a bit of a "fishing expedition" with several values of the parameter. Factor patterns with highly correlated factors can look quite pretty, but such factors will be highly redundant, by definition.

As was noted in the orthogonal case, one may wish to leave the first PC unrotated as a measure of " g " and rotate the remaining PCs obliquely. The first factor, i.e., PC_1 , will be orthogonal to all remaining factors, but the latter will be correlated among themselves, producing a *correlated* or *oblique bifactor solution*. Unfortunately, bifactor solutions are difficult to obtain directly with existing packages.

The Common Factor Model

Considerable space has been devoted to the component model which many investigators support, at least to some extent (e.g., Bernstein & Eveland, 1982; Bernstein & Garbin, 1985a, 1985b; Nunnally, 1978; Schoenemann & Wang, 1972). However, most researchers working in the field probably prefer a common factor model, although nothing was wasted in discussing the component model because most concepts covered earlier are also applicable to the common factor model. One reason for the greater popularity of the common factor model is that the component model does not contain a concept of error, and it is clear that some of the observed variance is nothing but garbage. Random error is granted an explicit status in a common factor model through the concept of *error variance*, which reflects the *unreliability* of the individual measures being factored. (See Chapter 12 for a detailed discussion of test reliability.) A second reason is that a variable may measure something quite systematically, but what it is measuring may be totally unrelated to what the other variables are measuring.⁹ Systematic variance in a variable that is not shared with other variables in the set is known as *specific variance*. Error and specific variance jointly and independently form the *unique variance* in the common factor model. One goal of common factor analysis is to separate

⁹If you are familiar with the relations among the MMPI's 14 standard scales, you might recognize Scales 7 and 5 (Mf or masculinity-femininity) as being relatively independent of the remaining major scales.

unique variance from common factor variance and prevent the former from affecting the factor structure.

Operationally, the common factor model involves a matrix whose diagonal elements (values of r_{ii}) are less than 1.0. One consequence is that the rank of the matrix will be reduced; hence, a given number of common factors will approximate the obtained correlations better than a like number of components. However, the gain in fit may be offset in part by the need to estimate the r_{ii} (communality) terms. The communalities need be distinguished from the *communality estimates* (h^2 values obtained from the diagonals of \mathbf{BS}'). Although communalities and communality estimates will usually be very similar, the two need not be identical. *From the user's standpoint, the most important feature of the common factor model is that it weights variables unequally; the component model ordinarily weights all variables equally, whether or not a given variable has anything in common with the other variables being investigated.*

One issue relevant to the common factor model, which some view as a problem, is that good arguments can be made for somewhat different communalities. In general, all common factor models employ an eigenanalysis of a matrix with communalities in the diagonal to obtain a primary solution. The same rotational concepts may then be applied to common factors as to a PC solution.

The term *principal axis* (PA) is used generically to describe an eigenvector of a correlation matrix containing communalities. In many cases, more specific names like "alpha factor" are used to describe specific ways that communalities estimates are determined. The more specific name would be preferred in that context.

An important consideration is that common factor analysis involves relations among *inferred* linear combinations. That is, common factors cannot be observed precisely in the data in the sense that components can be. Common factors are not simple linear transformations of observable variables, as will be shown, because they are not completely *determined*. In a great many situations, your interest may be limited to relations among observable variables, in which case a component model would be preferable.

There are three general ways to define communalities. One is by a formal, theoretically based definition. The second is by a direct estimation process. The third is to define them as a by-product of the operations defining other statistics, specifically the factor pattern (B).

The following are examples of the first approach:

1. Communalities (values of r_{ii}) can be defined as the total amount of *systematic* variation. This definition assumes that *specific* variance is, in principle, zero. In other words, if a variable measures something systematically, another variable could be found that would correlate perfectly with that variable. Conversely, all unique variance (lack of communality) is assumed

to arise from random error (Chapter 12) as the common factors obtained are similar to PCs unless one or The communalities obtain larger than the communalit

2. Conversely, one can use s_i as the given variable and all other s_j in Chapter 4, the SMCs ma corresponding elements of \mathbf{B} and \mathbf{S} limit analysis to that other variables in the data. A specific set of variables is true. SMCs will differ more from reliabilities because the SM In addition, the communality estimates.
3. An attempt can be made to estimate communalities, h^2 , by iteration (2), as initial values. These values. If the two sets of va the process stops. Otherwise, communalities. The matrix is th values) are again compared process stops at convergence the process does converge, the communality estimates,
4. *Image* analysis is an extensi is the portion of variance of variables in the set, i.e., 1 - observed correlations, the c implied by the name of the could be obtained directly t other variables as predictor resulting matrix, the actual nious algorithm.

Examples of direct estimatio

1. Various formulations of the kog, 1967; Rao, 1955). These of all parameters, i.e., \mathbf{B} , Φ , Λ as the fit provided by the gi
2. *Generalized least squares* (I alternative criterion to maxi

ance and prevent the former from

I involves a matrix whose diagonal consequence is that the rank of a number of common factors will be greater than a like number of components in part by the need to estimate communalities need be distinguished from obtained from the diagonals of \mathbf{BS}' . Reliability estimates will usually be very small. *From the user's standpoint, the most important model is that it weights variables equally, whether or not they are in common with the other variables being analyzed.*

vector model, which some view as a model for somewhat different commonalities. Some models employ an eigenanalysis of the covariance matrix to obtain a primary solution. This can be applied to common factors as to

erically to describe an eigenvector of the commonalities. In many cases, more specific methods may be used to describe specific ways that communalities and specific name would be preferred in

mon factor analysis involves relating the total variance of each variable to the variance explained by the common factors. That is, common factors cannot be identified directly from the observed variables, as will be shown below. Common factors are said to be *determined*. In a great many situations among observable variables, this approach would be preferable.

Commonalities. One is by a formal, indirect estimation process. Another is by a direct estimation process. Both of the operations defining other

approach:

defined as the total amount of systematic variance in the variables. The measure of systematic variance is, in principle, something systematically, that is, that would correlate perfectly with that variable. The lack of communality is assumed

to arise from random error (unreliability). Using the *reliability coefficient* (Chapter 12) as the communality follows from this definition. Empirically, common factors obtained using reliabilities as communalities will be quite similar to PCs unless one or more of the measures are extremely unreliable. The communalities obtained using reliability coefficients will generally turn larger than the communality estimates (values of h^2).

2. Conversely, one can use *squared multiple correlations* (SMCs) between a given variable and all other variables in the set as communalities. As noted in Chapter 4, the SMCs may be obtained as 1.0 minus the reciprocals of the corresponding elements of \mathbf{R}^{-1} , i.e., $1 - 1/r_{ii}$. Using SMCs as communalities limits analysis to that variance a given variable actually shares with other variables in the data set. Consequently, variance not shared with the specific set of variables is treated as unique. Common factors derived from SMCs will differ more from components than common factors derived from reliabilities because the SMCs will always be smaller than the reliabilities. In addition, the communalities will generally be *smaller* than the communality estimates.
3. An attempt can be made to *equate* the communalities, r_{ii} , to the communality estimates, h^2 , by *iteration*. The analysis usually begins with SMCs, as in (2), as initial values. These SMCs are then compared to the resulting h^2 values. If the two sets of values agree within tolerance (which is unlikely), the process stops. Otherwise, the h^2 values replace the SMCs as the communalities. The matrix is then refactored and the communalities (initial h^2 values) are again compared with the resulting communality estimates. The process stops at convergence or after a preset number of iterations. When the process does converge, which need not occur, the communalities *equal* the communality estimates, by definition. This method is very widely used.
4. *Image analysis* is an extension of method (2). The image of a given variable is the portion of variance of that variable that is predictable from the other variables in the set, i.e., $1 - 1/r_{ii}$ for variable i . Instead of factoring the observed correlations, the correlations among the images are factored, as implied by the name of the procedure (Guttman, 1953). Although images could be obtained directly by generating scores for each variable using the other variables as predictors, intercorrelating the scores, and factoring the resulting matrix, the actual computations follow from an extremely ingenious algorithm.

Examples of direct estimation include:

1. Various formulations of the *exploratory maximum likelihood* model (Jöreskog, 1967; Rao, 1955). These models provide maximum likelihood estimates of all parameters, i.e., \mathbf{B} , Φ , and \mathbf{U} . Consequently, r_{ii} will be as similar to h^2 as the fit provided by the given number of factors allows.
2. *Generalized least squares* (Bentler & Bonett 1980; Browne 1974) is an alternative criterion to maximum likelihood estimation and produces very

The desired matrix is therefore the right-hand pseudoinverse, $\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}$. As it is customary to present \mathbf{W} with variables as rows and factors as columns, the pseudoinverse is usually transposed.

The factor score weight matrix derived from a PC solution is quite simple to obtain. The relation is

$$\mathbf{W} = \mathbf{V}'\lambda^{-1} \quad (6-10)$$

Equation (6-10) is very similar to Eq. (6-5). To obtain a *pattern* (\mathbf{B}) matrix from PCs, one *multiplies* normalized eigenvectors by the square root of their associated eigenvalues; to obtain a *factor score weight* (\mathbf{W}) matrix from PCs, one *divides* normalized eigenvectors by the square root of their associated eigenvalues.

The above procedures are called "exact" because they will produce an \mathbf{F} matrix having the properties of the model *for that set of data*. However, the procedures are also subject to sampling error just as any other quantity is, so that the precision you get from using the \mathbf{W} matrix output from a computer package is more apparent than real. At the same time, it is often convenient to use the \mathbf{W} matrix, as it has already been calculated.

Estimation Procedures

A consequence of the second form of factor indeterminacy discussed previously and the subsequent lack of perfect multiple correlation between variables and any given factor is that there exist an infinite number of sets of factor scores that can fit the data equally well. Consequently, the "true" factor scores are unknowable. Perhaps the most common way to define \mathbf{W} is through the relation $\mathbf{R}^{-1}\mathbf{S}$. This definition reveals a relationship between how regression weights were obtained in Eq. (4-10) and how factor score weights are derived.¹⁰

Although *common factors* may be orthogonal, *estimates* of factor scores may not be, particularly when the regression method is used. One consequence of the second form of factor indeterminacy is that matrices describing relations among *factors*, i.e., \mathbf{B} , Φ , \mathbf{S} , etc., have corresponding, but not identical, counterparts that describe relations among *factor scores*. Likewise, factor scores need not be perfectly correlated with the factors they presumably measure and may actually be correlated with factors they are not intended to measure. The lack of correspondence will not be a major problem in a clearly defined solution in which the R^2 values that predict factors from variables is high.

In theory, it is not possible to achieve all the goals desired of common factor scores. Mulaik (1972), Harman (1976), and Gorsuch (1983) discuss the various methods possible. Most of the methods lead to rather complex and uninstruc-

¹⁰The result is just a special case of Eq. (4-13). Structure elements may be viewed as correlations between variables (predictors) and factors (criteria). \mathbf{R}^{-1} may therefore be viewed as the inverse of a predictor intercorrelation matrix.

tive equations, so they will not be allow alternative ways to generate defined, in the sense that its R^2 wit estimation methods will agree hig

Estimation procedures generally Their variance will ordinarily be t

Approximation Procedures

The many similarities between fact similar considerations in the defini that are much simpler than those of will suffice. Weight the salient vari the resulting sums, if desired (whic scaling assumptions. The most imp the properties of the derived variabl

In other words, if you wished to above, simply define one as $X_1 + 1$ but be sure to compute their correlat be explored further using the metl considered in Chapter 7.

The basic concepts of explorato duced. Perhaps the most importa different factoring methods. Starti discussed how rotations often made showed how a common factor m played by variables that had little i being factored. The chapter conclu factor scores could be obtained wh

In Chapter 7, I will deal with analysis. Part of my discussion deal how you can determine the similari with how you can evaluate the prop depend on those generated by satisf part of the next chapter deals with may be inferred from correlational

Addendum: Constructing C Desired Factor Structure

You may wish to generate a series of structure. The following SAS DAT measures used in the common facto

nd pseudoinverse, $\mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}$. As it
rows and factors as columns, the
om a PC solution is quite simple

(6-10)

5). To obtain a *pattern* (**B**) matrix
ctors by the square root of their
core weight (**W**) matrix from PCs,
e square root of their associated

because they will produce an F
for that set of data. However, the
r just as any other quantity is, so
matrix output from a computer
same time, it is often convenient
calculated.

r indeterminacy discussed previ-
ultiple correlation between vari-
in infinite number of sets of factor
sequently, the “true” factor scores
a way to define **W** is through the
elationship between how regres-
nd how factor score weights are

nal, *estimates* of factor scores may
hod is used. One consequence of
hat matrices describing relations
onding, but not identical, counter-
ores. Likewise, factor scores need
ey presumably measure and may
ot intended to measure. The lack
m in a clearly defined solution in
n variables is high.

re goals desired of common factor
orsuch (1983) discuss the various
to rather complex and uninstruc-

ture elements may be viewed as
ctors (criteria). \mathbf{R}^{-1} may therefore be
on matrix.

tive equations, so they will not be presented here. Many computer packages allow alternative ways to generate factor scores. In practice, if a factor is well defined, in the sense that its R^2 with the predictor variables is large, different estimation methods will agree highly with each other.

Estimation procedures generally *do not* produce scores with unit variance. Their variance will ordinarily be the R^2 between the factor and variables.

Approximation Procedures

The many similarities between factor analysis and multiple regression lead to similar considerations in the definition of a factor score. In general, weights that are much simpler than those obtained from an exact or estimation process will suffice. Weight the salient variables on each factor equally. Standardize the resulting sums, if desired (which usually is unnecessary), to conform to scaling assumptions. The most important thing to keep in mind is to *examine the properties of the derived variables empirically*.

In other words, if you wished to define factor scores for the two factors above, simply define one as $X_1 + X_2 + X_3$ and the other as $X_4 + X_5 + X_6$, but be sure to compute their correlation. The properties of the two sums may be explored further using the method of confirmatory factor analysis to be considered in Chapter 7.

The basic concepts of exploratory factor analysis have now been introduced. Perhaps the most important point was how different goals dictate different factoring methods. Starting with principal component analysis, I discussed how rotations often made the results more readily interpreted. I then showed how a common factor model might be used to minimize the role played by variables that had little in common with other variables in the set being factored. The chapter concluded with a discussion of the different ways factor scores could be obtained when they are needed.

In Chapter 7, I will deal with concepts related to *confirmatory* factor analysis. Part of my discussion deals with the essential topic of replication—how you can determine the similarity of two sets of factors. Another deals with how you can evaluate the properties of proposed factors and not simply depend on those generated by satisfying certain mathematical criteria. A final part of the next chapter deals with the controversial topic of how causality may be inferred from correlational data—path analysis.

Addendum: Constructing Correlation Matrices with a Desired Factor Structure

You may wish to generate a series of measures that produce a specified factor structure. The following SAS DATA step was used to construct the seven measures used in the common factor analysis:

```

DATA;
RETAIN SEED1 SEED2 1984 ERR1-ERR7 1984
DO I = 1 TO 500;
F1 = RANNOR (SEED1);
F2 = RANNOR (SEED2);
E1 = RANNOR (ERR1);
E2 = RANNOR (ERR2);
E3 = RANNOR (ERR3);
E4 = RANNOR (ERR4);
E5 = RANNOR (ERR5);
E6 = RANNOR (ERR6);
E7 = RANNOR (ERR7);
X1 = .96 * F1 + .28 * F2 + .33 * E1;
X2 = .95 * F1 + .31 * F2 + .33 * E2;
X3 = .98 * F1 + .20 * F2 + .33 * E3;
X4 = .31 * F1 + .95 * F2 + .33 * E4;
X5 = .28 * F1 + .96 * F2 + .33 * E5;
X6 = .20 * F1 + .98 * F2 + .33 * E6;
X7 = E7;
OUTPUT;
END;

```

Note: The program simply consists of a loop that is executed 500 times. The variables F1 and F2 are common factor scores that are the same for each of the seven observed variables (X1 to X7). The variables E1 to E7 are error scores that are unique to their respective observed variables. The common factor scores and the error scores are all sampled independently from standard normal distributions using the SAS random number generating (RANNOR) function. The RETAIN statement sets a random number “seed” which allows the sequence of random numbers to be reproduced. Changing the “seed” to a different number provides a new, independent replication sample. The quantities .96, .28, ..., are the weights used to define the factor scores. Note that the sums of the squared weights for each observed variable are 1.0 in all cases, e.g., $.96^2 + .28^2 + .33^2 = .95^2 + .31^2 + .33^2 = 1.0$. Because F1, F2, and each of the error scores are independent, each of the observed variables will also be in standard normal form. The principle may readily be extended to any number of variables and any number of factors. Simply make sure that the sums of squared coefficients for each variable adds to 1.0. Note that if, in addition, the coefficients for the error scores are .0, the data will fit a component model.

The same computations may be programmed in SPSSX as follows:

```

INPUT PROGRAM
LOOP #I = 1 TO 500
SET SEED = 987654321
COMPUTE F1 = NORMAL (1)

```

```

COMPUTE F2 =
COMPUTE E1 =
COMPUTE E2 =
COMPUTE E3 =
COMPUTE E4 =
COMPUTE E5 =
COMPUTE E6 =
COMPUTE E7 =
COMPUTE X1 =
COMPUTE X2 =
COMPUTE X3 =
COMPUTE X4 =
COMPUTE X5 =
COMPUTE X6 =
COMPUTE X7 =
END CASE
END LOOP
END FILE
END INPUT PROC
PEARSON CORR
STATISTICS ALL
LIST VARIABLE

```

ERR1-ERR7 1984

E1;
E2;
E3;
E4;
E5;
E6;

that is executed 500 times. The scores that are the same for each of the variables E1 to E7 are error observed variables. The common factor loadings are generated independently from standard normal number generating (RANNOR) command number "seed" which allows replication produced. Changing the "seed" to generate different replication sample. The factor loadings define the factor scores. Note that the observed variable are 1.0 in all three dimensions because $.33^2 = 1.0$. Because F1, F2, and F3 are orthogonal, each of the observed variables will have a unique factor loading. Simply make sure that the sum of the factor loadings for each observed variable adds to 1.0. Note that if, in the case where the factor loadings are .0, the data will fit a component perfectly.

ned in SPSSX as follows:

```
COMPUTE F2 = NORMAL (1)
COMPUTE E1 = NORMAL (1)
COMPUTE E2 = NORMAL (1)
COMPUTE E3 = NORMAL (1)
COMPUTE E4 = NORMAL (1)
COMPUTE E5 = NORMAL (1)
COMPUTE E6 = NORMAL (1)
COMPUTE E7 = NORMAL (1)
COMPUTE X1 = .96 * F1 + .28 * F2 + .33 * E1
COMPUTE X2 = .95 * F1 + .31 * F2 + .33 * E2
COMPUTE X3 = .98 * F1 + .20 * F2 + .33 * E3
COMPUTE X4 = .31 * F1 + .95 * F2 + .33 * E4
COMPUTE X5 = .28 * F1 + .96 * F2 + .33 * E5
COMPUTE X6 = .20 * F1 + .98 * F2 + .33 * E6
COMPUTE X7 = E7
END CASE
END LOOP
END FILE
END INPUT PROGRAM
PEARSON CORR X1 TO X7
STATISTICS ALL
LIST VARIABLE = X1 TO X7
```