Exploratory Factor Analysis

Ledyard R Tucker
and
Robert C. MacCallum

Ledyard Tucker is Professor Emeritus of Psychology at the University of Illinois. Robert MacCallum is Professor of Psychology at Ohio State University.

Table of Contents

CHAPTER I
GENERAL CONCEPTS AND OBJECTIVES OF FACTOR ANALYSIS

1.0. General Introduction

The field of factor analysis involves the study of order and structure in multivariate data. The field includes both theory about the underlying constructs and dynamics which give rise to observed phenomena, and methodology for attempting to reveal those constructs and dynamics from observed data. The purpose of this chapter is to provide an explanation of the central concepts of factor analytic theory and the objectives of factor analytic methodology. This will be done in an entirely non-mathematical fashion.

The material discussed in this chapter, as well as in the remainder of the book, will be presented in the context of the field of psychology. Often the more specific context of the study of mental abilities will be used for demonstrations, examples, and explanations of concepts. This context represents the field in which the primary development and use of factor analysis has occurred. Thus, we will generally consider problems where the entities under study are people, and the variables of interest are attributes of those people. It should be kept in mind, however, that factor analysis is valuable and applicable in a wide range of other settings. It has been used widely in other social sciences, education, business fields, biological sciences, etc. Thus, though much of the theory and methodology has developed within psychology, factor analysis can be applied beneficially to subject matter from many other disciplines.

1.1. Introduction to Factor Analytic Theory

To begin to develop an understanding of the type of problem to which factor analysis can be applied, and the objectives of this approach, let us first consider a process which characterizes much research in psychology. Such research often includes a number of steps which, though they may not be explicitly stated, can be recognized clearly and which represent integral parts of the research process. The first step is the identification of the domain and population of interest. The domain can be defined as the range of phenomena of interest in the research. The domain may be very broad, such as mental abilities, or more narrow, such as verbal abilities. Other examples of domains could include attitude, personality, physical abilities, etc. In addition to defining the domain, the researcher also defines the population, which is the entire set of entities of interest. In psychology the entities are very often people, but may be other things such as animals, groups, etc. For discussion purposes, we will assume that the entities of interest are people. Populations of entities are usually very large (e.g., 12-year-old children), almost always too large to allow for each entity to be measured.

Given a domain and population of interest, the researcher selects from the domain variables which are to be measured. We will refer to these variables as surface attributes; a surface attribute is any one of the many attributes of people that may be observed and measured. For instance, in the domain of mental abilities, tests could be constructed to measure distinct kinds of attributes. One test could contain addition problems; another could test spelling; a third could contain memory problems; etc. Each of these tests represents a surface attribute. Clearly, in any given domain, it would be possible to identify and measure a vast array of surface attributes. A set of surface attributes measured in a given study will be referred to as a battery of surface attributes.

In addition to selecting surface attributes to be measured, a researcher also must select a sample of individuals from the population. Measurements then are obtained for each individual in the sample on each surface attribute. When the researcher views such data, several things usually are quite apparent. First, there normally will be substantial variation among individuals in terms of their measures on the surface attributes. That is, individual differences will occur. In addition, there normally will be substantial covariation (i.e., correlation) between different surface attributes. For instance, an addition test may show a high correlation with a multiplication test. Given more than a few surface attributes, there likely will be a wide range of correlations among the attributes, some correlating highly and others quite low. This observed variation and covariation on the surface attributes may seem chaotic. Viewed at this level, it generally will be difficult to understand and account for in any simple manner. The theory and methods of factor analysis are designed explicitly for this purpose. Factor analysis involves a set of techniques designed to identify order and structure in such data by providing a parsimonious and meaningful explanation for the observed variation and covariation in surface attributes.

The cornerstone of factor analytic theory is the postulate that there exist internal attributes. An internal attribute is taken to be an unobservable characteristic of people on which people differ in extent or degree. Internal attributes are more fundamental than surface attributes. For instance, in the domain of mental abilities, internal attributes might be such things as numerical ability, verbal ability, etc. These attributes cannot be directly measured, but their effects are reflected when one obtains measures on surface attributes. The internal attributes commonly are referred to as factors, or latent variables. These terms will be used interchangeably.

Internal attributes also can be thought of as hypothetical constructs. They are not necessarily taken to be completely real and concrete. Rather they are constructs which, as will be seen, can be used to understand and account for observed phenomena. The central role of this concept of unobserved variables in factor analysis should not be viewed as a basis for criticism of factor analytic theory. The use of hypothetical constructs is routine in many fields of science. For

instance, physicists make use of a multitude of constructs (e.g., gravity, magnetism, etc.) to account for observed phenomena. These forces actually represent unobserved constructs whose existence has been hypothesized based on observed events. These constructs, along with theories which make use of them, then serve to account for a vast array of observed phenomena. In much the same way, factor analytic theory postulates that there exist unobservable internal attributes which account for observed variation and covariation across a wide range of surface attributes.

A central theoretical concept in factor analysis involves the relation of surface attributes to internal attributes. The basic principle is that internal attributes influence surface attributes in a systematic fashion. This implies that when a researcher obtains a measurement for an individual on a surface attribute, that measurement is, at least in part, the result of the influence of underlying internal attributes. For instance, an individual's score on an addition test (surface) is influenced by numerical ability (internal). According to the traditional representations of factor analytic theory, these influences of internal attributes on surface attributes are taken to be linear.

At this point it is useful to distinguish between two types of internal attributes, or factors. One type is called a common factor, which is defined as an internal attribute which affects more than one of the surface attributes in the selected set, or battery. For example, if the selected battery of surface attributes includes more than one which is influenced by numerical ability (e.g., both an addition test and a multiplication test), then numerical ability is a common factor. The second type of internal attributes is specific factors, each of which influence only one of the surface attributes in any given battery. There may be a number of specific factors for any given surface attribute; however, their influences can be viewed as being combined into a single specific factor. With a change of the battery of measured surface attributes one of the common factors may become a specific factor for one of the surface attributes, or one of the specific factors for one surface attribute may become a common factor. For example, consider that all surface attributes that are influenced by numerical ability are removed from a battery except for a test of addition; then numerical ability becomes a specific factor for the test of addition. The reverse can be true when a battery is enlarged; putting back into the battery the other tests which are influenced by numerical ability will make this ability a common factor again. For now, suffice it to say that there exist common factors, which are those that affect more than one attribute in the battery, and specific factors, each of which affects only one surface attribute in the battery.

In addition to the two types of internal attributes represented by common and specific factors, there exists a third influence on the surface attributes. This third influence is errors of measurement in observing each surface attribute. These errors of measurement are represented in factor theory as additional factors, though they do not correspond to internal attributes as defined above; i.e., they do not represent unobserved characteristics of individuals. Rather, errors of

measurement factors arise from transient, unsystematic events which influence the measurement of the surface attributes. Note the direct relation between errors of measurement factors and the reliability of the measures of the surface attributes. Higher reliability implies lower errors of measurement. Thus, the use of more or less reliable tests in a battery would influence the errors of measurement factors, but would not necessarily affect the common or specific factors. Note also that altering the test battery by removing or adding tests may affect the common and specific factors, as described above, but would not influence the errors of measurement factors since reliability is not affected by such changes in the battery.

For any particular battery the specific factors and errors of measurement factors may be combined into unique factors. There will be one unique factor for each surface attribute, with each unique factor defined as the combination of the specific and errors of measurement factors for the corresponding surface attribute. Note that unique factors will change as a result of any changes in specific or errors of measurement factors. Thus, changes in the battery which affect specific factors will in turn affect the corresponding unique factors, and the use of more or less reliable tests would affect errors of measurement factors and, in turn, the unique factors.

Given these basic definitions of factors, together with the principle that the factors influence the surface attributes, it can be understood that the factors combine to account for an individual's degree or level on a surface attribute. That is, any individual's level on a surface attribute can be viewed as arising from that individual's level on the relevant factors. For instance, an individual's score on an addition test can be viewed as being attributable to the individual's level on the internal attribute of numerical ability, along with his or her level on the specific factor associated with the addition test. Also, errors of measurement contribute to the measure of an individual's level on the addition test. By simple extension of this notion, it can be stated furthermore that individual differences, or variance, on surface attributes can be attributed to the underlying factors. That is, individuals differ on surface attributes because they differ on the internal attributes which influence those surface attributes, as well as due to the presence of errors of measurement. For instance, individual scores on a spelling test vary, in part, because those individuals vary with respect to their level on an underlying internal attribute, verbal ability. Further variation in the test scores can be attributed to individual differences on the specific factor associated with the spelling test, as well as the presence of errors of measurement.

The distinction drawn earlier between common, specific, and errors of measurement factors is very relevant to this view of how factors account for variation on surface attributes. It is very important to note that the observed variance on a given surface attribute in a battery can, according to factor analytic theory, be recognized as arising from three sources. Some of the variance is due to the influence of the common factors. This portion is referred to as the common variance, or communality. A second portion of the variance on the surface attributes arises from

the specific factor. This portion is termed specific variance, or specificity. A third portion of the variance on the attribute arises from errors of measurements, and is termed error of measurement variance. The specific variance frequently is combined with the error of measurement variance to form the unique variance, or uniqueness. To state this another way, the observed variation on a given surface attribute in a given battery is due in part to factors which influence other surface attributes in the battery, and in part to factors which influence only the given surface attribute. In this sense, the common, specific, and errors of measurement factors are taken to account for the observed variation on each surface attribute in the battery. Equivalently, it can be said that the observed variation on each surface attribute in the battery can be attributed to common and unique factors, since the unique factors represent the combined influences of the specific and errors of measurement factors.

By a further extension of these concepts, factor analytic theory also can be viewed as accounting for covariation between surface attributes. It is commonly observed that surface attributes within any given domain tend to be correlated with each other to varying degrees. According to factor analytic theory, such correlation is due to the influence of the common factors. That is, a correlation between two surface attributes is due to the dependence of those attributes on one or more of the same common factors. For instance, an addition test and a multiplication test will exhibit a fairly high positive correlation because both depend on the numerical ability factor. By contrast, an addition test and a spelling test will exhibit a much lower correlation because they are influenced by different common factors. Thus, the degree of intercorrelation between surface attributes is taken to be the result of the degree to which those attributes are influenced by the same internal attributes.

It is of utmost importance to understand that it is only the common factors which account for correlations of surface attributes. That is, unique factors do not give rise to such correlations. Equivalently, specific factors and errors of measurement factors do not give rise to correlations among surface attributes. By definition, they cannot because they each influence only a single surface attribute.

To summarize these basic principles of factor analytic theory, it is hypothesized that, in a given domain, there will exist a small number of common factors which influence the potentially vast array of surface attributes. Variation on the surface attributes is attributable, in part, to variation on the common factors. The remainder is due to unique factors, or, alternatively, to specific factors and errors of measurement factors. Covariation of surface attributes is attributable to the dependence of surface attributes on some of the same common factors. Thus, the basic tenets of this theory allow for the explanation of covariation on a potentially large number of surface attributes in terms of a much smaller number of internal attributes, or common factors.

We will now present a series of figures designed to provide a schematic representation of the theoretical concepts developed to this point. These figures are called path diagrams and were devised by Wright (1921). They were first used to represent factor analytic theory by Tucker (1940). Within these diagrams squares represent surface attributes and circles represent internal attributes and other types of factors. Uni-directional arrows represent directional linear influences, indicating a linear effect of one variable on another, via some type of causation or process. Bi-directional arrows represent correlation, where two variables are correlated, but where no directional influence of one on the other is hypothesized.

Figure 1.1 illustrates the relationships between surface attributes and three types of factors defined above. This figure shows four surface attributes on the right. These could be thought of as four mental tests, for example. On the left are three types of factors. The two internal attributes at the top each are represented as influencing more than one of the surface attributes. Thus, these satisfy the definition of common factors. According to the pattern of effects represented in the diagram, the first two surface attributes would be correlated with each other because both are influenced by the first factor. The last three surface attributes would also be intercorrelated because they are influenced by the second factor. Note that bi-directional correlational arrows among the surface attributes are not included in Figure 1.1. This is because the correlations among the surface attributes are seen as arising from the effects represented in the figure. That is, the effects of the common factors on the surface attributes imply that the surface attributes will be intercorrelated, so those intercorrelations are not explicitly represented in the figure.

It is important to note that one of the surface attributes in Figure 1.1 is represented as being influenced by both common factors. This is emphasized here in an effort to dispel a common misconception in applied factor analysis; i.e., that each surface attribute should be associated with only one factor. In fact it is entirely permissible and, indeed, quite common for a surface attribute to be influenced by more than one common factor. For example, if the two common factors in Figure 1.1 were numerical ability and verbal ability, a test composed of mathematical "story problems," requiring an ability to read and understand the problem as well as to solve it mathematically, would be dependent on both factors.

Another important aspect of Figure 1.1 is that the two common factors are represented as being intercorrelated; note that there is a bi-directional path connecting these two internal attributes. This illustrates the point that common factors may be related to each other, and, in fact, are often found to be so in practice. For instance, factors such as numerical ability and spatial ability may be correlated rather than uncorrelated constructs. Just as with surface attributes, correlations between internal attributes indicate a degree of linear relationship between measures for individuals on these attributes.

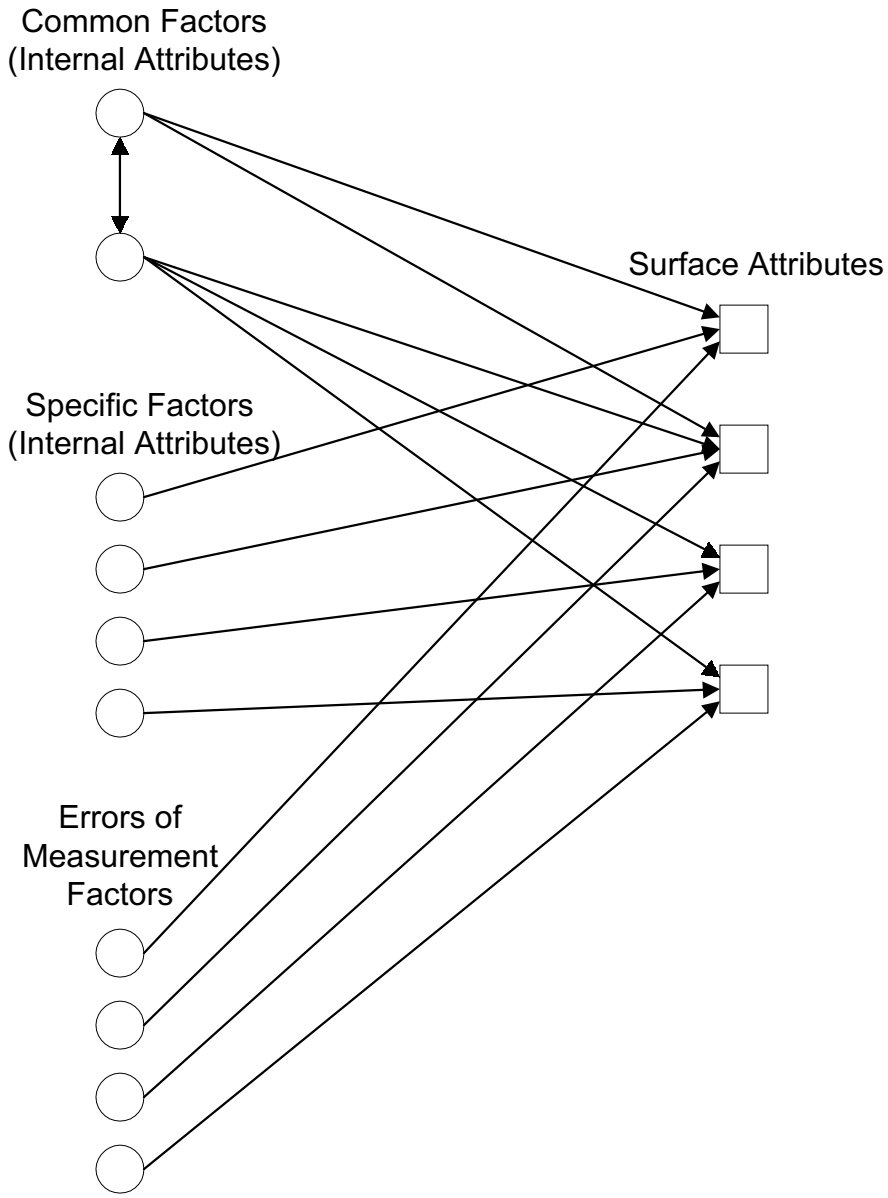

Figure 1.1: Basic Schemata for Common Factor Theory

Below the common factors in Figure 1.1 are shown four specific factors, which also represent internal attributes. There is one specific factor for each surface attribute. Note that each specific factor is shown as influencing only a single surface attribute. Below the surface attributes in Figure 1.1 are shown four errors of measurement factors. These factors are not designated as internal attributes since they do not correspond to unobserved characteristics of the individuals, but rather arise from unsystematic, transient events. As shown in Figure 1.1, there is one errors of measurement factor for each surface attribute, with each such factor influencing only a single surface attribute.

Figure 1.1 then illustrates a number of fundamental points. Considering each surface attribute separately, the variation in a given surface attribute is seen as arising from the effects of common, specific, and errors of measurement factors. As described earlier, the variance arising from the common factors is termed the communality of a surface attribute; the part of the variance on a surface attribute arising from the specific factor is termed the specificity of the surface attribute; and the part of the variance due to errors of measurement is termed the error of measurement variance of the surface attribute. Thus, Figure 1.1 indicates how the variance on a surface attribute may be partitioned into three sources: communality, specificity, and error of measurement variance. Inclusion of errors of measurement within factor analytic theory is a very important point which is overlooked by many practitioners in factor analysis. With regard to the separate issue of relations between surface attributes, Figure 1.1 illustrates how the relation between any two surface attributes is a result of the degree to which those attributes are influenced by the same common factors.

The representation of common factor theory in Figure 1.1 is expanded in Figure 1.2 to show how the specific and errors of measurement factors can be viewed as combining to form unique factors. The common factor portion of Figure 1.2 is the same as that shown in Figure 1.1. In Figure 1.2 four unique factor are shown. Each unique factor is represented as the combination of the specific and errors of measurement factor for a given surface attribute. In this view of common factor theory, each surface attribute is taken to be affected by some number of common factors and a single unique factor. According to this framework, the covariation of surface attributes is still attributed to the influences of common factors, and the variation in a given surface attribute is seen as arising from the effects of common and unique factors. In terms of variance, it was stated earlier that specificity and error of measurement variance are combined to form the uniqueness of the surface attribute. Thus, Figure 1.2 indicates how the variance on each surface attribute may be partitioned into communality and uniqueness.

The presentation of basic factor analytic theory is often limited to the view represented by Figures 1.1 and 1.2. However, it is important to develop an expanded view which considers the internal attribute in more depth. This expanded view will provide a further understanding of the

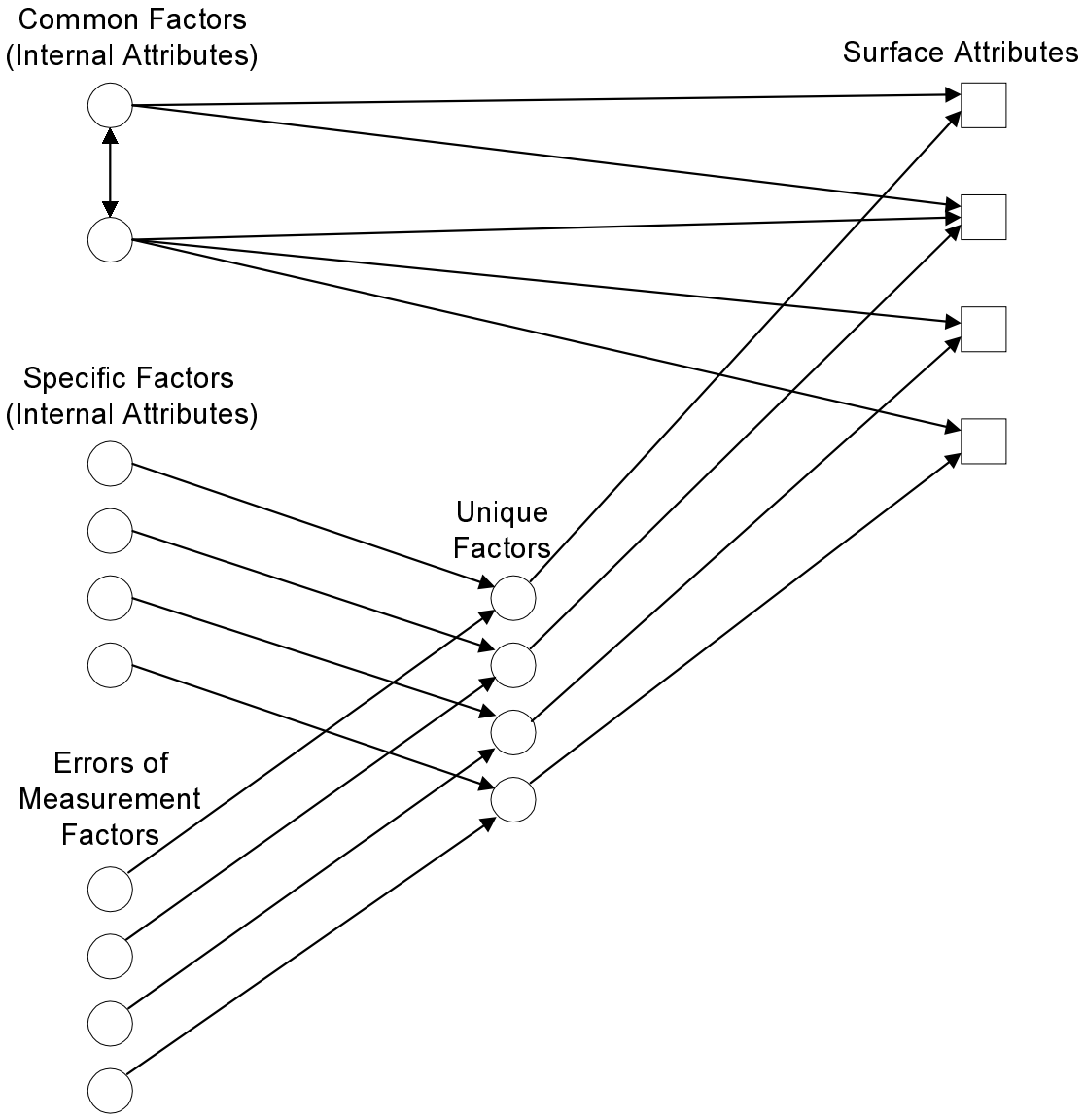

Figure 1.2: Expanded Schemata for Common Factor Theory
Showing Unique Factors

nature of these attributes and the influences which give rise to them, and thus yield a fuller conceptualization of the theory. Figure 1.3 provides an expanded path diagram which illustrates a more complete representation of factor analytic theory. The right half of Figure 1.3 is identical to Figure 1.2, with one exception: the intercorrelation of the two common factors no longer is represented explicitly by a bi-directional path. The reason for this change is given in the next paragraph.

The far left side of Figure 1.3 represents the sources and dynamics which give rise to common, specific, and errors of measurement factors. This portion of Figure 1.3 contains two different types of influences. At the top is shown a group of effects designated "Basic Characteristics and External Influences." These are intended to represent (a) basic characteristics of individuals, such as innate abilities, age, etc., and (b) external influences, such as education, experience, cultural effects, etc. These characteristics are shown as influencing both the common and specific factors. The common factors can thus be thought of as constructs which are made up of combinations of basic characteristics and experiential elements of individuals. The common factors are reflections of the coalescing of such elements as they influence surface attributes. The broad arrows representing the effects of the basic characteristics and external influences on the common factors are intended to represent this general coalescing, which might include nonlinear relations. As implied by Figure 1.3, different common factors may be influenced by some of the same basic characteristics and external influences. Such a phenomenon would give rise to correlated common factors. Thus, correlations among common factors are a result of the influences represented in Figure 1.3, and so need not be represented explicitly by bi-directional paths. This is the same reasoning which eliminates the need to represent correlations among surface attributes explicitly, since such correlations are accounted for by other effects represented in the figures. In general in path diagrams, relations which are accounted for by effects in the model are not represented explicitly.

The basic characteristics and external influences also are shown as affecting specific factors. This simply means that, just as those effects can coalesce to yield a common factor (which affects more than one surface attribute), they can also combine to yield a factor which has a systematic effect on only one surface attribute.

In the lower left section of Figure 1.3 are shown "Transient Effects." These are effects which influence the performance of the individuals on the surface attributes, but which are unstable, or transient. The most common example of such effects would be errors of measurement, which are unsystematic and transient, yet influence observed measures. Each surface attribute will be influenced by some transient effects; thus, Figure 1.3 contains a transient effect for each surface attribute. These effects can be conceived of as giving rise to the errors of measurement factors, as shown in Figure 1.3.

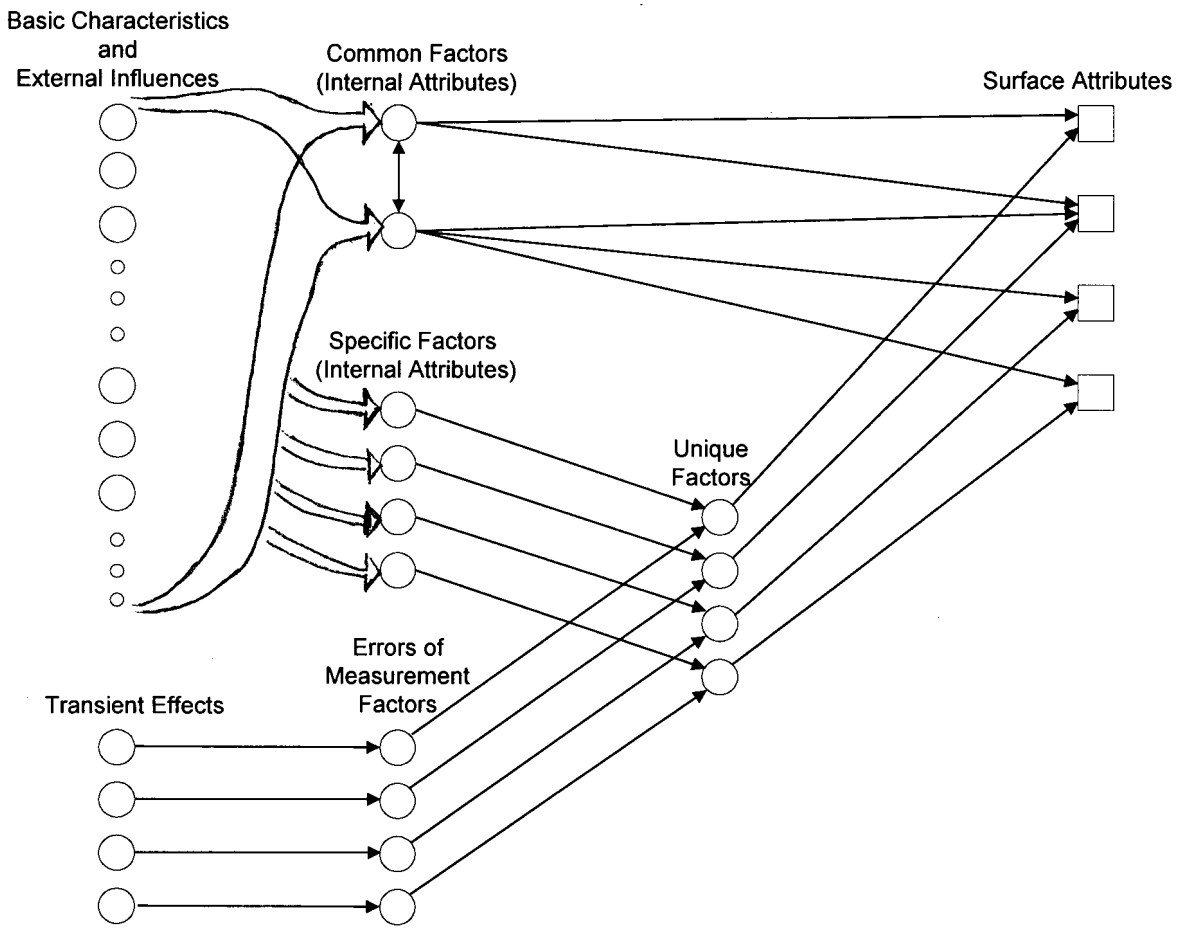

Figure 1.3: Full Schemata for Common Factor Theory

It is important to keep in mind that the factor analytic theory as described here and represented in Figure 1.3 is most assuredly a theory. It is a theory about the influences and dynamics which give rise to observed variation and covariation on surface attributes. Whenever a theory of any kind is used in science, a central issue becomes the degree of correspondence between the theory and the real world. That in, how well does the theory account for the observed phenomena under study? The degree of success of factor analytic theory will vary from one application to another, and it should never be assumed that the theory is a correct representation of the real world in any particular study. The issue of goodness of fit of the theory to the observed data will have to be considered as a routine part of any application of factor analysis.

In terms of the conceptualization of the theory developed here, an interesting way to view the issue of correspondence to the real world is to conceive of each surface attribute as being composed of two parts. One part arises from formal factor analytic theory, as portrayed in Figure 1.3. The other part is the remainder, which cannot be represented in the given framework. In terms of the theory, the former part is systematic and can be accounted for, while the latter part is unsystematic and cannot be accounted for. The unsystematic portion may arise from any number of sources, such as a multitude of minor common factors (i.e., common factors which have real effects on the surface attributes, but whose effects are very minor compared to the major common factors represented in Figure 1.3), and nonlinear effects in places where the model postulates linear effects. Regardless of the source, this lack of perfect correspondence between the theory and the real world must be acknowledged and accepted as an inherent aspect of the use of factor analysis. Furthermore, as with any scientific theory, the degree of fit of the theory to observed data must be evaluated in practice.

1.2. Objectives of Factor Analysis

Of primary interest in factor analytic theory is the nature of the common factors. Therefore, the primary objective in factor analysis methodology is to determine the number and nature of those factors, and the pattern of their influences on the surface attributes.

In very general terms, this is accomplished by making use of the implications of factor analytic theory as discussed earlier. Since the influence of common factors on surface attributes gives rise to correlations among those attributes, the observation of correlated surface attributes can be taken as an indication that common factors are operating. Given correlations among a number of surface attributes drawn from a particular domain, it may be possible to determine, from the pattern of correlations, the number of common factors operating and something about their nature. The methodology of factor analysis is designed to achieve this goal. Based on the intercorrelations among surface attributes, it is possible to estimate the number of common factors and to obtain numerical coefficients representing the degree of effect of each common

factor on each surface attribute. Based on this information, it will be seen that it is possible to attempt to interpret the nature of the common factors themselves, as well as to obtain estimates of the amount of common and unique variance in each surface attribute. In practice, when factor analysis is conducted on a single battery in a single sample with no further information, only the common and unique variances can be determined. That is, it will not be possible to partition the uniqueness into portions due to specific and errors of measurement factors without further information about the reliabilities of the measures in the battery. However, this is a tangential issue in factor analysis. The central objective of factor analysis is focused on the common factors -- to gain an understanding of their nature and the dynamics of their relationships to each other and to the surface attributes.

We wish to emphasize that we do not view this as an attempt to determine measures for the individuals on the common factors. The purpose of factor analysis is not to measure individual differences on the common factors, or to obtain such measures for further analysis. Such measures are referred to as common factor scores. Though it is possible to estimate such scores for individuals on the common factors, and though we will cover this issue in detail in Chapters 15 and 16, this endeavor should not be viewed as a primary objective of factor analysis. Rather, the goal is to account for variation and covariation on the surface attributes by identifying relevant common factors. This is the end toward which factor analytic theory and methodology are primarily directed.

An important final point is that it generally is not possible to achieve these objectives in a single study. Rather, a succession of studies, each one building on knowledge gained from preceding studies, is necessary. The purpose of the several studies should be enhancement of the understanding of the internal attributes. Repeated studies with the same battery of measured attributes using new samples of entities may be beneficial but are not sufficient. Rather, the battery of measured attributes should be changed in successive studies by some elimination of uninformative attributes as well as, most importantly, by the addition of new measures constructed from interpretation of the factors. A desirable feature would be the prediction of the factorial composition of these new measures. For instance, attempts might be made to construct new measures which would load in a predictable fashion on two or three factors. These predictions, then, could be checked in a new study. Each success in construction of new measures with predictable characteristics would strengthen the interpretation of the factors. Even when predictions are not borne out, the interpretation of the factors could be revised to a stronger position. Following such a plan of a succession of studies would lead to better understanding of the domain being investigated. Factor analysis research must not be viewed as involving single studies. Further discussion of this view of factor analysis research is provided in Chapter 6.

1.3. Illustration of Factor Analysis

In an effort to clarify the concepts and objectives discussed to this point, a simple illustration of an application of factor analysis will be presented. In presenting this illustration, we will avoid explicit discussion of methodology and focus simply on the potential for factor analysis to achieve the objectives described in the previous section. To accomplish this, it is necessary first to describe (a) the type of data to which factor analysis is applied, and (b) the most important results which are obtained from a factor analysis and how those results are interpreted. This discussion is somewhat oversimplified, but will serve the purpose of introducing the reader to these points and providing a framework for the subsequent illustration.

In most applications of factor analysis, the data analyzed consist of sample correlations or covariances among the surface attributes. Issues differentiating the analysis of correlations vs. covariances will be discussed in subsequent chapters. The present illustration employs correlations. These typically are presented in the form of a correlation matrix, where the rows and the columns represent the surface attributes and each element is a correlation for a given pair of attributes. When such data are subjected to factor analysis a considerable amount of information can be generated. For present purposes, the most important results consist of common factor weights, common factor intercorrelations, and communalities.

The common factor weights also are referred to as common factor loadings. These loadings normally are arranged in the form of a matrix, where the rows represent the surface attributes, the columns represent the common factors, and each element is a factor loading representing the effect of a given factor on a given surface attribute. In Figures 1.1, 1.2, and 1.3, the loadings can be thought of as numerical coefficients corresponding to the directional paths connecting the common factors to the surface attributes. The loadings provide the basis for attempting to interpret the nature of the common factors. In very basic terms, the substantive meaning of each factor is interpreted by examining the coefficients in the corresponding column of the factor loading matrix. High loadings represent surface attributes which are influenced strongly by the factor, and low loadings represent surface attributes which are influenced weakly by the factor. By examining the pattern of loadings for a given factor, the researcher attempts to identify a construct whose effects on the surface attributes correspond to the pattern of loadings. That is, the construct should be one that plays a strong role in the surface attributes which have high loadings and a weak role in those that have low loadings. This process will be illustrated below.

The other fundamental results of a factor analysis mentioned above are common factor intercorrelations and communalities. The former provide measures of the degree of relationship among the constructs which have been identified, and the latter indicate how much of the variance in each surface attribute is accounted for by the common factors.

Our example involves selected data from a study by Thurstone and Thurstone (1941) which consisted of analyses of two large batteries of tests given to seventh and eighth grade students in Chicago. For the current example, we will consider data on nine of these tests from the eighth grade students only. The sample consisted of 710 students. A brief description of the nine tests follows:

(1) Addition -- Items involved adding columns of numbers.

(2) Multiplication -- Items involved multiplying two numbers.

(3) Three-Higher -- Each item consisted of a series of numbers; the task was to identify each number in the series that was exactly three more than the number just before it.

(4) Figures -- Each item consisted of a series of figures, such as letters, rotated or reflected in various ways; the task was to identify each figure in the series which could be rotated to match the first figure.

(5) Cards -- Each item consisted of a series of cards representing geometric shapes, rotated or reflected in various ways; the task was to identify each card in the series which could be rotated to match the first card.

(6) Flags -- Each item consisted of two flags; the task was to determine whether the two flags were the same, within a simple rotation.

(7) Identical Numbers -- Each item consisted of a column of numbers; the task was to mark each number in the column that was identical to the first number.

(8) Faces -- Each item consisted of a set of three faces, where two of the three were identical; the task was to mark the one that was different.

(9) Mirror Reading -- Each item consisted of a typed word, followed by four other words which were printed backward; the task was to identify which of the four was the same as the original word.

The nine tests were administered to the sample of students, and correlations among the tests were obtained. The correlation matrix is shown in Table 1.1. Note that the nine tests correspond to surface attributes, and the primary objective of factor analysis is to identify internal attributes which account for the relationships among these tests. Factor analysis was applied to the correlation matrix in Table 1.1. Discussion of the exact methodology employed is not practical at this point. Suffice it to say that three common factors were identified, and the resulting factor loadings, communalities, and factor intercorrelations are presented in Table 1.2. One can attempt to interpret the substantive nature of the common factors by examining the factor loadings, as described above. Considering the first column of loadings in the factor matrix, it is clear that this factor has a strong influence on the first three tests, but essentially is absent in the remaining tests. Since the first three tests explicitly involve numerical calculations, while none of the remaining tests require such activity, this first factor can be interpreted as a numerical

Table 1.1

Correlations among Nine Mental Tests (Sample: 710 Eighth Grade Students)

Attribute	1	2	3	4	5	6	7	8	9
1. Addition	1.000								
2. Multiplication	.499	1.000							
3. Three-Higher	.394	.436	1.000						
4. Figures	.097	.007	.292	1.000					
5. Cards	.126	.023	.307	.621	1.000				
6. Flags	.085	.086	.328	.510	.623	1.000			
7. Ident. Number	.284	.467	.291	.044	.114	.086	1.000		
8. Faces	.152	.235	.309	.319	.376	.337	.393	1.000	
9.Mirror Reading	.232	.307	.364	.213	.276	.271	.431	.489	1.000

Table 1.2

Factor Analysis Solution for Nine Mental Tests

Factor Weights				
	Factor			Communalities
Attribute	1	2	3	
1	.66	.05	-09	.38
2	.67	-.05	.11	.53
3	.52	.32	.01	.43
4	.00	.72	-.04	.51
5	-.02	.80	.03	.63
6	.02	.71	.02	.51
7	.24	.03	.49	.44
8	-.06	.40	.53	.45
9	.08	.28	.52	.44

Factor Intercorrelations

Factor	1	2	3
1	1.00		
2	.15	1.00	
3	.57	.11	1.00

calculations factor. Considering the second column of factor loadings, it is clear that this factor has a strong influence on tests 4, 5, and 6, and a moderate to weak influence on the remaining tests. Since those three tests all require an ability to recognize relations between shapes which are in different spatial orientations, this factor could be thought of as spatial relations. This interpretation is consistent with the finding that this factor has a mild effect on some of the other tests (e.g., mirror reading). The third factor is characterized by high loadings on the last three tests and very low loadings on the other six. The last three tests all require the subject to rapidly examine simple stimuli and identify those which are the same. This factor could be thought of as a perceptual speed factor.

The factor intercorrelations shown in Table 1.2 reveal that these three factors are not completely independent. Specifically, there is a substantial positive correlation between the numerical calculations and perceptual speed factors. The spatial relations factor is correlated only weakly with the other two. The communalities show how such of the variance in each of the tests is accounted for by the three factors. These values reveal that each of the tests has at least a moderate amount of variance which is not accounted for by the common factors; i.e., which is due to unique factors.

This illustration demonstrates the potential for factor analysis to achieve its objective of identifying internal attributes which account for variation and correlation on surface attributes. In the demonstration, three common factors are found which can be shown to account quite well for the relationships among the nine surface attributes, as well as for a portion of the variance on each surface attribute. Furthermore, these factors were found to be easily interpretable through their relationships with the surface attributes. As a result, the factors can be viewed as substantively meaningful constructs which have been found to be related to surface attributes in a systematic way. These constructs and their relationships to each other and to the surface attributes serve to account for observed phenomena; i.e., the complex pattern of variation and covariation of the surface attributes. However, this illustration should not be viewed as an isolated study which achieves all of the objectives of factor analysis. As noted at the end of the previous section, a study such as this would be one in a series of studies. Preceding studies led to the development of the battery of nine tests which represented the three common factors identified. Subsequent studies could be conducted in which new tests could be added to the battery so as to further verify and refine the understanding of the internal attributes.

Finally, we wish to emphasize that we do not consider factors such as those identified in the illustration just presented to be equivalent to fundamental, unitary mental abilities. Factor analysis is not intended to identify factors which correspond to real, fundamental, underlying attributes of individuals. The factors are constructs which, as described in the previous section, can be viewed as a coalescing of basic characteristics and experiential elements of individuals.

This view of factors as constructs should not, however, be taken to imply that factors are artifacts. They are not. Factor analysis does not create factors, but rather reveals them based on patterns of intercorrelations among surface attributes. Each factor indicates the presence of some systematic influence operating on the surface attributes. The factors vary with respect to the degree with which they can be understood in substantive terms. The purpose of factor analysis methodology is to extract numerical information about these factors from the data, and the responsibility of the researcher is to make use of factor analytic theory and knowledge of the surface attributes to achieve an understanding of the meaning of the factors and the dynamics by which they operate.

CHAPTER 3
THE COMMON FACTOR MODEL IN THE POPULATION

From
Exploratory Factor Analysis
Ledyard R Tucker
and
Robert C. MacCallum

CHAPTER 3
THE COMMON FACTOR MODEL IN THE POPULATION

3.0. Introduction

In Chapter 1 we presented a conceptual, non-mathematical view of common factor theory. The purpose of the present chapter is to develop a detailed mathematical framework for this approach. Such a framework will serve the dual purpose of (a) providing an explicit representation of common factor theory as a mathematical model, so that its properties and implications can be understood; and (b) providing a basis for the solution of the problem, to be addressed in Chapter 7, of fitting the model to observed data.

The common factor model will be developed in this chapter in the context of a population. That is, we will express the model in terms of population parameters rather than sample statistics. Issues of sampling and parameter estimation will be treated in later chapters. So, for now, the reader should conceive of the case where the theoretical population of observations is available. The development of the model and the discussion of related topics and issues will require the use of rather extensive mathematical notation. In general, we will adhere to a notational system where population parameters are represented by Greek letters and italicized English letters. Sample statistics (though none are used in this chapter) will be presented by normal (un-italicized) English letters. A glossary of notation is presented for the reader's use in Appendix A. The approach taken in this mathematical presentation employs linear algebra, or matrix algebra. Readers must be comfortable with matrix terminology and manipulations in order to follow easily the developments to be presented. A basic review of matrix operations is presented in Appendix B. In addition, we will make use of a number of operations and basic theorems involving linear transformations of attributes. These operations and theorems are reviewed in Appendix C.

3.1. Algebraic Representation of the Common Factor Model

To begin, let us recall the basic concepts of common factor theory as presented in Chapter 1. For a given battery of surface attributes, there will exist a set of factors. These factors will be of three different types of common, specific, and error of measurement. The common factors are those that affect more than one attribute in the set, and specific and error of measurement factors each affect a single attribute. Each surface attribute is, to a degree, linearly dependent on these underlying factors. The variation in a surface attribute is accounted for in part by its dependence on common, specific, and error of measurement factors. The covariation between any two surface attributes is accounted for in part by their joint dependence on one or more of the same common factors. Since common factor theory cannot be expected to represent

the real world precisely, it is not expected that the variation and covariation of the surface attributes will be exactly explained by the underlying common, specific, and error of measurement factors.

To begin to express this theory in formal mathematical terms, let us first define n as the number of surface attributes to be observed. We next define a row vector $\underline{\boldsymbol{y}}$, of order *n*, as a generalized row vector of measures on the surface attributes. This generalized row vector can be thought of as containing measures on the surface attributes for any individual sampled from the population.

A critical point made in Chapter 1, and reiterated above, is that the common factor model is not expected to account precisely for the variation and covariation of the surface attributes in the population. Alternatively, it could be stated that each surface attribute can be conceived of as being composed of two parts: one part that is accounted for by common factor theory, and the remainder, which is not accounted for by the theory. In conjunction with this conception of the surface attributes, we define two more generalized row vectors. Vector $\underline{\boldsymbol{z}}$, of order *n*, contains that part of vector $\underline{\boldsymbol{y}}$ which is accounted for by common factor theory. Vector $\underline{\ddot{\boldsymbol{z}}}$, also of order *n*, contains that part of vector $\underline{\boldsymbol{y}}$ which is not accounted for by common factor theory. Given these definitions, the relationship among the three generalized vectors is given by:

$$\underline{\boldsymbol{y}} = \underline{\boldsymbol{z}} + \underline{\ddot{\boldsymbol{z}}} \tag{3.1}$$

This in turn implies that

$$\underline{\ddot{\boldsymbol{z}}} = \underline{\boldsymbol{y}} - \underline{\boldsymbol{z}} \tag{3.2}$$

Eq. (3.2) simply shows that the non-systematic portion of the surface attributes can be defined as the difference between the observed attributes and the systematic portion. In almost every discussion of factor analytic theory, no distinction is drawn between entries in $\underline{\boldsymbol{y}}$ and $\underline{\boldsymbol{z}}$. That is, the model is presented as exactly representing the surface attributes in the population. However, it is well understood that this will not be the case in the real world. The model, in fact, should be thought of as exactly representing the systematic portion of the surface attributes, as given in $\underline{\boldsymbol{z}}$, and only imperfectly representing the surface attributes themselves, as given in $\underline{\boldsymbol{y}}$. Therefore, we feel it is important to maintain explicitly the distinction between the surface attributes and the systematic portion of the surface attributes. This distinction will be seen to have important philosophical and practical consequences. For the purpose of maintaining efficient terminology, we will continue to refer to the attributes represented in $\underline{\boldsymbol{y}}$ as the surface attributes, and we will refer to the attributes in $\underline{\boldsymbol{z}}$, which represent that portion of the surface attributes that are

accounted for by the common factor model, as modeled attributes. In addition, we will refer to entries in $\ddot{\underline{z}}$ as errors of fit; again, these values represent that portion of the surface attributes not accounted for by the model.

As discussed earlier, factor analytic theory postulates the existence of underlying internal attributes, or factors. Let $\underline{x}$ represent a generalized row vector of measures on the factors. The vector $\underline{x}$ will be of order *m*, where *m* is the total number of factors of all three types. Thus, this vector contains the (unobservable) measures for an individual on the common, specific, and error of measurement factors. Next, let us define a matrix Ω containing weights which represent the effects of the factors on the modeled attribute. This matrix will be of order $n \times m$, where the rows represent the modeled attributes and the columns represent the factors. An entry ω_{jk} is a weight, analogous to a regression weight, representing the effects of factor *k* on modeled attribute *j*. Given these definitions, the common factor model can be represented simply as

$$\underline{z} = \underline{x}\Omega' \tag{3.3}$$

This equation simply defines each modeled attribute as a linear combination of the measures on the factors. This representation of the model can be stated in an expanded fashion by incorporating the distinction among common, specific and error of measurement factors. The vector $\underline{x}$ containing measures on the factors can be conceived of as containing measures on the types of factors. Let *r* be the number of common factors, and recall that *r* should be much less than *n*, the number of attributes. We then can define $\underline{x}_\beta$ as a row vector of order *r* containing measures on the common factors. In similar fashion, let $\underline{x}_\xi$ contain measures on the specific factors and let $\underline{x}_\varepsilon$ contain measures on the error of measurement factors. Both $\underline{x}_\xi$ and $\underline{x}_\varepsilon$ will be of order *n*, since there is one specific factor and one error of measurement factor for each attribute. For the present, the measures on all three types of factors will be defined as being given in standardized form. As a result of the way in which the three types of factors have been defined, the measures on the factors can be seen to have some interesting properties. In particular, since the specific factors represent systematic portions of each modeled attribute which are unique to each attribute, the specific factors must be uncorrrelated with each other. Similarly, since the error of measurement factors represent transient aspects of each modeled attribute, they also must be uncorrelated with each other. Finally, the definitions of these two types of factors also require that they be uncorrelated with each other and with the common factors. The vectors of measures on the three types of factors can be adjoined horizontally to form the super-vector $\underline{x}$, as follows :

$$\underline{x} = [\underline{x}_\beta, \underline{x}_\xi, \underline{x}_\varepsilon] \tag{3.4}$$

The order of $\underline{x}$, defined as *m* above, now can be seen more clearly. The vector $\underline{x}$ will contain

measures on the *r* common factors, the *n* specific factors, and the *n* error of measurement factors. Thus, the total number of factors, *m*, is given by

$$m = r + 2n \tag{3.5}$$

Just as the vector of measures on the factors can be seen as containing separate sections corresponding to the three types of factors, so also can the factor weight matrix $\boldsymbol{\Omega}$ be partitioned. We can define weight matrices for each of the three types of factors. Let $\boldsymbol{B}$ be a matrix of order $n \times r$, where the rows represent the modeled attributes, the columns represent the common factors, and each entry β_{jk} represents the weight for common factor *k* on modeled attribute *j*. Let $\boldsymbol{\Xi}$ be a matrix of order $n \times n$, where the rows represent the modeled attributes, the columns represent the specific factors, and each diagonal entry ξ_j represents the weight for specific factor *j* on modeled attribute *j*. Note that matrix $\boldsymbol{\Xi}$ will be diagonal, since each specific factor affects only one modeled attribute. Similarly, let $\boldsymbol{E}$ be an $n \times n$ matrix, with rows representing modeled attributes, columns representing error of measurement factors, and diagonal entries ϵ_j representing the weight for error of measurement factor *j* on modeled attribute *j*. This matrix of error of measurement factor weights must also be diagonal, since each error of measurement factor affects only one modeled attribute. These three matrices of factor weights then can be adjoined horizontally to form the full matrix $\boldsymbol{\Omega}$ of factor weights:

$$\boldsymbol{\Omega} = [\boldsymbol{B}, \boldsymbol{\Xi}, \boldsymbol{E}] \tag{3.6}$$

Thus, $\boldsymbol{\Omega}$ can be seen to be a super-matrix containing the three separate matrices of factor weights representing the three types of factors.

The representations of $\underline{\boldsymbol{x}}$ and $\boldsymbol{\Omega}$ given in Eqs. (3.4) and (3.6) respectively can be substituted into Eq. (3.3) to provide an expanded representation of the common factor model, as follows:

$$\underline{\boldsymbol{z}} = [\underline{\boldsymbol{x}}_{\boldsymbol{\beta}}, \underline{\boldsymbol{x}}_{\boldsymbol{\xi}}, \underline{\boldsymbol{x}}_{\boldsymbol{\varepsilon}}] \begin{bmatrix} \boldsymbol{B}' \\ \boldsymbol{\Xi}' \\ \boldsymbol{E}' \end{bmatrix} = \underline{\boldsymbol{x}}_{\boldsymbol{\beta}}\boldsymbol{B}' + \underline{\boldsymbol{x}}_{\boldsymbol{\xi}}\boldsymbol{\Xi}' + \underline{\boldsymbol{x}}_{\boldsymbol{\varepsilon}}\boldsymbol{E}' \tag{3.7}$$

This representation of the model shows more clearly that each modeled attribute is defined as a linear combination of the common, specific, and error of measurement factors. Furthermore, given that $\boldsymbol{\Xi}$ and $\boldsymbol{E}$ are diagonal, it can be seen that each modeled attribute is represented as a linear combination of *r* common factors, one specific factor, and one error of measurement factor.

An interesting aspect of this representation is that error of measurement factors contribute to the modeled attributes. Recognition of this fact should help the reader distinguish between errors of measurement, which give rise to error of measurement factors, and errors of fit, which

were defined in Eq. (3.2). Errors of measurement are represented explicitly in the model and contribute to the modeled attributes. Errors of fit represent the lack of fit of the model to the real world, or the lack of correspondence between the modeled attributes and the surface attributes.

As defined to this point, the common factor model represents the underlying structure of each modeled attribute. Recall from Chapter 1 that the general objective of factor analysis is to account for the variation and covariation of the attributes. To achieve this, it is necessary to express the model in terms of the variances and covariances of the modeled attributes. This can be accomplished fairly easily by making use of theorems presented in Appendix C. We are dealing with a situation where we have two sets of attributes (the modeled attributes and the factors) which are related via a linear transformation as defined in Eq. (3.3). In this case it is possible to define a functional relationship between the covariances of the two sets of attributes, as shown in Appendix C. Let us define $\mathbf{\Sigma}_{zz}$ as a population covariance matrix for the modeled attributes. The matrix $\mathbf{\Sigma}_{zz}$ will be of order $n \times n$, with diagonal entries σ_{jj} representing the population variance of modeled attribute *j*, and off-diagonal entries σ_{hj} representing the population covariance of modeled attributes *h* and *j*. Let us next define $\mathbf{\Sigma}_{xx}$ as a population covariance matrix for the factors. Matrix $\mathbf{\Sigma}_{xx}$ will be of order $m \times m$ and will contain entries σ_{kl} representing the population variances (on the diagonal) and covariances (off the diagonal) of the factors. The relationship between $\mathbf{\Sigma}_{zz}$ and $\mathbf{\Sigma}_{xx}$ can be obtained by direct application of Corollary 3 of Theorem 4 in Appendix C. This yields

$$\mathbf{\Sigma}_{zz} = \mathbf{\Omega}\,\mathbf{\Sigma}_{xx}\,\mathbf{\Omega}' \tag{3.8}$$

This equation represents the common factor model in terms of covariances of modeled attributes and factors. It states that the variances and covariances of the modeled attributes are a function of the factor weights and the variance and covariances of the factors.

It is informative to consider a more detailed version of Eq. (3.8), which can be obtained by incorporating the distinction among common, specific, and error of measurement factors. First, let us consider more closely the nature of $\mathbf{\Sigma}_{xx}$. It will be shown that this matrix has a very simple and interesting form. Recall that we are defining the measures on the factors to be in standardized form. Therefore, the matrix $\mathbf{\Sigma}_{xx}$ will take the form of a correlation matrix. Recall also that there exist three different types of factors (common, specific, and error of measurement), as defined in Eq. (3.6). This implies that $\mathbf{\Sigma}_{xx}$ can be thought of as a super-matrix, with sub-matrices containing correlations within or between types of factors as follows:

$$\mathbf{\Sigma}_{xx} = \begin{bmatrix} \mathbf{\Sigma}_{\beta\beta} & \mathbf{\Sigma}_{\beta\xi} & \mathbf{\Sigma}_{\beta\varepsilon} \\ \mathbf{\Sigma}_{\xi\beta} & \mathbf{\Sigma}_{\xi\xi} & \mathbf{\Sigma}_{\xi\varepsilon} \\ \mathbf{\Sigma}_{\varepsilon\beta} & \mathbf{\Sigma}_{\varepsilon\xi} & \mathbf{\Sigma}_{\varepsilon\varepsilon} \end{bmatrix} \tag{3.9}$$

For each sub-matrix in $\boldsymbol{\Sigma}_{xx}$, the first subscript represents the factors defined by the rows of the sub-matrix, and the second subscript represents the factors defined by the columns of the sub-matrix. For example, the sub-matrix $\boldsymbol{\Sigma}_{\xi\beta}$ will be of order $n \times r$ and will contain correlations between specific factors (rows) and common factors (columns). Given this representation of $\boldsymbol{\Sigma}_{xx}$, it is very useful to consider the nature of the sub-matrices. Considering the diagonal submatrices in Eq. (3.9), note that these contain correlations among the factors within each given type (common, specific, error of measurement). Since the common factors may be correlated with each other, let us define

$$\boldsymbol{\Phi} = \boldsymbol{\Sigma}_{\beta\beta} \tag{3.10}$$

Matrix $\boldsymbol{\Phi}$ will be of order $r \times r$ and will contain intercorrelations among the common factors. It was noted earlier that specific factors will be mutually uncorrelated, as will error of measurement factors. Therefore, it can be stated that

$$\boldsymbol{\Sigma}_{\xi\xi} = \boldsymbol{I} \tag{3.11}$$

and

$$\boldsymbol{\Sigma}_{\epsilon\epsilon} = \boldsymbol{I} \tag{3.12}$$

The identity matrices shown in Eqs. (3.11) and (3.12) will be of order n. Considering the off-diagonal sections of $\boldsymbol{\Sigma}_{xx}$ shown in Eq. (3.9), it can be seen that these submatrices will contain all zeroes. Each of these sub-matrices represents a matrix of intercorrelations among factors of two different types (common and specific, common and error of measurement, specific and error of measurement), and, as was explained earlier in this section, factors of different types are, by definition, uncorrelated in the population. Given these properties of the sub-matrices in Eq. (3.9), we can rewrite that equation as

$$\boldsymbol{\Sigma}_{xx} = \begin{bmatrix} \boldsymbol{\Phi} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{I} \end{bmatrix} \tag{3.13}$$

A very important representation of the common factor model can now be obtained by substitution of the super-matrix form of $\boldsymbol{\Omega}$, from Eq. (3.6), and $\boldsymbol{\Sigma}_{xx}$, from Eq. (3.13), into the model as shown in Eq. (3.8). This yields

$$\boldsymbol{\Sigma}_{zz} = [\boldsymbol{B}, \boldsymbol{\Xi}, \boldsymbol{E}] \begin{bmatrix} \boldsymbol{\Phi} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{B}' \\ \boldsymbol{\Xi}' \\ \boldsymbol{E}' \end{bmatrix} \tag{3.14}$$

$$= \boldsymbol{B\Phi B'} + \boldsymbol{\Xi^2} + \boldsymbol{E^2} \tag{3.15}$$

This equation shows that the population covariances for the modeled attributes are functions of the weights and intercorrelatlons for the common factors, the squared specific factor weights, and the squared error of measurement factor weights. When the common factor model is written in this form, the mathematical representation of the model can be seen to be consistent with the basic principles of common factor theory discussed in Chapter 1. In particular, it was stated that the common factors alone account for the relationships among the modeled attributes. This is revealed in Eq. (3.15) by the fact that the off-diagonal elements of $\boldsymbol{\Sigma}_{zz}$ are a function of only the parameters in the term $\boldsymbol{B\Phi B'}$. Thus, the common factor weights and intercorrelations combine to account for the relationships among the modeled attributes. Another principle of common factor theory is that the variation in the modeled attributes is explained by all three types of factors. In mathematical terms, this is seen in Eq. (3.15) in that the diagonal elements of $\boldsymbol{\Sigma}_{zz}$ are a function of the common factor parameters in $\boldsymbol{B\Phi B'}$, as well as the diagonal elements in $\boldsymbol{\Xi^2}$ and $\boldsymbol{E^2}$. Thus, since $\boldsymbol{\Xi^2}$ and $\boldsymbol{E^2}$are diagonal, the specific and error of measurement factors can be seen as contributing only to the variances of the modeled attributes (i.e., the diagonal elements of $\boldsymbol{\Sigma}_{zz}$), and not the covariances.

This view implies that it should be possible to define separate components of the variance of each modeled attribute; i.e., portions due to common, specific, and error of measurement factors. This is, in fact, quite simple and will be seen to provide useful information. Let us define a diagonal matrix $\boldsymbol{H^2}$as

$$\boldsymbol{H^2} = \boldsymbol{Diag}(\boldsymbol{B\Phi B'}) \tag{3.16}$$

Matrix $\boldsymbol{H^2}$ will be diagonal and of order $n \times n$, with diagonal elements h_j^2 representing the amount of variance in modeled attribute *j* that is accounted for by the common factors. This will be referred to as the common variance of modeled attribute *j*. It can be seen that

$$\boldsymbol{Diag}(\boldsymbol{\Sigma}_{zz}) = \boldsymbol{H^2} + \boldsymbol{\Xi^2} + \boldsymbol{E^2} \tag{3.17}$$

In terms of the variance of a given modeled attribute *j*, this implies that

$$\boldsymbol{\sigma_{jj}} = \boldsymbol{h_j^2} + \boldsymbol{\xi_j^2} + \boldsymbol{\epsilon_j^2} \tag{3.18}$$

Thus, the variance of each modeled attribute is made up of a common portion, a specific portion, and an error of measurement portion. It is interesting to convert these variance components into proportions by dividing each side of Eq. (3.18) by σ_{jj}. This yields

$$\sigma_{jj}/\sigma_{jj} = h_j^2/\sigma_{jj} + \xi_j^2/\sigma_{jj} + \boldsymbol{\epsilon_j^2/\sigma_{jj} = 1} \tag{3.19}$$

If we define

$$\widetilde{\boldsymbol{h}_j^2} = \boldsymbol{h}_j^2 / \boldsymbol{\sigma}_{jj} \tag{3.20}$$

$$\widetilde{\xi_j^2} = \boldsymbol{\xi}_j^2 / \boldsymbol{\sigma}_{jj} \tag{3.21}$$

$$\widetilde{\boldsymbol{\epsilon}_j^2} = \boldsymbol{\epsilon}_j^2 / \boldsymbol{\sigma}_{jj} \tag{3.22}$$

we then can note that for any modeled attribute, we obtain

$$\widetilde{\boldsymbol{h}_j^2} + \widetilde{\boldsymbol{\xi}_j^2} + \widetilde{\boldsymbol{\epsilon}_j^2} = \mathbf{1} \tag{3.22}$$

These three proportions of variance are important characteristics of the modeled attributes. The first, $\widetilde{h_j^2}$, represents the proportion of variance in modeled attribute *j* that is due to the common factors. This quantity is called the communality of modeled attribute *j*. The second, $\widetilde{\xi_j^2}$, represents the proportion of variance in modeled attribute *j* that is due to the specific factor for that attribute. This quantity is called the specificity for modeled attribute *j*. The third, $\widetilde{\epsilon_j^2}$, represents the proportion of variance in modeled attribute *j* that is due to the error factor for that attribute This quantity is called the error of measurement variance for modeled attribute *j*. Note that we can define diagonal matrices $\widetilde{\boldsymbol{H}^2}$, $\widetilde{\boldsymbol{\Xi}^2}$, and $\widetilde{\boldsymbol{E}^2}$, containing these proportions of variance for the *n* modeled attributes. Let us define a diagonal matrix $[\boldsymbol{\Sigma}_d]_{zz}$ containing the variances of the modeled attributes. That is,

$$[\boldsymbol{\Sigma}_d]_{zz} = \boldsymbol{Diag}\left(\boldsymbol{\Sigma}_{zz}\right) \tag{3.24}$$

Following Eqs. (3.20)-(3.22), we then can define matrices $\widetilde{\boldsymbol{H}^2}$, $\widetilde{\boldsymbol{\Xi}^2}$, and $\widetilde{\boldsymbol{E}^2}$ as follows:

$$\widetilde{\boldsymbol{H}^2} = \boldsymbol{H}^2 [\boldsymbol{\Sigma}_d]_{zz}^{-1} \tag{3.25}$$

$$\widetilde{\boldsymbol{\Xi}^2} = \boldsymbol{\Xi}^2 [\boldsymbol{\Sigma}_d]_{zz}^{-1} \tag{3.26}$$

$$\widetilde{\boldsymbol{E}^2} = \boldsymbol{E}^2 [\boldsymbol{\Sigma}_d]_{zz}^{-1} \tag{3.27}$$

According to Eq. (3.23), the sum of these matrices would be an identity matrix:

$$\widetilde{\boldsymbol{H}^2} + \widetilde{\boldsymbol{\Xi}^2} + \widetilde{\boldsymbol{E}^2} = \boldsymbol{I} \tag{3.28}$$

There are two interesting variance coefficients which can be derived from the three types of variance defined in the common factor model. The first is the reliability of a modeled attribute. Reliability is defined in mental test theory as unity minus the error of measurement variance for standardized measures. In the present context, this can be written

$$r_j = 1 - \widetilde{\xi}_j^2 \tag{3.29}$$

where r_j is the reliability of modeled attribute *j*. Based on Eq. (3.23), we can rewrite Eq. (3.29) as

$$r_j = \widetilde{h}_j^2 + \widetilde{\xi}_j^2 \tag{3.30}$$

This equation implies that, in factor analytic terms, reliability is the sum of the variances due to the common factors and the specific factor, or, in other words, the sum of the communality and specificity. Since reliability is defined as true score variance, this result means that true score variance can be represented as the sum of the two types of systematic variance in the common factor model: that due to the common factors and that due to the specific factor.

An alternative combination of components of variance is to sum the specific and error portions for each modeled attribute. This sum would yield a portion which is called unique variance; i.e., that due to factors (both systematic and error of measurement) which affect only one attribute. In unstandardized form this would be written

$$\boldsymbol{U^2} = \boldsymbol{\Xi^2} + \boldsymbol{E^2} \tag{3.31}$$

where $\boldsymbol{U^2}$ is a diagonal matrix with entries u_j^2 representing the variance in modeled attribute *j* due to specific and error of measurement factors. In terms of the variance expressed as proportions, we would write

$$\widetilde{\boldsymbol{U}}^{\boldsymbol{2}} = \widetilde{\boldsymbol{\Xi}}^{\boldsymbol{2}} + \widetilde{\boldsymbol{E}}^{\boldsymbol{2}} \tag{3.32}$$

where $\widetilde{\boldsymbol{U}}^{\boldsymbol{2}}$ is a diagonal matrix with entries $\widetilde{u}_j^2$ representing the proportion of variance in modeled attribute *j* due to specific and error of measurement factors. This proportion is called the uniqueness of modeled attribute *j*, and is defined as the sum of the specificity and error of measurement variance of that attribute. The relation between the uniquenesses of the modeled attributes, defined in Eq. (3.32), and the unique variances, defined in Eq. (3.31), can be obtained by substitution from Eqs. (3.26) and (3.27) into Eq. (3.32). This yields the following:

$$\widetilde{\boldsymbol{U}}^{\boldsymbol{2}} = \boldsymbol{U^2}[\boldsymbol{\Sigma}_d]_{zz}^{-1} \tag{3.33}$$

This relation follows those given in Eqs. (3.25)-(3.27) for other variance components and merely represents the conversion of unique variances into proportions called uniqueness. A final interesting relation is obtained by substituting from Eq. (3.32) into Eq. (3.28), yielding

$$\widetilde{\boldsymbol{H}}^2 + \widetilde{\boldsymbol{U}}^2 = \boldsymbol{I} \tag{3.34}$$

This equation simply implies that the communality plus the uniqueness for each modeled attribute will equal unity. In other words, the variance of a modeled attribute can be viewed as arising from two sources: (a) the common factors, or influence which are common to other attributes in the battery; and (b) the unique factors, or influences which are unique to the given attribute.

A most important equation is produced by substituting from Eq. (3.31) into the expression for the common factor model given in Eq. (3. 15). This yields

$$\boldsymbol{\Sigma}_{zz} = \boldsymbol{B}\boldsymbol{\Phi}\boldsymbol{B}' + \boldsymbol{U}^2 \tag{3.35}$$

Alternatively, we could write

$$\boldsymbol{\Sigma}_{zz} - \boldsymbol{U}^2 = \boldsymbol{B}\boldsymbol{\Phi}\boldsymbol{B}' \tag{3.36}$$

In either form, these equations provide a very fundamental expression of the common factor model. They show the functional relationship between the population variances and covariances of the modeled attributes, given in $\boldsymbol{\Sigma}_{zz}$, and critical parameters of the model contained in $\boldsymbol{B}$, the matrix of common factor weights, $\boldsymbol{\Phi}$, the matrix of common factor intercorrelations, and $\boldsymbol{U}^2$, the diagonal matrix of unique variances. As will be seen in subsequent chapters, this expression of the model provides a basis for some procedures for estimating model parameters. Thus, Eq. (3.35) should be viewed as a fundamental equation of common factor theory, expressing the common factor model in terms of the variance and covariances of the modeled attributes. It is important to understand, however, that this version of the model is a derived statement, i.e., derived from the model as given in Eq. (3.3), which defined the structure of the modeled attributes themselves. That is, based on the model defined in Eq. (3.3), the model in Eq. (3.35), defining the structure of the covariances of the modeled attributes as a function of correlated common factors, is derived. Correlated common factors are also called oblique common factors, and the model represented by Eq. (3.35) is called the oblique common factor model.

It is interesting and useful to recognize that the model given by Eq. (3.35) could have been derived based on definitions of $\underline{\boldsymbol{x}}$ and $\boldsymbol{\Omega}$ slightly different than those given in Eqs. (3.4) and (3.6), respectively. This alternative approach involves combining the specific and error of measurement factor portions of the model into a unique factor portion at the outset. That is, rather than define $\underline{\boldsymbol{x}}$ as containing scores on common, specific, and error of measurement factors,

as in Eq. (3.4), let us define it as containing scores on the r common factors and n unique factors. There would be a unique factor for each attribute, representing that portion of the attribute not accounted for by the common factors. To differentiate this representation from that given above, let us define a vector $\underline{\boldsymbol{x}}_{\beta\mu}$ whose contents are given by

$$\underline{\boldsymbol{x}}_{\beta\mu} = [\ \underline{\boldsymbol{x}}_{\beta}\ ,\ \underline{\boldsymbol{x}}_{\mu}\ ] \tag{3.37}$$

where $\underline{\boldsymbol{x}}_{\beta}$ and $\underline{\boldsymbol{x}}_{\mu}$ are vectors of scores on common and unique factors, respectively. In a similar fashion, we could define a factor weight matrix $\boldsymbol{\Omega}_{\beta u}$whose contents are given by

$$\boldsymbol{\Omega}_{\beta\mu} = [\,\boldsymbol{B},\,\boldsymbol{U}\,] \tag{3.38}$$

where $\boldsymbol{B}$ is the n x r matrix of common factor weights and $\boldsymbol{U}$ is an n x n diagonal matrix of unique factor weights. Following the form of Eq. (3.3), the common factor model could then be written as

$$\begin{aligned}\underline{\mathbf{z}} &= \underline{\boldsymbol{x}}_{\beta\mu}\boldsymbol{\Omega}'_{\beta u} \\ &= [\underline{\boldsymbol{x}}_{\beta}, \underline{\boldsymbol{x}}_{\mu}]\begin{bmatrix}\boldsymbol{B}' \\ \boldsymbol{U}'\end{bmatrix} \\ &= \underline{\boldsymbol{x}}_{\beta}\boldsymbol{B}' + \underline{\boldsymbol{x}}_{\mu}\boldsymbol{U}' \end{aligned} \tag{3.39}$$

This model represents the modeled attributes as linear combinations of common and unique factors. The model then could be stated in terms of covariances of modeled attributes, following the form of Eq. (3.8). This requires that we first define a covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{x}_{\beta\mu}\boldsymbol{x}_{\beta\mu}}$ for the common and unique factors. This matrix would take the form

$$\boldsymbol{\Sigma}_{\boldsymbol{x}_{\beta u}\boldsymbol{x}_{\beta u}} = \begin{bmatrix}\boldsymbol{\Sigma}_{\beta\beta} & \boldsymbol{\Sigma}_{\beta\mu} \\ \boldsymbol{\Sigma}_{\mu\beta} & \boldsymbol{\Sigma}_{\mu\mu}\end{bmatrix} = \begin{bmatrix}\boldsymbol{\Phi} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}\end{bmatrix} \tag{3.40}$$

since we define all factors as being in standardized form and since unique factors would, by definition, be uncorrelated with each other and with the common factors. The covariance form of the model then becomes

$$\begin{aligned}\boldsymbol{\Sigma}_{zz} &= \boldsymbol{\Omega}_{\beta\mu}\boldsymbol{\Sigma}_{\boldsymbol{x}_{\beta\mu}\boldsymbol{x}_{\beta\mu}}\boldsymbol{\Omega}'_{\beta\mu} \\ &= [\boldsymbol{B}, \boldsymbol{U}]\begin{bmatrix}\boldsymbol{\Phi} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{B}' \\ \boldsymbol{U}'\end{bmatrix} \\ &= \boldsymbol{B}\boldsymbol{\Phi}\boldsymbol{B}' + \boldsymbol{U}^2 \end{aligned} \tag{3.41}$$

Obviously, this expression is identical to that given in Eq. (3.35), thus verifying that these alternative definitions of the factor score vector and factor weight matrix yield the same oblique common factor model.

3.2. The Common Factor Model with Uncorrelated Common Factors

The common factor model as presented in the previous section represents the case where the common factors were allowed to be intercorrelated. Recall that the correlations among the common factors are contained in matrix $\boldsymbol{\Phi}$, defined in Eq. (3.10) and playing a prominent role in the model given by Eq. (3.35). Let us now consider a special case of this model, where the common factors are defined as being mutually uncorrelated. It will be shown that such a solution can be obtained from a solution with correlated common factors.

Suppose that a solution has been obtained in terms of correlated common factors, and that matrices $\boldsymbol{B}$ and $\boldsymbol{\Phi}$ are given. Let $\boldsymbol{T}$ be any non-singular matrix of order $r \times r$ such that

$$\boldsymbol{\Phi} = \boldsymbol{T}\boldsymbol{T}' \tag{3.42}$$

As will be seen as a result of procedures to be discussed in Chapter 9, it is routinely the case that, for a given $\boldsymbol{\Phi}$, an infinite number of such matrices $\boldsymbol{T}$ will exist. Substituting from Eq. (3.42) into Eq. (3.36) yields

$$\boldsymbol{\Sigma}_{zz} - \boldsymbol{U}^2 = \boldsymbol{B}\boldsymbol{T}\boldsymbol{T}'\boldsymbol{B}' \tag{3.43}$$

If we define a matrix $\boldsymbol{A}$ as

$$\boldsymbol{A} = \boldsymbol{B}\boldsymbol{T} \tag{3.44}$$

then Eq. (3.43) can be rewritten as

$$\boldsymbol{\Sigma}_{zz} - \boldsymbol{U}^2 = \boldsymbol{A}\boldsymbol{A}' \tag{3.45}$$

if we then rewrite Eq. (3.45) as

$$\boldsymbol{\Sigma}_{zz} - \boldsymbol{U}^2 = \boldsymbol{A}\boldsymbol{I}\boldsymbol{A}' \tag{3.46}$$

It becomes clear that we have "transformed" the original solution containing correlated common factors into a new solution containing uncorrelated, or orthogonal, common factors. In the orthogonal solution, matrix $\boldsymbol{A}$ is an $n \times r$ matrix of common factor weights; entries a_{jk} represent the weight for modeled attribute *j* on factor *k*. The $\boldsymbol{\Phi}$ matrix in the model as given in Eq. (3.36) has become an identity matrix of order *r*, indicating that the factors represented by the $\boldsymbol{A}$ matrix are uncorrelated. Eq. (3.45) can be thought of as representing the orthogonal common factor model. Thus, such an orthogonal solution can be obtained by transformation of the original oblique solution by defining a non-singular matrix $\boldsymbol{T}$ which satisfies Eq. (3.42), and then obtaining the matrix of factor weights via Eq. (3.44).

Let us now consider the opposite case. That is, instead of being given an oblique solution and transforming it into an orthogonal solution, we are given an orthogonal solution, represented by a matrix $\boldsymbol{A}$ which satisfies Eq. (3.45). It would be straightforward to transform that solution

into an oblique solution. Let $\boldsymbol{T}$ be any non-singular matrix with elements in each row scaled so that the sum of squares in each row is unity. That is,

$$\boldsymbol{Diag(TT')} = \boldsymbol{I} \tag{3.47}$$

An infinite number of such matrices will exist. By solving Eq. (3.44), it then can be seen that weights for transformed oblique factors can be obtained by

$$\boldsymbol{B} = \boldsymbol{AT'} \tag{3.48}$$

Correlations among the transformed factors would be obtained via Eq. (3,42). An interesting special case of this transformation process would arise if a matrix $\boldsymbol{T}$ were defined such that $\boldsymbol{TT'}$=$\boldsymbol{I}$. In such a case, the transformed solution would also contain orthogonal factors, and the transformation would have converted one orthogonal solution into another. Thus, an orthogonal solution can be transformed into an oblique solution or another orthogonal solution by defining a matrix **T** having the appropriate properties.

As will be seen in subsequent chapters, the orthogonal common factor model and the capability of transforming orthogonal solution into oblique solutions, or other orthogonal solutions, form the basis for much factor analysis methodology. Most factor analytic procedures work by first seeking an orthogonal solution which fits a given covariance matrix as well as possible. This initial orthogonal solution represents an arbitrary solution. Then a transformation is determined which transforms this initial solution, by the equations given above, into a new solution such that the resulting factor weights and intercorrelations are substantively interpretable. For example, the three-factor oblique solution presented in Chapter 1 for the nine mental tests was in fact obtained by transforming an initial orthogonal solution. This process is commonly referred to as the rotation, or transformation problem, and is discussed in depth in Chapter 9.

It is interesting to note some simple properties of orthogonal common factor solutions. For instance, it can be seen that for orthogonal factors, Eq. (3.16) can be re-written as

$$\boldsymbol{H^2} = \boldsymbol{Diag(AA')} \tag{3.49}$$

This implies that the common portion of the variance of modeled attribute *j* can be defined as

$$h_j^2 = \sum_{f=1}^{r} a_{jf}^2 \tag{3.50}$$

Thus, the common portion of the variance of modeled attribute j can be obtained by summing the squared entries in row *j* of $\boldsymbol{A}$. This definition for this quantity is the one commonly used in the

factor analytic literature. The definition given in Eq. (3.46) is more general in that it applies to both correlated and uncorrelated common factors. Note that summing the squares of the entries in a row of a correlated factor matrix, $\boldsymbol{B}$, does not necessarily yield the communality of the modeled attribute. Given an initial orthogonal solution, the common variances for the modeled attributes can be obtained via Eq. (3.50), and the unique variances could in turn be obtained by

$$\boldsymbol{U^2 = Diag(\Sigma_{zz}) - H^2} \tag{3.51}$$

which is implied by Eqs. (3.17) and (3.31). Thus, the common and unique variances of the modeled attributes can be obtained from the factor weights of an initial orthogonal solution. These values will not be affected by transformation of that solution.

Another simple property of an orthogonal factor solution involves the relationship between the off-diagonal entries in $\boldsymbol{\Sigma_{zz}}$ and the factor weights in $\boldsymbol{A}$. Eq. (3.45) implies that the following relation will hold, for h not equal to *j*:

$$a_{hj} = \sum_{f=1}^{r} a_{hf} a_{jf} \tag{3.52}$$

According to this equation, the covariance between modeled attributes *h* and *j* can be obtained by summing the products of the factor weights in rows *h* and *j* of $\boldsymbol{A}$. This provides a simple method for determining the relations between modeled attributes, given an orthogonal factor solution.

It is important to recognize that the quantities defined in Eqs. (3.50) and (3.52) are not altered when the orthogonal solution is transformed. That is, the representation by the common factors of the variances and covariances of the modeled attributes is not altered when those factors are transformed according to the procedure described earlier in this section. This can be seen by noting that the elements of the matrix ($\boldsymbol{\Sigma_{zz} - U^2}$) are not affected by the transformation defined in Eqs. (3.42)-(3.45), nor by the transformation given in Eqs. (3.47)-(3.48).

3.3. A Geometric Representation of the Common Factor Model

It is very interesting and useful to recognize that there exists a geometric representation of the common factor model, which corresponds to the algebraic representation presented in the previous two sections. This geometric framework will be described here. It will serve to aid in the understanding of the common factor model. It will also provide the basis for defining several additional types of information which are quite useful in representing a complete common factor solution.

Suppose we are given an orthogonal common factor solution characterized by a matrix $\boldsymbol{A}$ of factor weights. Consider a geometric representation of this solution defined by a Euclidean space, where the axes in the space correspond to the common factors. Thus, the space will be of

dimensionality *r*. Let each modeled attribute be represented by a vector in the space, with the coordinates of the end-points of each factor weights for the corresponding modeled attribute. Thus, there will be n vectors in the space. Figure 3.1 provides an illustration of such a space for a case of two factors, designated A_1 and A_2, and two modeled attributes, designated z_h and z_j. The reader should keep in mind that this illustration is over-simplified in the sense that real world cases would involve more attributes (vectors), and very often more factors (dimensions). Nevertheless, this illustration will serve to demonstrate a number of important points. Note that the axes are orthogonal, meaning that the common factors are uncorrelated. The coordinates of the end-points of the two attribute vectors represent the factor weights for those attributes. These values would correspond to the entries in rows *h* and *j* of the factor weight matrix $\boldsymbol{A}$. Such a matrix, along with other information referred to below, is shown in Table 3.1.

It is interesting to consider some simple geometric properties of these modeled attribute vectors. For instance, the squared of such a vector could be obtained by

$$L_j^2 = \sum_{f=1}^{r} a_{jf}^2 \tag{3.53}$$

This equation can be recognized as equivalent to Eq. (3.50), which defines the common portion of the variance of each modeled attribute. Thus, it is seen that the common variance of each modeled attribute corresponds to the squared length of the corresponding vector. For the two attributes represented in Figure 3.1, these values are obtained as follows:

$$\begin{aligned} h_h^2 &= a_{h1}^2 + a_{h2}^2 \\ &= (.3)^2 + (.5)^2 \\ &= .34 \end{aligned}$$

$$\begin{aligned} h_j^2 &= a_{j1}^2 + a_{j2}^2 \\ &= (.8)^2 + (.4)^2 \\ &= .80 \end{aligned}$$

Recall that if the modeled attributes are taken to be standardized, then these values are communalities, representing the proportion of variance in each modeled attribute due to the common factors. For the unstandardized case, these values are actual variances due to the common factors. An obvious interpretation of these values is that modeled attributes which are more strongly associated with the common factors will be represented by longer vectors in the geometric model. Another interesting aspect of the attribute vectors is the scalar product. For each pair of vectors the scalar product is defined as

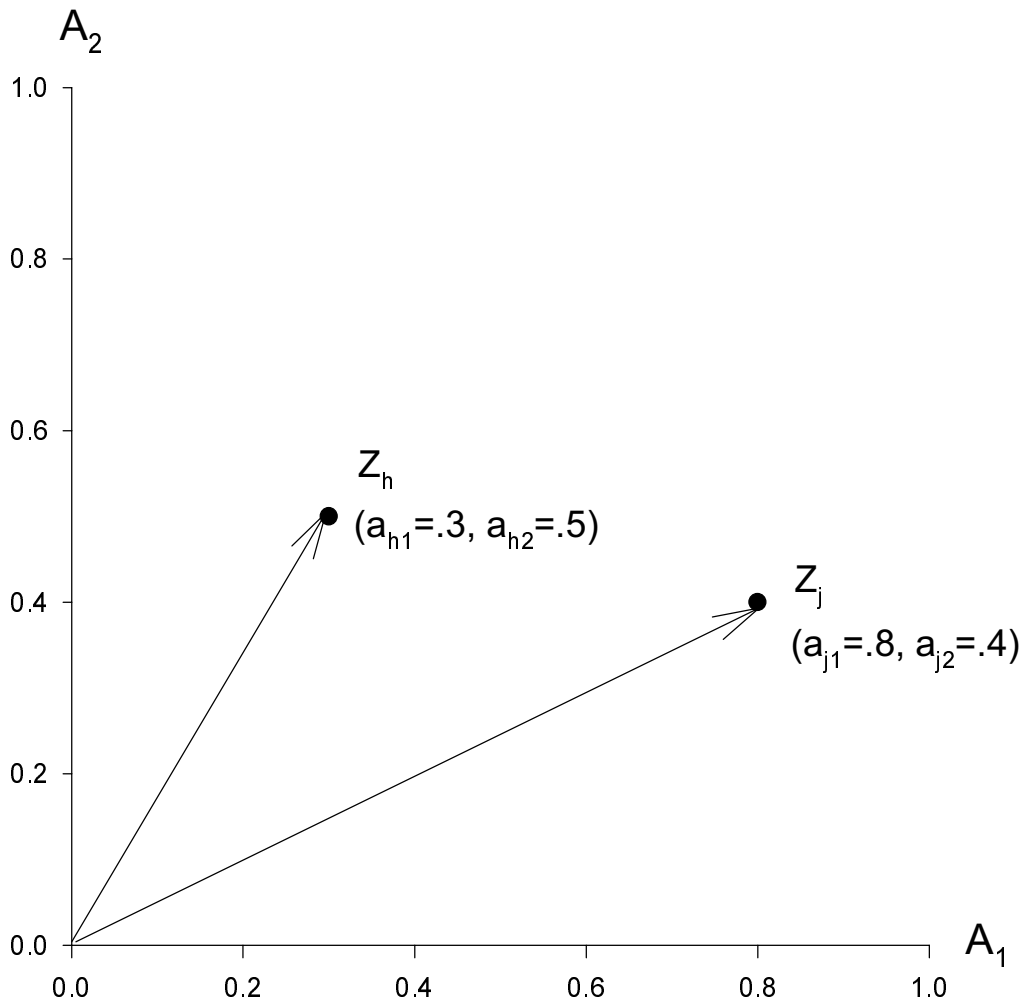

Figure 3.1: Geometric representation of two attribute vectors in a two-factor space

Table 3.1
Matrices for Illustration of Transformation of Two-factor Solution

Matrix A

	A_1	A_2
Z_1		
.		
Z_h	.30	.50
	.80	.40
Z_j		
Z_n		

Matrix T

	A_1	A_2
T_1	.96	.28
T_2	.60	.80

Matrix Φ=TT'

	T_1	T_2
T_1	1.00	.80
T_2	.80	1.00

Matrix T^{-1}

	T_1	T_2
A_1	1.33	−.47
A_2	−1.00	1.60

Matrix $B=AT^{-1}$

	T_1	T_2
Z_1		
.		
Z_j	.66	.26
.		
.		
Z_n		

Matrix Q=AT'

	T_1	T_2
Z_1		
.		
Z_j	.88	.80
.		
.		
Z_n		

Matrix $D=(Diag(T'^{-1}T^{-1})^{1/2}$

	F_1	F_2
T_1	.60	.00
T_2	.00	.60

Matrix $F=DT'^{-1}$

	A_1	A_2
F_1	.80	−.60
F_2	−.28	.96

Matrix G=AF'

	F_1	F_2
Z_1		
.		
Z_j	.40	.16
.		
.		
Z_n		

$$s_{hj} = \sum_{f=1}^{r} a_{hf} a_{jf} \tag{3.54}$$

This equation can be recognized as being equivalent to Eq. (3.52), which expressed the covariance between modeled attributes h and j as a function of the factor weights. Thus, it is seen that the scalar product of two attribute vectors in the factor space corresponds to the covariance (or correlation, for standardized modeled attributes) between the two modeled attributes. For the example shown in Figure 3. 1, this value is obtained by

$$\begin{aligned} \sigma_{hj} &= a_{h1} a_{j1} + a_{h2} a_{j2} \\ &= (.8)(.3) + (.4)(.5) \\ &= .44 \end{aligned}$$

Again, for the case of standardized modeled attributes, this value is a correlation. For the unstandardized case, it is a covariance.

Let us now consider the geometric representation of the transformation process defined in the previous section. Recall that the transformation of an orthogonal solution, such as that shown in Figure 3.1, to another solution (oblique or orthogonal) is based on first defining a non-singular matrix $\boldsymbol{T}$ of order $r \times r$, satisfying Eq. (3.47). It is possible to represent such a matrix directly in the geometric model being developed here. Consider the columns of $\boldsymbol{T}$ as corresponding to the original factors, and let each row of $\boldsymbol{T}$ be represented by a vector in the factor space, where the coordinates of the end-points of the vectors are given by the entries in $\boldsymbol{T}$. Each entry t_{kl} represents the coordinate for vector k on original factor l. These vectors will be called trait vectors, or simply traits. They correspond to transformed factors, and these terms will be used interchangeably. Conceptually, the objective in factor analysis is to determine trait vectors which represent substantively meaningful internal attributes. The matrix $\boldsymbol{T}$, whose row correspond to these trait vectors, will be referred to as a trait matrix. For illustrative purposes, such a matrix is show in Table 3.1. Figure 3.2 shows these trait vectors inserted into the factor space. The two vectors, designated $\boldsymbol{T_1}$ and $\boldsymbol{T_2}$ correspond to the rows of $\boldsymbol{T}$. It is worth noting that, because the trait matrix satisfies Eq. (3.47), the trait vectors will be unit length vectors. Such vectors are referred to as normalized; similarly, it is said that the rows of $\boldsymbol{T}$ are normalized.

An interesting aspect of any pair of trait vectors is their scalar product. The scalar products of all pairs of trait vectors can be obtained by taking the matrix product $\boldsymbol{T}\boldsymbol{T}'$. Eq. (3.42) shows that the resulting scalar products will correspond to correlations among the transformed factors, or traits. For the traits represented in the illustration, the resulting matrix $\boldsymbol{\Phi}$ is given in Table 3.1. The off-diagonal entry of .80 represents the correlation between the two transformed common factors, or traits. Since the vectors are unit length, this value also corresponds to the

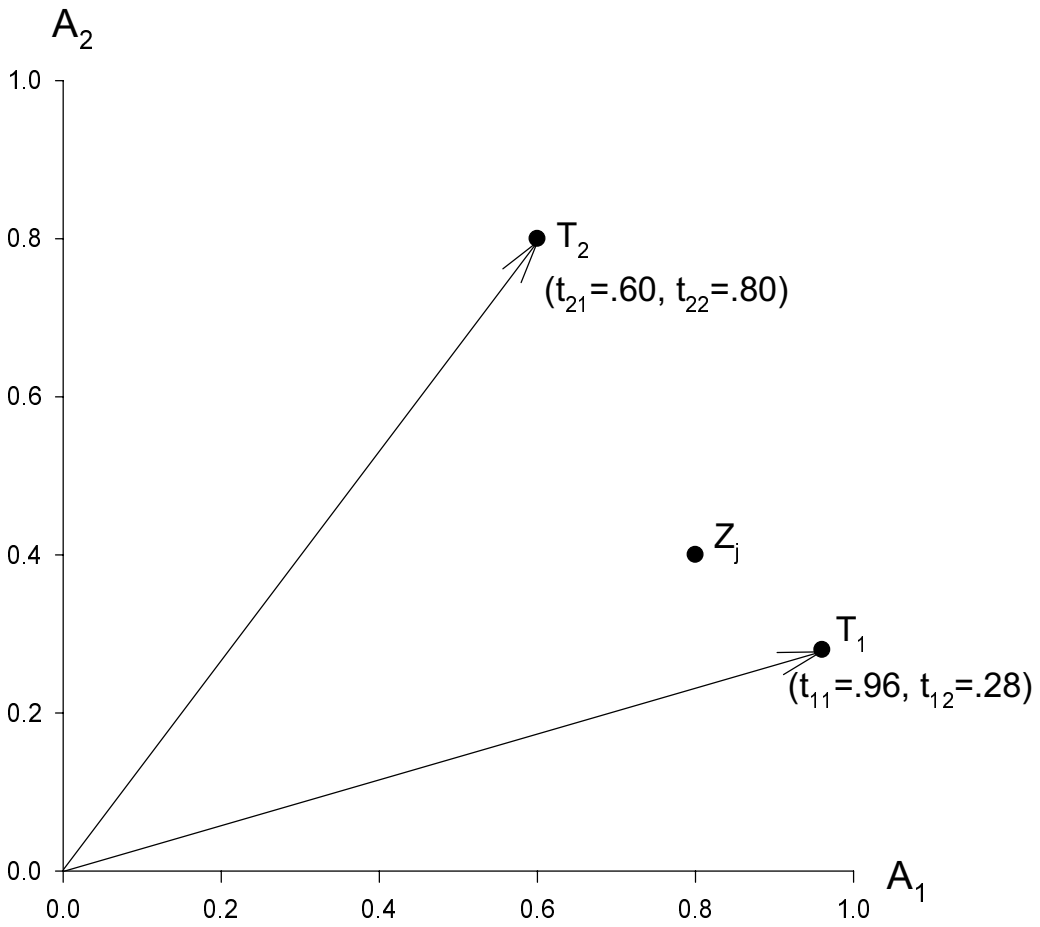

Figure 3.2: Geometric representation of two attribute vectors in a two-factor space

cosine of the angle between the two trait vectors in Figure 3.2.

We will next consider how to represent the relationships between the trait vectors and the modeled attributes. As we will show, this can be approached in a number of different ways. One approach has, in effect, already been defined. Eq. (3.48), repeated here for convenience, defines a transformation of factor weights from the initial orthogonal factors to the transformed oblique factors, or traits:

$$\boldsymbol{B} = \boldsymbol{A}\boldsymbol{T}'$$

The resulting matrix $\boldsymbol{B}$ contains weights for the transformed factors on the modeled attributes. In the geometric model, these weights correspond to the Cartesian projections of the attribute vectors on the trait vectors. Each such weight β_{jk} is analogous to a partial regression coefficient, representing the unique effect on modeled attribute *j* of factor *k*. The computation of these weights for one of the attributes is given in Table 3.1 for attribute *j*, and the projections are shown in Figure 3.3.

An alternative representation of the relationships of the traits to the modeled attributes can be obtained by determining scalar product, of the attribute vectors and the trait vectors. Algebraically, the scalar products would be given by

$$\boldsymbol{Q} = \boldsymbol{A}\boldsymbol{T}' \tag{3.55}$$

Matrix $\boldsymbol{Q}$ will have *n* rows representing the modeled attributes and *r* columns representing the traits. Each entry q_{jk} will represent a scalar product, or covariance, between a modeled attribute and a trait. (Again, note that for the case of standardized modeled attributes, these values will be correlations between modeled attributes and traits). Geometrically, these values correspond to perpendicular projections of modeled attribute vectors on trait vectors. The computation of these values for one of the attribute vectors is shown in Table 3.1. and the corresponding projections are shown in Figure 3.3. These projections are analogous to simple measures of covariance (or correlation) between a modeled attribute and a trait, ignoring the fact that other traits are present.

Continuing further with the issue of evaluating relationships between modeled attributes and traits, it has been found very useful to define a second type of vector associated with the trait vectors. This second type of vector is based on the notion of a hyperplane. Consider a solution containing *r* correlated factors. For any given factor, the remaining factors can be conceived of as defining a space of (*r*-1) dimensions. This subspace is referred to as a hyperplane. For each factor, there exists a corresponding hyperplane made up of the other (*r*-1) factors. For each such hyperplane, let us define a unit-length vector orthogonal to that hyperplane. These vectors are called normals to hyperplanes, or more simply, just normals. These normals are equivalent to what Harman (1 967) and others call "reference vectors". Note that for each trait there will be a

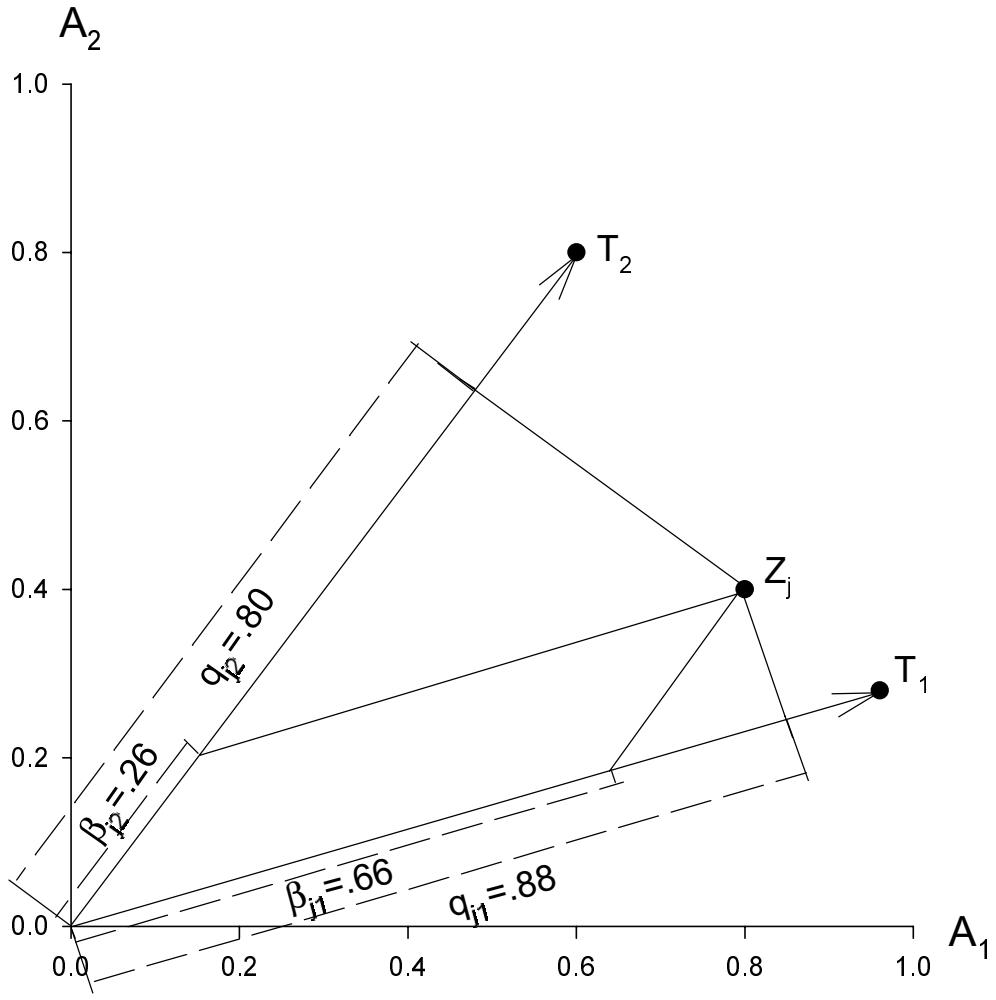

Figure 3.3: Geometric representation of projections of one attribute vector on two trait vectors

corresponding normal, defined as a vector orthogonal to the corresponding hyperplane. These normals have been used extensively in representing oblique common factor solutions.

Let us define a matrix $\boldsymbol{F}$' which contains these vectors as rows. Thus, the r columns of $\boldsymbol{F}$ correspond to the original orthogonal factors and the r rows represent the normals. An entry f_{kl} represents the coordinate of the end-point of normal $\boldsymbol{F}_k$ on original factor A_l. It is useful to consider the relationship between these normals and the traits, represented by the rows of $\boldsymbol{T}$. Note that each normal is defined as orthogonal to all but one trait. Therefore, if we obtain the scalar products of the normals with the traits, as given by

$$\boldsymbol{T}\boldsymbol{F}' = \boldsymbol{D} \tag{3.56}$$

the resulting product matrix $\boldsymbol{D}$, of order $r \times r$, must be diagonal. This will be the case because each normal is orthogonal to all but one trait, and vice versa. This interesting relationship can be employed to determine a method to obtain the matrix $\boldsymbol{F}$ of normals. Since the rows of $\boldsymbol{F}$ are normalized vectors, this implies

$$Daig(\boldsymbol{F}\boldsymbol{F}') = \boldsymbol{I} \tag{3.57}$$

Solving Eq. (3.56) for $\boldsymbol{F}$' yields

$$\boldsymbol{F}' = \boldsymbol{T}^{-1}\boldsymbol{D} \tag{3.58}$$

Substituting from Eq. (3.58) into Eq. (3.57) yields

$$Diag(\boldsymbol{D}\boldsymbol{T}'^{-1}\boldsymbol{T}^{-1}\boldsymbol{D}) = \boldsymbol{I} \tag{3.59}$$

Solving Eq. (3.59) for $\boldsymbol{D}$ gives us

$$\begin{aligned} \boldsymbol{D} &= (Diag(\boldsymbol{T}'^{-1}\boldsymbol{T}^{-1}))^{-\frac{1}{2}} \\ &= (Daig(\boldsymbol{T}\boldsymbol{T}')^{-1})^{-\frac{1}{2}} \\ &= (Diag\,\boldsymbol{\Phi}^{-1})^{-\frac{1}{2}} \end{aligned} \tag{3.60}$$

This result is important because it provides a method to obtain matrix $\boldsymbol{D}$, the diagonal matrix containing relations between traits and normals. The diagonal elements of $\boldsymbol{D}$ actually are equivalent to constants which serve to normalize the columns of $\boldsymbol{T}^{-1}$, as shown in Eq. (3.58). Given the trait matrix $\boldsymbol{T}$, $\boldsymbol{D}$ can be obtained according to Eq. (3.60). It can then be used to obtain $\boldsymbol{F}$' according to Eq. (3.58). Matrix $\boldsymbol{F}$ then contains the normals as rows. These normals can be conceived of as an additional set of vectors inserted into the factor space. The calculation of $\boldsymbol{D}$ and $\boldsymbol{F}$' is illustrated in Table 3.1, and the geometric representation of the normals, designated $\boldsymbol{F}_1$, and $\boldsymbol{F}_2$, is shown in Figure 3.4. Note in Figure 3.4 that normal $\boldsymbol{F}_1$, is orthogonal to all traits other than $\boldsymbol{T}_1$; i.e., it is orthogonal to $\boldsymbol{T}_2$ in this limited illustration. Also, normal $\boldsymbol{F}_2$ is orthogonal to all

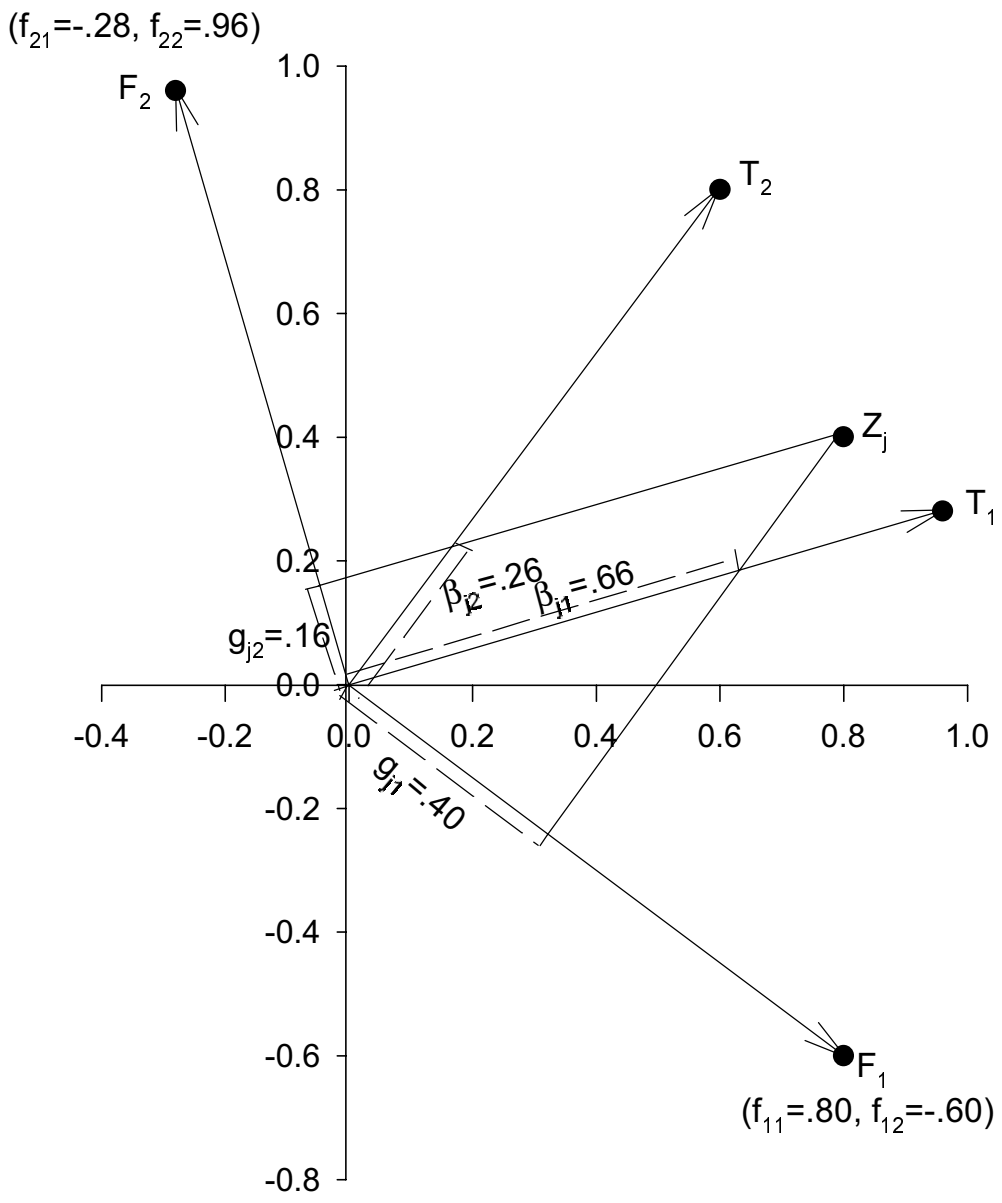

Figure 3.4: Geometric representation of projections of one attribute vector on two normals

traits other than $\boldsymbol{T}_2$; i.e., it is orthogonal to $\boldsymbol{T}_1$.

The utility of these normals can be seen when we consider the relationship of the modeled attributes to the normals. Let us define the scalar products of the modeled attributes with the normals as

$$\boldsymbol{G} = \boldsymbol{A}\boldsymbol{F}' \tag{3.61}$$

Matrix $\boldsymbol{G}$ will be of order $n \times r$, with rows representing modeled attributes and columns corresponding to the normals. Each entry g_{jk} can be interpreted as a measure of a partialled relationship. Such an entry represents the partial covariance (or correlation in the standardized case) between modeled attribute *j* and trait *k*, with the effects of the other (*r*-1) traits partialled out. This interpretation is based on the definition of normals as being orthogonal to hyperplanes. As such, each normal represents a partialled trait; i.e., that portion of a given trait which remains after the effects of the other (*r*-1) traits are partialled out. The computation of these values is illustrated for a single attribute in Table 3.1. Geometrically, these values correspond to perpendicular projections of attribute vectors on normals. These projections for a single attribute are shown in Figure 3.4.

A very interesting relationship exists between the factor weights in matrix $\boldsymbol{B}$ and the projections on the normals, given in matrix $\boldsymbol{G}$. Substitution from Eq. (3.58) into Eq. (3.61) produces

$$\boldsymbol{G} = \boldsymbol{A}\boldsymbol{T}^{-1}\boldsymbol{D} \tag{3.62}$$

Then, by substituting into this equation from Eq. (3.48), we obtain

$$\boldsymbol{G} = \boldsymbol{B}\boldsymbol{D} \tag{3.63}$$

According to this equation, the columns of $\boldsymbol{G}$ and $\boldsymbol{B}$ will be proportional. The constants of proportionality will be the diagonal elements of $\boldsymbol{D}$. Thus, the factor weights for a given factor will be proportional to the projections on the normals for the corresponding factor. An examination of Figure 3.4 reveals the geometric basis for this relationship. Note that the projection of a given modeled attribute on a normal in this example is obtained by "continuing" the projection from the attribute, through the trait, to the normal. The Cartesian projections of the attribute vectors on the trait vectors were illustrated originally in Figure 3.3. These same projection are shown also in Figure 3.4, which illustrates how those projections are related to the projections of the attribute vectors on the normals. This gives rise to the proportionality defined in Eq. (3.63)

To summarize the developments presented in this section, begin with the assumption that an orthogonal solution characterized by matrix $\boldsymbol{A}$ has been obtained. Given such a solution, a

transformation then can be carried out by defining a trait matrix $\boldsymbol{T}$, the rows of which represent the traits, or transformed factors. One then can obtain, by equations presented above, matrices $\boldsymbol{\Phi}$, containing the correlations of the factors; $\boldsymbol{B}$, containing the factor weights; $\boldsymbol{Q}$, containing the covariance of the modeled attributes with the factors; and $\boldsymbol{G}$, containing the partialled covariances of the modeled attributes with the factors. In the geometric representation of the model, the modeled attributes are represented by vectors in a space whose axes correspond to the initial orthogonal factors. The transformed factors are represented by trait vectors, defined by the rows of matrix $\boldsymbol{T}$. The correlations among the trait vectors, contained in matrix $\boldsymbol{\Phi}$, corresponds to cosines of angles between the trait vectors. The factor weights in matrix $\boldsymbol{B}$ correspond to Cartesian projections of the attribute vectors on the trait vectors. The covariances of the modeled attributes, with the trait vectors, contained in matrix $\boldsymbol{Q}$, correspond to perpendicular projections of the attribute vectors on the trait vectors. Finally, the partialled covariance of the modeled attributes with the trait vectors, contained in matrix $\boldsymbol{G}$, correspond to projections of the attribute vectors on normals to the hyperplane defined by each trait vector. These matrices define an oblique common factor solution.

It is important to recognize the properties of such a solution when the traits themselves are orthogonal. That is, consider the case in which

$$\boldsymbol{T}\boldsymbol{T}' = \boldsymbol{I} \tag{3.64}$$

Under this condition, $\boldsymbol{T}'$ is equivalent to $\boldsymbol{T}$. As a result, matrix $\boldsymbol{B}$ as defined in Eq. (3.48) will be equivalent to matrix $\boldsymbol{Q}$ as defined in Eq. (3.55). In addition, the matrix $\boldsymbol{D}$ as defined in Eq. (3.60) will be an identity matrix. As a result, Eq. (3.63) shows that $\boldsymbol{B}$ and $\boldsymbol{G}$ will be equivalent. The implication of these relationships is that when the traits are orthogonal, there will be no distinction among matrices $\boldsymbol{B}$, $\boldsymbol{Q}$, and $\boldsymbol{G}$. All of the types of information contained in these matrices will be equivalent in this case. In geometric terms, this phenomenon is revealed by the fact that when the traits are orthogonal, there will be no distinction between Cartesian projections (as given in $\boldsymbol{B}$) and perpendicular projections (as given in $\boldsymbol{Q}$) of the attribute vectors on the trait vectors. Furthermore, each normal will be equivalent to the corresponding trait vector, since the trait vectors themselves are orthogonal to the hyperplanes. Thus, the projections on the normals (given in $\boldsymbol{G}$) will be equivalent to the projections on the trait vectors. This all demonstrates that an orthogonal solution is, in some senses, simpler than an oblique solution, since there are no trait correlations to consider and since the types of relationships between attributes and traits are all equivalent. However, as will be discussed in Chapter 9, this relative simplicity should not be considered to represent an advantage of orthogonal solutions over oblique solutions. The critical issue is to determine traits which correspond to substantively meaningful internal attributes, whether or not they are or orthogonal.

3.4. Effects of Factor Transformations on Factor Scores

A recurring theme in the previous three sections has been the transformation of factors. Mathematical and geometric frameworks have been described for representing the transformation of common factor weight matrices. It is very important that the reader recall the role of these factor weights in the common factor model. According to the model as defined in Eq. (3.7) or Eq. (3.39), these weights are weights which are applied to scores on the common factors. The resulting linear combination of common factor scores, combined with specific factor and error of measurement factor terms, yield the modeled attributes. Let us focus at this point on the common factor contribution to the modeled attributes. Given the nature of the model, it is clear that any transformation of the common factor weights must also imply a corresponding transformation of the common factor scores. To put this more simply, if some operation is carried out to alter the common factor weights in model expressed in Eq. (3.7) or Eq. (3.39), then the common factor scores must also be altered in order that the model still produce the same modeled attributes. In this section we will present a framework for representing this phenomenon, and we will consider a number of related issues of both theoretical and practical importance.

To begin, let us consider Eqs. (3.7) and (3.39). In these expressions of the model, the common factor contributions to the modeled attribute scores in $\underline{z}$ are contained in the product $\underline{x}_\beta \boldsymbol{B}'$. Let us define a vector $\underline{z}_c$. as containing these contributions. That is

$$\underline{z}_c = \underline{x}_\beta \boldsymbol{B}' \tag{3.65}$$

As described in sections 3.2 and 3.3, let us consider $\boldsymbol{B}$ to be a transformed matrix of factor weights. Suppose that an initial orthogonal weight matrix $\boldsymbol{A}$ had been obtained, and that $\boldsymbol{B}$ is the result of transforming $\boldsymbol{A}$ according to Eq. (3.48). Substituting from Eq. (3.48) into Eq. (3.65) yields

$$\underline{z}_c = \underline{x}_\beta \boldsymbol{T}'^{-1} \boldsymbol{A}' \tag{3.66}$$

Let us define a vector $\underline{x}_\alpha$ as follows:

$$\underline{x}_\alpha = \underline{x}_\beta \boldsymbol{T}'^{-1} \tag{3.67}$$

Substituting from this equation into Eq. (3.66) yields

$$\underline{z}_c = \underline{x}_\alpha \boldsymbol{A}' \tag{3.68}$$

A comparison of Eqs. (3.65) and (3.68) indicates that these provide two different representations of the $\underline{z}_c$ scores, based on two different factor weight matrices. Thus, it can be seen that $\underline{x}_\beta$

contains the common factor scores corresponding to the weights given in $\boldsymbol{B}$, and $\underline{\boldsymbol{x}}_\alpha$ contains the common factor scores corresponding to the weights given in $\boldsymbol{A}$. Furthermore, Eq. (3.67) represents the transformation between the two sets of common factor scores.

It is interesting to consider also the covariance among the factor scores. The covariance among the factor scores in $\underline{\boldsymbol{x}}_\beta$ are given by $\boldsymbol{\Sigma}_{\beta\beta}$, which is defined in Eq. (3. 10) as matrix $\boldsymbol{\Phi}$, the intercorrelation matrix for the common factors. In the present section, we will employ an additional level of subscripts and will designate this matrix as $\boldsymbol{\Sigma}_{x_\beta x_\beta}$. Eq. (3.67) defines a linear transformation relating $\underline{\boldsymbol{x}}_\beta$ to $\underline{\boldsymbol{x}}_\alpha$. Based on such a transformation, the relation between the covariance matrices for these two sets of common factor scores can be obtained by employing Corollary 3 of Theorem 4 in Appendix C. This yields the following:

$$\boldsymbol{\Sigma}_{x_\alpha x_\alpha} = \boldsymbol{T}^{-1}\boldsymbol{\Sigma}_{x_\beta x_\beta}\boldsymbol{T}'^{-1} \tag{3.69}$$

Since $\boldsymbol{\Sigma}_{x_\alpha x_\alpha}$ is designated as $\boldsymbol{\Phi}$, and since $\boldsymbol{\Phi}$ is identified in Eq. (3.42) as being equal to $\boldsymbol{TT'}$, Eq. (3.69) can be rewritten as follows:

$$\boldsymbol{\Sigma}_{x_\alpha x_\alpha} = \boldsymbol{T}^{-1}\boldsymbol{T}\boldsymbol{T}'\boldsymbol{T}'^{-1} \tag{3.70}$$

This shows that the covariance matrix for the common factor scores corresponding to the orthogonal weight matrix $\boldsymbol{A}$ will be an identity matrix; i.e., these common factor scores will be uncorrelated.

A geometric representation of these relations is of interest. Let us consider a geometric framework of a different nature than that employed in section 3.3. In particular, let us define a factor score space in terms of the factor scores given in $\underline{\boldsymbol{x}}_\alpha$. The axes in this space represent the common factors, and each individual will be represented by a point in the space, with the coordinates of each individual's point being given by the common factor score vector, $\underline{\boldsymbol{x}}_\alpha$, for that individual. Equivalently, each individual can be thought of as being represented by a vector in the factor score space, with the coordinates of the endpoint of each individual's vector being defined by the scores for that individual on the common factors. Note that, according to Eq. (3.70) the distribution of points in the space will be such that the scores on the original axes will be standardized and uncorrelated. Let us next consider the representation of the scores on the transformed factors, given in $\underline{\boldsymbol{x}}_\beta$. Employing Eq. (3.67), we can obtain the following relation:

$$\underline{\boldsymbol{x}}_\beta = \underline{\boldsymbol{x}}_\alpha\boldsymbol{T}' \tag{3.71}$$

This equation indicates that the factors scores in $\underline{\boldsymbol{x}}_\beta$ are scalar products of the scores in $\underline{\boldsymbol{x}}_\alpha$ with the trait vectors in $\boldsymbol{T}$. Since the trait vectors are of unit length, these scalar products correspond to projections of the $\underline{\boldsymbol{x}}_\alpha$ vectors on the trait vectors. The correlations among the factor scores in

$\underline{\boldsymbol{x}}_{\beta}$, which are given in $\boldsymbol{\Phi}$, are cosines of angles between the trait vectors. A point of considerable importance is that the relations among the factors are defined in this factor score space.

An interesting special case of the transformation problem is the transformation from one uncorrelated solution to a second such solution. Let us examine the impact of such a transformation on the factor scores and their interrelationships, employing the mathematical framework developed above. In this case, $\boldsymbol{T}$ will be defined such that

$$\boldsymbol{T}\boldsymbol{T}^{'} = \boldsymbol{I} \tag{3.72}$$

This implies that

$$\boldsymbol{T}' = \boldsymbol{T}^{-1} \tag{3.73}$$

Let $\boldsymbol{A}_1$ be the weight matrix for the first uncorrelated solution and $\boldsymbol{A}_2$ be the weight matrix for the second such solution. By Eq. (3.48)

$$\boldsymbol{A}_2 = \boldsymbol{A}_1\boldsymbol{T}' \tag{3.74}$$

Let the factor scores corresponding to $\boldsymbol{A}_1$ be given by $\underline{\boldsymbol{x}}_{\alpha 1}$, then by using Eqs. (3.68), (3.73), and (3.74), we can obtain the following:

$$\underline{\boldsymbol{z}}_c = \underline{\boldsymbol{x}}_{\alpha 1}\boldsymbol{A}_1' = \underline{\boldsymbol{x}}_{\alpha 1}\boldsymbol{T}'\boldsymbol{A}_2' \tag{3.75}$$

Let the scores on the second solution be given by $\underline{\boldsymbol{x}}_{\alpha 2}$, which according to Eq. (3.71), would be given by

$$\underline{\boldsymbol{x}}_{\alpha 2} = \underline{\boldsymbol{x}}_{\alpha 1}\boldsymbol{T}' \tag{3.76}$$

Employing Eqs. (3.75) and (3.76), we obtain the following:

$$\underline{\boldsymbol{z}}_c = \underline{\boldsymbol{x}}_{\alpha 1}\boldsymbol{A}_1' = \underline{\boldsymbol{x}}_{\alpha 2}\boldsymbol{A}_2' \tag{3.77}$$

This equation simply shows that both sets of common factor scores and weights produce the same $\underline{\boldsymbol{z}}_c$ values. It also can be shown that the covariance matrices for the two sets of factor scores have the expected form. Making use of the linear transformation defined in Eq. (3.74) along with Corollary 3 of Theorem 4 in Appendix C, we obtain the following:

$$\boldsymbol{\Sigma}_{x_{a2}x_{a2}} = \boldsymbol{T}\boldsymbol{\Sigma}_{x_{a1}x_{a1}}\boldsymbol{T}' \tag{3.78}$$

Since the initial solution $\boldsymbol{A}_1$ is defined as orthogonal, we can write

$$\boldsymbol{\Sigma}_{x_{a1}x_{a1}} = \boldsymbol{I} \tag{3.79}$$

Substituting from Eqs. (3.79) and (3.72) into Eq. (3.78), we obtain

$$\boldsymbol{\Sigma}_{x_{a1}x_{a2}} = \boldsymbol{T}\boldsymbol{T}' = \boldsymbol{I} \tag{3.80}$$

Thus, both solutions are characterized by factor scores which are standardized and uncorrelated.

We wish to consider next a very interesting issue which often gives rise to substantial confusion in empirical applications of factor analysis. The issue concerns the nature and meaning of product matrices obtained by pre-multiplying factor weight matrices by their respective transposes. A common misconception is that the relations among the common factors are represented in such matrices. Let us first consider this issue in the context of the special case just described--the transformation of one orthogonal solution $\boldsymbol{A}_1$ into a second such solution $\boldsymbol{A}_2$. Let us define product matrices $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ as follows:

$$\boldsymbol{P}_1 = \boldsymbol{A}_1'\boldsymbol{A}_1 \tag{3.81}$$

$$\boldsymbol{P}_2 = \boldsymbol{A}_2'\boldsymbol{A}_2 \tag{3.82}$$

Substituting from Eq. (3.74) into Eq. (3.82) yields

$$\boldsymbol{P}_2 = \boldsymbol{T}\boldsymbol{P}_1\boldsymbol{T}' \tag{3.83}$$

The misconception involves the issue of whether the relations among the common factors are indicated by the entries in $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$. Suppose $\boldsymbol{P}_1$ is diagonal and the diagonal entries are unequal (this is a property of a particular type of solution, called a principal factors solution, to be described in Chapter 7). According to Eq. (3.83), $\boldsymbol{P}_2$ generally will not be a diagonal matrix. That is, even though the transformed solution represented by $\boldsymbol{A}_2$ is an orthogonal solution, the product matrix $\boldsymbol{P}_2$ generally will not be diagonal. This is a point which troubles many practitioners of factor analysis. How can $\boldsymbol{A}_2$ be an orthogonal factor matrix when its columns are not "orthogonal"? The resolution of this apparent paradox was given earlier in this section: the orthoganality of the factors is defined in the factor score space, and is not defined by product matrices such as $\boldsymbol{P}_2$. That is, the factor scores for the factors defined by $\boldsymbol{A}_2$ are uncorrelated, as shown in Eq. (3.80), even though matrix $\boldsymbol{P}_2$ is not diagonal.

Let us consider this issue in the more general context of transformations from uncorrelated factors to correlated factors. Eq. (3.71) defines the impact of such a transformation on the factor scores. The area of concern here is with the product matrices, which will be designated as follows:

$$\boldsymbol{P}_\alpha = \boldsymbol{A}'\boldsymbol{A} \tag{3.84}$$

$$\boldsymbol{P}_{\beta} = \boldsymbol{B}'\boldsymbol{B} \tag{3.85}$$

Employing Eq. (3.48), we can rewrite Eq. (3.85) as follows:

$$\boldsymbol{P}_{\beta} = \boldsymbol{T}'^{-1}\boldsymbol{A}'\boldsymbol{A}\boldsymbol{T}^{-1} \tag{3.86}$$

Substituting from Eq. (3.84) into Eq. (3.86) yields

$$\boldsymbol{P}_{\beta} = \boldsymbol{T}'^{-1}\boldsymbol{P}_{\alpha}\boldsymbol{T}^{-1} \tag{3.87}$$

This equation gives the transformation of $\boldsymbol{P}_{\alpha}$ to $\boldsymbol{P}_{\beta}$. The inverse transformation is implied. The critical aspect of Eq. (3.87) is that it shows that there is no necessary, direct relation of $\boldsymbol{P}_{\beta}$ to $\boldsymbol{\Phi}$. That is, matrix $\boldsymbol{\Phi}$ defines the relations among the common factors, and the relations are not indicated in any way in matrix $\boldsymbol{P}_{\beta}$.

An illustration of the distinction between $\boldsymbol{P}_{\beta}$ and $\boldsymbol{\Phi}$ involves a type of solution called an independent cluster solution. In this case, each modeled attribute has a non-zero weight on one and only one factor in B. In such a case, $\boldsymbol{P}_{\beta}$ would be diagonal since the sums of products between columns of **B** would be zero. However, matrix $\boldsymbol{\Phi}$ would not necessarily be diagonal. The independent clusters of attributes might be intercorrelated so that $\boldsymbol{\Phi}$ would have non-zero off-diagonal entries.

To complete the discussion of this issue, let us consider four possible situations defined by combining the possibilities of $\boldsymbol{\Phi}$ being diagonal or not diagonal. When both are diagonal, the solution would be a principal factors solution; this type of solution was mentioned earlier in this section and will be discussed in detail in Chapter 7. The other cases can be thought of as arising from various possible transformations of such a solution. The case where $\boldsymbol{\Phi}$ is diagonal and $\boldsymbol{P}_{\beta}$ is not diagonal would be produced by an orthogonal transformation of a principal factor solution. This is a legitimate view since any factor solution could be transformed into a principal factors solution. The case where $\boldsymbol{\Phi}$ is not diagonal and $\boldsymbol{P}_{\beta}$ is diagonal would be produced by certain special transformations from a principal factors solution, including a transformation to an independent cluster solution. Finally, the case where $\boldsymbol{\Phi}$ is not diagonal and $\boldsymbol{P}_{\beta}$ is not diagonal represents the more general case of transformation to a general oblique solution.

This completes our discussion of the effects of factor transformations on factors scores. Critical points to keep in mind are that transformations of factor weights imply corresponding transformations of factor scores., and that the relations among the factors are defined in the factor score space. It is a misconception to consider the relations among the factors as being defined by the factor weights themselves.

3.5. Correspondence Between the Model and the Real World

An issue emphasized in Chapter 1 and in the first section of the present chapter involves the correspondence between the common factor model and the real world. It is recognized that we do not expect the model to provide an exact and complete accounting for the variance and covariances of the surface attributes. This fact is represented in the mathematical framework by differentiating between the surface attributes (vector $\underline{\boldsymbol{y}}$) and that portion of the surface attributes that is consistent with the common factor model (the "modeled attributes" in vector $\underline{\boldsymbol{z}}$). According to the mathematical representation of the model defined in this chapter, the model does account for the variances and covariances of the modeled attributes. That is, the covariance matrix $\boldsymbol{\Sigma}_{zz}$ is a function of the parameters of the model, as shown in Eq. (3.35).

It is important to represent explicitly the fact that, though the model accounts for the variances and covariance of the modeled attributes, it will not necessarily do the same for the surface attributes themselves. This can be seen by defining a population covariance matrix $\boldsymbol{\Sigma}_{yy}$ for the surface attributes. Matrix $\boldsymbol{\Sigma}_{yy}$ will be of order $n \times n$, with entries σ_{hj} representing the population covariance for surface attributes *h* and *j*. It is especially important to consider the relation between $\boldsymbol{\Sigma}_{yy}$ and $\boldsymbol{\Sigma}_{zz}$. This can be done by making use of the fact, as defined in Eq. (3.1), that the surface attributes in $\underline{\boldsymbol{y}}$ are actually sums of the component portions in $\underline{\mathbf{z}}$ and $\underline{\ddot{\mathbf{z}}}$. Appendix C treats the case where one vector of measures is defined as the sum of two other vectors of measures, and shows the relationship of the covariances of the summed measures to those of the component portions. By applying the relationship shown in Theorem 5 of Appendix C, we can write, for the present case, the following:

$$\boldsymbol{\Sigma}_{yy} = \boldsymbol{\Sigma}_{zz} + \boldsymbol{\Sigma}_{z\ddot{z}} + \boldsymbol{\Sigma}_{\ddot{z}z} + \boldsymbol{\Sigma}_{\ddot{z}\ddot{z}} \tag{3.88}$$

From this equation it can be seen clearly that the population covariance matrix for the surface attributes is not, in general, equivalent to the population covariance matrix for the modeled portions of the attributes. More specifically, if we define a matrix $\boldsymbol{\Delta}_{\Sigma}$ as

$$\boldsymbol{\Delta}_{\Sigma} = \boldsymbol{\Sigma}_{\ddot{z}z} + \boldsymbol{\Sigma}_{z\ddot{z}} + \boldsymbol{\Sigma}_{\ddot{z}\ddot{z}} \tag{3.89}$$

then Eq. (3.88) can be re-written as

$$\boldsymbol{\Sigma}_{yy} = \boldsymbol{\Sigma}_{zz} + \boldsymbol{\Delta}_{\Sigma} \tag{3.90}$$

The matrix $\boldsymbol{\Delta}_{\Sigma}$ then represents that portion of the population covariances of the surface attributes that cannot be accounted for by the common factor model. It can then be seen that only when all entries in $\boldsymbol{\Delta}_{\Sigma}$ are zero will $\boldsymbol{\Sigma}_{yy}$ be equivalent to $\boldsymbol{\Sigma}_{zz}$. In that situation, the model will exactly represent the population covariances for the surface attributes. But if some entries in $\boldsymbol{\Delta}_{\Sigma}$ are not zero, then $\boldsymbol{\Sigma}_{yy}$ will not be equivalent to $\boldsymbol{\Sigma}_{zz}$, and the model will not hold exactly in the

population. As the entries in $\boldsymbol{\Delta}_{\Sigma}$ deviate further from zero, the correspondence between the model and the population becomes weaker. This lack of correspondence between the model and the real world is a very important concept which is often overlooked by neglecting the distinction between $\boldsymbol{y}$ and $\underline{\boldsymbol{z}}$. We will refer to such lack of correspondence as model error, meaning that the representation of the real world by the model is in error to some degree. It must be understood that model error is present in the population and is completely separate from lack of fit arising from sampling. That is, even if there is no model error, meaning that the model holds exactly in the population, there still will be sampling error. That is, the model probably would not fit a sample covariance matrix exactly. This issue will be dealt with further in Chapters 4 and 5.

A final step in developing the common factor model in the population now can be achieved by substituting from Eq. (3.35) into Eq. (3.90). This yields

$$\boldsymbol{\Sigma}_{yy} = \boldsymbol{B}\boldsymbol{\Phi}\boldsymbol{B}' + \boldsymbol{U}^2 + \boldsymbol{\Delta}_{\Sigma} \tag{3.91}$$

This equation represents the milestone of expressing the common factor model in terms of the population covariances of the surface attributes. This expression represents these covariances as a function of the parameters in $\boldsymbol{B}$, $\boldsymbol{\Phi}$, and $\boldsymbol{U}^2$, plus a term representing model error. Most presentations of the model ignore the need to explicitly include model error.

The presence of model error can serve as a basis for introducing the concept of fitting the model to a covariance matrix. Considering the hypothetical case where a population covariance matrix $\boldsymbol{\Sigma}_{yy}$ is available, suppose we wish to obtain a solution for the model as represented in Eq. (3.91). Recognizing that $\boldsymbol{\Delta}_{\Sigma}$ represents the lack of fit of the model to $\boldsymbol{\Sigma}_{yy}$, the objective would be to obtain a solution for the model which, in some sense, minimizes the entries in $\boldsymbol{\Delta}_{\Sigma}$. Such a solution would be optimal in the sense of providing the most accurate accounting of the population covariances of the surface attributes. As it will be seen in Chapter 7, there are different ways to define the optimal solution; i.e., different ways to define a criterion representing an optimal $\boldsymbol{\Delta}_{\Sigma}$. These alternative definitions of an optimal solution in turn lead to alternative techniques for obtaining a solution to the model, and these alternative techniques can yield different solutions. In other words, there may be no single correct solution. Rather, different definitions of an optimal solution yield different $\boldsymbol{B}$, $\boldsymbol{\Phi}$, and $\boldsymbol{U}^2$ matrices, and, in turn, different $\boldsymbol{\Sigma}_{zz}$ and $\boldsymbol{\Delta}_{\Sigma}$ matrices. Only when a solution can be found for which all entries in $\boldsymbol{\Delta}_{\Sigma}$ are zero can it be said that the model is consistent with the real world; in that case, there would be no model error.

Surely, however, almost all acceptable models will not fit the real world perfectly in the population. To insist on a perfect fit at all times is a route to chaos by the inclusion of many small, trivial factors. To be sure, we should avoid missing small factors which can be made large and important with special studies. All experimenters should be alert to this possibility. The

distinction between trivia and important small influences is a matter for experimenter insight and judgment. However, the complexities of factor analysis make it imperative that a distinction be made. Inclusion of a number of very small factors in an analysis results in unmanageably large dimensionality of the common factor space. Great care is required of an experimenter in making the decision between possibly meaningful factors and trivia. There should be no doubt but that some trivia will exist.

3.6. The Common Factor Model for a Population Correlation Matrix

The development of the common factor model in this chapter has been carried out in the context of covariance matrices. That is, the fundamental reputation of the model given in Eq. (3.91) was stated in terms of a population covariance matrix for the surface attributes. Since many applications of factor analysis are conducted using correlation matrices rather than covariance matrices, it is important to consider how the expression of the model is affected when the data are correlations rather than covariances. This can be achieved easily by employing the simple relation between a covariance matrix and a correlation matrix. Given a population covariance matrix for the surface attributes, $\boldsymbol{\Sigma}_{yy}$, let us define a diagonal matrix $[\boldsymbol{\Sigma}_d]_{yy}$ containing the population variances of the surface attributes. That is,

$$[\boldsymbol{\Sigma}_d]_{yy} = Diag(\boldsymbol{\Sigma}_{yy}) \tag{3.92}$$

We can then define a population correlation matrix for the surface attributes, $\boldsymbol{R}_{yy}$, as follows:

$$\boldsymbol{R}_{yy} = [\boldsymbol{\Sigma}_d]_{yy}^{-\frac{1}{2}} \boldsymbol{\Sigma}_{yy} [\boldsymbol{\Sigma}_d]_{yy}^{-\frac{1}{2}} \tag{3.93}$$

The effect of this standardization of the surface attributes on the common factor model can then be seen by substituting from Eq. (3.91) into Eq. (3.93), yielding

$$\begin{aligned} \boldsymbol{R}_{yy} &= [\boldsymbol{\Sigma}_d]_{yy}^{-\frac{1}{2}} (\boldsymbol{B}\boldsymbol{\Phi}\boldsymbol{B}' + \boldsymbol{U}^2 + \boldsymbol{\Delta}_{\Sigma}) [\boldsymbol{\Sigma}_d]_{yy}^{-\frac{1}{2}} \\ &= [\boldsymbol{\Sigma}_d]_{yy}^{-\frac{1}{2}} \boldsymbol{B}\boldsymbol{\Phi}\boldsymbol{B}' [\boldsymbol{\Sigma}_d]_{yy}^{-\frac{1}{2}} + [\boldsymbol{\Sigma}_d]_{yy}^{-\frac{1}{2}} \boldsymbol{U}^2 [\boldsymbol{\Sigma}_d]_{yy}^{-\frac{1}{2}} \\ &\quad + [\boldsymbol{\Sigma}_d]_{yy}^{-\frac{1}{2}} \boldsymbol{\Delta}_{\Sigma} [\boldsymbol{\Sigma}_d]_{yy}^{-\frac{1}{2}} \end{aligned} \tag{3.94}$$

To simplify this expression, let us define the following matrices:

$$\boldsymbol{B}^* = [\boldsymbol{\Sigma}_d]_{yy}^{-\frac{1}{2}} \boldsymbol{B} \tag{3.95}$$

$$\boldsymbol{U}^{*^2} = [\boldsymbol{\Sigma}_d]_{yy}^{-\frac{1}{2}} \boldsymbol{U}^2 [\boldsymbol{\Sigma}_d]_{yy}^{-\frac{1}{2}} \tag{3.96}$$

$$\boldsymbol{\Delta}_R = [\boldsymbol{\Sigma}_d]_{yy}^{-\frac{1}{2}} \boldsymbol{\Delta}_\Sigma [\boldsymbol{\Sigma}_d]_{yy}^{-\frac{1}{2}} \tag{3.97}$$

Substituting from these three equations into Eq. (3.94) yields

$$\boldsymbol{R}_{yy} = \boldsymbol{B}^* \boldsymbol{\Phi} \boldsymbol{B}^{*\prime} + \boldsymbol{U}^{*^2} + \boldsymbol{\Delta}_R \tag{3.98}$$

This is an expression for the common factor model in terms of a population correlation matrix for the surface attributes. It has the same general form as Eq. (3.91), which expressed the model in terms of a population covariance matrix, but some of the terms in the equation have been rescaled. Note that, according to Eq. (3.95), the common factor weights obtained from a population covariance matrix would be rescaled in that each row of $\boldsymbol{B}$ would be divided by the population standard deviation of the corresponding surface attribute to yield the rows of $\boldsymbol{B}^*$. Matrix $\boldsymbol{B}^*$ would contain the common factor weights obtained from a population correlation matrix or, equivalently, from surface attributes standardized in the population. Note that the common factor intercorrelations, given in $\boldsymbol{\Phi}$, are not affected by this standardization. The unique variances, though, do undergo a rescaling, as shown in Eq. (3.96). The unique variances resulting from this rescaling represent population unique variances that would be obtained from a factor analysis of a population correlation matrix. Based on Eq. (3.16), corresponding population common variance, would be defined as follows:

$$\boldsymbol{H}^{*^2} = Diag(\boldsymbol{B}^* \boldsymbol{\Phi} \boldsymbol{B}^{*'}) \tag{3.99}$$

Employing Eqs. (3.16) and (3.95), this could be rewritten as follows:

$$\boldsymbol{H}^{*^2} = [\boldsymbol{\Sigma}_d]_{yy}^{-\frac{1}{2}} \boldsymbol{H}^2 [\boldsymbol{\Sigma}_d]_{yy}^{-\frac{1}{2}} \tag{3.100}$$

This shows how the population common variances are rescaled when they are obtained from a factor analysis of a population correlation matrix rather than a covariance matrix. It is important to understand the distinction between two sets of matrices: $\boldsymbol{U}^{*^2}$and $\boldsymbol{H}^{*^2}$ defined in Eqs. (3.96) and (3.100), versus $\widetilde{\boldsymbol{U}^2}$ and $\widetilde{\boldsymbol{H}^2}$ defined in Eqs. (3.33) and (3.25). Both sets of matrices represent rescaling of $\boldsymbol{U}^2$ and $\boldsymbol{H}^2$, which contain the population unique and common variances, respectively. However, the rescalings are slightly different. Matrices $\widetilde{\boldsymbol{U}^2}$ and $\widetilde{\boldsymbol{H}^2}$ are obtained by dividing entries in $\boldsymbol{U}^2$ and $\boldsymbol{H}^2$ by the population variances of the modeled attributes, thus yielding the population uniquenesses and communalities. On the other hand, matrices $\boldsymbol{U}^{*^2}$ and $\boldsymbol{H}^{*^2}$ are obtained by dividing entries in $\boldsymbol{U}^2$ and $\boldsymbol{H}^2$ by the population variances of the surface attributes. This operation merely rescales the variance components so that they represent surface attributes standardized in the population. It is interesting to note that in general $\boldsymbol{U}^{*^2}$ and $\boldsymbol{H}^{*^2}$ will not sum to an identity matrix because some of the population variance of the standardized surface attributes will arise from error of fit of the model. Finally, to complete the discussion of

Eq. (3.98) above, the matrix $\boldsymbol{\Delta}_R$ represents model error in the population correlation matrix and is defined by Eq. (3.97) via a rescaling of $\boldsymbol{\Delta}_\Sigma$.

In sum, these developments show that the analysis of population correlation matrices rather than population covariance matrices results in a rescaling of factor weights, unique variances, and the model error term, but does not affect common factor intercorrelations. Further theoretical and practical implications of the issue of analysis of correlation vs. covariance matrices will be discussed in subsequent chapters. We will continue to present major developments in the context of covariance matrices and to consider the analysis of correlation matrices as a special case.

3.7. The Components Analysis Model

In the previous sections in this chapter we have presented the algebraic and geometric representation of the common factor model. There exists a variation of this model which is commonly used in practice. This alternative approach is called components analysis. Though components analysis is often viewed as a variety of factor analysis, and in fact is often called factor analysis in the applied literature, it is important to recognize it as a distinct model and to understand how it is different from the common factor model.

In mathematical terms, the components analysis model can be defined by eliminating specific and error of measurement factors from the common factor model. Thus, if we delete the specific and error of measurement factors from Eq. (3.7), we obtain

$$\underline{\boldsymbol{z}} = \underline{\boldsymbol{x}}_\beta \boldsymbol{B}' \tag{3.101}$$

In terms of covariance matrices, we could write an oblique components model by modifying Eq. (3.35) to obtain

$$\boldsymbol{\Sigma}_{zz} = \boldsymbol{B}\boldsymbol{\Phi}\boldsymbol{B}' \tag{3.102}$$

In a similar fashion, Eq. (3.45) could be modified to yield an orthogonal components model as follows:

$$\boldsymbol{\Sigma}_{zz} = \boldsymbol{A}\boldsymbol{A}' \tag{3.103}$$

As represented in Eq. (3.72) the components model states that modeled attributes are linear combinations of *r* underlying dimensions. These dimensions will be called components and must be distinguished from common factors. Recall that a common factor represents a substantively meaningful internal attribute which affect more than one attribute in the battery and which partially explains why surface attributes are correlated. It is essential to recognize that the mathematical representation of common factors requires that specific and error of measurement

factors be represented separately, since the latter represent real phenomena which influence the attributes in a manner different from the common factors. Since the components model does not provide for these specific and error of measurement factors, the components in this model cannot be considered to be conceptually the same thing as common factors. This is due to the fact that the components will represent not only common variance, but also specific and error of measurement variance. Furthermore, it will not be possible to differentiate these various influences in a components solution.

Nevertheless, components analysis has often been used in practice and viewed as equivalent to, or at least an acceptable substitute for, common factor analysis. Supporters of this position tend to base their view on several points. The relative simplicity of the components model often is considered to be an advantage over factor analysis. The absence of specific and error of measurement factors from the model simplifies the process of fitting the model to the data in that it eliminates the need to estimate the unique influences. In addition, it is often argued that components analysis solutions are often found to be quite similar to common factor solutions. Finally, the major argument used for support of the components model is that it is possible to determine exactly the measures for the individuals on the components (i.e., the entries in $\underline{\boldsymbol{x}}$ in Eq. (3.101). This is not the case in the common factor model; that is, the measures for the individuals on the common factors, contained in vector $\underline{\boldsymbol{x}}$ in Eq. (3.7), cannot be determined exactly. These scores are said to be "indeterminate", and can only be estimated. Component scores, on the other hand, are determinate. These arguments have led many researchers to view components analysis as holding advantages over common factor analysis.

Our view, however, is that these arguments carry very little impact. Let us consider them in turn. First, though it is true that the components model is simpler, it must be recognized that this simplicity is gained at the cost of the realism of the model. The common factor model offers a much more realistic representation of attributes in that it explicitly recognizes and incorporates the existence of specific and error factors. The fact that such influences are ignored in the components model yields a simpler but very unrealistic model. Furthermore, this simpler model yields dimensions which, as pointed out above, cannot be conceived of as representing substantively meaningful common factors. Considering the argument that components solutions often closely approximate common factor solutions, this will be the case under certain conditions. It can be seen easily by comparing Eqs. (3.102) and (3.103) to Eqs. (3.36) and (3.45), respectively, that as the unique variances of the attributes become smaller the two models become more similar. When unique variances are zero, the two models are equivalent. However, when substantial unique variances are present, solutions obtained from the two models may not be very similar. Thus, the argument that solutions from the two models are quite similar will not hold under all conditions. Finally, considering the issue of indeterminacy, we do not consider this

as a fatal flaw in the common factor model for two reasons. First, as emphasized in Chapter 1, the objective of obtaining measures for the individuals on the common factors is not, in our view, a primary objective of factor analysis. The primary objective is to obtain an understanding of the nature of the common factors and the manner in which they influence the attributes. Given this view, the fact that the measures on the common factors are indeterminate is not a serious problem. Furthermore, as will be shown in Chapters 15 and 16, when a researcher does have a legitimate need for the scores on the common factors, there are procedures for estimating those scores, and to estimate them in such a way as to negate some of the effects of indeterminacy.

In summary, our view is that the components model is unrealistic and is not an acceptable substitute for the common factor model. Though there is an indeterminacy problem in the common factor model, this problem is not serious and can be dealt with when common factor scores are desired. This problem does not involve the central objective of factor analysis. In general then, we strongly oppose the use of components model and do not recommend it in practice. We support the use of the common factor model because it is based on a more realistic representation of the structure of the attributes and their relationships and because it offers a better chance to obtain an undergoing of the nature of the common factors and their effects on the attributes.

3.8. The Issue of Linearity

The common factor model is commonly referred to as a linear model. It is most important to achieve an understanding of the sense in which the model is linear. Let us consider three types of relationships which are relevant to the model: relationships of modeled attributes to each other; relationships of factors to each other; and relationships of modeled attributes to factors.

As defined in Eq. (3.7), the model states that the modeled attributes are linear combinations of the factors. If, in the real world, the surface attributes are non-linear functions of the factors, these relationships cannot be explicitly represented by the linear common factor model. In such a situation, the population covariance matrix for the surface attributes, $\boldsymbol{\Sigma}_{yy}$, could not be perfectly represented by the model. That is, given $\boldsymbol{\Sigma}_{yy}$, it would not be possible to find a solution such that all entries in matrix $\boldsymbol{\Delta}_{\Sigma}$, defined in Eq. (3.89), would be zero.

It is important to recognize, however, that the model makes no statement, either implicit or explicit, about the linearity of the relationships of the attributes to each other, or the factors to each other. In fact, the relationships can be non-linear. It is possible for the attributes to have non-linear relationships to each other, and for the factors to have non-linear relationships to each other, and for the model still to exactly represent the population covariance matrix for the surface attributes. This would simply require that the surface attributes be linear functions of the factors.

Two demonstrations will be presented to illustrate the points made here. The general procedure used in these demonstrations involves first constructing artificial data representing a population. That is, a population covariance matrix for surface attributes will be constructed such that the underlying relationships among attributes and factors have the desired properties of linearity or non-linearity. A common factor solution then will be obtained (by procedures to be described in Chapter 7) which fits the population covariance matrix as well as possible. It is not necessary for the reader to understand, at this point, the methodology for obtaining such a solution. The key issue in these demonstrations will be to determine whether or not the obtained solution fits the population covariance matrix perfectly; i.e., whether the matrix Δ_Σ contains all zeroes. If this is the case, then the linear factor model holds in the population; if not, then the relationships among the surface attributes cannot be represented exactly by the linear factor model.

The first demonstration is designed to show the effects of curvelinear relations among factors and among attributes. A population distribution of measures on two common factors was defined such that the relationship between the factors was curvelinear. This distribution is shown in Figure 3.5. In this figure, the axes represent two common factors. Each individual in the population was defined as falling at one of 25 discrete points, represented by the centers of the circles in the space. The values in the circles represent relative frequencies of observations at each corresponding point. The resulting distribution represents a population distribution of measures on two factors which are related in a curvelinear fashion. Scores on both factors had a mean of .644 and a standard deviation of .302; note that these original factors are unstandardized. The correlation between these factors was -.762. Each observation draw from this distribution would provide a vector $\underline{\boldsymbol{x}}$ of measures on common factors.

Next a population factor weight matrix $\boldsymbol{B}$ was defined. This matrix is presented in Table 3.2. Matrix $\boldsymbol{B}$ contains weights for the two unstandardized factors on 10 artificially defined attributes. Table 3.2 also includes the factor weight matrix which would be applied to standardized factor scores. These weights are obtained by multiplying each column of original factor weights by the standard deviation of the factor (.302). It is important to note that the weights were defined in such a way that attribute 1 is equivalent to factor 1 and attribute 10 is equivalent to factor 2. As a result, the relationship of attributes 1 and 10 will be identical to that represented in Figure 3.5. That is, in addition to there being a curvelinear relationship between the factors, there also will be a curvelinear relationship between attributes 1 and 10 (as well as other pairs of attributes).

Measures on two common factors for a hypothetical population of 100 individuals were generated so as to follow the distribution shown in Figure 3.5. Population measures on the surface attributes then were obtained according to Eq. (3.7). For simplicity, no specific or error

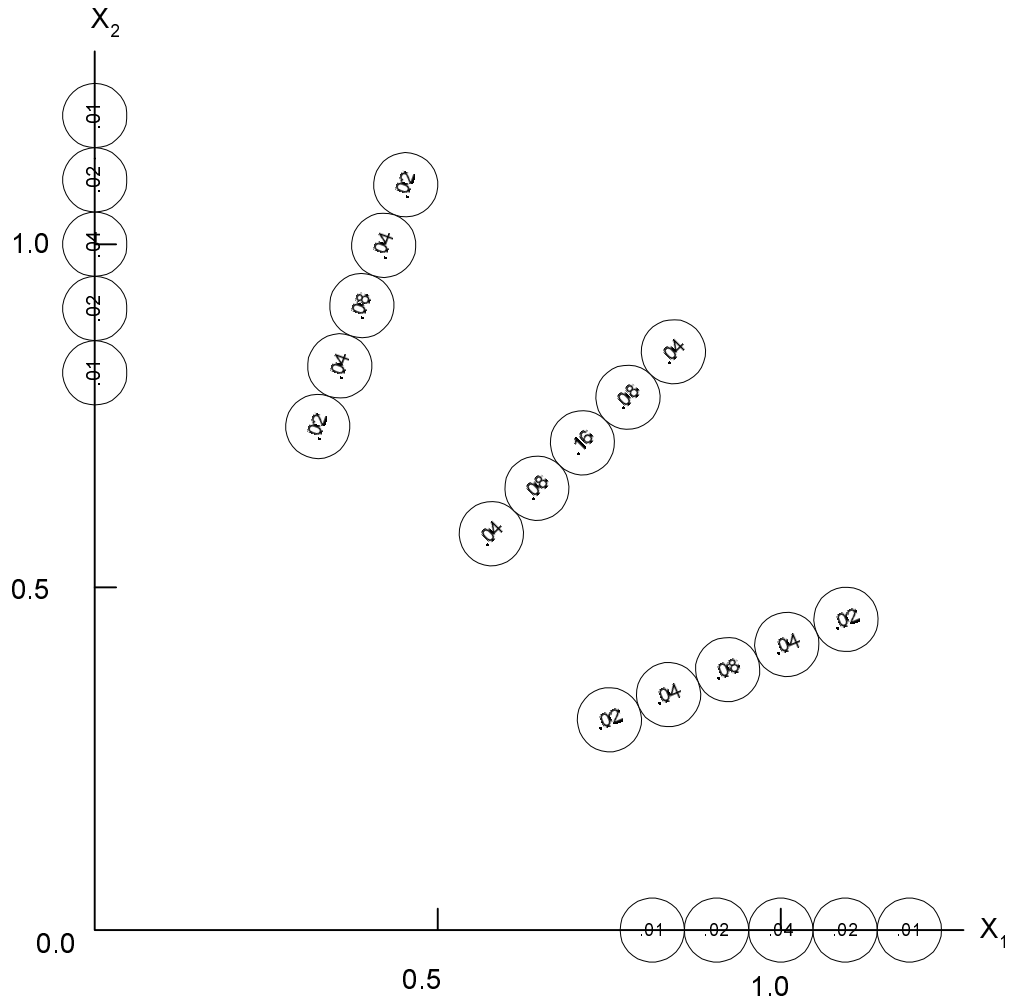

Figure 3.5: Population distribution of measures on two factors showing curvelinear relation between factors

Table 3.2

Input Factor Weight Matrix

for Demonstration of Curvelinear Factor Relations

	Unstandardized Factors		Standardized Factors	
Attribute	1	2	1	2
1	1.000	.000	.302	.000
2	1.000	.000	.302	.000
3	.960	.280	.290	.084
4	.800	.600	.241	.181
5	.800	.600	.241	.181
6	.710	.710	.214	.214
7	.600	.800	.181	.241
8	.280	.960	.084	.290
9	.000	1.000	.000	.302
10	.000	1.000	.000	.302

factors were included in the demonstration. Thus, the surface attributes were defined as linear combinations of the common factors, with the weights given in matrix $\boldsymbol{B}$. Next, a population covariance matrix for the surface attributes was computed from the population of measures on those attributes. The resulting matrix $\boldsymbol{\Sigma}_{yy}$ is presented in Table 3.3. Note that this covariance matrix has been produced from a case where there exist curvelinear relations among factors and among attributes, but the relations of the attributes to the factors are linear.

A factor solution then was obtained (via procedures to be discussed in Chapter 7). This yielded a solution containing two factors. The resulting orthogonal factor weight matrix is shown in Table 3.4. Considering this as matrix $\boldsymbol{A}$, it was found that the resulting matrix $\boldsymbol{\Sigma}_{zz}$ produced via Eq. (3.45), recalling that unique variances are zero in this demonstration, was identical to the population covariance matrix $\boldsymbol{\Sigma}_{yy}$ in Table 3.3. In other words, the matrix $\boldsymbol{\Delta}_{\Sigma}$, whose entries represent lack of correspondence between the model and the population, contained all zeroes. Furthermore, it was found that the orthogonal solution represented by matrix $\boldsymbol{A}$ in Table 3.4 could be transformed to match the original input factor weights given in Table 3.2. Note that since the obtained factor weights represent standardized factors, the transformation of the obtained solution should yield the input weights for standardized factors. Following the procedures in section 3.3 above, this transformation was achieved by passing the first trait vector and the second trait vector through the last attribute vector. The resulting trait matrix $\boldsymbol{T}$ is shown in Table 3.4. Based on this trait matrix, we then obtained the intercorrelation matrix $\boldsymbol{\Phi}$ and the weight matrix $\boldsymbol{B}$ for the transformed factors by Eqs. (3.42) and (3.44), respectively. The resulting matrices are shown in Table 3.4 and can be seen to be identical to the corresponding input information. These findings of perfect fit and exact recovery of the input factors indicate that the linear factor model holds exactly in this population, despite the presence of curvelinear relations among factors and among attributes. The fact that the relations of the attributes to the factors are linear allows for this perfect correspondence between the factor model and the population.

The second demonstration involves the case of curvelinear relations of attributes to factors. This demonstration makes use of "binary" attributes; i.e., attributes which can take on only two possible values. The most common example of such an attribute is a test item, where responses to the item are scored as correct (1) or incorrect (0). Such items are very commonly used, and are normally considered to be surface measures of underlying common factors. For instance, a single addition item on a test could be considered as a binary measure of an underlying continuous internal attribute called numerical facility. An interesting aspect of this view is that relations of binary surface attributes to common factors must be considered to be curvelinear. This is due to the nature of the binary attributes. A linear relationship of such an attribute to a common factor would necessarily result in values of the binary attribute outside the

Table 3.3

Population Covariance Matrix for 10 Attributes

for Demonstration of Curvelinear Factor Relations

Attribute	1	2	3	4	5	6	7	8	9	10
1	.091									
2	.091	.091								
3	.068	.068	.054							
4	.031	.031	.030	.024						
5	.031	.031	.030	.024	.024					
6	.015	.015	.019	.022	.022	.022				
7	-.001	-.001	.008	.018	.018	.022	.024			
8	-.041	-.041	-.020	.008	.008	.019	.030	.054		
9	-.069	-.069	-.041	-.001	-.001	.015	.031	.068	.091	
10	-.069	-.069	-.041	-.001	-.001	.015	.031	.068	.091	.091

Table 3.4

Obtain Factor Solution for Demonstration of Curvelinear Factor Relation

Weight Matrix A			Trait Matrix T			Tranformed Weight Matrix $B=AT^{-1}$		
	1	2		1	2		1	2
1	.301	-.021	1	.997	-.071	1	.302	.000
2	.301	-.021	2	-.071	.699	2	.302	.000
3	.229	.039				3	.290	.084
4	.111	.110				4	.241	.181
5	.111	.110				5	.241	.181
6	.061	.135	Factor Intercorrelation			6	.214	.214
7	.008	.156	Matrix $\Phi=TT'$			7	.181	.214
8	-.123	.197		1	2	8	.084	.290
9	-.216	.211	1	1.000	-.762	9	.000	.302
10	-.216	.211	2	-.762	1.000	10	.000	.302

permissible values (e.g., 0 and 1). Thus, the view that a binary surface attribute is a measure of an underlying continuous factor implies that the relationship of the attribute to the factor must be curvelinear.

To demonstrate the effects of such relationships, an artificial population covariance matrix was constructed for 10 binary items. The items were defined as varying in difficulty, where difficulty is defined as the proportion of the population which provides the correct response to the item. Thus, lower values of this index represent more difficult items. Item difficulties for the ten items are shown in Table 3.5. Each of these items was considered to be a measure of a single common factor. Thus, the theoretical weights for the factor on the items were all unity. This implies that the theoretical correlations among these 10 items would all be unity. Given that these items would form a perfect Guttman scale (i.e., if an individual produces the correct response to a given item, the individual will also produce the correct response to all easier items), it is possible to obtain the actual population correlation matrix for the items. This matrix, shown in Table 3.5, contains phi-coefficients representing the actual relationship for each pair of binary items.

The correlation matrix in Table 3.5 represents a standardized version of a covariance matrix $\mathbf{\Sigma}_{yy}$ for surface attributes. Factor analysis procedures were applied to this matrix to determine whether a solution could be obtained which fit the covariances among these items perfectly. Despite the fact that the items are defined as being measures of a single factor, a one-factor solution does not provide perfect fit to $\mathbf{\Sigma}_{yy}$. In fact, $\mathbf{\Sigma}_{yy}$ could not be fit perfectly using any small number of factors. For illustrative purposes, a two-factor orthogonal solution is shown in Table 3.6. This solution fits fairly well; that is, the elements of the resulting $\mathbf{\Delta}_{\Sigma}$ matrix are not large. However, the important point is that it does not fit perfectly. The reason for this is that the linear model cannot account for the relationships represented in $\mathbf{\Sigma}_{yy}$ because these relationships arise from an underlying structure where the effects of the factors on the surface attributes are curvelinear.

It is interesting to note the pattern of weights found in the solution shown in Table 3.6. Factor 1 seems to differentiate items of medium difficulty from items of extreme difficulty (i.e., either very difficult or very easy). Factor 2 shows a continuum of weights from low to high according to the difficulty of the items. Such factors are routinely observed when factor analyzing binary data in practice, and have often been called "difficulty factors". For some time they have been thought to be artifacts related to the presence of items of varying difficulty (Gourlay, 1951; Guilford, 1941; Wherry & Gaylord, 1951).. However, they have been recognized more recently as arising from non-linearities in the relationships of the items to the factors (Gibson, 1959, 1960; McDonald, 1965a). The general point to recognize is that when

Table 3.5
Item Difficulties and Intercorrelations for Demonstration of
Curvelinear Relations of Attributes to Factors

Item Difficulties

Item	1	2	3	4	5	6	7	8	9	10
Difficulties	.05	.15	.25	.35	.45	.55	.65	.75	.85	.95

Item Intercorrelations

Item	1	2	3	4	5	6	7	8	9	10
1	1.000									
2	.546	1.000								
3	.397	.728	1.000							
4	.313	.572	.787	1.000						
5	.254	.464	.638	.811	1.000					
6	.208	.380	.522	.664	.818	1.000				
7	.068	.308	.424	.538	.664	.811	1.000			
8	.132	.243	.333	.424	.522	.638	.787	1.000		
9	.096	.176	.243	.308	.380	.464	.572	.728	1.000	
10	.053	.096	.132	.168	.208	.254	.313	.397	.546	1.000

Table 3.6
Two-Factor Solution for Demonstration of
Curvelinear Relations of Attributes to Tactors

	Factor	
Item	1	2
1	.345	-.330
2	.590	-.489
3	.742	-.486
4	.813	-.318
5	.840	-.109
6	.840	.106
7	.814	.315
8	.744	.483
9	.591	.487
10	.346	.329

there exist curvelinear relations of attributes to factors, the linear factor model is not appropriate. If used, it will yield artifactual factors arising from non-linearities in the relations of attributes to factors.

To summarize the issue of linearity in the common factor model, it should be understood that the model assumes that the relations of the attributes to the factors are linear, but allows for non-linear relations of attributes to each other, or of factors to each other. The model can be used effectively when non-linear relations exist among attributes or among factors, as long as relations of the attributes to the factors are linear. Of course, in the real world the relations of the attributes to the factors in a given domain almost surely would not be exactly linear. It is likely that such relations would be at least slightly nonlinear in many cases. However, as long as these relations are approximately linear, the linear common factor model generally would provide an adequate and useful representation of the relations. But when the relations of the attributes to the factors are severely nonlinear, problems arise in the use of the model. As a result of the recognition of this phenomenon, non-linear varieties of factor analysis have been developed (e.g., McDonald, 1967). However, such developments are beyond the scope of the present text, which focuses on the linear common factor model.

CHAPTER 4
THE COMMON FACTOR MODEL IN THE SAMPLE

From
Exploratory Factor Analysis
Ledyard R Tucker
and
Robert C. MacCallum

CHAPTER 4
THE COMMON FACTOR MODEL IN THE SAMPLE

4.0. Introduction

In Chapter 3, we presented a mathematical representation of common factor theory in the context of a population. As was noted at the time, this was a theoretical presentation, which did not treat issues involved in sampling. Obviously, the application of factor analysis in practice requires that the model be expressed in the context of a sample, and that methods be developed for fitting the model to sample data. The purpose of the present chapter is to show how the common factor model can be represented in a sample. The mathematical framework to be developed for this purpose will serve as the basis for the treatment in later chapters of the problem of estimating the parameters of the model.

4.1. Representation of the Common Factor Model in a Sample

The representation of the common factor model in a sample will be developed by employing concepts and relationships presented in Chapter 3, which showed how the model is defined in a population. We will examine how these concepts and relationships are affected when we attempt to apply the same model to a sample drawn from the population. Given this objective, it will be useful to begin by briefly reviewing a few important equations and issues discussed in Chapter 3. A number of equations from Chapter 3 will be re-stated here for convenience and review. Considering first the expression of the model in terms of modeled attributes, recall that the population covariance matrix for the modeled attributes, $\boldsymbol{\Sigma}_{zz}$, can be defined as having the following structure:

$$\boldsymbol{\Sigma}_{zz} = \boldsymbol{\Omega}_{\beta u}\boldsymbol{\Sigma}_{xx}\boldsymbol{\Omega}'_{\beta u} \tag{4.1}$$

In this equation, $\boldsymbol{\Omega}_{\beta u}$ is the population factor weight matrix containing submatrices $\boldsymbol{B}$, whose entries are the weights for the common factors, and $\boldsymbol{U}$, whose entries are the weights for the unique factors. The representation of $\boldsymbol{\Omega}_{\beta u}$ as a supermatrix is given by

$$\boldsymbol{\Omega}_{\beta u} = [\boldsymbol{B}, \boldsymbol{U}] \tag{4.2}$$

The matrix $\boldsymbol{\Sigma}_{xx}$ is the covariance matrix for the common and unique factors, and can be represented as a supermatrix of the following form:

$$\boldsymbol{\Sigma}_{xx} = \begin{bmatrix} \boldsymbol{\Sigma}_{\beta\beta} & \boldsymbol{\Sigma}_{\beta u} \\ \boldsymbol{\Sigma}_{u\beta} & \boldsymbol{\Sigma}_{uu} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I} \end{bmatrix} \tag{4.3}$$

As discussed in Chapter 3, given the nature of the factors and the imposed condition that they are in standardized form in the population, matrix $\boldsymbol{\Phi}$ will be a correlation matrix for the common factors, and the unique factors will be uncorrelated with each other and with the common factors. Substituting from Eqs. (4.2) and (4.3) into (4. 1) yields the oblique common factor model:

$$\boldsymbol{\Sigma}_{zz} = \boldsymbol{B\Phi B}' + \boldsymbol{U}^2 \tag{4.4}$$

The orthogonal common factor model can be represented as

$$\boldsymbol{\Sigma}_{zz} = \boldsymbol{AA}' + \boldsymbol{U}^2 \tag{4.5}$$

Chapter 3 also defined the framework for the transformation of an orthogonal solution into an oblique solution. Considering only the weight matrices for present purposes, the transformation process requires defining a population trait matrix $\boldsymbol{T}$, and the relationships between the orthogonal and oblique solutions are given as follows:

$$\boldsymbol{B} = \boldsymbol{AT}^{-1} \tag{4.6}$$

$$\boldsymbol{\Phi} = \boldsymbol{TT}' \tag{4.7}$$

Another important issue emphasized in Chapter 3 involved the distinction between modeled attributes, contained in vector $\underline{\mathbf{z}}$, and surface attributes, contained in vector $\underline{\boldsymbol{y}}$. Recall that the relationship is given by

$$\underline{\boldsymbol{y}} = \underline{\mathbf{z}} + \ddot{\underline{\mathbf{z}}} \tag{4.8}$$

where $\ddot{\underline{\mathbf{z}}}$ represents that part of the surface attributes that is not consistent with the common factor model. In terms of covariance matrices, the following relationship was developed in Chapter 3:

$$\boldsymbol{\Sigma}_{yy} = \boldsymbol{\Sigma}_{zz} + \boldsymbol{\Delta}_{\Sigma} \tag{4.9}$$

$$\boldsymbol{\Delta}_{\Sigma} = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{zz} \tag{4.10}$$

Finally, by substitution from Eq. (4.4) into Eq. (4.9) we obtain

$$\boldsymbol{\Sigma}_{yy} = \boldsymbol{B\Phi B}' + \boldsymbol{U}^2 + \boldsymbol{\Delta}_{\Sigma} \tag{4.11}$$

This expression states the relationship between the population variances and covariances for the surface attributes, in $\boldsymbol{\Sigma}_{yy}$, and the parameters of the common factor model, in $\boldsymbol{B}$, $\boldsymbol{\Phi}$, and $\boldsymbol{U}^2$. As explained in Chapter 3, the entries in $\boldsymbol{\Delta}_{\Sigma}$ represent model error, in the sense that they reflect the degree to which the population covariances among the surface attributes cannot be accounted for by the common factor model. The reader also should recall the relationship of these developments to the issue of fitting the common factor model to data. If the model were being fit

to $\boldsymbol{\Sigma}_{yy}$ as represented in Eq. (4.11), the objective would be to find a solution which would yield an optimal $\boldsymbol{\Delta}_{\Sigma}$. Alternative approaches to defining what is an optimal $\boldsymbol{\Delta}_{\Sigma}$ would yield different solutions; i.e., different $\boldsymbol{B}$, $\boldsymbol{\Phi}$, and $\boldsymbol{U}^2$ matrices, and in turn different $\boldsymbol{\Sigma}_{zz}$ and $\boldsymbol{\Delta}_{\Sigma}$ matrices, obtained by Eqs. (4.4) and (4.10), respectively. Thus, it is not the case that there is a single true $\boldsymbol{\Sigma}_{zz}$ matrix, and corresponding $\boldsymbol{\Delta}_{\Sigma}$ matrix; rather, different $\boldsymbol{\Sigma}_{zz}$ and $\boldsymbol{\Delta}_{\Sigma}$ matrices would be defined for each different approach to optimizing $\boldsymbol{\Delta}_{\Sigma}$.

Though this is an important issue, it should be recognized that, at this point, it is stated in terms of fitting the model to a population covariance matrix. Obviously, to conduct factor analysis in practice it is necessary to represent the model in terms of a sample covariance matrix and to solve the problem of how to fit the model to such data. We will now consider the problem of expressing the model in terms of a sample covariance matrix. We wish to emphasize that this process is not as simple and straightforward as it is often taken to be. In our view, there are a number of subtle issues inherent in this problem which are often oversimplified, or completely overlooked. Of greatest importance are the distinctions between model error and sampling error, and between modeled attributes and surface attributes. We will attempt to present a framework which allows for explicit representation of these issues, along with a careful treatment of the distinction between populations and samples.

To begin, let us assume that a sample of N observations has been obtained, and that a vector $\boldsymbol{y}$ containing measures on surface attributes has been obtained from each individual. Each such vector still can be conceived of as arising from two sources, as defined in Eq. (4.8); i.e., a portion consistent with the common factor model, and a portion inconsistent with the model. This distinction between modeled attributes and surface attributes is as valid in the sample as it is in the population. Therefore, we can define a sample covariance matrix $\boldsymbol{C}_{zz}$ whose entries represent sample variances and covariances for the modeled attributes. Keep in mind that $\boldsymbol{C}_{zz}$ is not directly observable, but that it represents that portion of the sample variances and covariances of the surface attributes which can be accounted for by the common factor model. Following the form of Eq. (4.1), we can express the common factor model in terms of $\boldsymbol{C}_{zz}$ as follows:

$$\boldsymbol{C}_{zz} = \boldsymbol{\Omega}_{\beta\mu}\boldsymbol{C}_{xx}\boldsymbol{\Omega}'_{\beta\mu} \tag{4.12}$$

In this equation, $\boldsymbol{C}_{xx}$ represents a sample covariance matrix for the factors; i.e., this is the sample matrix corresponding to $\boldsymbol{\Sigma}_{xx}$ and would take the following form:

$$\boldsymbol{C}_{xx} = \begin{bmatrix} \boldsymbol{C}_{\beta\beta} & \boldsymbol{C}_{\beta\mu} \\ \boldsymbol{C}_{\mu\beta} & \boldsymbol{C}_{\mu\mu} \end{bmatrix} \tag{4.13}$$

A comparison of Eqs. (4.12) and (4.13) to the corresponding equations for the population, given in (4.1) and (4.3) is of central important. It is essential to recognize and understand that the

weight matrix for the factors, given by $\boldsymbol{\Omega}_{\beta\mu}$, is shown as being the same in the sample as in the population. A consideration of what these factor weights represent will verify that this must be the case. Recall that the fundamental common factor model, as shown in Eq. (3.3), states that the entries in any sampled vector $\underline{\boldsymbol{z}}$ will be linear combinations of the entries in a vector $\underline{\boldsymbol{x}}$, which contains the measures on the factors. The weights defining these linear combinations, which are the entries in $\boldsymbol{\Omega}_{\beta\mu}$, will be the same for all such vectors. Thus, the weight matrix relating factors to modeled attributes is the same in a sample as in the full population. This relationship, however, does not imply that a common factor weight matrix obtained from an analysis of sample data will be identical to the population common factor weight matrix. In fact, as will be seen in subsequent developments in this section, several effects preclude exact determination of population common factor weights from a sample. The relationship stated here simply indicates that the sampling process per se is not the problem. The sampling process does however have an effect on the variances and covariances of the factors, both common and unique. Due to the chance characteristics of the sample, the variances and covariances of the common and unique factors in the sample will almost surely be different than the corresponding values in the population. As a result, the matrix $\boldsymbol{C}_{xx}$ will be different from its population counterpart $\boldsymbol{\Sigma}_{xx}$.

A very important representation of the model in a sample can be obtained by substituting from Eqs. (4.2) and (4.13) into (4.12). This yields the following:

$$\begin{aligned}\boldsymbol{C}_{zz} &= [\boldsymbol{B}, \boldsymbol{U}]\begin{bmatrix}\boldsymbol{C}_{\beta\beta} & \boldsymbol{C}_{\beta\mu}\\ \boldsymbol{C}_{\mu\beta} & \boldsymbol{C}_{\mu\mu}\end{bmatrix}\begin{bmatrix}\boldsymbol{B}'\\ \boldsymbol{U}'\end{bmatrix}\\ &= \boldsymbol{B}\boldsymbol{C}_{\beta\beta}\boldsymbol{B}' + \boldsymbol{B}\boldsymbol{C}_{\beta\mu}\boldsymbol{U}' + \boldsymbol{U}\boldsymbol{C}_{\mu\beta}\boldsymbol{B}' + \boldsymbol{U}\boldsymbol{C}_{\mu\mu}\boldsymbol{U}' \quad \mathbf{(4.14)}\end{aligned}$$

This expression can be regarded as representing a full statement of the common factor model in the sample. However, this is not the model actually employed in practice. The model employed in practice is obtained as a result of some simplifying assumptions applied to the expression in Eq. (4.14). To see how this is accomplished, let us consider the hypothetical case in which the factor scores in the sample have some of the properties which characterize the factor scores in the population. In particular, suppose that in the sample the unique factors were uncorrelated with each other and with the common factors. This would result in all entries in matrices $\boldsymbol{C}_{\beta\mu}$ and $\boldsymbol{C}_{\mu\beta}$ being zero, and matrix $\boldsymbol{C}_{\mu\mu}$ being diagonal. Furthermore, let us suppose that the unique factors are scaled so that they have unit variances, thus yielding a matrix $\boldsymbol{C}_{\mu\mu}$ which is an identity matrix. Under these conditions, the model in Eq. (4.14) would become

$$\boldsymbol{C}_{zz} = \boldsymbol{B}\boldsymbol{C}_{\beta\beta}\boldsymbol{B}' + \boldsymbol{U}^2 \tag{4.15}$$

If this model were a valid representation of the true structure of $\boldsymbol{C}_{zz}$, then it would imply that the sampling process has no effect on the common factor weights or unique variances, and only

affects the common factor covariances. However, it is critical to recognize that the hypothetical conditions which yielded this equation will almost certainly not hold in practice. That is, the process of sampling will almost surely give rise to a sample in which correlations of the unique factors with each other and with the common factors are not exactly zero. As a result, matrices $\boldsymbol{C}_{\beta\mu}$ and $\boldsymbol{C}_{\mu\beta}$ will not actually be zero, and matrix $\boldsymbol{C}_{\mu\mu}$ will not actually be diagonal. This in turn implies that matrix $\boldsymbol{C}_{zz}$ cannot be exactly represented by a model which is based on these assumptions; i.e., the model in Eq. (4.15). This context is helpful in moving toward an understanding of how factor analysis is conducted in practice, and of what sources contribute to a lack of perfect fit of the model to real data. In effect, and this is a very important point, the representation of the common factor model in the sample is defined in practice as in Eq. (4.15). The implication is that in practice we treat the sampling issue as if unique factors were uncorrelated with each other and with the common factors in the sample. Since this will not in fact be the case in the real world, the matrix $\boldsymbol{C}_{zz}$ cannot actually be fit exactly by the model in Eq. (4.15). To represent this mathematically, it is useful to define a matrix $\boldsymbol{\Delta}_z$ whose entries represent a lack of fit in that model. The model could then more appropriately be written as

$$\boldsymbol{C}_{zz} = \boldsymbol{B}\boldsymbol{C}_{\beta\beta}\boldsymbol{B}' + \boldsymbol{U}^2 + \boldsymbol{\Delta}_z \tag{4.16}$$

The entries in $\boldsymbol{\Delta}_z$ can be viewed as reflecting a primary source of sampling error. That is, if the characteristics of the sample were identical to those of the population, then all entries in $\boldsymbol{\Delta}_z$ would be zero, and the model in Eq. (4.15) would be valid. However, the phenomenon of sampling error gives rise to samples with properties not identical to the population. In the present context, this causes the model in Eq. (4.15) to be inexact, with the degree of error reflected in the entries in $\boldsymbol{\Delta}_z$ in Eq. (4.16). In other words, the model in Eq. (4.15) is based on simplifying assumptions which generally do not hold in practice, and this fact gives rise to a primary source of sampling error in the model, as represented in Eq. (4.16). If the assumptions did hold exactly in a sample, then there would be no sampling error of this type; in that case, all entries in $\boldsymbol{\Delta}_z$ in Eq. (4.16) would be zero and the simplified model in Eq. (4.15) would be exactly correct. This framework thus shows explicitly the manner in which this source of sampling error impacts on the common factor model. Other potential source of sampling error will be considered later in this chapter and in Chapter 5.

A remaining aspect of this representation of the model which requires attention is the issue of the standardization of the factors. As expressed in Eq. (4.16) the model is stated in terms of unstandardized common factors; that is, though we have considered the common factors to be standardized in the population, they almost surely will not be such in the sample. This can be easily alleviated by converting the covariance matrix $\boldsymbol{C}_{\beta\beta}$ in Eq. (4.16) into a correlation matrix. In fact, this standardization of the common factors in the sample is the most common approach

taken to resolving a critical problem arising in the process of fitting the common factor model to real data. This problem, called the identification problem, involves the fact that when the model is expressed as in Eq. (4.16), the scales of the common factors are completely arbitrary. In order to estimate the parameters of the model, it is necessary to establish a scale, or unit of measurement, for the factors in the sample. The simplest and most common way to achieve this is to define the common factors as being standardized in the sample. This entire issue will be discussed in more detail in Chapters 8 and 9. For present purposes, let us define a mathematical representation of the model where the common factors are standardized in the sample.

Let us first define a diagonal matrix $[\boldsymbol{C_d}]_{\beta\beta}$ containing the variances of the common factors in the sample; that is,

$$[\boldsymbol{C_d}]_{\beta\beta} = \boldsymbol{Diag}(\boldsymbol{C_{\beta\beta}}) \tag{4.17}$$

We can then define a matrix $\widetilde{\boldsymbol{C}}_{\beta\beta}$ which is the correlation matrix for the common factor in the sample as follows:

$$\widetilde{\boldsymbol{C}}_{\beta\beta} = [\boldsymbol{C_d}]_{\beta\beta}^{-\frac{1}{2}} \boldsymbol{C_{\beta\beta}} [\boldsymbol{C_d}]_{\beta\beta}^{-\frac{1}{2}} \tag{4.18}$$

Let us next define a matrix $\widetilde{\boldsymbol{B}}$ as follows:

$$\widetilde{\boldsymbol{B}} = \boldsymbol{B}[\boldsymbol{C_d}]_{\beta\beta}^{\frac{1}{2}} \tag{4.19}$$

Note that $\widetilde{\boldsymbol{B}}$ contains population factor weights which have been re-scaled so as to correspond to common factors standardized in the sample. That is, the columns of $\boldsymbol{B}$, the matrix of population common factor weights, are each multiplied by the sample standard deviation of the corresponding common factor to produce the weights in $\widetilde{\boldsymbol{B}}$. Given these definitions, it can be seen that

$$\begin{aligned} \widetilde{\boldsymbol{B}}\, \widetilde{\boldsymbol{C}}_{\beta\beta} \widetilde{\boldsymbol{B}}' &= \boldsymbol{B}[\boldsymbol{C_d}]_{\beta\beta}^{\frac{1}{2}} [\boldsymbol{C_d}]_{\beta\beta}^{-\frac{1}{2}} \boldsymbol{C_{\beta\beta}} [\boldsymbol{C_d}]_{\beta\beta}^{-\frac{1}{2}} [\boldsymbol{C_d}]_{\beta\beta}^{\frac{1}{2}} \boldsymbol{B}' \\ &= \boldsymbol{B C_{\beta\beta} B}' \end{aligned} \tag{4.20}$$

This relationship is important because it allows us to re-write the model in Eq. (4.16) in terms of common factors standardized in the sample. That is, by substituting from Eq. (4.20) into Eq. (4.16) we obtain

$$\boldsymbol{C_{zz}} = \widetilde{\boldsymbol{B}}\, \widetilde{\boldsymbol{C}}_{\beta\beta} \widetilde{\boldsymbol{B}}' + \boldsymbol{U^2} + \boldsymbol{\Delta}_z \tag{4.21}$$

Note that $\boldsymbol{U^2}$ and $\boldsymbol{\Delta}_z$ are the same in Eqs. (4.21) and (4.16); the distinction between the two expressions is that the common factors in the latter are defined as standardized in the sample.

Let us now consider the final step in expressing the model in terms of observable data; i.e., the step from modeled attributes to surface attributes. The preceding developments were set in the context of modeled attributes, with the common factor model being stated in terms of the sample covariance matrix for the modeled attributes, $\boldsymbol{C_{zz}}$. As noted above, the objective is to express the model in terms of the sample covariance matrix for the surface attributes, which will be designated $\boldsymbol{C_{yy}}$. We must first consider the relationship between $\boldsymbol{C_{zz}}$ and $\boldsymbol{C_{yy}}$. This relationship follows the same pattern as the relationship between $\boldsymbol{\Sigma_{zz}}$ and $\boldsymbol{\Sigma_{yy}}$, which was derived in
Chapter 3 and is shown in Eq. (3.88). By following the same procedure, we can state the following relationship:

$$\boldsymbol{C_{yy} = C_{zz} + C_{z\ddot{z}} + C_{\ddot{z}z} + C_{\ddot{z}\ddot{z}}} \tag{4.22}$$

Following the developments presented in Chapter 3 for the population, let us define a matrix $\boldsymbol{\Delta_c}$ as follows:

$$\boldsymbol{\Delta_c = C_{z\ddot{z}} + C_{\ddot{z}z} + C_{\ddot{z}\ddot{z}}} \tag{4.23}$$

Substituting from Eq. (4.23) into Eq. (4.22), we can write

$$\boldsymbol{C_{yy} = C_{zz} + \Delta_c} \tag{4.24}$$

The matrix $\boldsymbol{\Delta_c}$ could alternatively be defined as

$$\boldsymbol{\Delta_c = C_{yy} - C_{zz}} \tag{4.25}$$

The elements in $\boldsymbol{\Delta_c}$ can be viewed as reflecting a lack of correspondence between $\boldsymbol{C_{yy}}$ and $\boldsymbol{C_{zz}}$. It is very interesting to examine the source of this lack of correspondence. The entries in $\boldsymbol{\Delta_c}$ can be thought of as arising from model error. That is, since the matrix $\boldsymbol{C_{zz}}$ represents that portion of the sample variances and covariances of the surface attributes that can be accounted for by the model, the entries in $\boldsymbol{\Delta_c}$ arise from lack of fit of the model to the sample data. Thus, matrix $\boldsymbol{\Delta_c}$, representing model error in the sample, is analogous to matrix $\boldsymbol{\Delta_\Sigma}$, defined in Eq. (4.10), which represents model error in the population.

It is now possible to combine preceding developments to achieve a unified view of model error and sampling error, and to produce a representation of the common factor model in terms of a sample covariance matrix for surface attributes. By substituting from Eq. (4.21) into Eq. (4.24), we obtain

$$\boldsymbol{C_{yy} = (\widetilde{B}\,\widetilde{C}_{\beta\beta}\widetilde{B}' + U^2 + \Delta_z) + \Delta_c} \tag{4.26}$$

If we define a matrix $\boldsymbol{\Delta_y}$ according to

$$\boldsymbol{\Delta}_y = \boldsymbol{\Delta}_z + \boldsymbol{\Delta}_c \tag{4.27}$$

and substitute from Eq. (4.27) into Eq. (4.26), we obtain

$$\boldsymbol{C}_{yy} = \widetilde{\boldsymbol{B}}\,\widetilde{\boldsymbol{C}}_{\beta\beta}\widetilde{\boldsymbol{B}}' + \boldsymbol{U}^2 + \boldsymbol{\Delta}_y \tag{4.28}$$

Given the interpretation of $\boldsymbol{\Delta}_z$, defined in conjunction with Eq. (4.16), as reflecting the impact of sampling error, and of $\boldsymbol{\Delta}_c$ as reflecting the impact of model error in the sample, it can be seen that the entries in $\boldsymbol{\Delta}_y$ reflect both sources of error. That is, the lack of fit of the model arises from two sources: (a) the fact that the simplifying assumptions made in conjunction with Eq. (4.14) will not hold exactly in practice, thus giving rise to sampling error; and (b) the fact that the common factor model is not expected to fit exactly the covariances of the surface attributes, thus giving rise to model error. It is not possible in practice to actually separate sampling error from model error. This would require knowing true values of population parameters. However, the developments just presented do serve to separate these two sources of error in theoretical terms, and to provide an explanation of their separate and combined impact on the representation of the model in a sample.

Eq. (4.28) is of central importance in understanding and solving the problem of fitting the common factor model to sample data. An objective of Chapters 3 and 4 has been to reach this point of representing the structure of a sample covariance matrix for surface attributes in terms of the parameters of the common factor model. This is achieved by Eq. (4.28). The equation expresses the sample covariance matrix for the surface attributes as a function of common factor weights, common factor intercorrelations, unique variances, and a "lack of fit" term. Eq. (4.28) is the sample equivalent of Eq. (4.11), which expresses the population covariance matrix for the surface attributes as a function of the model parameters. The representation of the model in the sample is the basis for developing methods for fitting the model to sample covariance matrices. In basic terms, the problem is as follows: given matrix $\boldsymbol{C}_{yy}$, we wish to obtain coefficient matrices $\widetilde{\boldsymbol{B}}$, $\widetilde{\boldsymbol{C}}_{\beta\beta}$, and $\boldsymbol{U}^2$ so that the entries in $\boldsymbol{\Delta}_y$ are in some sense made optimally small. There is a close relation between the sense in which $\boldsymbol{\Delta}_\Sigma$ is optimal and the sense in which $\boldsymbol{\Delta}_y$ is optimal. Remember that the matrices $\boldsymbol{B}, \boldsymbol{\Phi}$, and $\boldsymbol{U}^2$ depend on the way in which $\boldsymbol{\Delta}_\Sigma$ is optimal (see Eq. (4.11) and ensuing discussion). A consequence is that the population parameters which are estimated by matrices $\widetilde{\boldsymbol{B}}$, $\widetilde{\boldsymbol{C}}_{\beta\beta}$, and $\boldsymbol{U}^2$ determined from $\boldsymbol{C}_{yy}$ depend on the sense in which $\boldsymbol{\Delta}_\Sigma$ and $\boldsymbol{\Delta}_y$ are optimal. In the following discussion, the sense in which optimality is defined will be taken to be fixed.

Several important aspects of this problem must be carefully considered. The first involves the nature of the parameters to be estimated. Note that entries in $\widetilde{\boldsymbol{B}}$, $\widetilde{\boldsymbol{C}}_{\beta\beta}$, and $\boldsymbol{U}^2$ are not all

pure population parameters unaffected by characteristics of the sample. Specifically, the entries in $\widetilde{\boldsymbol{B}}$ are population common factor weights which have been rescaled to represent common factors standardized in the sample. An obtained estimate of $\widetilde{\boldsymbol{B}}$ can be viewed as an estimate of $\boldsymbol{B}$, with some error of estimation arising from the fact that the factors are standardized in the sample rather than the population. In a similar fashion, the entries in $\widetilde{\boldsymbol{C}}_{\beta\beta}$ are actually sample correlations among the common factors. However, these sample values cannot be observed or determined exactly because they are defined in terms of the population common factor weights in $\boldsymbol{B}$. An obtained estimate of $\widetilde{\boldsymbol{C}}_{\beta\beta}$ can be viewed as an estimate of $\boldsymbol{\Phi}$; again, some error of estimation of $\boldsymbol{\Phi}$ arises from the fact that the factors are standardized in the sample rather than the population.

This view helps to achieve an understanding of what is being estimated when the model given in Eq. (4.28) is fit to sample data. Though the parameters in that model are represented as $\widetilde{\boldsymbol{B}}$, $\widetilde{\boldsymbol{C}}_{\beta\beta}$, and $\boldsymbol{U}^2$, that representation incorporates the fact that the common factors are standardized in the sample. That standardization will most likely be at least slightly different from standardization in the population and is thus the source of some degree of sampling error. This sampling error occurs in addition to that defined by $\boldsymbol{\Delta}_z$ in Eq. (4.16) and incorporated into $\boldsymbol{\Delta}_y$ in Eq. (4.27). Thus, when the model in Eq. (4.28) is fit to sample data, an optimal solution for the values in $\widetilde{\boldsymbol{B}}$, $\widetilde{\boldsymbol{C}}_{\beta\beta}$, and $\boldsymbol{U}^2$ can be seen to provide estimates of the parameters in $\boldsymbol{B}$, $\boldsymbol{\Phi}$, and $\boldsymbol{U}^2$. Error in those estimates arises from the sampling error and model error included in $\boldsymbol{\Delta}_y$ as well as from the standardization of the factors in the sample as defined by Eq. (4.18) and (4.19).

A second important aspect of this problem involves the fact that, as has been noted in other similar contexts previously, there are alternative possible definitions of what would be an optimal $\boldsymbol{\Delta}_y$. Each alternative definition would yield an alternative method for estimating the model parameters, and, in turn, a different set of estimates of those parameters. While alternative methods for defining optimal fit and estimating parameters will be discussed in Chapter 7, it is important at the present time to understand that parameter estimates can be obtained in different ways, and to recognize how this fact impacts on the developments presented in this section. Let us define matrices of parameter estimates obtained by any desired method. Matrix $\boldsymbol{B}$ will be an $n \times r$ matrix of estimated common factor weights, with element b_{jk} representing the estimated weight for factor k on attribute j. Matrix $\boldsymbol{R}_{\beta\beta}$ will be an $r \times r$ matrix of estimated common factor intercorrelations, with entry r_{kl}, representing the estimated correlation of factors k and l. Finally, matrix $\boldsymbol{U}^2$ will be an $n \times n$ diagonal matrix, with diagonal entry u_{jj} representing the estimated unique variance for attribute j. For any given solution, the resulting matrices $\boldsymbol{B}$, $\boldsymbol{R}_{\beta\beta}$, and $\boldsymbol{U}^2$

can be employed to obtain a covariance matrix for the modeled attributes as represented by that solution. Let such a matrix be designated $\boldsymbol{C}_{zz}^{+}$. This matrix can be obtained as follows:

$$\boldsymbol{C}_{zz}^{+} = \boldsymbol{B}\boldsymbol{R}_{\beta\beta}\boldsymbol{B}' + \boldsymbol{U}^{2} \tag{4.29}$$

It is important to understand the distinction between $\boldsymbol{C}_{zz}$ and $\boldsymbol{C}_{zz}^{+}$**.** The matrix $\boldsymbol{C}_{zz}$ is the modeled attribute covariance matrix which is defined by the full common factor model in the sample, as given in Eq. (4.14) or (4.16). Recall that simplifying assumptions are made regarding that full model, resulting in the simplified model given in Eq. (4.15). Matrix $\boldsymbol{C}_{zz}^{+}$ is thus a modeled attribute covariance matrix constructed from a solution by substituting parameter estimates into Eq. (4.15). The matrix $\boldsymbol{C}_{zz}$ cannot be exactly determined in practice because it is defined in terms of true parameter values. However, matrix $\boldsymbol{C}_{zz}^{+}$ is a very useful matrix which can be obtained from any common factor solution. It represents the covariance matrix for the modeled attributes as represented by that solution. The greater the correspondence between $\boldsymbol{C}_{zz}^{+}$ and $\boldsymbol{C}_{yy}$, the more closely the model is found to fit the sample data. An important point is that there is no single "true" $\boldsymbol{C}_{zz}^{+}$ matrix. Rather, there is a different $\boldsymbol{C}_{zz}^{+}$ for each different method of fitting the model to $\boldsymbol{C}_{yy}$. That is, different definitions of an optimal solution to the model in Eq. (4.28) will yield different $\boldsymbol{B}$, $\boldsymbol{R}_{\beta\beta}$, and $\boldsymbol{U}^{2}$ matrices, and, in turn, different $\boldsymbol{C}_{zz}^{+}$ matrices. Recall that the same phenomenon was discussed earlier in this section in the context of the population. The present development merely shows how this phenomenon occurs in the sample.

When a factor solution is obtained and the corresponding $\boldsymbol{C}_{zz}^{+}$ matrix is calculated, it is then possible to determine a matrix representing the lack of fit of that solution to sample data. This matrix will be designated as $\boldsymbol{\Delta}_{y}^{+}$ will be defined as follows:

$$\boldsymbol{\Delta}_{y}^{+} = \boldsymbol{C}_{yy} - \boldsymbol{C}_{zz}^{+} \tag{4.30}$$

This matrix represents the optimal solution for $\boldsymbol{\Delta}_{y}$ obtained as a result of the application of some specific method for estimating the parameters of the model. The entries in $\boldsymbol{\Delta}_{y}^{+}$ arise from a combination of model error, which is affected by the method employed to fit the model to the data, and sampling error.

A final point about the parameter estimation problem is that some additional complexities will arise when model fitting procedures are considered in Chapters 8 and 9. Though the model represented by Eq. (4.28) provides a useful framework for the model fitting problem, it will be seen that this framework requires slight modification for some methods. Since this modification will involve a rescaling of the parameters, it is not the case that all model fitting procedures are providing estimates of the exactly the same parameters. This complication will be discussed in detail in Chapters 8 and 9.

We wish to close this section by noting two important points which will be treated at length in subsequent chapters. First, though the model given in Eq. (4.28) is expressed in terms of oblique common factors, most fitting procedures obtain solutions in which the common factors are orthogonal. This will be discusses in Chapters 8 and 9. Second, these obtained orthogonal common factors then can be transformed into oblique common factors. This transformation procedure was discussed in detail in Chapter 3 in the context of the population. For present purposes, suffice it to say that this freedom to transform or "rotate" factors exists in obtained sample solutions as well. The mathematical framework for this process, along with methods for seeking meaningful transformed factors, will be presented in Chapters 10 and 11.

4.3. The Common Factor Model for a Sample Correlation Matrix

The developments in the previous section were carried out in the context of a sample covariance matrix. The model given by Eq. (4.28) defines the factorial structure of such a matrix. An issue which has been raised previously in this book involves the distinction between factor analyzing covariance vs. correlation matrices. This issue was treated in Chapter 3 in the context of the population, and will be considered here in the context of the sample. Following the developments in Section 3.6, let us begin by representing the conversion of a sample covariance matrix, $\boldsymbol{C_{yy}}$, into a sample correlation matrix, $\boldsymbol{R_{yy}}$**.** This is achieved by first defining a diagonal matrix $[\boldsymbol{C_d}]_{yy}$ containing the sample variance of the surface attributes. That is,

$$[\boldsymbol{C_d}]_{yy} = \boldsymbol{Diag}(\boldsymbol{C_{yy}}) \tag{4.31}$$

We then can define a sample correlation matrix, $\boldsymbol{R_{yy}}$, for the surface attributes as follows:

$$\boldsymbol{R_{yy}} = [\boldsymbol{C_d}]_{yy}^{-\frac{1}{2}} \boldsymbol{C_{yy}} [\boldsymbol{C_d}]_{yy}^{-\frac{1}{2}} \tag{4.32}$$

Effects of this standardization of the surface attributes on the common factor model in the sample can be seen by substituting from Eq. (4.28) into Eq. (4.32). This yields:

$$\begin{aligned} \boldsymbol{R_{yy}} &= [\boldsymbol{C_d}]_{yy}^{-\frac{1}{2}} (\widetilde{\boldsymbol{B}}\, \widetilde{\boldsymbol{C}}_{\beta\beta} \widetilde{\boldsymbol{B}}' + \boldsymbol{U^2} + \boldsymbol{\Delta}_y) [\boldsymbol{C_d}]_{yy}^{-\frac{1}{2}} \\ &= [\boldsymbol{C_d}]_{yy}^{-\frac{1}{2}} \widetilde{\boldsymbol{B}}\, \widetilde{\boldsymbol{C}}_{\beta\beta} \widetilde{\boldsymbol{B}}' [\boldsymbol{C_d}]_{yy}^{-\frac{1}{2}} + [\boldsymbol{C_d}]_{yy}^{-\frac{1}{2}} \boldsymbol{U^2} [\boldsymbol{C_d}]_{yy}^{-\frac{1}{2}} \\ &\quad + [\boldsymbol{C_d}]_{yy}^{-\frac{1}{2}} \boldsymbol{\Delta}_y [\boldsymbol{C_d}]_{yy}^{-\frac{1}{2}} \end{aligned} \tag{4.33}$$

To simplify this expression, we define the following matrices:

$$\widetilde{\widetilde{\boldsymbol{B}}} = [\boldsymbol{C_d}]_{yy}^{-\frac{1}{2}} \widetilde{\boldsymbol{B}} \tag{4.34}$$

$$\widetilde{\widetilde{\boldsymbol{U^2}}} = [\boldsymbol{C_d}]_{yy}^{-\frac{1}{2}} \boldsymbol{U^2} [\boldsymbol{C_d}]_{yy}^{-\frac{1}{2}} \tag{4.35}$$

$$\widetilde{\widetilde{\Delta}}_y = [C_d]_{yy}^{-\frac{1}{2}} \Delta_y [C_d]_{yy}^{-\frac{1}{2}} \tag{4.36}$$

These matrices contain common factor weights, unique variances, and lack of fit terms, respectively, which have been rescaled as a result of the standardization of the attributes in the sample. Substituting from these equations into Eq. (4.33) yields the following model:

$$R_{yy} = \widetilde{\widetilde{B}}\, \tilde{C}_{\beta\beta} \widetilde{\widetilde{B}}' + \widetilde{\widetilde{U}}^2 + \widetilde{\widetilde{\Delta}}_y \tag{4.37}$$

This equation expresses the common factor model in terms of a sample correlation matrix for the surface attributes. It has the same general form as Eq. (4.28), which expressed the model in terms of a sample covariance matrix, but some of the terms have been rescaled. Another important comparison is to note that Eq. (4.37) has the same form as Eq. (3.98); the latter expressed the model in terms of a population correlation matrix for the surface attributes. Thus, Eq. (4.37) is the sample equation analogous to the population equation given in Eq. (3.98). When the model in Eq. (4.37) is fit to a sample correlation matrix, R_{yy}, we would obtain a solution consisting of common factor weights, common factor intercorrelations, and unique variances. To represent necessary distinctions between these values and those which would be obtained from analysis of a covariance matrix, as defined in the previous section, let us define the following matrices: $\widetilde{\widetilde{B}}$ contains sample common factor weights obtained from analysis of the correlation matrix; $R_{\beta\beta}$ contains corresponding sample common factor intercorrelations; and $\widetilde{\widetilde{U}}^2$ contains corresponding sample unique variances. The fact that no new notation is used for $R_{\beta\beta}$ reflects the fact that this matrix is not affected by the analysis of a correlation matrix rather than a covariance matrix; this can be seen by comparing Eqs. (4.28) and (4.37). The sample values obtained in $\widetilde{\widetilde{B}}$, $R_{\beta\beta}$, and $\widetilde{\widetilde{U}}^2$ can be viewed as estimates of the parameters in B^*, Φ , and U^{*^2} represented in Eq. (3.98). That is, the population parameters representing the factorial structure of a population correlation matrix are estimated by obtaining a factorial solution for a sample correlation matrix.

An interesting aspect of this issue can be seen by comparing Eqs. (3.95)-(3.96) to Eqs. (4.34)-(4.35). Note that the former represent a rescaling of population parameters by population standard deviations of the attributes, while the latter represent rescaling of those same parameters by sample standard deviations of the attributes. This distinction reveals a source of sampling error which influences the accuracy of the sample values in $\widetilde{\widetilde{B}}$ and $\widetilde{\widetilde{U}}^2$ as estimates of the parameters in B^* and U^{*^2}. This sampling error is present in addition to that contained in $\widetilde{\widetilde{\Delta}}_y$. A more detailed discussion and demonstration of this phenomenon will be presented in Chapter 5.

To complete this presentation of the common factor model for a sample correlation matrix, note that we can follow the development given in Eq. (4.29) by defining a "reconstructed" correlation matrix , $\boldsymbol{R}_{zz}^{+}$, as follows:

$$\boldsymbol{R}_{zz}^{+} = \widetilde{\boldsymbol{B}}\, \boldsymbol{R}_{\beta\beta} \widetilde{\boldsymbol{B}}' + \widetilde{\boldsymbol{U}^2} \tag{4.38}$$

This matrix is a correlation matrix for the modeled attributes constructed from the obtained factor solution. It provides the correlations among the modeled attributes as represented by that solution. We can also obtain a matrix representing the lack of fit of that solution to the sample data. This matrix will be designated $\widetilde{\boldsymbol{\Delta}}_{y}^{+}$ and will be defines as follows:

$$\widetilde{\boldsymbol{\Delta}}_{y}^{+} = \boldsymbol{R}_{yy} - \boldsymbol{R}_{zz}^{+} \tag{4.39}$$

This matrix represents the optimal solution for $\widetilde{\boldsymbol{\Delta}}_{y}$ obtained as a result of the application of some specific method for estimating the parameters of the model in Eq. (4.37). As noted in other similar contexts, it should be recognized that alternative fitting methods would yield different solutions for $\boldsymbol{B}$, $\boldsymbol{R}_{\beta\beta}$, and $\boldsymbol{U}^2$, and thus different $\boldsymbol{R}_{zz}^{+}$ and $\widetilde{\boldsymbol{\Delta}}_{y}^{+}$ matrices.

The important general points to recognize from the developments in this section are that (a) the common factor model for a sample correlation matrix is a special case of the model for a sample covariance matrix, and (b) the standardization of the attributes in the sample introduces an additional source of sampling error to be discussed further in Chapter 5.

4.4. Sources of Error in Fitting the Common Factor Model to Sample Data

Special attention should be given to the fact that developments in this chapter have served to identify a number of sources of error which affect solutions obtained when the common factor model is fit to sample data. These effects cause parameter estimates to be in error and unstable to some degree, and also give rise to lack of fit of the model to the data. A primary source of such error has been designated as model error. The occurrence of model error simply reflects the fact that the common factor model will generally not precisely account for all variance and covariance of the surface attributes. This may be due to such phenomena as nonlinear relations of factors to attributes, the presence of a large number of minor factors not represented in the model, etc. Algebraically, model error is represented in the population by matrix $\boldsymbol{\Delta}_{\Sigma}$ in Eq. (4.11), and in the sample by matrix $\boldsymbol{\Delta}_{c}$ in Eq. (4.26).

Another general source of error is sampling error, which includes phenomena whereby chance characteristics of a sample influence parameter estimates and the fit of the model to the data. Developments in this chapter have revealed that there exist several separate sources of sampling error. A primary source arises from the fact that the assumption that unique factors are

uncorrelated with each other and with common factors will generally not hold exactly in a sample. This gives rise to the lack of fit represented in matrix $\boldsymbol{\Delta}_z$, which was discussed in conjunction with Eqs. (4.15) and (4.16). This source of sampling error would affect estimates of all parameters of the model. A second type of sampling error arises from the common procedure of standardizing the common factors in the sample. Since this standardization will tend to be at least slightly different from that in the population, this gives rise to some error in estimation of the parameters in $\boldsymbol{B}$ and $\boldsymbol{\Phi}$. This was discussed in conjunction with Eq. (4.28). An important aspect of this source of sampling error is that, as discussed previously in this chapter, standardization of the factors in the sample resolves the identification problem in parameter estimation. Since this problem will be dealt with in order to fit the model to data, some sampling error will arise no matter how it is resolved. A third source of sampling error arises when the attributes are standardized in the sample. This occurs when a sample correlation matrix rather than a sample covariance matrix is analyzed. It was shown in the previous section that this standardization introduces a source of error which would affect estimates of common factor weights and unique variances.

In practice these various sources of error cannot be separated. However, we feel it is important to have a theoretical framework in which their separate effects can be represented and understood. Some of these effects will be demonstrated and described further in Chapter 5, and implications for practice will be discussed in Chapter 6.

4.5. Illustration of Fitting the Common Factor Model to Sample Data

Rather than present a completely new illustration of the developments described in this chapter, we will briefly describe how an illustration presented in a previous chapter can be viewed in the present context. In Chapter 1 we presented results of factor analysis of data drawn from a study by Thurstone and Thurstone (1941). The data which were factor analyzed consisted of correlations among nine mental tests, based on measures obtained from a sample of 710 students. The correlation matrix shown in Table 1.1 corresponds to matrix $\boldsymbol{R}_{yy}$. The model represented by Eq. (4.37) was fit to this matrix (by methods described in Chapter 7), and the resulting solution is shown in Table 1.2. The factor weights in Table 1.2 correspond to matrix $\widetilde{\widetilde{\boldsymbol{B}}}$ and the factor intercorrelations correspond to matrix $\boldsymbol{R}_{\beta\beta}$. The communalities in Table 1.2 were obtained by subtracting the unique variances in $\widetilde{\widetilde{\boldsymbol{U}^2}}$ from unity. We could define these communalities as entries in a diagonal matrix $\widetilde{\widetilde{\boldsymbol{H}^2}}$, where

$$\widetilde{\widetilde{\boldsymbol{H}^2}} = \boldsymbol{I} - \widetilde{\widetilde{\boldsymbol{U}^2}} \tag{4.40}$$

According to developments in the previous section, values given in Table 1.2 represent estimates of the parameters in $\boldsymbol{B}^*$, $\boldsymbol{\Phi}$, and $\boldsymbol{H}^{*^2}$, respectively. From this solution, one could compute a correlation matrix for the modeled attributes according to Eq. (4.38) and a matrix of lack of fit terms according to Eq. (4.39). We leave this as an exercise for the reader.

The reader should keep in mind a number of influence which affect the solution obtained in this illustration. If a covariance matrix rather than a correlation matrix had been analyzed, the effects of standardization of the attributes would not have been present and the illustration would have fit the framework described in Section 4.2 rather than that in Section 4.3. The effect of this would have been a rescaling of the factor weights and unique variances, as indicated by Eqs. (4.34) and (4.35). In addition, if a different method had been used to fit the model to the data, a slightly different solution would have been obtained. Further consideration of this issue, along with additional illustrations, will be presented in Chapters 8 and 9.

4.6. Conclusion

This completes the development of the representation of the common factor model in the sample. We wish to re-emphasize that the framework presented here serves to provide an explicit treatment of a number of important issues inherent in factor analytic theory: (a) the distinction between the population and the sample; (b) the distinction between surface attributes and modeled attributes; and (c) the distinction between model error and various types of sampling error. The last two issues in particular are almost totally ignored in most presentations of factor analysis. We believe, however, that a complete theoretical understanding of the model requires that these aspects be considered very explicitly. The framework developed here will be employed in Chapters 8 and 9 to resolve the problem of fitting the model to a sample covariance matrix. Prior to that, however, some important issues involved in selecting observations and attributes in a factor analysis study will be considered in Chapter 6.

CHAPTER 5
ISSUES IN SELECTING OBSERVATIONS AND ATTRIBUTES

From
Exploratory Factor Analysis
Ledyard R Tucker
and
Robert C. MacCallum

CHAPTER 5
ISSUES IN SELECTING OBSERVATIONS AND ATTRIBUTES

5.0. Introduction

The previous chapter provided a detailed presentation of the common factor model in a sample. The collection of sample data requires that a researcher do two things: (a) select a sample of observations from the population; and (b) select a battery of attributes from the domain of interest. The objective of the present chapter is to closely examine a number of critical issues involved in the process of selecting observations and attributes to be measured. It will be shown that there is a number of aspects of this process which can have a substantial impact on the results obtained in a factor analytic study.

5.1. Effects of Random Sampling

Let us first consider the process of selecting a sample from a population, and how factor analytic solutions may be affected by certain aspects of this process. As shown in the previous chapter, when the common factor model is applied to a sample, a degree of sampling error will be present and will influence the obtained factor solution. In examining the nature of this influence, it is useful to consider two distinct cases defined in terms of the type of sampling which is conducted. The first case is random sampling, where a sample is drawn by some random process from the population. In random sampling, each individual has an equal probability of being included in the sample. The second case is selective sampling, where observations are selected for the sample according to their level on one or more attributes; i.e., the probability of a given individual being included in the sample depends on the individual's score on one or more attributes. We will examine these two cases in turn.

In the following subsections we will consider three factors which influence the effect of sampling error under random sampling: (a) sample size; (b) the presence of unique factors; and (c) the standardization of the attributes in the sample.

5.1.1 Effects of Sample Size

Factor analysis is similar to other multivariate statistical methods in that the effects of sampling error are reduced as sample size increases. This phenomenon can be understood most easily by considering some equations developed in the previous chapter. In Eq. (4.11), the model defines population variances and covariances for the surface attributes as a function of the parameters in $\boldsymbol{B}$, $\boldsymbol{\Phi}$, and $\boldsymbol{U^2}$. Lack of fit of this model, as represented in $\boldsymbol{\Delta_\Sigma}$, arises only from model error. In Eq. (4.28), the model defines sample variances and covariances for the surface attributes as a function of the parameters in $\widetilde{\boldsymbol{B}}$, $\widetilde{\boldsymbol{C}}_{\beta\beta}$, and $\boldsymbol{U^2}$. Lack of fit of this model, as

represented in $\boldsymbol{\Delta}_y$, arises from a combination of model error and a primary source of sampling error. Recall that the presence of sampling error here is due to the fact that simplifying assumptions are made about the sample (i.e., that the unique factors are uncorrelated with each other and with the common factors in the sample), and that these assumptions almost surely do not hold exactly. Clearly, however, as sample size becomes very large these assumptions will hold more closely. That is, since these properties of unique factors do hold by definition in the population, they will tend to hold more closely in larger samples than in smaller samples. As a result, the impact of sampling error is reduced in larger samples, and lack of fit of the model to $\boldsymbol{C}_{yy}$, as represented in $\boldsymbol{\Delta}_y$, becomes primarily attributable to model error.

The impact of two other sources of sampling error described in Chapter 4 will also be reduced in large samples. The sources of sampling error were identified as arising from the effects of standardization of factors and of attributes in the sample. However, in large samples such a standardization would be very similar to the corresponding standardization in the population, thus indicating a reduced effect of these types of sampling error also.

It can also be recognized that solutions obtained from analyses of large samples will tend to be more similar to population parameter values. This can be understood by comparing Eqs. (4.1 1) and (4.28). When sample size becomes very large, the sample covariance matrix $\boldsymbol{C}_{yy}$ will become more similar to the population covariance matrix $\boldsymbol{\Sigma}_{yy}$. In addition, the lack of fit term in the sample, $\boldsymbol{\Delta}_y$, arising in large samples primarily from model error, will correspond closely to $\boldsymbol{\Delta}_\Sigma$, representing model error in the population. As a result, obtained sample solutions represented by matrices $\boldsymbol{B}$, $\boldsymbol{R}_{\beta\beta}$, and $\boldsymbol{U}^2$ will tend to be more similar to solutions for population parameters in $\boldsymbol{B}$, $\boldsymbol{R}_{\beta\beta}$, and $\boldsymbol{U}^2$ obtained by corresponding methods. All of these tendencies become stronger as sample size becomes larger. In addition, sampling variability of these sample solutions will be reduced as sample size increases; i.e., solutions will be more stable over repeated sampling in large samples. The mathematical basis for this phenomenon will be developed in later chapters.

In general, the effects of random sampling on obtained factor solutions are greatly reduced in large samples. This point will be emphasized and discussed in both theoretical and practical terms at several points in this and subsequent chapters.

5.1.2. The Presence of Unique Factors

A very interesting phenomenon is that the primary effect of sampling error in the common factor model depends on whether or not unique factors are present. In examining this phenomenon, we will make use of the model as expressed in terms of covariances of modeled attributes (Eqs. (4.4) and (4.14)), since we have shown that the primary impact of sampling error is revealed in that context (Eq. (4.16)). Equivalently, we will examine the issue by employing a

context where no model error is present. Let us first consider the case where no unique factors are present. In this case, Eq. (4.4) for the population can be rewritten as follows:

$$\boldsymbol{\Sigma}_{zz} = \boldsymbol{B}\boldsymbol{\Phi}\,\boldsymbol{B}' \tag{5.1}$$

In addition, Eq. (4.14), which represents the complete expression of the common factor model in the sample, would be greatly simplified. Since no unique factors are present, the last three terms in that equation would vanish, yielding

$$\boldsymbol{C}_{zz} = \boldsymbol{B}\boldsymbol{C}_{\beta\beta}\boldsymbol{B}' \tag{5.2}$$

A comparison of Eqs. (5.1) and (5.2) clearly shows that the common factor weights in the sample will be the same as the population common factor weights in this situation. However, the common factor covariance matrix for the sample will not match the corresponding matrix from the population. Thus, when no unique factors are present, sampling error will not influence the common factor weights but will influence the common factor variances and covariances.

Two aspects of this phenomenon should be pointed out. The first is that sample factor solutions are generally represented in a form where the common factors are standardized in the sample. This issue was discussed in Chapter 4, and the mathematical framework for this standardization was presented in Eqs. (4.17)-(4,20). Substituting from Eq. (4.20) into Eq. (5.2) yields the following representation of $\boldsymbol{C}_{zz}$ in terms of parameters corresponding to common factors standardized in the sample:

$$\boldsymbol{C}_{zz} = \widetilde{\boldsymbol{B}}\,\widetilde{\boldsymbol{C}}_{\beta\beta}\widetilde{\boldsymbol{B}}'\,' \tag{5.3}$$

In this context where no model error and no unique factors are present, matrix $\widetilde{\boldsymbol{B}}$ would correspond to an obtained sample common factor weight matrix $\boldsymbol{B}$. The relationship between this obtained solution and the population common factor weight matrix $\boldsymbol{B}$ is given by Eq. (4.19). Rewriting this equation to suit the present context yields

$$\boldsymbol{B} = \boldsymbol{B}[\boldsymbol{C}_d]^{\frac{1}{2}}_{\beta\beta} = \widetilde{\boldsymbol{B}} \tag{5.4}$$

This relationship further isolates one effect of sampling on common factor weights discussed in Chapter 4: when the common factors are standardized in the sample, the obtained common factor weights will equal the population common factor weights for standardized factors multiplied by the standard deviations of the common factors in the sample. It must be kept in mind that this simple relationship is defined in the particular context where model error and unique factors are not present. When model error and unique factors are present, the multiplicative effect given by Eq. (5.4) will still occur, but the relationship between $\boldsymbol{B}$ and $\widetilde{\boldsymbol{B}}$ will be influenced by these additional effects.

A second important aspect of the relationship between obtained and population common factor weights in the present context involves the fact that most methods for obtaining a factor solution provide an initial solution characterized by orthogonal common factors. Thus an initial sample common factor weight matrix $\boldsymbol{A}$ would be obtained such that

$$\boldsymbol{C}_{zz} = \boldsymbol{A}\boldsymbol{A}' \tag{5.5}$$

Matrix A will be of order $n \times r$ and will contain entries a_{jk} representing the sample weight for factor *k* on attribute *j*. Matrix $\boldsymbol{B}$ in Eq. (5.4) would then actually correspond to a transformed solution obtained by transforming matrix $\boldsymbol{A}$. That is, following the framework developed in Chapter 3, a sample trait matrix $\boldsymbol{T}$ will exist such that

$$\boldsymbol{A}\boldsymbol{T}^{-1} = \boldsymbol{B} \tag{5.6}$$

Since, as noted above, $\boldsymbol{B}$ is equivalent to $\widetilde{\boldsymbol{B}}$ in the present context, we could also write

$$\boldsymbol{A}\boldsymbol{T}^{-1} = \widetilde{\boldsymbol{B}} \tag{5.7}$$

This indicates that when no unique factors are present the obtained matrix of common factor weights can be transformed into the population weights rescaled by sample common factor standard deviation. In addition, the intercorrelations of the transformed factors in the sample would be given by

$$\widetilde{\boldsymbol{C}}_{\beta\beta} = \boldsymbol{T}\boldsymbol{T}' \tag{5.8}$$

Thus, the relationship defined in Eq. (5.4) would be observable only after an appropriate transformation of an initial orthogonal solution.

It must be kept in mind that the relationships just defined represent the case of no model error and no unique factors. Let us now consider the case where unique factors are present. In effect, the impact of sampling error in this situation has already been described in the context of Eq. (4.14). That equation represents the complete explanation of the common factor model in a sample. That model is then simplified by assuming that unique factors are uncorrelated with each other and with the common factors in the sample, thus yielding Eq. (4.15). When that equation was introduced, it was explained that the presence of sampling error gives rise to non-zero correlations of unique factors with each other and with common factors, meaning that the simplifying assumptions generally do not hold in practice. This in turn gives rise to the matrix $\boldsymbol{\Delta}_z$ in Eq. (4.16), representing a lack of fit arising from this of sampling error. Solutions obtained via factor analytic methods which attempt to optimize fit will thus be influenced by the presence of sampling error. That is, obtained common factor weights in matrix $\boldsymbol{A}$ will be influenced by the presence of sampling error when unique factors are present. As a result, it will not be possible to

transform the obtained solution $\boldsymbol{A}$ to a matrix $\boldsymbol{B}$ satisfying Eq. (5.4), as it is when no unique factors are present.

This influence of sampling error on obtained common factor weights when unique factors are present can be expected to increase as the importance of the unique factors increases. Consider Eqs. (4.14) - (4.16), it can be seen that if the influence of the unique factors is very slight, the contribution of the last three terms in Eq. (4.14) will be minor, thus yielding only a slight influence of sampling error as represented in $\boldsymbol{\Delta}_z$ in Eq. (4.16). As the importance of the unique factors increases, this influence of sampling error will also increase. This effect of sampling error will also increase as sample size is reduced. As explained in section 5.1.1, in a large sample the properties of unique factors in the population (uncorrelated with each other and with the common factors) will hold more closely in the sample, and the influence of sampling error will be reduced. In a small sample, there is greater likelihood of substantial correlations among unique factors and between unique and common factors, thus causing a greater impact of sampling error. Thus, the influence of sampling error on obtained factor solutions is reduced when sample size is large and the influence of unique factors is small.

One final interesting point regarding this issue of the impact of unique factors on sampling error concerns the components analysis model, described in section 3.5. The components model can be viewed as being obtained by deleting unique factors from the common factor model. In the present context, it can thus be seen that sampling error would not affect weights obtained in the components model (except for the multiplicative effect given in Eq. (5.4)), but would affect variances and covariances of the components in a random manner.

5.1.3. The Effect of Standardization of Attributes

It is of interest to consider the effect of sampling error when attributes have been standardized. (The reader must be careful here to distinguish between standardization of attributes and standardization of factors. One issue treated in the previous section was standardization of factors; standardization of attributes is a completely separate issue.) To this point in this chapter, we have dealt with the case where attributes are unstandardized in both the population and sample. Let us now examine the influence of standardization in both of those groups. We will still employ the context of modeled attributes, and we will make use of the case where no unique factors are present, so as to allow us to examine the effect of standardization when no other sources of sampling error is operating. This represents a more detailed and isolated study of this effect than was presented in Chapter 4.

Considering the population first, we have defined a population covariance matrix $\boldsymbol{\Sigma}_{zz}$ for modeled attributes. This can be converted to a correlation matrix (i.e., a covariance matrix for standardized attributes) by making use of a diagonal matrix $[\boldsymbol{\Sigma}_d]_{zz}$ containing variances of the modeled attributes on the diagonal. That is

$$[\Sigma_d]_{zz} = \boldsymbol{Diag}(\Sigma_{zz}) \tag{5.9}$$

We can then define a correlation matrix $\boldsymbol{R}_{zz}$ for the modeled attributes according to

$$\boldsymbol{R}_{zz} = [\Sigma_d]_{zz}^{-\frac{1}{2}} \Sigma_{zz} [\Sigma_d]_{zz}^{-\frac{1}{2}} \tag{5.100}$$

substituting from Eq. (5.1) into Eq. (5.10) yields

$$\boldsymbol{R}_{zz} = [\Sigma_d]_{zz}^{-\frac{1}{2}} \boldsymbol{B\Phi}\ \boldsymbol{B}' [\Sigma_d]_{zz}^{-\frac{1}{2}} \tag{5.11}$$

Let us next define a matrix $\boldsymbol{B}^*$ containing population common factor weights for standardized attributes as follows

$$\boldsymbol{B}^* = [\Sigma_d]_{zz}^{-\frac{1}{2}} \boldsymbol{B} \tag{5.12}$$

We can then substitute from Eq. (5.12) into Eq. (5.11) to obtain

$$\boldsymbol{R}_{zz} = \boldsymbol{B}^* \boldsymbol{\Phi}\ \boldsymbol{B}^* \tag{5.13}$$

This equation expresses the common factor model for standardized attributes in the population, for the case of no unique factors.

Let us now consider a similar framework for the sample. The sample covariance matrix for the modeled attributes can be converted to a correlation matrix by making use of a diagonal matrix $[\boldsymbol{C}_d]_{\boldsymbol{zz}}$ containing sample variances of the modeled attributes on the diagonal. That is

$$[\boldsymbol{C}_d]_{\boldsymbol{zz}} = \boldsymbol{Diag}(\boldsymbol{C}_{zz}) \tag{5.14}$$

The desired correlation matrix for the modeled attributes will be designated $\boldsymbol{R}_{zz}$ and can be defined as follows:

$$\boldsymbol{R}_{zz} = [\boldsymbol{C}_d]_{zz}^{-\frac{1}{2}} \boldsymbol{C}_{zz} [\boldsymbol{C}_d]_{zz}^{-\frac{1}{2}} \tag{5.15}$$

Substituting from Eq. (5.5), which represents an obtained orthogonal common factor solution for $\boldsymbol{C}_{zz}$, into Eq. (5.15) gives us

$$\boldsymbol{R}_{zz} = [\boldsymbol{C}_d]_{zz}^{-\frac{1}{2}} \boldsymbol{AA}' [\boldsymbol{C}_d]_{zz}^{-\frac{1}{2}} \tag{5.16}$$

Let us next define a matrix $\boldsymbol{A}^*$ containing obtained weights for orthogonal common factors on standardized modeled attributes as follows:

$$\boldsymbol{A}^* = [\boldsymbol{C}_d]_{zz}^{-\frac{1}{2}} \boldsymbol{A} \tag{5.17}$$

We can then substitute from Eq. (5.17) into Eq. (5.16) to obtain

$$\boldsymbol{R}_{zz} = \boldsymbol{A}^{*}\boldsymbol{A}^{*} \tag{5.18}$$

We now have equations representing the model for standardized attributes in the population (Eq. (5.13)) and in the sample (Eq. (5.18)), Considering the two expressions, the issue of interest here is whether $\boldsymbol{A}^{*}$ can be transformed to obtain $\boldsymbol{B}^{*}$. If that is possible, then standardization has no negative impact. If such a transformation is not possible, however, then standardization of the attributes has introduced another source of error. To resolve this issue, note first that it has already been shown that in the case of unstandardized attributes the obtained matrix $\boldsymbol{A}$ can be transformed into $\widetilde{B}$, as given in Eq. (5.7). By making use of relationships defined in Eqs. (5.4) and (5.12) we can obtain the following:

$$\boldsymbol{B} = [\boldsymbol{\Sigma}_d]_{zz}^{\frac{1}{2}}\boldsymbol{B}^{*}[\boldsymbol{C}_d]_{\beta\beta}^{\frac{1}{2}} \tag{5.19}$$

Substituting from this equation into Eq. (5.7) yields

$$\boldsymbol{A}\boldsymbol{T}^{-1} = [\boldsymbol{\Sigma}_d]_{zz}^{\frac{1}{2}}\boldsymbol{B}^{*}[\boldsymbol{C}_d]_{\beta\beta}^{\frac{1}{2}} \tag{5.20}$$

Employing the relationship given in Eq. (5,17), we can rewrite Eq. (5.20) as follows:

$$[\boldsymbol{C}_d]_{\boldsymbol{zz}}^{\frac{1}{2}}\boldsymbol{A}^{*}\boldsymbol{T}^{-1} = [\boldsymbol{\Sigma}_d]_{zz}^{\frac{1}{2}}\boldsymbol{B}^{*}[\boldsymbol{C}_d]_{\beta\beta}^{\frac{1}{2}} \tag{5.21}$$

Finally, solving for $\boldsymbol{B}^{*}$ on the right side yields

$$[\boldsymbol{\Sigma}_d]_{zz}^{-\frac{1}{2}}[\boldsymbol{C}_d]_{\boldsymbol{zz}}^{\frac{1}{2}}\boldsymbol{A}^{*}\boldsymbol{T}^{-1}[\boldsymbol{C}_d]_{\beta\beta}^{-\frac{1}{2}} \tag{5.22}$$

This equation is important because it shows that there is not a simple direct transformation of $\boldsymbol{A}^{*}$ to $\boldsymbol{B}^{*}$. That is, there is no trait matrix $\boldsymbol{T}$ which would provide this transformation using the framework defined in section 3.3. Rather, the rows of $\boldsymbol{A}^{*}$ would have to be rescaled using population and sample standard deviations as shown in Eq. (5.22) before such a transformation would be possible. It can be seen that the adverse impact of standardization will be reduced as sample sizes increases. With very large samples, the sample standard deviations in $[\boldsymbol{C}_d]_{\boldsymbol{zz}}$ will be very similar to the population standard deviations in $[\boldsymbol{\Sigma}_d]_{zz}$. In this case, the product of the first two terms in Eq. (5.22) will be very nearly an identity matrix, thus allowing the transformation of $\boldsymbol{A}^{*}$ to more closely approximate $\boldsymbol{B}^{*}$.

Regardless of the reduced effect in large samples, the fact remains that the standardization of attributes in a sample introduces an additional source of sampling error which further hinders the degree to which the population common factor weights can be recovered. This source of error can be easily eliminated via the use of covariance matrices in factor analytic studies.

5.1.4. Summary and Demonstration of Effects of Sampling Error Under Random Sampling

We have described three factors which influence the degree of effects of sampling error on obtained factor solutions under the condition of random sampling. The factors are (a) sample size, (b) the role of unique factors, and (c) the standardization of attributes. Sampling error will be reduced when sample size is large, since parameter estimates obtained in large samples are more stable and provide more accurate estimates of true parameter values. Sampling error will have only a multiplicative influence on common factor weights for each factor when unique factors are absent, but will randomly influence common factor variances and covariances. The multiplicative effect arises from the standardization of the factors in the sample. When unique factors are present, estimates of all of these parameters will be affected by sampling error, with the magnitude of the effect being reduced when sample size is large and the influence of the unique factors is small. Finally, standardization of attributes in a sample introduces an additional source of sampling error in that common factor weights obtained for such attributes cannot be transformed into population factor weights, even when no unique factors are present.

We will now offer a demonstration of these phenomena. The demonstration was carried out using a $2 \times 2 \times 2$ design. The factors represented in this design were (a) 2 levels of sample size; (b) two levels of communality, representing cases where unique factors were either absent or present; and (c) two types of matrices analyzed, covariance or correlation. An artificial sample data set was generated for each of the eight conditions in this design. This was accomplished by first constructing two population factor weight matrices containing weights for two factors on ten attributes. These two matrices, shown in Table 5.1, contain weights for standardized, uncorrelated factors on attributes considered to be standardized in the population. The first such matrix represents the case where are no unique factors; this is revealed by the fact that the communalities for the attributes, as given by the sum of the weights in each row, are all equal to unity. The second weight matrix represents the case where substantial unique factors are present; the weights in this matrix were obtained by multiplying all weights in the first matrix by .7, thus yielding a case where the communalities of all attributes were .49. Each of these two weight matrices can be considered to be a population factor weight matrix $\boldsymbol{B}$.

Next a sample of scores on factors was generated for each of two sample sizes, N=25 and N=1000. Scores for the individuals in these samples were generated on two common factors and ten unique factors by sampling from a multidimensional normal population with mean zero and variance unity. Based on these two samples of factor scores, a factor covariance matrix was computed for each sample. Each of the two population factor weight matrices were then applied to each of the two sample factor covariance matrices, as in Eq. (4.12), to yield four sample covariance matrices. Each of these could be considered to be a $\boldsymbol{C}_{zz}$ matrix. Each of these covariance matrices was then standardized to obtain a correlation matrix; each of the resulting

Table 5.1

Population Factor Weight Matrices*

Population Communalities=1.00				Population Communalities=.49			
	Weights on Factors				Weights on Factors		
	1	2	Unique		1	2	Unique
1	1.000	.000	.000	1	.700	.000	.714
2	1.000	.000	.000	2	.700	.000	.714
3	1.000	.000	.000	3	.700	.000	.714
4	.960	.280	.000	4	.672	.196	.714
5	.800	.600	.000	5	.560	.420	.714
6	.600	.800	.000	6	.420	.560	.714
7	.600	.800	.000	7	.420	.560	.714
8	.000	1.000	.000	8	.000	.700	.714
9	.000	1.000	.000	9	.000	.700	.714
10	.000	1.000	.000	10	.000	.700	.714

* In the population the factor scores were standardized, uncorrelated.

four correlation matrices could be considered to be a $\boldsymbol{R}_{zz}$ matrix. This yielded eight sample data matrices, one for each cell defined by the $2 \times 2 \times 2$ design.

Each of these eight matrices was then analyzed by factor analysis procedures to be discussed in Chapter 7. As with previous demonstrations, it is not necessary at this point to understand the methodology employed. All that is necessary to understand at this point is that for each of the eight constructed matrices a factor solution was obtained which, according to a specific criterion, optimally fit the matrix in question. In terms of the developments in Chapter 4, this was achieved by considering each of the constructed $\boldsymbol{C}_{zz}$ and $\boldsymbol{R}_{zz}$ matrices to be equivalent to $\boldsymbol{C}_{yy}$ and $\boldsymbol{R}_{yy}$ matrices, respectively, and fitting the models given by Eqs. (4.28) and (4.37). Consideration of the constructed $\boldsymbol{C}_{zz}$ and $\boldsymbol{R}_{zz}$ matrices as $\boldsymbol{C}_{yy}$ and $\boldsymbol{R}_{yy}$ matrices is appropriate in the present context since no model error is incorporated into this demonstration. Of primary importance in the present context is the relationship between these obtained solutions and the known population parameters. It should be pointed out that in order to evaluate this relationship it was necessary to transform each of the obtained orthogonal common factor weight matrices to approximate the appropriate population common factor weight matrix. This was achieved by a method called canonical congruence transformation. The mathematical framework for this transformation will be presented in Chapter 9. For present purposes, it is sufficient to understand that each obtained orthogonal weight matrix $\boldsymbol{A}$ was transformed to an oblique solution $\boldsymbol{B}$, such that $\boldsymbol{B}$ would be maximally similar to the corresponding population weight matrix, $\boldsymbol{B}$.

Based on the developments in the previous section, we can state the results that we would expect. First, we would expect that the variability of obtained statistics from population parameters will be more evident for the small sample than for the large sample. In the analysis of covariance matrices, we expect that the columns of obtained common factor weights will equal the corresponding population common factor weights, multiplied by the sample standard deviations of the corresponding common factor scores. This relationship, as defined in Eq. (5.4), will be exact when unique factor weights are zero but not when unique factor weights are greater than zero. In this case, additional variability in attribute factor weights is brought on by unique factor variances and covariance. Thus, when communalities are less than unity, we expect that the relationship between population and obtained common factor weights will be randomly perturbed from that defined in Eq. (5.4). Finally, we expect that solutions obtained from the analysis of correlation matrices will be characterized by differential effects on the common factor weights for the attributes, as implied by Eq. (5.22) and arising from the effects of standardization of the attributes in the sample.

Major results are shown in the figures. Figure 5.1 and 5.2 are for the analyses of covariance matrices, Figures 5.3 and 5.4 are for the analyses of correlation matrices. Figures 5.1

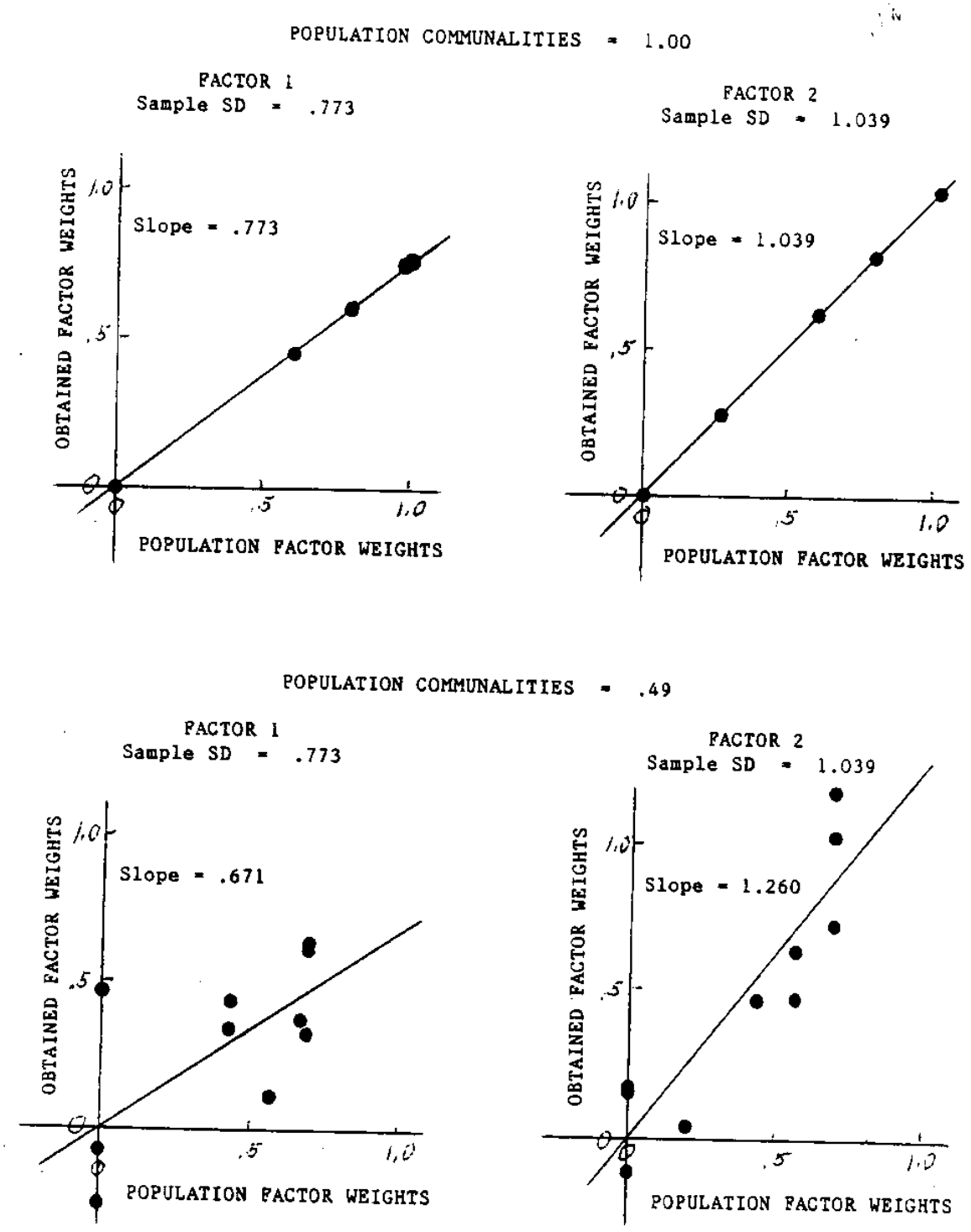

Figure 5.1: Examples of Effects of Random Sampling; Comparison of Obtained Factor Weights with Population Factor Weights; Sample Covariance Matrix Analysed; Sample Size = 25

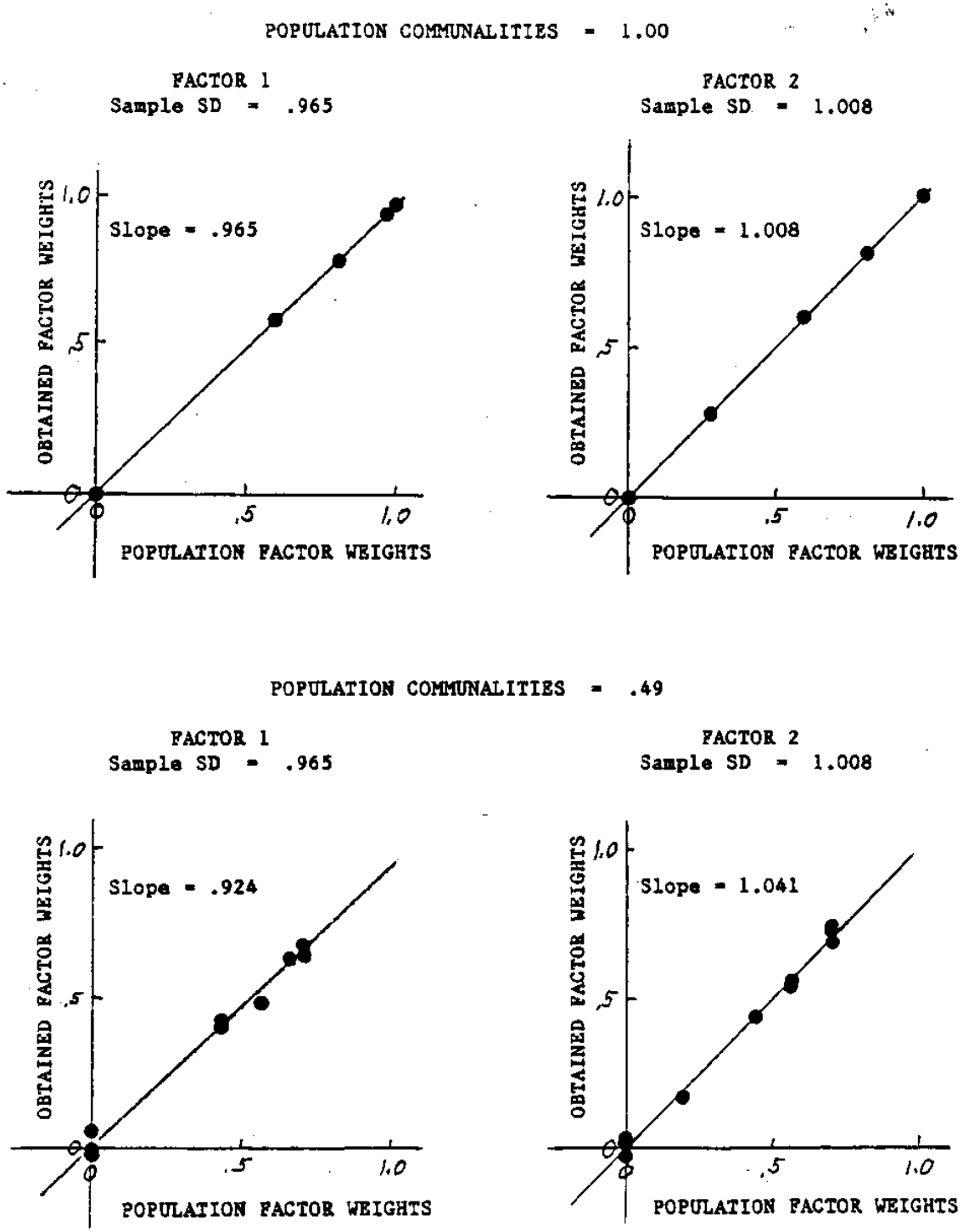

Figure 5.2: Examples of Effects of Random Sampling; Comparison of Obtained Factor Weights with Population Factor Weights; Sample Covariance Matrix Analysed; Sample Size = 1000

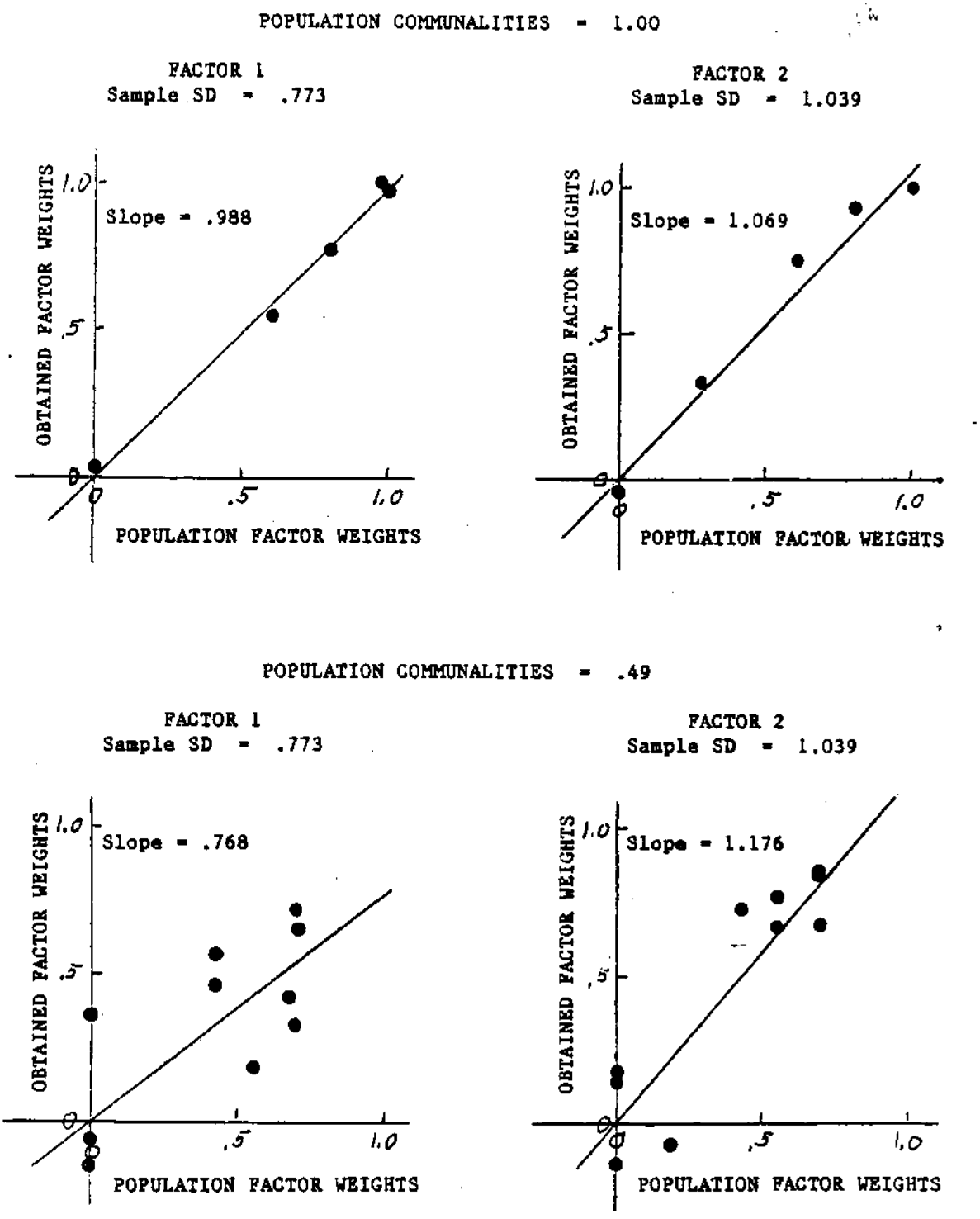

Figure 5.3: Examples of Effects of Random Sampling; Comparison of Obtained Factor Weights with Population Factor Weights; Sample Covariance Matrix Analysed; Sample Size = 25

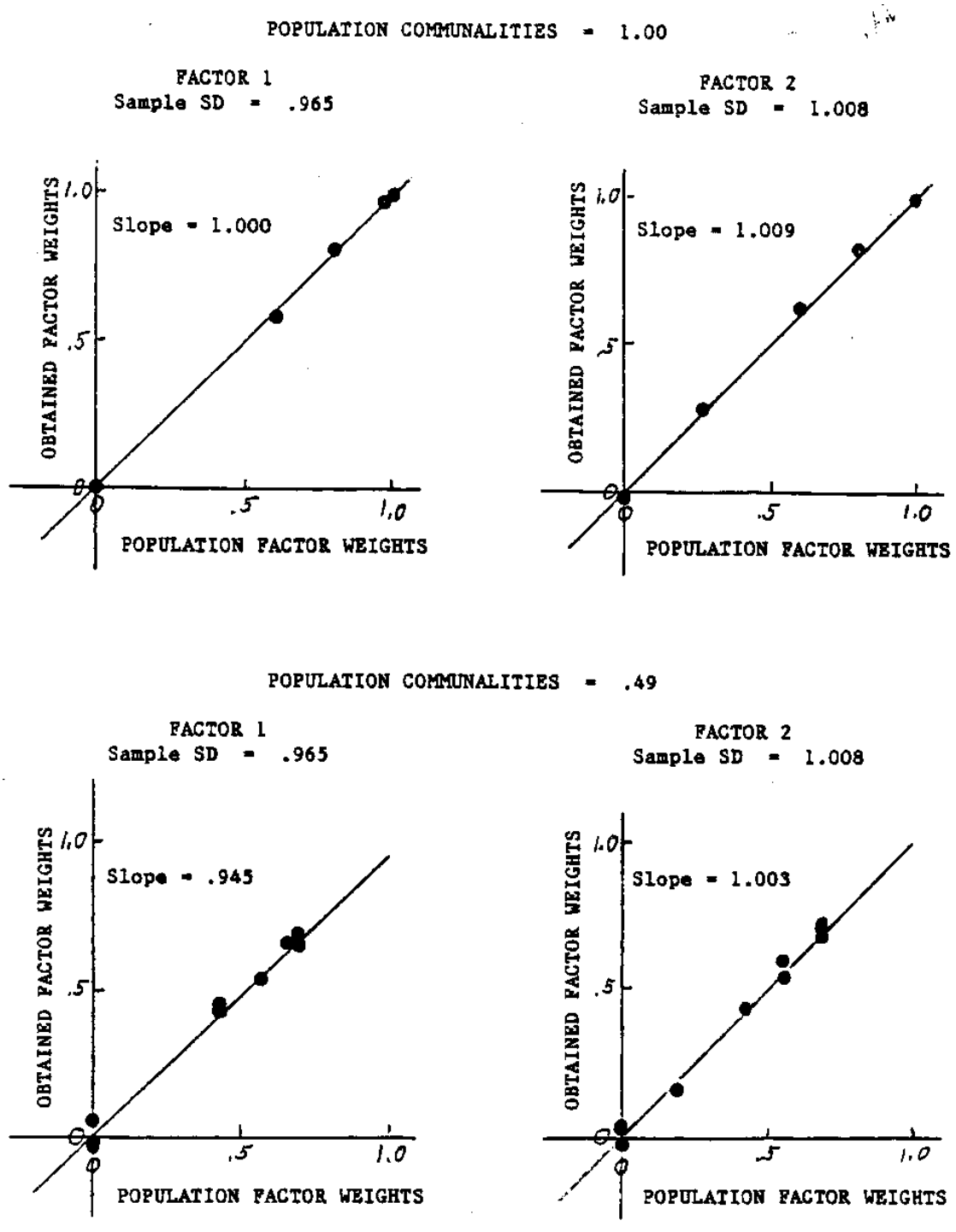

Figure 5.4: Examples of Effects of Random Sampling; Comparison of Obtained Factor Weights with Population Factor Weights; Sample Covariance Matrix Analysed; Sample Size = 1000

and 5.3 are for the sample size of 25 and Figures 5.2 and 5.4 are for the sample size of 1000. The top pair of graphs in each figure show results for the population factor matrix having communalities equal to unity; the lower pair of graphs show results for the population factor matrix having communalities equal to .49. For each figure, the left graph shows results for factor 1 while the right graph shows results for factor 2. Each graph plots obtained factor weights against corresponding population factor weights. There is a plotted point for each attribute. In several instances, however, several of the attributes plot on the same point. A line of best fit to the points, with discrepancies measured vertically, is drawn on each graph.

Consider Figure 5.1 and 5.2 for the analysis of covariance matrices. The upper pair of graphs in each figure are for the matrices having population communalities equal to 1.00. On both graphs on both figures, the points are precisely on the lines of best fit. Note also that the slopes of these lines equal the standard deviations of the factor scores in the sample. Thus, random sampling results in the population factor weights being multiplied by the sample standard deviations of the factor scores. This is a systematic effect for each factor and corresponds to the relationship defined in Eq. (5.4). The size of the effect depends on the sampling variability of the factor score standard deviation, this variability being larger for small samples than for larger samples. This is demonstrated by the fact that the slopes in the upper pair of graphs in Figure 5.2 are closer to 1.00 (the population standard deviation of the factors) than are the corresponding slopes in Figure 5. 1.

Consider the lower pair of graphs in Figures 5.1 and 5.2; these represent the case of population communalities equal to .49, meaning that substantial unique factors are present. It is seen that the points in the lower graphs in Figure 5.1, where sample size is 25, have considerable divergence from the line of best fit. As discussed above, this is brought on by the sampling variances and covariances of the unique factors. Increasing sample size to 1000 (see Figure 5.2) reduces the divergence of the points from the line of best fit. A further effect is that the lines of best fit have slopes which differ from the sample factor score standard deviations. Thus, random sampling has two effects when the unique loadings do not equal zero: (a) there is a general multiplicative effect for each factor, with population factor weights being multiplied by sample factor standard deviations; and (b) there is perturbation of individual attribute weights due to the effects of unique factor variances and covariances.

Consider next Figures 5.3 and 5.4, representing results from the analysis of correlation matrices. For the case of population communalities equal to 1.00, the points no longer are precisely on the lines of best fit. This is brought on by differential effects between attributes arising from standardization of attribute scores in the sample, as implied by Eq. (5.20). These effects are more pronounced in the small sample than in the large sample, as seen by comparing the upper pair of graphs in Figure 5.3 to those in Figure 5.4. Further, the slopes of the lines of

best fit no longer equal the factor score standard deviations in the samples. A further comment about the slopes will be made in the discussion of Table 5.3. Considering the lower graphs in Figures 5.3 and 5.4, representing the case of population communalities equal to .49, considerable divergence of the points from the line of best fit remains for the small sample in Figure 5.3. For the large sample represented in Figure 5.4, the points are close to the line of best fit.

Table 5.2 presents statistics which summarize the goodness of fit of the lines of best fit to the points in the graphs. The statistic presented in this table is called a coefficient of congruence. This is a measure of correspondence between two factors. In the present context, if we consider factor *k* from obtained weight matrix $\boldsymbol{B}$ and factor 1 from population weight matrix $\boldsymbol{B}$, the coefficient of congruence, g_{kl}, between these two factors would be defined as follows:

$$g_{kl} = \frac{\sum_{j=1}^{n} b_{jk}\beta_{jl}}{\underline{\nu}(\sum_{j=1}^{n} b_{jk}^2)(\sum_{j=1}^{n} \beta_{jl}^2)} \qquad (5.23)$$

Geometrically, this value represents the cosine of the angle between the factors when they are plotted in the same factor space. When two factors are exactly congruent, the coefficient of congruence will equal 1.00. As factors become less congruent, the value of the coefficient decreases. Table 5.2 provides coefficients of congruence between corresponding obtained and population common factors for each of the eight analyses conducted.

For the analysis of the covariance matrices and population communalities of 1.00, the coefficients of congruence equal 1.000 indicating a very good fit of the line of best fit to the points. This result is expected from theory and agrees with results represented in Figures 5.1 and 5.2. Continuing to consider the analysis of the covariance matrices, the coefficients of congruence for the communalities of .49 with the small sample are quite low, especially for factor 1. The influence of sample size is indicated by the quite high coefficients of congruence when the sample size is increased to 1000.

In the analyses of the correlation matrices, for the case of population communalities of 1.000 in the small sample, the coefficients of congruence decreased from the corresponding values for the analysis of covariance matrices. There was a similar decrease in the coefficients of congruence for the large sample, but beyond the three decimal places used in the table. For the case of population communality equal to .49, there was a general increase in the coefficients of congruence from analysis of correlation matrices over the values from analysis of covariance matrices.

This somewhat unexpected event was found to occur fairly consistently in a Monte Carlo study involving 100 samples of the same type analyzed here. For factor 1, 81 of the 100

Table 5.2

Coefficients of Congruence: Obtained Factor Weights with Population Factor Weights

Population Communalities	Sample Size	Coefficients of Congruence	
		Factor 1	Factor 2
	Sample Covariance Matrix Analyzed		
1.00	25	1.000	1.000
.49	25	.849	.962
1.00	1000	1.000	1.000
.49	1000	.998	.999
	Sample Correlation Matrix Analyzed		
1.00	25	.999	.995
.49	25	.888	.962
1.00	1000	1.000	1.000
.49	1000	.999	.999

replications yielded a higher coefficient of congruence in the solution obtained from analysis of a correlation matrix over that obtained from analysis of a covariance matrix. For factor 2, 78 of the 100 replications followed the same trend. Both of these divisions were significantly different from chance.

Table 5.3 summarizes results for the slopes of the lines of best fit. In the analysis of the covariance matrices and population communalities of unity, the slopes equal the sample factor score standard deviations. This is the expected result. Comparison of the slopes of the best fitting lines for the analyses of the covariance matrices and the analyses of the correlation matrices indicates that there is a strong tendency for the slopes for the analyses of the correlation matrices to change from the analyses of the covariance matrices toward unit slopes. This could be interpreted that the standardization of the attributes in the sample compensated for the effects of considering standardized factors in the sample. This compensation is not completely accurate.

Table 5.4 presents the correlations between the factors in two different contexts: (a) for the input factor scores in the sample; (b) for the factors obtained from the factor analysis of the sample data. In the analysis of the covariance matrices and population communality of 1.00, the correlations obtained from the analyses equal the factor score correlations in the sample. This equivalence does not hold in any other case. The foregoing equivalencies are the expected results from theory. One other observation is that the correlations for the large samples more nearly equal the zero correlation between the factors in the population than do the correlations for the small samples. This is an expected result, since the accuracy of these estimates of population parameters from sample statistics are dependent on sample size.

These examples were undertaken to illustrate the source of variation in factor weights produced by random sampling of entities in samples. Since, in the factor analytic model, the attribute measures are dependent on the factor scores, sampling effects on the attribute scores are dependent on the sampling variabilities of the factor score statistics. In these examples the population distribution of the factor scores was taken to be multidimensional normal. Thus, sampling effects may be considered in terms of sample means and sample variances and covariances. The effects of sample means are eliminated by the use of deviation measures in the samples (as implied by the computation of covariance and correlation matrices). Thus, the remaining sample effects arise from the sample factor score variance and covariances.

Factors may be divided into the common factors and the unique factors. Each of these classes of factors will have different sampling effects on the factor weights obtained from analyses. Major effects from the common factor variances arise from standardization of the factor scores in the sample, a process inherent in the factoring methods. The restandardization of the factor scores in the sample results in the population factor weights being multiplied by the sample standard deviations of the common factor scores. This effect is illustrated by the slopes of

Table 5.3

Comparisons between Sample Factor SD's and Slopes of Best Fitting Lines:

Obtained Factor Weights on Population Factor Weights

Population Communalities	Sample Size	Factor 1		Factor 2	
		Sample SD	Slope	Sample SD	Slope
Sample Covariance Matrix Analyzed					
1.00	25	.773	.773	1.039	1.039
.49	25	.773	.671	1.039	1.260
1.00	1000	.965	.965	1.008	1.008
.49	1000	.965	.924	1.008	1.014
Sample Correlation Matrix Analyzed					
1.00	25	.773	.988	1.039	1.069
.49	25	.773	.768	1.039	1.176
1.00	1000	.965	1.000	1.008	1.009
.49	1000	.965	.945	1.008	1.003

Table 5.4

Correaltion between Factors (Population Correlation=.000)

Population Communalities	Sample Size	Coefficients of Congruence	
		Factor 1	Factor 2
Sample Covariance Matrix Analyzed			
1.00	25	-.206	-.206
.49	25	-.206	-.389
1.00	1000	-.033	-.033
.49	1000	-.033	-.047
Sample Correlation Matrix Analyzed			
1.00	25	-.206	-.166
.49	25	-.206	-.325
1.00	1000	-.033	-.026
.49	1000	-.033	-.023

the lines of best fit in Figures 5.1 and 5.2 being equal to sample standard deviations when population communalities equal unity. This effect applies to the analysis of matrices of covariance when the attribute scores are expressed in the population unit of measures. Standardization of the attribute scores in the sample by analysis of the sample correlation matrix introduces sampling variability for each attribute. This restandardization of the scores on each attribute may partially adjust for the restandardization of the factor scores. However, this adjustment is incomplete as shown in the examples in Figures 5.3 and 5.4. Since the unique loadings are zero for the case when the population communalities equal unity, there are no effects of the unique factor score variances and covariances.

When the population communalities are less than unity and, thus, the unique factor weights are greater than zero, sampling variation is introduced for individual attribute factor weights by the sampling variation in the unique factors. These variations of the factor weights for the individual attributes are most pronounced in the results for the small sample. See the lower pair of graphs in Figures 5.1 and 5.3. These sampling results are greatly reduced for the large sample.

Finally, as expected, all sampling variations are substantially reduced from the small to the large sample.

5.2. Effects of Selective Sampling

Let us now turn our attention to effects of sampling error on factor analysis under selective sampling. The previous section focused on the effects of sampling error under random sampling, where each individual in the population has a probability of being included in the sample. By contrast, selective sampling involves the case where each individual does not have an equal probability of being selected; rather, the selection of a given individual for a sample is dependent at least in part on that individual's level or score on one or more attributes. These attributes will be referred to as selection variables. If a given individual's scores on the selection variables satisfy the selection criteria, then that individual may be included in the sample; if the scores do not satisfy the criteria, then the individual may not be included in the sample.

Two basic facts about selective sampling must be recognized: (a) selective sampling occurs routinely in practice; and (b) selective sampling may affects results obtained in factor analysis. To consider the first point, samples of individuals are influenced not only by the actions of the experimenter but also by actions of many others and of institutions, as well as by actions of the prospective members of the sample. The experimenter may choose among sources of subjects for experiments, thus imposing a type of selection. In addition, other individuals and institutions may affect the pool from which an experimenter draws samples. For instance, the student body at a highly prestigious college has undergone a stringent system of selection. This system of

selection not only raises the level of talent of the student body, but also tends to reduce the variability of talent and, thus, of the correlations between measures of attributes. Some colleges of lower prestige suffer restriction on both ends of the range of talent. Top applicants are drained away to the more prestigious colleges while the college of lower prestige rejects less qualified applicants. These effects also reduce the range of talent and reduce the variability and correlations among attributes. Thus, samples drawn from student bodies at colleges surely are characterized by various forms of selection. To consider another example, the body of enlistees in the armed forces tend not only to be of lower talent, but also of lower variability. Any sample drawn from such sources will be affected by the selection that has occurred. Finally, one must keep in mind the potential impact of self-selection. That is, in many situations, prospective members of a sample must take some action (e.g., sign up for an experiment, apply to a college) which influences whether or not they may be included in a sample. Considering all of these potential sources of selection, it is clear that most research sample will have been subjected to some type of selective sampling.

To consider the second point made above (i.e., that selective sampling will affect results of factors analysis), we need only recognize some basic influences of selection. As illustrated in the previous paragraph, selected samples will tend to be characterized by low variability on the selection variables, and in turn on other attributes related to the selection variables. This restriction of the range of scores on these attributes will affect not only the variances of those attributes, but also the correlations (and, thus, covariances) of those attributes. Since factor analysis solutions are defined so as to optimally account for the variation and covariation of the attributes in question, it is clearly possible that any selection process which influences the variances and covariances of those attributes will in turn influence results of factor analyses.

To demonstrate these potential affects more clearly, let us consider two simple illustrations. Some effects of selective sampling by an institution are illustrated in a study by French, Tucker, Newman, and Bobbitt (1952). This project involved a study of the system for selection of Reserved Officer candidates for the United States Coast Guard. Some results are shown in Table 5.5. The large pool of applicants was subjected to three stages of selection, where more information was used at each subsequent state. A quantitative test and a verbal test were used at each stage, and some relevant statistics for these tests are shown in Table 5.5. Note that the mean scores increased while the standard deviations decreased in general from stage to stage. In addition, there was a drastic effect on the correlation between the two tests. This effect on the correlation is in part a result of the two tests being tested directly in the selection process. However, some reduction of the correlation would still have occurred if parallel forms of the tests had been used in making the selections. Note that the final entering class included only

Table 5.5
Sequence of Statistics During Selection for a Quantitative Test and a Verbal Test*

	Applicants	After Initial Selection	After Second Selection	Entering Class
N	2253	968	374	128
Quantitative				
M	38.6	42.6	46.2	48.7
D	9.8	9.3	9.0	9.0
Verbal				
M	54.5	64.8	70.1	75.4
SD	14.5	12.5	12.2	10.9
Correlation	.50	.38	.32	.12

* These data are for the class of 1949 at the United States Coast Guard Academy, see French, et al (1952).

French, John W., Tucker, Ledyard R, Newman, Sidney H., and Bobbitt, Joseph M. A factor analysis of aptitude and achievement entrance tests and course grades at the United States Coast Guard Academy. *Journal of Educational Psychology*, 1952, 43, 65-80.

about $5\frac{1}{2}\%$ of the applicants. Any study using this class of cadets must take into account the effects of the considerable selection which occurred.

A different form of selective sampling is illustrated by some results drawn from the study of intelligence by Thurstone and Thurstone (1941). Results are shown in Table 5.6. These results involve two samples which differ in heterogeneity as to semesters in high school; students in sample 1 were from a single half-grade, while students in sample 2 were from a span of one and one-half grades. To illustrate the effects of this difference in heterogeneity, six tests were selected from the two batteries included in the studies by Thurstone and Thurstone (1941). These six tests were similar in the two batteries, though some editing did occur from the tests given to sample 1 to the test given to sample 2. The two correlation matrices for the six tests were taken from the larger matrices published by Thurstone and Thurstone (1941) and are shown in Table 5.6. Since sample 2 is more heterogeneous than sample 1, we would expect that the correlations for corresponding pairs of tests would be higher for sample 2. With one exception, this pattern holds. This effect would in turn lead us to expect difference in the factor solutions obtained from the two samples. In particular, we would expect two things to occur: (a) that the loadings for the more heterogeneous sample would be larger, since they must account for larger correlations among the tests; and (b) that the factor intercorrelations would be larger, since the correlations among the surface attributes are larger. To determine whether these predictions would hold, factor analyses were performed for the two correlation matrices yielding the factor weights given in Table 5.6. Factors V and W correspond to the Thurstones' interpretation of a verbal factor and a word fluency factor. There is seen to be a tendency toward the weights being larger for the more heterogeneous sample, though the strength of this tendency is mediocre. Reasons for the weakness of this tendency will be explored later in this section. The expectation of a larger interfactor correlation in sample 2 is borne out quite distinctly.

A point of considerable importance is that individuals conducting factor analytic studies should be aware of the effects which may be generated by the nature of the samples they use. It must be recognized that selection of some type occurs almost universally in obtaining samples for such studies, and it is clear that such selection can impact on the nature of the factor analysis results obtained.

5.2.1. Mathematical Framework for Theory of Linear Selective Sampling

Given this state of affairs, we will next develop a conceptual and mathematical framework for systematically examining the effects of selective sampling in factor analysis. The development of this framework involves some concepts additional to the common factor model as it has been to this point. A primary addition is a domain of variables involved directly in selection among individuals in a population. These variables, as noted above, will be called

Table 5.6

Comparison of Statistics for Two Samples Differing in Heterogeneity as to School Grade of Individuals*

Sample 1: 710 Students in Grade VIIIB

	Correlation Matrix						Factor Weights		
Test	1	2	3	4	5	6		V	M
1 Completion	1.00	.68	.72	.35	.37	.31	1	.73	.14
2 Sentences	.68	1.00	.77	.23	.29	.22	2	.87	-.07
3 Vocabulary	.72	.77	1.00	.24	.30	.30	3	.87	-.02
4 First Letters	.35	.23	.24	1.00	.47	.53	4	-.04	.71
5 Four-Letter Words	.37	.29	.30	.47	1.00	.43	5	.09	.57
6 Suffixes	.31	.22	.30	.53	.43	1.00	6	-.01	.66

Correlation Between Factors=.50

Sample 2: 437 Students in Grade VIIA, VIIIB, VIIIA

	Correlation Matrix						Factor Weights		
Test	1	2	3	4	5	6		V	M
1 Completion	1.00	.77	.78	.43	.35	.43	1	.84	.01
2 Sentences	.77	1.00	.83	.42	.36	.41	2	.92	-.05
3 Vocabulary	.78	.83	1.00	.47	.42	.48	3	.86	.06
4 First Letters	.43	.42	.47	1.00	.65	.56	4	.03	.76
5 Four-Letter Words	.35	.36	.42	.65	1.00	.51	5	-.04	.77
6 Suffixes	.43	.41	.48	.56	.51	1.00	6	.13	.59

Correlation Between Factors=.62

* These data were collected in 1939 by L. L. Thurstone and Thelma G. Thurston from the studies in the Chicago Schools. At that time advancement in school was by half grade. The correlation matrices were taken from: Thurston, L. L. & Thurston , Thelma G. Factor Studies of Intelligence. Psychometric Monograph No. 2. Chicago, University of Chicago Press, 1941.

selection variables. Selection variables may be of many types. Material used in selection among applicants for admission to a college, or for a job may be characterized as selection variables. Less concrete selection variables are the variety of matters considered by individuals in selecting whether or not to apply. These are variables of selfselection. To consider some other examples, some claim may be made that selective mating though years may have increased the variance on some attributes and led to higher correlations among these attributes. An experimenter may choose among several sources of subjects for experiments. All of these possibilities will be included here under the concept of selection variables.

Another important notion in the present framework is that of a subpopulation. For present purposes, selection will be represented as operating at the level of populations. That is, when selection is applied to some population, those individuals who satisfy the selection criteria comprise a subpopulation. The effects of selection in factor analysis can be examined by determining how the factors in the general population are affected by selection of a subpopulation. In practice one would actually be observing a sample drawn from the subpopulation, and the results of factor analysis of that sample would be subject to the additional effects of random sampling discussed in the previous section. Thus, to examine the effects of selection, it is necessary to employ a framework defined at the level of populations and subpopulations.

A schemata for a theory of selective sampling is shown in Figure 5.5. This schemata is a modification of the schemata shown in Figure 1.2, which represented the basic common factor model and which was discussed in Chapter 1. The common factor model involves internal attributes (or factors) pictured on the left, and surface attributes pictured on the right. A domain of selection variables is pictured in the lower middle. Solid lines with arrowheads are drawn from the internal attributes to the surface attributes to indicate possible causal relations. These effects correspond to the common and unique factor weights. It is important to understand the nature of the relations of the selection variables to the other attributes included in this framework. No causal relations are assumed for the selection variables. Rather, the selection variables are assumed to have associational relations with the internal attributes. These non-zero associational relations are represented in the schemata by dashed lines with double arrowheads. Note also that selection variables may be associated with each other. In the schemata three selection variables are pictured which have associational relations with the two common factors. In addition, one association is indicated between one of the selection variables and one of the unique factors. Of course, other associations are possible between selection variables and all unique factors; however, only one is shown. Relations of unique factors with selection variables are special cases and result in complex effects to be described later. Primary attention will be given to effects of relations of selection variables with common factors. It is these relations which underlie the

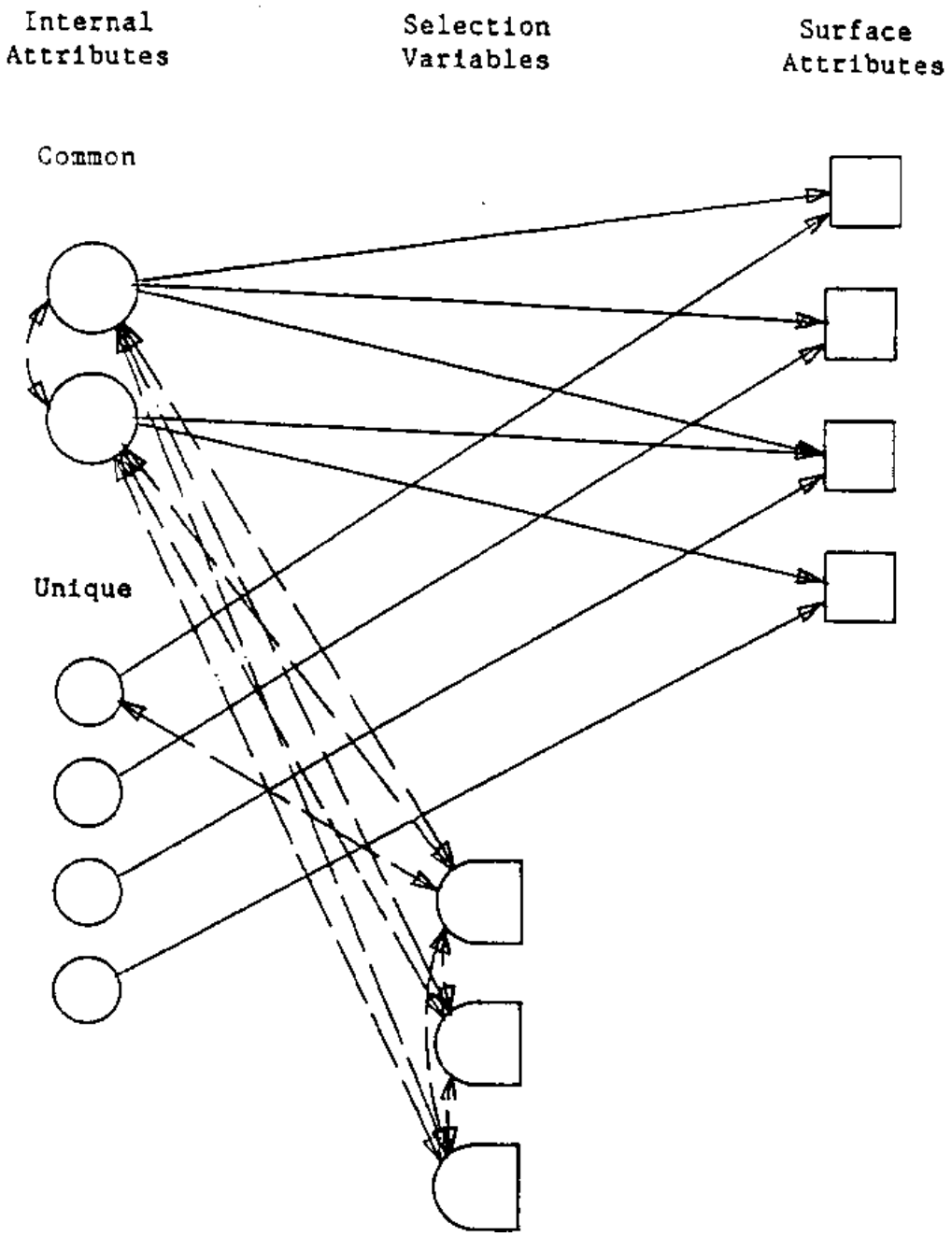

Figure 5.5: Schemata for Theory of Selective Sampling

effects of selection on the observed data. That is, according to the schemata, a selection process based on the selection variables will yield a subpopulation in which the underlying factors have been influenced by selection. Since the factors directly influence the surface attributes, the statistical properties of the surface attributes will also have been influenced by selection. Thus, the relations of the surface attributes to the selection variables are taken to be mediated by relations of the factors to the selection variables.

We now begin the development of a mathematical framework to represent these concepts and relations. The objective is to determine how the parameters of the common factor model, primarily the common factor weights and intercorrelations, are affected by selective sampling; i.e., how the parameters in a subpopulation are different from those in the general population, and how the parameters in different subpopulations are different from each other. We will continue to employ notation used in this and previous chapters for the common factor model, and we will extend this notation and define new notation as necessary to incorporate the selection variables and subpopulations into this framework. A vector of scores on the selection variables is designated $\underline{\boldsymbol{v}}$. In the case considered here, the factor scores are taken to be linearly related to the selection variables. There is a close similarity of the situation to linear, homoscedastic, multiple regression of the factor scores on the selection variables. However, no causal relation is implied; only a relation of association is taken to exist. Even so, the factor scores may be divided into two components: one component of composite measures from the selection variables and a second component of discrepancies. Stated differently, the factor scores may be divided into a component which can be accounted for by their linear association with the selection variables, and a component which cannot be accounted for by that association, The vector of scores on the discrepancy component will be designated $\underline{\ddot{\boldsymbol{v}}}$. These discrepancy scores are taken to be unrelated to the selection variables.

Given these concepts, the relation of a factor score vector, $\underline{\boldsymbol{x}}$, to the selection variables vector, $\underline{\boldsymbol{v}}$, is given by:

$$\underline{\boldsymbol{x}} = \underline{\boldsymbol{v}}\boldsymbol{W}' + \underline{\ddot{\boldsymbol{v}}} \tag{5.23}$$

where $\boldsymbol{W}$ is a weight matrix representing weights applied to the selection variables to produce that part of the factor scores which is linearly associated with the selection variables. In further developing this approach, it will be necessary to employ notation to differentiate among subpopulations. The reader should keep in mind that, as noted above, different selection processes or criteria would yield different subpopulations. The letter *p* will be used as an index for subpopulation. Given this, we define $\boldsymbol{\Sigma}_{xxp}$ as the subpopulation covariance matrix among the factor scores. The assumption stated in the previous paragraph that the discrepancies are unrelated to the selection variables implies that:

$$\boldsymbol{\Sigma}_{v\ddot{v}p} = \mathbf{0} \tag{5.24}$$

for all *p*. A further assumption of homoscedaticity implies that the covariance matrix for the discrepancies is constant across subpopulations. That is,

$$\boldsymbol{\Sigma}_{\ddot{v}\ddot{v}p} = \boldsymbol{\Sigma}_{\ddot{v}\ddot{v}} \tag{5.25}$$

for all *p*.

Let us briefly consider relations between mean vectors on the factors and selection variables. The mean factor score vector, $\boldsymbol{\mu}_{xp}$, for subpopulation *p*, is given, according to Eq. (5.23), by:

$$\boldsymbol{\mu}_{xp} = \boldsymbol{\mu}_{vp}\boldsymbol{W} + \boldsymbol{\mu}_{\ddot{v}p} \tag{5.26}$$

where $\boldsymbol{\mu}_{vp}$ is the mean score vector for the selection variables and $\boldsymbol{\mu}_{\ddot{v}p}$ is the mean discrepancy score vector. No assumption is made in the present context as to the constancy of the mean vectors. Though some assumptions may be made in particular contexts about the mean vectors, emphasis here is on the covariance and correlation relations. In the linear system the covariance and correlation relations are independent of the mean vectors.

Fundamental covariance relations are considered next. The relation of the factor score covariance matrix $\boldsymbol{\Sigma}_{xxp}$ in subpopulation *p* to the selection variable covariance matrix $\boldsymbol{\Sigma}_{vvp}$ in subpopulation *p* and the covariance matrix $\boldsymbol{\Sigma}_{\ddot{v}\ddot{v}}$ is developed from Eq. (5.23) with modifications indicated by Eqs. (5.24) and (5.25). This yields:

$$\boldsymbol{\Sigma}_{xxp} = \boldsymbol{W}\boldsymbol{\Sigma}_{vvp}\boldsymbol{W}' + \boldsymbol{\Sigma}_{\ddot{v}\ddot{v}} \tag{5.27}$$

Another fundamental covariance relation is represented by the common factor model, which is modified in the present context to apply to each subpopulation *p*. The model would then be written as follows:

$$\boldsymbol{\Sigma}_{zzp} = \boldsymbol{\Omega}\boldsymbol{\Sigma}_{xxp}\boldsymbol{\Omega}' \tag{5.28}$$

It is especially important to note that the factor weight matrix, $\boldsymbol{\Omega}$, is not changed in the modification; i.e., there is no subscript *p* on $\boldsymbol{\Omega}$ to indicate a different factor weight matrix for each subpopulation. This is a very strong assumption: the basic attribute score equation is invariant over selective sampling. However, this assumption follows quite naturally from common factor theory, where it is assumed that the basic model defining the linear relations between the modeled attributes and the internal attributes (see Eq. (3.3)) applies to every individual in a population; this would not change when individuals are selected into subpopulations. It is interesting to note here that any particular individual may be a member of several subpopulations; from that view, it would be difficult to argue for different weight

matrices in different subpopulations. However, in some contexts objections may be raised that factor patterns may change from one subpopulation to another, such as for males vs. females. However, some aspects of the current framework may account for such differences. Such difference will be important in factor analysis in multiple populations, which will be discussed later in this section, as well as in Chapter 19.

It will be useful to define a basic, or reference subpopulation. This will be a subpopulation for which the variances and covariances of the selection variables are zero. Such a subpopulation would be the result of "absolute" selection; i.e., in order for an individual to be selected, the individual would have to have exactly a particular set of scores on the selection variables. Let this subpopulation be designated as subpopulation o with $p = o$. We could then write

$$\boldsymbol{\Sigma}_{vvo} = \mathbf{0} \tag{5.29}$$

Combining this definition with the relationship given by Eq. (5.27), we obtain

$$\boldsymbol{\Sigma}_{xxo}=\boldsymbol{\Sigma}_{\ddot{v}\ddot{v}} \tag{5.30}$$

This indicates that in the case of absolute selection, the covariance matrix for the factors in the subpopulation will equal the covariance matrix for the discrepancies. Combining this result with Eq. (5.28), we obtain

$$\boldsymbol{\Sigma}_{zzo} = \boldsymbol{\Omega}\, \boldsymbol{\Sigma}_{xxo} \boldsymbol{\Omega}' = \boldsymbol{\Omega} \boldsymbol{\Sigma}_{\ddot{v}\ddot{v}} \boldsymbol{\Omega}' \tag{5.31}$$

This equation shows the representation of the common factor model in the subpopulation in the case of absolute selection. According to this result, the factor weights are unchanged, and the factor covariances will equal the covariances of the discrepancies.

To this point we have not differentiated among common, specific, and error factors in the development of this mathematical framework. However, since different assumptions may be made and different relations may hold for the different classes of factors, it is useful to do so. The factor weight matrix $\boldsymbol{\Omega}$ and covariance matrix $\boldsymbol{\Sigma}_{\ddot{v}\ddot{v}}$ may be expressed in terms of sections for the common, specific, and error factors. For the factor weight matrix, the following partitioning was presented in Eq. (3.6):

$$\boldsymbol{\Omega} = [\boldsymbol{B}, \boldsymbol{\Xi}, \boldsymbol{E}] \tag{5.32}$$

To distinguish different classes of factors for $\boldsymbol{\Sigma}_{\ddot{v}\ddot{v}}$, it is necessary to employ a second level of subscripts. These are β for common factors, ξ for the specific factors, and ϵ for the error factors. The matrix $\boldsymbol{\Sigma}_{\ddot{v}\ddot{v}}$ can be represented as follows:

$$\boldsymbol{\Sigma}_{\ddot{v}\ddot{v}} = \begin{bmatrix} \boldsymbol{\Sigma}_{\ddot{v}_\beta \ddot{v}_\beta} & \boldsymbol{\Sigma}_{\ddot{v}_\beta \ddot{v}_\xi} & \boldsymbol{\Sigma}_{\ddot{v}_\beta \ddot{v}_\epsilon} \\ \boldsymbol{\Sigma}_{\ddot{v}_\xi \ddot{v}_\beta} & \boldsymbol{\Sigma}_{\ddot{v}_\xi \ddot{v}_\xi} & \boldsymbol{\Sigma}_{\ddot{v}_\xi \ddot{v}_\epsilon} \\ \boldsymbol{\Sigma}_{\ddot{v}_\epsilon \ddot{v}_\beta} & \boldsymbol{\Sigma}_{\ddot{v}_\epsilon \ddot{v}_\xi} & \boldsymbol{\Sigma}_{\ddot{v}_\epsilon \ddot{v}_\epsilon} \end{bmatrix} \tag{5.33}$$

Let us examine the submatrices within defined in this equation. Based on Eq. (5.27), matrix $\boldsymbol{\Sigma}_{\ddot{v}_\beta \ddot{v}_\beta}$ is the covariance matrix among the common factors in the basic subpopulation. Let us consider the common factors to be standardized in that subpopulation, and let us define matrix $\boldsymbol{\Phi}_o$ as the correlation matrix among the common factors in the basic subpopulation. That is,

$$\boldsymbol{\Phi}_o = \boldsymbol{\Sigma}_{\ddot{v}_\beta \ddot{v}_\beta} \tag{5.34}$$

Applying the conventional definitions of factor theory to the basic subpopulation, covariances of the specific factors with the common factors are assumed to equal zero, as are the covariances of the error factors with the common factors and the specific factors. Thus, we can write

$$\boldsymbol{\Sigma}_{\ddot{v}_\xi \ddot{v}_\beta} = \boldsymbol{\Sigma}_{\ddot{v}_\epsilon \ddot{v}_\beta} = \boldsymbol{\Sigma}_{\ddot{v}_\xi \ddot{v}_\epsilon} = \mathbf{0} \tag{5.35}$$

Covariances among the specific factors and among the error factors are also assumed to equal zero. Considering these to be standardized in the reference subpopulation allows us to write

$$\boldsymbol{\Sigma}_{\ddot{v}_\xi \ddot{v}_\xi} = \boldsymbol{\Sigma}_{\ddot{v}_\epsilon \ddot{v}_\epsilon} = \boldsymbol{I} \tag{5.36}$$

Applying the foregoing definition and assumptions to Eq. (5.33), the matrix $\boldsymbol{\Sigma}_{\ddot{v}\ddot{v}}$ can be represented as follows:

$$\boldsymbol{\Sigma}_{\ddot{v}\ddot{v}} = \begin{bmatrix} \boldsymbol{\Phi}_o & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{I} \end{bmatrix} \tag{5.37}$$

Recall that the representation of the common factor model in the basic subpopulation was given in Eq. (5.31). A more detailed representation of this model can now be obtained by substitution from Eqs. (5.32) and (5.37) into (5.31). This yields the following:

$$\boldsymbol{\Sigma}_{zzo} = \boldsymbol{B}\boldsymbol{\Phi}_o\boldsymbol{B}' + \boldsymbol{\Xi}^2 + \boldsymbol{E}^2 \tag{5.38}$$

When the specific factors and the error factors have zero associations with the selection variables, the specific variances and error variances given in $\boldsymbol{\Xi}^2$ and $\boldsymbol{E}^2$, respectively, may be combined into unique variances. (As previously noted, when specific factors have nonzero relations with the selection variables, considerable complexities arise, which will be discussed later in this section.) The combining of specific and error variances to form unique variances was defined in Eq. (3.27), which is repeated here for convenience:

$$\boldsymbol{U^2 = \Xi^2 + E^2} \tag{5.39}$$

Substituting from this equation into Eq. (5.38) yields the following:

$$\boldsymbol{\Sigma_{zzo} = B\Phi_o B' + U^2} \tag{5.40}$$

This equation provides a more detailed (than Eq. (5.31)) expression of the common factor model in the basic subpopulation. It indicates that the common factor weights and unique variances in that subpopulation will be the same as in the general population, but the factor intercorrelations, given in $\boldsymbol{\Phi_o}$, may be different.

Given this representation for the basic subpopulation, let us now consider further examination of the covariance relations defined in Eqs. (5.27) and (5.28) for any subpopulation *p*. It is useful to define partitioned forms of some of the matrices in those equations. In particular, the weight matrix, $\boldsymbol{W}$, expressing the relations of the factors to the selection variables, can be defined as having sections corresponding to the several types of factors. This would be represented as follows:

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{W_\beta} \\ \boldsymbol{W_\xi} \\ \boldsymbol{W_\epsilon} \end{bmatrix} \tag{5.41}$$

It will be assumed that the weights for the error factors are zero; i.e., there is no association between the selection variables and the error of measurement factors. Logically, the errors of measurement in the surface attributes should be independent of the selection variables. Thus, we can write

$$\boldsymbol{W_\epsilon = 0} \tag{5.42}$$

Considering the relation of the specific factors to the selection variables, when these associations are zero then section $\boldsymbol{W_\xi}$ will equal zero. For present purposes, this will be taken to be the case. However, this submatrix will be treated later in this section as being, possibly, nonzero, to represent the case of associations of specific factors to the selection variables.

The preceding developments provide a basis for consideration of the effects of selective sampling on the factor covariance matrix. This necessitates the sectioning of matrix $\boldsymbol{\Sigma_{xxp}}$, the factor covariance matrix for subpopulation *p*, into sections representing the several types of factors as follows:

$$\boldsymbol{\Sigma_{xxp}} = \begin{bmatrix} \boldsymbol{\Sigma_{x_\beta x_\beta p}} & \boldsymbol{\Sigma_{x_\beta x_\xi p}} & \boldsymbol{\Sigma_{x_\beta x_\epsilon p}} \\ \boldsymbol{\Sigma_{x_\xi x_\beta p}} & \boldsymbol{\Sigma_{x_\xi x_\xi p}} & \boldsymbol{\Sigma_{x_\xi x_\epsilon p}} \\ \boldsymbol{\Sigma_{x_\epsilon x_\beta p}} & \boldsymbol{\Sigma_{x_\epsilon x_\xi p}} & \boldsymbol{\Sigma_{x_\epsilon x_\epsilon p}} \end{bmatrix} \tag{5.43}$$

It will be useful to examine relations between sections of $\boldsymbol{\Sigma}_{xxp}$ and matrices $\boldsymbol{W}$, $\boldsymbol{\Sigma}_{vvp}$, and $\boldsymbol{\Sigma}_{\ddot{v}\ddot{v}}$. This can be achieved via Eqs. (5.27), (5.33), and (5.43), and results in the following equations:

$$\boldsymbol{\Sigma}_{x_\beta x_\beta p} = \boldsymbol{W}_\beta \boldsymbol{\Sigma}_{vvp} \boldsymbol{W}_\beta' + \boldsymbol{\Phi}_o \tag{5.44}$$

$$\boldsymbol{\Sigma}_{x_\xi x_\beta p} = \boldsymbol{W}_\xi \boldsymbol{\Sigma}_{vvp} \boldsymbol{W}_\beta' \tag{5.45}$$

$$\boldsymbol{\Sigma}_{x_\epsilon x_\beta p} = \boldsymbol{W}_\epsilon \boldsymbol{\Sigma}_{vvp} \boldsymbol{W}_\beta' \tag{5.46}$$

$$\boldsymbol{\Sigma}_{x_\xi x_\xi p} = \boldsymbol{W}_\xi \boldsymbol{\Sigma}_{vvp} \boldsymbol{W}_\xi' + \boldsymbol{I} \tag{5.47}$$

$$\boldsymbol{\Sigma}_{x_\xi x_\epsilon p} = \boldsymbol{W}_\xi \boldsymbol{\Sigma}_{vvp} \boldsymbol{W}_\epsilon' \tag{5.48}$$

$$\boldsymbol{\Sigma}_{x_\epsilon x_\epsilon p} = \boldsymbol{W}_\epsilon \boldsymbol{\Sigma}_{vvp} \boldsymbol{W}_\epsilon' + \boldsymbol{I} \tag{5.49}$$

With the assumption of Eq. (5.42) that the weights for the error of measurement factors equal zero, the covariance of the error factors with the other factors, given by Eqs. (5.46) and (5.48) remain zero. That is,

$$\boldsymbol{\Sigma}_{x_\epsilon x_\beta p} = \boldsymbol{\Sigma}_{x_\epsilon x_\xi p} = \boldsymbol{0} \tag{5.50}$$

The covariance matrix among the error of measurement factors remains an identity matrix:

$$\boldsymbol{\Sigma}_{x_\epsilon x_\epsilon p} = \boldsymbol{I} \tag{5.51}$$

More complex is the status of the covariances involving the specific factors; as shown in Eq. (5.45) and (5.47), these covariances will depend upon the specific factors weight matrix, $\boldsymbol{W}_\xi$. As noted above, we will assume for present purposes that $\boldsymbol{W}_\xi$ equals zero, meaning that there is no relation between the selection variables and the specific factors.

A very important and interesting topic for mathematical consideration is that of standardization of factor scores and of attribute scores in subpopulations. Usually, factor analytic procedures define the factor scores to be standardized. When factor analysis is applied to a correlation matrix, the attribute scores have also been standardized. It will be shown that these standardizations affect comparisons between results obtained from studies in different subpopulations. That is, factor analysis results obtained from factor analyzing data drawn from different subpopulations may be different simply as a result of the standardizations of factors and attributes within those subpopulations.

Let us consider first the standardization of common factor scores in a subpopulation. Let us define a diagonal matrix $[\boldsymbol{\Sigma}_d]_{x_\beta x_\beta p}$ which contains the variances of the common factors in

subpopulation *p*. That is,

$$[\Sigma_d]_{x_\beta x_\beta p} = Diag(\Sigma_{x_\beta x_\beta p}) \tag{5.52}$$

Then the correlation matrix among the common factors in subpopulation *p* can be defined as follows:

$$\Phi_p = [\Sigma_d]^{-\frac{1}{2}}_{x_\beta x_\beta p} \Sigma_{x_\beta x_\beta p} [\Sigma_d]^{-\frac{1}{2}}_{x_\beta x_\beta p} \tag{5.53}$$

A compensating rescaling of the common factor weight matrix involves definition of a matrix $\widetilde{B}_p$ as follows:

$$\widetilde{B}_p = B[\Sigma_d]^{-\frac{1}{2}}_{x_\beta x_\beta p} \tag{5.54}$$

From the definitions given in Eqs. (5.53) and (5.54) we can write

$$B\Sigma_{x_\beta x_\beta p} B' = \widetilde{B}_p \Phi_p \widetilde{B}_p{}' \tag{5.55}$$

When the specific factor weights in W_ξ are zero, the factor equation for covariance matrix Σ_{zzp}, obtained from Eqs. (5.28), (5.32), (5.39), and (5.43), is

$$\Sigma_{zzp} = B\Sigma_{x_\beta x_\beta p} B' + U^2 \tag{5.56}$$

The usual equation for factor analysis of a covariance matrix can then be obtained by combining equations (5.55) and (5.56), yielding the following:

$$\Sigma_{zzp} = \widetilde{B}_p \Phi_p \widetilde{B}_p{}' + U^2 \tag{5.57}$$

Note that matrices $\widetilde{B}_p$ and Φ_p represent the common factor weights and intercorrelations in subpopulation *p*, and that they are particularly affected by the standardization of the common factor scores within the subpopulation, as given in Eqs. (5.53) and (5.54).

Now that we have developed a representation of the common factor model within a given subpopulation *p*, an important topic concerns the relations between the factors obtained in the subpopulation and factors obtained from a different subpopulation. Researchers often wish to study differences in factor structures across populations. It will now be shown that certain types of such differences can be represented using the present framework of subpopulations. Let *q* designate a second subpopulation. Eq. (5.54) may be written for subpopulation q as follows:

$$\widetilde{B}_q = B[\Sigma_d]^{-\frac{1}{2}}_{x_\beta x_\beta q} \tag{5.58}$$

We wish to determine the relation between $\widetilde{B}_q$ and $\widetilde{B}_p$; i.e., between the factor weights obtained from the two subpopulations. This relation can be found from Eqs. (5.54) and (5.58) to be

$$\widetilde{B}_q = \widetilde{\boldsymbol{B}}_p [\boldsymbol{\Sigma}_d]^{-\frac{1}{2}}_{x_\beta x_\beta p} [\boldsymbol{\Sigma}_d]^{\frac{1}{2}}_{x_\beta x_\beta q} \tag{5.59}$$

Since both $[\boldsymbol{\Sigma}_d]_{x_\beta x_\beta p}$ and $[\boldsymbol{\Sigma}_d]_{x_\beta x_\beta q}$ are diagonal matrices, their product is also a diagonal matrix which may be designated $\boldsymbol{\Psi}_{pq}$:

$$\boldsymbol{\Psi}_{pq} = [\boldsymbol{\Sigma}_d]^{-\frac{1}{2}}_{x_\beta x_\beta p} [\boldsymbol{\Sigma}_d]^{\frac{1}{2}}_{x_\beta x_\beta q} \tag{5.60}$$

Substituting from this equation into Eq. (5.59) yields

$$\widetilde{B}_q = \widetilde{\boldsymbol{B}}_p \boldsymbol{\Psi}_{pq} \tag{5.61}$$

This is an important equation because it indicates that the effect of standardization of common factors within subpopulations is to make subpopulation factor weight matrices proportional by columns. This has implications for a number of subsequent topics such as comparing results from different factor analytic studies and factor analysis in multiple populations.

Let us next consider the second type of standardization which affects results obtained from subpopulations: standardization of attribute scores. Such standardization involves converting the covariance matrix among attribute scores into a correlation matrix. Let us define a matrix $[\boldsymbol{\Sigma}_d]_{zzp}$ containing variances of the modeled attributes:

$$[\boldsymbol{\Sigma}_d]_{zzp} = \boldsymbol{Diag}(\boldsymbol{\Sigma}_{zzp}) \tag{5.62}$$

The population correlation matrix among the attribute scores is then given by

$$\boldsymbol{R}_{zzp} = [\boldsymbol{\Sigma}_d]^{-\frac{1}{2}}_{zzp} \boldsymbol{\Sigma}_{zzp} [\boldsymbol{\Sigma}_d]^{-\frac{1}{2}}_{zzp} \tag{5.63}$$

This transformation must be applied to the common factor weight matrix to yield a common factor weight matrix for analysis of the correlation matrix. This can be seen by applying the transformation shown in Eq. (5.63) to both sides of Eq. (5.57). This leads us to define the common factor weight matrix for the analysis of the correlation matrix as follows:

$$\widetilde{\widetilde{\boldsymbol{B}}}_p = [\boldsymbol{\Sigma}_d]^{-\frac{1}{2}}_{zzp} \widetilde{\boldsymbol{B}}_p \tag{5.64}$$

The unique variances must also be transformed, thus yielding

$$\widetilde{\boldsymbol{U}^2_p} = [\boldsymbol{\Sigma}_d]^{-\frac{1}{2}}_{zzp} \boldsymbol{U}^2 [\boldsymbol{\Sigma}_d]^{-\frac{1}{2}}_{zzp} \tag{5.65}$$

Substitution from Eqs. (5.64) and (5.65) into Eq. (5.57) yields the usual factor equation for analysis of a correlation Matrix:

$$\boldsymbol{R}_{zzp} = \widetilde{\widetilde{\boldsymbol{B}}}_p \boldsymbol{\Phi}_p \widetilde{\widetilde{\boldsymbol{B}}}_p' + \widetilde{\widetilde{\boldsymbol{U}}}^2 \tag{5.66}$$

Note that the correlation matrix, $\boldsymbol{\Phi}_p$, for the common factors in subpopulation p is not transformed when the attributes are standardized. However, matrices $\widetilde{\widetilde{\boldsymbol{B}}}_p$ and $\widetilde{\widetilde{\boldsymbol{U}}}_p^2$, which represent the usual results from factoring a correlation matrix, have been affected by this standardization.

To complete the discussion of this issue of standardization of attributes, let us consider how the relation between solutions from two different subpopulations is affected. Let us define a factor weight matrix $\widetilde{\widetilde{\boldsymbol{B}}}_q$ for subpopulation q. Equation (5.64) allows us to write

$$\widetilde{\widetilde{\boldsymbol{B}}}_q = [\boldsymbol{\Sigma}_d]_{zz\mathbf{q}}^{-\frac{1}{2}} \widetilde{B}_q \tag{5.67}$$

Substituting into this equation from Eq. (5.61) yields

$$\widetilde{\widetilde{\boldsymbol{B}}}_q = [\boldsymbol{\Sigma}_d]_{zz\mathbf{q}}^{-\frac{1}{2}} \widetilde{\boldsymbol{B}}_p \boldsymbol{\Psi}_{pq} \tag{5.68}$$

Finally, substituting from Eq. (5.64) into this equation yields

$$\widetilde{\widetilde{\boldsymbol{B}}}_q = [\boldsymbol{\Sigma}_d]_{zz\mathbf{q}}^{-\frac{1}{2}} [\boldsymbol{\Sigma}_d]_{zzp}^{\frac{1}{2}} \widetilde{\widetilde{\boldsymbol{B}}}_p \boldsymbol{\Psi}_{pq} \tag{5.69}$$

This equation shows that when attributes are standardized within subpopulations, the relation between common factor weight matrices obtained from the subpopulations will be affected by that differential standardization. In particular, there will be a proportionality of rows in the two matrices. Combining this result with the effect of standardization of common factors, also shown in Eq. (5.69), it can be seen that standardization of factors and attributes within subpopulations causes differences in factor weights obtained when factor analysis is applied to data from those subpopulations. Specifically, there is a proportional rescaling of both rows and columns of the common factor weight matrix. This result has considerable implications for the matching of factor matrices from different studies. When these studies involve analysis of correlation matrices, the factor matrices would have to be rescaled by rows so that there is a common unit of measure between the two studies for each attribute.

5.2.2. Demonstration of Effects of Selective Sampling

We will now describe and report results of a simulated demonstration of selective sampling designed to illustrate various features of the theory of linear selective sampling. This example was designed to parallel in a number of aspects the results given in Table 5.6 from the studies by Thurstone and Thurstone (1941). This demonstration includes no effects of random sampling; that is, it shows the effects of selection in the context of subpopulations as described above. Further, no model error was incorporated into the demonstration; i.e., there is no lack of

fit of the factor model to the simulated data. Consequently, theoretical relations are fitted precisely. This demonstration reveals possible unobservable relations and intermediate results not available in the results given in Table 5.6 for the Thurstone and Thurstone (1941) data.

Parameters for the demonstration are given in Table 5.7. The basic common factor matrix, $\boldsymbol{B}$, representing the basic subpopulation, and uniquesses, $\boldsymbol{U^2}$, are given at the upper left. Diagonal entries in $\boldsymbol{U^2}$ are listed in a column. There were six attributes and two common factors. Three of the attributes had high weights on the first factor and zero weights on the factor. The reverse was true for the last three attributes, having zero weights on the first factor and high weights on the second factor. The basic covariance matrix among common factors, $\boldsymbol{\Phi}_o$, is given at the upper right. The foregoing matrices were used to compute the basic covariance matrix among attributes, $\boldsymbol{\Sigma}_{zzo}$, by Eq. (5.42). This matrix is given in the center of Table 5.7. One selection variable was used in the demonstration. This variable might be thought of as general mental development. To parallel the Thurstone and Thurstone (1941) studies, the standard deviation on this selection variable was taken to equal 1.00 (for one semester for the subjects in sample 1 of the Thurstone's studies) for subpopulation 1, and equal to 3.00 (for three semesters for the subjects in sample 2 of the Thurstone's studies) for subpopulation 2. These standard deviations and associated variance are given next to the bottom of Table 5.7. The covariance matrix, $\boldsymbol{\Sigma}_{vvp}$, is 1 $\times$ 1 and contains for each subpopulation the variance of the selection variable for the subpopulation. The weight matrix, $\boldsymbol{W}_\beta$, for the common factors is given at the bottom of Table 5.7.

Covariance relations in the two subpopulations are given in Table 5.8. The top half of this table is for subpopulation 1 and the bottom half is for subpopulation 2. For both subpopulations, the covariance matrix for the common factors is at the upper left of the halves for the subpopulations. The standard deviations are given below each covariance matrix. These covariance matrices are $\boldsymbol{\Sigma}_{x_\beta x_\beta 1}$ and $\boldsymbol{\Sigma}_{x_\beta x_\beta 2}$. They were computed by Eq. (5.44). Note that the variances and covariances have increased from the values in the basic covariance matrix in Table 5.7. These increases are greater for subpopulation 2 than for subpopulation 1, this being the result of the variance on the selection variable being greater for subpopulation 2 than for subpopulation 1.

The correlation matrices among the common factors are given to the right of the covariance matrices. These are matrices $\boldsymbol{\Phi}_1$, and $\boldsymbol{\Phi}_2$; they were computed by Eq. (5.53). Note that the correlations of .50 and .64 between the common factors are greater than the corresponding correlation of .47 for the basic population given in Table 5.7. These two values correspond closely to the correlations for the Thurstone's results given in Table 5.6.

Covariance matrices among the attributes are given in Table 5.8 for both subpopulations.

Table 5.7
Parameters for Demonstration of Effects of Selective Sampling in Factor Analysis

Basic Factor Weight Matrix

	Factor		Uniqueness	
Attribute	1	2	Loading	Variance
1	.90	.00	.44	.19
2	.70	.00	.71	.51
3	.50	.00	.87	.75
4	.00	.60	.80	.64
5	.00	.70	.71	.51
6	.00	.80	.60	.36

Basic Covariance Matrix Among Common Factors

Factor	1	2
1	1.00	.47
2	.47	1.00

Basic Covariance Matrix Among Attributes

Attribute	1	2	3	4	5	6
1	1.00					
2	.63	1.00				
3	.45	.35	1.00			
4	.25	.20	.14	1.00		
5	.30	.23	.16	.42	1.00	
6	.34	.26	.19	.48	.56	1.00

Subpopulation SD and Variance on Selection Variable

Subpopulation	SD	Variance
1	1.00	1.00
2	3.00	9.00

Weight Matrix of Common Factors on Selection Variable

Common Factor	Weight
1	.20
2	.30

These are matrices $\boldsymbol{\Sigma}_{zz1}$, and $\boldsymbol{\Sigma}_{zz2}$. They were obtained by Eq. (5.56) using the basic factor matrix $\boldsymbol{B}$ and uniqueness $\boldsymbol{U^2}$ given in Table 5.7 and the covariance matrices among the factors in Table 5.8, which were described previously. The variances and covariances are larger for subpopulation 2 than for subpopulation 1. The factor matrices for the covariance matrices were obtained by Eq. (5.54). These are matrices $\widetilde{\boldsymbol{B}}_{\mathbf{1}}$ and $\widetilde{B}_{\mathbf{2}}$. These factor matrices along with the correlation matrices among the common factors and the uniqueness reproduce the covariance matrices according to Eq. (5.57). Note that the factor loadings are larger for subpopulation 2 than for subpopulation 1. The relation between the factor weights is given by Eq. (5.61). This relation will be discussed further in connection with the graphs in subsequent figures.

Table 5.9 presents the correlation relations in the subpopulations. The correlation matrices are $\boldsymbol{R}_{zz1}$ and $\boldsymbol{R}_{zz2}$ which were computed by Eq. (5.63). As in the Thurstone data given in Table 5.6, the correlations are greater for subpopulation 2 than for subpopulation 1. The common factor matrices obtained by factor analyzing these correlation matrices are given on the right of each correlation matrix. These are matrices $\widetilde{\boldsymbol{B}}_{\mathbf{1}}$ and $\widetilde{\boldsymbol{B}}_{\mathbf{2}}$. These factor matrices are related to the factor matrices in Table 5.8 for covariance matrices by Eq. (5.64). The factor weights are greater for subpopulation 2 than for subpopulation 1. The factor weights for the correlation matrices will be compared graphically.

Figures 5.6 and 5.7 present graphical comparisons of factor weights in two samples or two subpopulations. A comparison of the factor weights for the Thurstone studies is given in Figure 5.6. The graph on the left is for factor V (verbal ability) and the graph on the right is for factor W (word fluency). In each of these graphs there is a point for each attribute with coordinates equal to the factor loadings in the two samples, A line of mutual relation is drawn on each graph. The slopes of these lines of relation are greater than unity indicating that the factor weights are larger for sample 2 than for sample 1. This corresponds to the greater heterogeneity of sample 2 than of sample 1 with respect to the grades in school. The dispersion of the points from the lines of relation may be due to both sampling variability and to editing of the tests between the two experiments.

It is interesting to compare these results to those shown in Figure 5.7 which presents factor weights for the two subpopulations in the simulation study. The upper pair of graphs are for the factor weights obtained from the covariance matrices while the lower pair of graphs are for the factor weights obtained from the correlation matrices. These factor weights themselves are shown in Table 5.8 and 5.9, respectively. As in Figure 5.6, there is a point on each graph for each attribute with coordinates equal to the factor weights in the two subpopulations. The graphs on the left are for factor 1 and the graphs on the right are for factor 2. A line of relation is drawn on each graph with slopes as indicated.

Table 5.8
Covariance Relations in Subpopulations

SUBPOPULATION 1

Covariance and Correlation Matrices Among Common Factors

Covariance Matrix			Correlation Matrix		
Factor	1	2	Factor	1	2
1	1.04	.53	1	1.00	.50
2	.53	1.09	2	.50	1.00
SD	1.02	1.04			

Covariance Matrix Among Attributes							Factor Matrix	
Attribute	1	2	3	4	5	6	1	2
1	1.03						1 .92	.00
2	.66	1.02					2 .71	.00
3	.47	.36	1.01				3 .51	.00
4	.29	.22	.16	1.03			4 .00	.63
5	.33	.26	.19	.46	1.04		5 .00	.73
6	.38	.30	.21	.52	.61	1.06	6 .00	.84
SD	1.02	1.01	1.00	1.02	1.02	1.03		

SUBPOPULATION 2

Covariance and Correlation Matrices Among Common Factors

CovarianceMatrix			Correlation Matrix		
Factor	1	2	Factor	1	2
1	1.36	1.01	1	1.00	.64
2	1.01	1.81	2	.64	1.00
SD	1.17	1.35			

Covariance Matrix Among Attributes							Factor Matrix	
Attribute	1	2	3	4	5	6	1	2
1	1.29						1 1.05	.00
2	.86	1.18					2 .82	.00
3	.61	.48	1.09				3 .58	.00
4	.55	.42	.30	1.29			4 .00	.81
5	.64	.49	.35	.76	1.40		5 .00	.94
6	.73	.57	.40	.87	1.01	1.52	6 .00	1.08
SD	1.14	1.08	1.00	1.14	1.18	1.23		

Table 5.9
Correlation Relations in Subpopulations

SUBPOPULATION 1

	Correlation Matrix Among Attributes							Factor Matrix	
Attribute	1	2	3	4	5	6		1	2
1	1.00						1	.90	.00
2	.64	1.00					2	.71	.00
3	.46	.36	1.00				3	.51	.00
4	.28	.22	.16	1.00			4	.00	.62
5	.32	.25	.18	.44	1.00		5	.00	.72
6	.37	.29	.21	.50	.58	1.00	6	.00	.81

Correlation Between Factors=.50

SUBPOPULATION 2

	Covariance Matrix Among Attributes							Factor Matrix	
Attribute	1	2	3	4	5	6		1	2
1	1.00						1	.92	.00
2	.70	1.00					2	.75	.00
3	.52	.42	1.00				3	.56	.00
4	.42	.34	.26	1.00			4	.00	.71
5	.47	.39	.29	.57	1.00		5	.00	.80
6	.52	.42	.31	.62	.70	1.00	6	.00	.87

Correlation Between Factors=.64

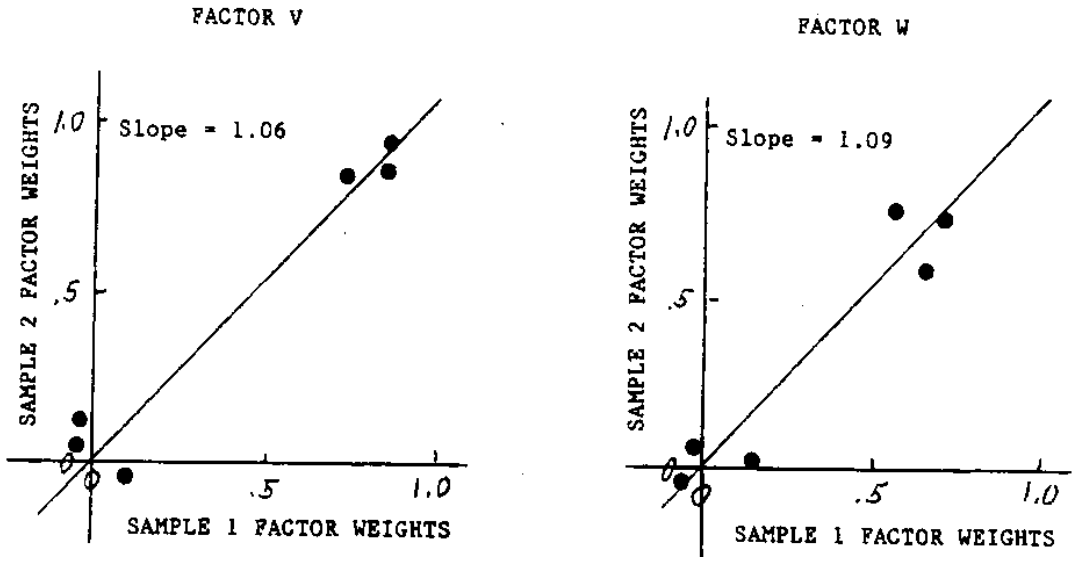

Figure 5.6: Comparison of Factor Weights for Two Samples Differing in Heterogenity as to School Grade of Individuals

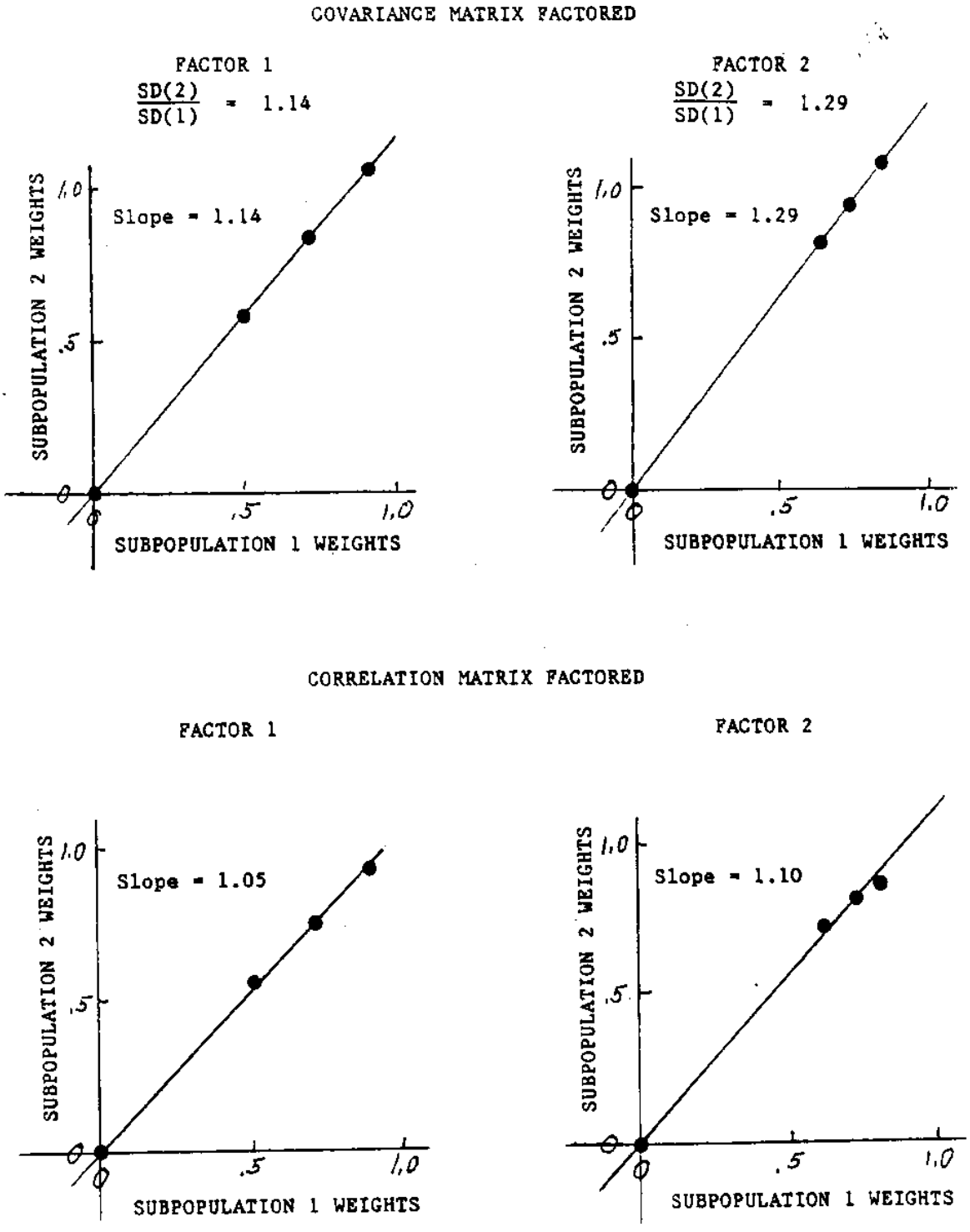

Figure 5.7: Comparison of Factor Weights for Two Subpopulaions in the Selective Sampling Demonstration. SD(1) is the Standard Deviation of Factor Scores in Subpopulation 1; SD(2) is the Standard Deviation of Factor Scores in Subpopulation 2.

For the covariances matrices graphs the ratios of the common factor scores are indicated. Note that the slopes of the lines of relations for these graphs equal the factor score ratios of standard deviations. Note, also, that the points fall precisely on the lines of relation. These results are consistent with and predicted by Eq. (5.61): the covariance matrices factor weights in two rations are strictly proportional by columns.

For the correlation matrices graphs in Figure 5.7, the slopes of the lines of best fit are greater than unity but less than the slopes for the covariance matrices graphs. A point of interest is that these lines of best fit for the correlation matrices have slopes quite close to the slopes of the lines for the Thurstone studies shown in Figure 5.6. Standardizing the attribute scores in subpopulations does not make the factor weights equal between the two subpopulations. There is only a partial adjustment toward equality from the relation for factor weights for the covariance matrices. Further, note that the points in the lower graphs in Figure 5.7 vary a small amount from the lines of best fit. The source of this variation can be understood by considering Eq. (4.69). The product of matrices $[\boldsymbol{\Sigma}_d]_{zz\mathbf{q}}^{-1}$ and $[\boldsymbol{\Sigma}_d]_{zzp}$ is generally not an identity matrix, nor even a constant. Thus, in matching factors from different subpopulations, an assumption of equality for factor weights from correlation matrices is not justified. Nor is it justifiable to assume that a proportionality exists between such weights and that the proportionality represents the ratio of standard deviations on a selection variable.

To conclude this discussion of the effects of selective sampling, let us consider the case in which selection variables are related to specific factors. All previous discussion has assumed that these relations were zero; a variety of complexities occurs when the specific factors have nonzero relations with the selection variables. In these cases the weight matrix $\boldsymbol{W}_\xi$ is not zero, which results in the specific factors not having zero covariances with the common factors. Further, the covariance matrix among the specific factors no longer is an identity matrix. These results are shown in Eqs. (5.45) and (5.47). A number of situations exist which lead to some simplifications; however, none of the special cases will be treated here.

Individuals using factor analysis should recognize that a fairly common type of situation occurs which can result in selection variables being related to specific factors. Frequently, data are available on tests used in selection. These tests may be admissions examinations such as the SAT or the ACT. When results of such tests are incorporated into the battery of attributes being analyzed, there are relations of the selection variables with the specific factors of these tests. In fact, there are relations of the error of measurement factors with the selection variables. Use of parallel tests to the selection tests eliminates the relations of the error of measurement factors to the selection variables, but does not eliminate the relations of the specific factors with the selection variables. If there is only one selection test in the battery, the loadings of this test will

be altered. If there are two or more selection tests in the battery, not only will the loadings of these tests be altered, but also one or more common factors will be added with some negative loadings. Experimenters should be aware of these possibilities.

5.3. Effects of Selection of Attributes

In the preceding sections we have examined how obtained factor solutions can be influenced by various aspects of the selection of a sample of individuals from the population. It is also important to recognize that obtained factor solutions can be influenced by another selection process: the selection of attributes from the domain of interest. A consideration of the nature of common factors themselves will provide a basis for understanding the critical role that selection of attributes plays in affecting the obtained factor solution. A common factor is defined as a factor which is common to more than one attribute in the battery. Unique factors, on the other hand, represent those influences which operate on only a single attribute. (Let us keep in mind that unique factors are composed of both specific and error portions, which correspond to systematic and random influences on single variables, respectively.) Given these definitions, it should be clear that the obtained factor solution is determined by the attributes selected for inclusion in the battery to be factor analyzed. In particular, the obtained common factors will be determined to a great extent by the selection of the attributes.

One way to gain a more explicit understanding of this phenomenon is to consider the effects of altering a given battery of attributes by deleting or adding attributes. Suppose that we have a battery of attributes and have obtained a factor solution that is characterized by a given number of common factors. Let us consider the impact of deleting attributes from the battery. If we were to delete all attributes representing a given common factor and conduct the factor analysis on the reduced battery, the given common factor would disappear completely from the obtained solution. An interesting variation of this phenomenon occurs if we consider the result of deleting all but one of the attributes representing a given factor. In that case, the factor would not be obtained as a common factor, but its influence would still be present in the form of a specific factor affecting the one remaining attribute. In such a case, the remaining attribute would have a much higher unique variance in the reduced battery than it did in the original battery.

It is interesting to consider the converse of this process; i.e., adding attributes to a given battery. Clearly, such a process can give rise to one or more additional common factors in the expanded battery, though this would not have to occur. To simplify the discussion, suppose two new attributes are added to an existing battery. These two new attributes may have in common a factor which was not present at all in the original battery. Thus, one additional common factor would be obtained in the expanded battery. Alternatively, the two new attributes may be influenced by the same common factors already represented in the battery, in which case no

additional common factors would be obtained. Another possible case is that one or both of the new attributes may be influenced by a factor which was a specific factor in the original battery; thus, the expanded battery would yield an additional common factor which involves both new and original attributes.

While a variety of other cases could be pointed out, the critical issue is to recognize that the modification of an existing battery of attributes by adding or deleting attributes can have a substantial impact on obtained factor solutions. Let us consider a simple demonstration of this phenomenon. This demonstration will involve examining the effects of modifying the battery of attributes employed in the illustration presented in Section 1.3. Recall that that battery contained nine attributes, with three attributes representing each of three common factors. The common factors, with corresponding attributes listed in parentheses, were numerical calculation; (addition, multiplication, three-higher), spatial relations (figures, cards, flags), and perceptual speed (identical numbers, faces, mirror reading). The three-factor solution for this battery was presented in Table 1.2. Suppose this battery were modified by deleting the last two tests; i.e., the faces test and the mirror reading test. Based on the discussion above, we can predict the impact of this modification on the obtained factor solution. The perceptual speed factor would no longer be present as a common factor, since it would influence only one remaining attribute. Thus, we would expect to obtain only two common factors. In addition, we would expect that the remaining test representing the perceptual speed factor (i.e., identical numbers) would have a much higher unique variance in the solution obtained from the reduced battery.

A factor solution was obtained for the reduced battery (by applying factor analysis methodology to the 7×7 correlation matrix obtained by deleting the last two attributes from the matrix shown in Table 1.1). A two-factor solution was found to fit these data very well, and the weights and intercorrelations for the two factors, along with the communalities of the attributes, are shown in Table 5.10. The characteristics of this solution correspond to the predicted results. The two obtained factors correspond to numerical calculations and spatial relations, and no perceptual speed factor is obtained. The correlation between the two factors (.16) is very similar to the correlation between those two factors in the three-factor solution (.15). The communality of the identical numbers test is substantially reduced in the two-factor solution (.28) from what it was in the three-factor solution (.44), reflecting the fact that some variance which was accounted for by the three-factor solution has now become unique variance in that attribute. The only reason that the communality has not dropped to nearly zero is that the identical numbers test does have a moderate loading on the numerical calculations test. Other demonstrations using the original nine-test battery could be carried out in order to examine other types of attribute selection on obtained factor solutions. We leave this as an exercise for the reader.

Another issue inherent in the process of attribute selection involves the number of

Table 5.10

Factor Solution for Seven Mental Tests

	Factor Weights		Communalities
Attribute	1	2	
1. Addition	.60	.05	.37
2. Multiplication	.75	-.06	.55
3. Three-Higher	.52	.32	.42
4. Figures	-.03	.71	.50
5. Cards	.00	.79	.63
6. Flags	.03	.71	.51
7. Ident. Numbers	.53	.02	.28

Factor Intercorrelations

Factor	1	2
1	1.00	.16
2	.16	1.00

attributes to be included in the battery which is to be analyzed. For scientifically convincing results, an obtained common factor structure should be overdetermined above a minimum necessary to satisfy mathematical conditions. By overdetermination, we mean that the structure represented in the obtained solution should be uniquely determined and highly constrained, rather than arbitrary. This overdetermination will be shown to depend on a relationship between the number of common factors in the domain being investigated with the number of attributes in the observed battery. A coefficient of overdetermination will be introduced, and an inequality will be developed between the number of attributes to be included in a battery and the number of common factors in the domain.

This development follows and extends Thurstone's (1935, 1947) work on the number of independent common factors. Inequalities developed by Thurstone have been reproduced in texts on factor analysis (e.g., Mulaik, 1972). Further investigation of this problem was reported by Ledermann (1937).

Let r represent the number of common factors expected from the analysis, and let n represent the number of attributes in the battery. It is useful to consider the number of correlations among the attributes; this would be the number on one side of the diagonal of the $n \times n$ correlation matrix, which would be $n(n\text{-}1)/2$. This value represents the number of intercorrelations, which are the values employed to obtain a factor solution. It is also useful to consider the number of independent factor weights being estimated in a factor analysis. The number of weights in the factor weight matrix would be nr. However, an important fact here is that it is always possible to transform a given factor weight matrix so that a number of entries in the weight matrix will be zero. It will be shown in Chapter 9 that this number is $r(r\text{-}1)/2$. Thus, the number of independent factor weights in an obtained solution can be defined as the total number of weights, minus the number that can be set to zero by transformation of the factors. This would yield $nr\text{-}r(r\text{-}1)/2$. The basic inequality is that the number of correlations (i.e., the values being used to obtain parameter estimates) should be equal to or greater than the number of independent factor weights (i.e., the parameters being estimated). This would be written as follows:

$$n(n-1)/2 \geq nr - r(r-1)/2 \tag{5.70}$$

In order for the results to be overdetermined, the number of correlations should be somewhat greater than the number of independent factor weights. A coefficient of overdetermination will be designated as g and may be defined such that the number of correlations is at least g times the number of independent factor weights. This leads to a revision of Eq. (5.70) to yield

$$n(n-1)/2 \geq \mathbf{g}[nr - r(r-1)/2] \tag{5.71}$$

Solution of this inequality for *n* is facilitated by defining n_c which satisfies the equality of Eq. (5.71) and letting

$$n \geq n_c \tag{5.72}$$

The definition of n_c allows us to write

$$n_c(n_c\text{-}1)/2 = \mathbf{g}[n_c r\text{-}r(r\text{-}1)/2] \tag{5.73}$$

Eq. (5.73) can be manipulated to obtain

$$n_c^2 - (1 + 2\mathbf{g}r)n_c + \mathbf{g}r(r-1) = 0 \tag{5.74}$$

Eq. (5.74) may be solved for n_c via the application of the quadratic formula. This yields the following:

$$n_c = (\mathbf{g}r + 1/2) + \underline{\nu}(\mathbf{g}r + 1/2)^2 - \mathbf{g}\boldsymbol{r}(r-1) \tag{5.75}$$

Eq. (5.75) coupled with Eq. (5.72) provides the desired relation of battery size *n* to the estimated number of factors *r*. The minimum value of *n* would be the least integer equal to or greater than n_c. These formulae allow one to calculate the minimum value of *n* for any given number of factors *r* and coefficient of overdetermination *g*. Values of minimum *n* for selected values of *g* in the range of 1 to 4 by steps of 0.5 and the numbers of factors from 1 to 20 are shown in Table 5.11. The column for coefficient of overdetermination 1.0 corresponds to values presented by Thurstone (1947). Thurstone's (1947) suggestion that "In order for a factor analysis to be stable and scientifically significant and convincing, the number of tests must be two or three times greater than those which are shown in the table" corresponds to use of a coefficient of overdetermination in the range of 1.5 to 2.0. In early studies in a domain the number of common factors may be subject to crude estimation and a larger coefficient of overdetermination should be used. In later studies in a domain the number of common factors may be better known and a smaller coefficient of overdetermination would be justified. In any case, using a number of attributes that is too small will result in poorly defined factors and unstable results.

Table 5.11
Minimum Number of Attributes for Given Number of Factors

	Coefficient of Overdermination						
Number of Factors	1.0	1.5	2.0	2.5	3.0	3.5	4.0
1	3	4	5	6	7	8	9
2	5	7	9	11	13	15	17
3	6	9	12	15	18	21	24
4	8	12	16	20	24	28	32
5	9	14	19	24	29	34	39
6	10	17	23	29	35	41	47
7	12	19	26	33	40	47	54
8	13	21	30	38	46	54	62
9	14	24	33	42	51	60	69
10	15	26	36	47	57	67	77
11	17	29	40	51	62	73	84
12	18	31	43	55	68	80	92
13	19	33	47	60	73	86	99
14	20	36	50	64	78	93	107
15	21	38	54	69	84	99	114
16	23	40	57	73	89	105	122
17	24	43	60	78	95	112	129
18	25	45	64	82	100	118	136
19	26	48	67	87	106	125	144
20	27	50	71	91	111	131	151

CHAPTER 6
DESIGNING AND CONDUCTING FACTOR ANALYTIC STUDIES

From
Exploratory Factor Analysis
Ledyard R Tucker
and
Robert C. MacCallum

CHAPTER 6
DESIGNING AND CONDUCTING FACTOR ANALYTIC STUDIES

6.0. Introduction

In previous chapters the common factor model has been presented in the context of both a population and a sample, and a number of issues have been discussed which can influence obtained factor solutions. These issues include such things as the effects of random sampling, selective sampling, sample size, the presence of unique factors, standardization of attributes, selection of attributes, and number of attributes. The discussion of these issues has been carried out in essentially a theoretical context. The purpose of the current chapter is to translate this discussion into principles and guidelines for practice. This should serve to aid researchers in designing and conducting factor analytic studies and in determining a general approach to take to the actual factor analysis of their data. Subsequent chapters will be based on an assumption that studies have been carried out and that data have been obtained, and that a general approach to the analysis has been determined.

6.1. General Approach to Factor Analytic Research

Recall that the general objective of factor analytic research, as stated in Chapter 1, is to determine the nature of the underlying factors and to develop an understanding of their relationships to the surface attributes and to each other. A very important fundamental point about factor analytic research is that it is virtually impossible to achieve this objective via a single study. That is, it is not feasible to attempt to carry out a single study in which a domain and population are defined, a set of attributes adequately representing the factors in the domain are constructed, measures on these attributes are obtained from a sample, and the results of a factor analysis are valid, stable, and clearly interpretable. In all likelihood, the results of a first attempt to achieve this will reveal a variety of problems; e.g., attributes which do not represent factors as intended, the occurrence of unanticipated factors, the absence of anticipated factors, etc. It is very unlikely that such phenomena can be avoided in an initial study in a domain, simply because of the lack of knowledge about the nature and dynamics of the underlying factors.

The implication of all of this is that the achievement of the general objective of factor analytic research will almost always require a series of studies. In the initial studies a domain is defined with little prior idea of the nature of the factors in the domain. Attributes are selected and/or constructed with an effort to achieve wide representation of the domain. There may be some prior notion of the existence of particular common factors, and attributes should be constructed so as to attempt to represent those factors. But the guiding principle should be to obtain wide coverage of the domain, so as not to miss any important common factors. Once data

are obtained, the factor analysis is conducted in an exploratory manner. Methodology for exploratory factor analysis is discussed in Chapters 7 through 13, and a more detailed discussion of this general approach is presented later in this chapter. For present purposes, the major point is that there is no testing of prior hypotheses; rather, the methodology is designed to explore the data to determine how many factors might be present, and to achieve some rough indication of, or conjecture as to, their nature and relationships to each other and to the surface attributes. Based on such results, the experimenter begins the task of refining the battery of attributes and designing subsequent studies to more closely investigate the factors in the domain under investigation. This process typically involves deleting attributes with undesirable properties (e.g., inadequate representation of factors), and adding new attributes so as to achieve a more complete understanding of the factors observed in the early studies. Distinctions between factors may be developed by addition of attributes which should "fan" between factors. Achievement of such distinctions should lead to better understanding of the nature of the factors. In these later studies, exploratory factor analysis methods would still be used because the objective is to continue to explore the data in search of information relevant to the impact of the modifications in the battery on the factor structure.

The succession of studies should be concern as aiding our understanding of the nature of the factors. In one sense, these studies involve changing samples of attributes included in the batteries used in the studies. This is a different question than the stability of results questions which would be approached by repeated studies with a constant battery of attributes having data collected for new samples of individuals.

The process of refinement of the battery, collection of additional data, and exploratory analysis may continue until the experimenter believe he or she has developed a clear and well-founded hypothesis about the number and nature of the factors in the domain under study, along with a battery of attributes which clearly reflect those factors. At this point, it would be most desirable to carry out a final study with the objective of testing the hypotheses about the factor structure in the domain. The study would employ the "final" version of the test battery, and the factor analysis methodology would be confirmatory rather than exploratory. This methodology is presented in Chapters 14 through 17, and a brief discussion of this approach is offered later in this chapter. For purposes of the present discussion, the major point is that confirmatory factor analysis provides for an explicit fitting and testing of the hypothesized factor structure, where the hypothesized structure is defined in terms of the number of factors and the hypothesized pattern of their relationships to each other and to the surface attributes. Results of the confirmatory analysis indicate the goodness of fit of the hypothesized structure to the data, and provide information to help evaluate what problems, if any, still exist. Clearly, the experimenter would hope that the hypotheses that were developed in the earlier studies would be strongly supported

by the confirmatory analysis in the final study in the series. Such an outcome would support a claim that the factor structure in the domain in question is well understood and well represented in the battery of attributes. This general description of a series of factor analytic studies should clarify the notion that the final objective probably cannot be achieved via a single study.

A brief description of an illustration of a series of factor analytic studies may help to further clarify the points made in the preceding paragraph. Suppose a researcher wishes to investigate the factorial structure in the domain of mental abilities. Such a project would begin with the assembly of a large battery of tests of a wide range of abilities. The experimenter probably would have some prior notions about basic internal attributes which might exist, such as mathematical, verbal, analytical, and spatial abilities, etc. The initial battery of attributes would include attributes that were thought to represent these factors, along with a variety of other attributes intended to represent the range of abilities in the domain. As stated above, the objective is to use a large and diverse initial battery so as not to overlook the existence of important common factors. Of course, the degree to which the prior notions are formed and supported will influence the size and diversity of the initial battery. However, even when prior hypotheses about the factors are present in early studies, they should be regarded as crude and subject to modification as research progresses. Regardless, the exploratory factor analysis in the initial study provides an initial indication of the number of common factors, along with estimates of the parameters of the model (factor weights, factor intercorrelations, and unique variances of the attributes). Based on this information, the researcher begins the process of refining the battery of attributes. Attributes which do not measure factors in a useful manner can be deleted. For instance, if there are a large number of attributes which measure the factor "numerical facility," some of those attributes can be deleted without sacrificing representation of this factor in the battery. More specifically, one may wish to delete those measures of the "numerical facility" factor which have the lowest communalities, as long as those attributes are not critical measures of other factors also. It may also be desirable to add attributes to the battery. For instance, if there is some indication of a factor which seems to represent "reading comprehension," additional attributes can be constructed and added to the battery which are intended to measure this factor directly. This would be an effort to more clearly determine this factor and to enhance its representation in the battery. Distinctions between related factors such as word fluency and verbal ability could be studied by the construction of several new tests which range between emphasizing word fluency and verbal ability. After such refinements are determined and carried out, a new sample of data could be collected and another exploratory factor analysis conducted. The results in this second study should be clearer, but may indicate the need for further modifications of the battery. Obviously, this process can continue until the researcher believes that a clear hypothesis about the factors in the domain of mental ability has been developed, and

that a battery of attributes has been defined which provides a clear representation of these factors. At that point, a final study could be conducted, with the data being subjected to confirmatory factor analysis. Positive results from this analysis would support the hypotheses about the factor structure in the domain of mental abilities, as well as the construct validity of the tests in the battery. Negative results would indicate problems of some type with the hypotheses and/or the data, necessitating further investigation. Specific techniques for evaluating these results will be discussed in the chapters on confirmatory factor analysis.

The critical point in this discussion is clearly that the achievement of the objective of factor analytic research requires a series of studies, proceeding from initial studies where hypotheses are only loosely formed and analyses are exploratory, to final studies where confirmatory analyses are conducted to test well-developed hypotheses. While the actual number of studies involved in a series will vary greatly from one domain to another, progress in understanding the factor structure in a domain is achieved in a step by step fashion through such a series of studies by one or more investigators.

Given this general view of factor analytic research, let us now turn our attention to the practical implications of a number of issues discussed in Chapter 5.

6.2. Selecting Observations from a Population

In Chapter 5 we discussed in a theoretical context some effects and issues involved in the process of sampling observations from a population. We will now consider these issues in a more practical context, and will follow the distinction employed in Chapter 5 between random and selective sampling.

6.2.1. Practical Implications Under Random Sampling

Let us begin by briefly reviewing how factor analytic results are affected under random sampling. As discussed in Chapters 4 and 5, two primary effects can be seen. First, random sampling affects the covariances among the factors; that is, the covariances among the factors in a sample will be affected by the random characteristics of that sample. This is represented in Eq. (4.13). Second, under random sampling the assumption that unique factors are uncorrelated with each other and with the common factors will generally be violated. As a result of the invalidity of this assumption, sampling error affects the common factor weights obtained in a sample. Additional sampling error effects arise as a result of the standardization of the common factors and of the attributes in the sample. These phenomena were discussed in Chapter 4 and demonstrated in Chapter 5.

An important point is that the magnitude of these effects of sampling error is dependent in part on several other characteristics of the data, as discussed in Chapter 5. First, the lower the unique factor weights, the less influence sampling error will have on the obtained common factor

weights. Second, standardization of attributes in a sample introduces additional sampling error affecting the common factor weights; i.e., the analysis of a correlation matrix rather than a covariance matrix adds additional error in the recovery of the population common factor weights. Third, the use of a large sample reduces the effects of sampling error described in the previous paragraph by improving the stability of the obtained solutions and the correspondence between obtained results and population parameters. Finally, it is very important to recognize that there is an interactive effect between sample size and the other two influence just mentioned-- the unique factor weights, and standardization of attributes. When sample size is large, the amount of sampling error arising from these influences is reduced.

These phenomena have direct implications for applied factor analytic research. In particular, they suggest some basic guidelines for practice. First, since the impact of sampling error increase as unique factor weights become larger, it is desirable to select attributes which will have small unique variances. When a succession of factor analytic studies is conducted, this is achieved by eliminating those attributes with low communalities and retaining those with high communalities. Alternatively, one might add attributes to a battery in order to more strongly represent certain common factors, thus increasing communalities of some of the original attributes. The general point is that one should seek to develop a battery containing attributes with high communalities. Not only will this help to provide clear representation of the common factors, but it will also reduce the impact of sampling error on the obtained results.

A second point is that it is desirable to apply factor analysis methods to covariance matrices rather than correlation matrices. This will eliminate the process of standardization of the attributes within the sample, and will thus avoid this type of sampling error. However, at the same time, it must be understood that standardization is often desirable in practice when measured attributes are characterized by radically different scales of measurement. Standardization of the attributes would then eliminate the influence of these highly disparate scales on the results of the factor analysis, thus rendering the solutions easier to interpret. In practice, then, the investigator must determine whether the scales of measurement of the attributes are sufficiently different to warrant standardization, thus implying a willingness to accept the additional source of sampling error introduced by this process.

Thirdly, it is most desirable to obtain a large sample. This will have two important effects. It will improve the stability of the results and the correspondence between the sample results and population parameters. It will also reduce the impact of sampling error arising from the presence of unique factors and the standardization of attributes. That is, a researcher can be less concerned about sampling error arising from those sources if the sample size is very large. This is an important point, especially with respect to the issue of unique factors. In early studies in a domain, it is likely that substantial unique factor weights may be present. Since results

obtained from early studies provide the basis for subsequent studies, it is important to do whatever possible to reduce the impact of sampling error. A large sample will thus be quite important in the early stages, both to enhance the stability of the obtained results and also to reduce the influence of sampling error arising from unique factors and possible standardization of attributes.

6.2.2. Practical Implications Under Selective Sampling

It is most important for researchers to understand that selective sample is the rule rather than the exception when samples are obtained in factor analytic research. As discussed in detail in Chapter 5, actions of the experimenter, of institutions, or of the subjects themselves routinely affect the nature of the obtained sample in such a way that it becomes more homogeneous with respect to certain variables. This occurs when subjects are included in the sample on the basis of their score on one or more variables, called selection variables. In some situations the experimenter may conduct selective sampling deliberately in order to obtain one or more samples which are homogeneous in certain ways; e.g., this would be the case when a researcher wished to compare children of different ages. In other situations, selective sampling may be intentional or unrecognized; e.g., when college students are employed as subjects in psychological research. Regardless of this issue, some degree of selection operates routinely in practice and researchers must be aware of its potential impact on obtained results, as well as of any procedures they can employ to reduce or control that impact.

As discussed in Chapter 5, the general effect of selective sampling is to increase the homogeneity of the sample and thus reduce variances and intercorrelations of attributes. This effect influences results of factor analysis applied to such samples. In Chapter 5 we developed a theoretical framework for linear selective sampling which postulated the dynamics of this phenomenon and provided a basis for studying the effects of selective sampling on obtained factor analysis solutions. The results of this approach showed that common factor weights and intercorrelations in a selected subpopulation will be affected by the selection process when the factors are standardized in that subpopulation. The usual effect will be for those weights and intercorrelations to be reduced due to the reduced variability in the subpopulation. Of greater interest were results involving the relations of solutions obtained from different selected subpopulations. These results showed that the common factor intercorrelations and weights will be different in the two groups. Furthermore, when covariance matrices are analyzed, the common factor weights in the two groups will be proportional by columns. However, when correlation matrices are analyzed, the rows of the weight matrices also will undergo a rescaling which results in an incomparability of the weights from the two groups. As a result of this effect, it is not advisable to analyze correlation matrices in such a setting. The entire issue of factor analysis in multiple groups will be considered in detail in Chapter 20.

The effects just described represent the case in which the selection variables are related only to the common factors, and not to the specific or error factors. When there are nonzero relations of the selection variables to the specific or error factors, considerable complexities arise with regard to the effects of selection on obtained factor solutions. This can occur fairly easily. For instance, whenever the selection variables themselves, or parallel tests to those variables, are included in the battery being factor analyzed, this type of relation will be present. In such a case, complex effects on the factor weights will occur, and additional common factors will arise and will be characterized by some negative weights. Given the complexity of such effects and the difficulty in recognizing and interpreting them in practice, we consider it highly advisable to avoid this situation whenever possible. That is, we recommend against including the selection variables themselves, or parallel tests to those variables, in the battery to be factor analyzed.

A final comment about the effects of selective sampling is that they will not occur in isolation in practice. Rather, they will occur in conjunction with other effects discussed in this and earlier chapters: the effects of random sampling and the effect of model error. These effects cannot be separated in practice. A major task for the researcher is to strive to design and conduct research in such a way that undesirable effects of these phenomena can be minimized, or at least recognized and understood.

6.3. Selecting Attributes from a Domain

In Chapter 4 we discussed and illustrated potential effects of attribute selection on the obtained factor solution. The major points made then were that (a) the deletion of attributes from a battery can cause common factors to vanish altogether or to become specific factors, and (b) the addition of attributes to a battery can cause specific factors to become common factors, or can give rise to new common factors altogether. It is important that applied researchers understand these phenomena because they are quite relevant to the process of constructing and modifying attribute batteries in a sequence of factor analytic studies.

As stated earlier, when an initial battery of attributes is constructed, attributes are selected so as to represent any hypothesized common factors, and also to provide wide coverage of the domain under study so as to enhance the opportunity to discover whatever important factors exist in the domain. In subsequent studies the initial battery is modified by deleting or adding attributes. This is done so as to enhance the representation of the common factors which are being identified. To achieve this goal, the researcher must understand and make use of the phenomena discussed above, regarding the effect of attribute selection on obtained factor solutions. To enhance the representation of a common factor, new attributes which are intended to measure that factor can be added to the battery, and attributes from the original battery which were designed to measure that factor but which failed to do so can be deleted. When an important

specific factor is suspected to exist (e.g., when an attribute has a very high reliability but a very low communality), new attributes can be added to the battery in an effort to bring out that factor as a common factor. In such ways, the researcher can carefully modify the battery so as to improve the representation of the common factors.

Another important issue regarding selection of attributes involves the number of attributes to include in a battery. In Chapter 4 this issue was discussed in the context of the concept of overdetermination of common factors. That is, it is necessary to have a sufficient number of attributes so as to achieve an adequate degree of mathematical overdetermination of the common factor weights. Table 4.10 provided information about the minimum number of attributes necessary to provide a given level of overdetermination for a given number of common factors. This information is very useful in the process of designing factor analytic studies. In early studies in a domain, the researcher may have only a crude estimate of the number of common factors present. In such a case, the number of attributes included in the battery should be sufficient to provide a high level of overdetermination of the common factors. As represented in Table 4.10, a coefficient of overdetermination of 3.0 or 3.5 may be desirable in such research. In later studies, when the number of common factors is known more accurately, a smaller coefficient of overdetermination would be sufficient; a value of 2.0 or 2.5 probably would be acceptable in most cases. In practical terms, the effect of including too few attributes in the battery to adequately overdetermine the factors would be to obtain unstable parameter estimates and poorly defined factors. Results would very likely be difficult to interpret, and these problems would have detrimental effects on the design and conduct of subsequent studies in the domain.

A final important issue in this area involves the interaction of sample size with the degree of overdetermination of the common factors. It has been shown (MacCallum, Widaman, and Lee, 1986) that a large sample is less necessary when common factors are highly overdetermined. An implication of this finding for practice is that when large samples are not easily obtained, an experimenter can enhance the quality of results by employing a battery of attributes which strongly overdetermines the common factors. That is, when the common factors are represented strongly by a number of attributes in the battery, results from relatively small samples may be quite stable and interpretable. Conversely, when the study is characterized by a low coefficient of overdetermination, the need for a large sample in order to obtain stable and interpretable results is greatly increased.

Given that the issue of sample size has been discussed in different contexts in this chapter, it is useful to briefly state a number of issues in combination. Combining the point made in the previous paragraph with points made in Section 6.2.1, it can be seen that a large sample is most important when unique factor weights are not small, attributes are standardized, and the coefficient of overdetermination is low. On the other hand, it is possible to obtain stable

and interpretable results with relatively small samples when unique factor weights are small, attributes are not standardized in the sample, and the coefficient of overdetermination is high. Under these conditions, the impact of sampling error arising from unique factors and standardization will be low, and the high coefficient of overdetermination will allow for stable representation of the common factors even in smaller samples.

6.4. Final Issues in Conducting Factor Analytic Studies

After a researcher has obtained data from a sample of observations on a battery of attributes, there are a few final issues which should be considered before factor analysis techniques are applied to the data. Three such issues will be discussed in the subsequent sections.

6.4.1. Standardization of Attributes and Factors

The issue of standardization of attributes and of factors has been discussed a number of times in various contexts. These developments have some important practical implications. Let us first consider the issue of standardization of attributes; i.e., the factor analysis of a correlation matrix rather than a covariance matrix. Though this is by far the most common approach to factor analysis in the applied literature, it is not necessarily the most desirable approach. It is an attractive approach in practice because it simplifies a number of aspects of the analysis. As data, correlations are simpler to interpret than covariances. In addition, factor weights are simpler to interpret and compare when attributes have been standardized. As noted earlier in this chapter, standardization eliminates influences of widely different raw scales of measurement of the attributes. However, as discussed and illustrated in Chapter 5, this standardization does introduce a source of at least a slight amount of additional error. Furthermore, potentially meaningful information about differences in variances of the attributes is lost. As a result, we encourage the analysis of covariance rather than correlation matrices when possible.

Most factor analysis is conducted under the imposed condition that the factors are standardized. Researchers must keep in mind that this is a separate and independent issue from standardization of attributes. Despite the fact that this technique introduces an additional sampling effect on the solution, it is generally acceptable. The sampling effects caused by this standardization are simple multiplicative effects on the weights for each common factor. Furthermore, standardization of the factors in the sample serves the dual purpose of simplifying interpretation of solutions (e.g., since they can be interpreted in terms of factor intercorrelations rather than intercovariances) as well as resolving the identification problem for the factors. This problem, discussed briefly in Chapter 4, involves the fact that it is necessary to establish a scale for the factors in order to estimate parameters involving those factors. If this problem is not resolved via standardization, then some other step must be taken to resolve it. While alternative

approaches will be discussed in some subsequent chapters, standardization of the factors is clearly the simplest and most common way to solve the identification problem.

A final point about standardization of attributes and factors involves the matter of selective sampling. It was shown in Chapter 5 that when solutions are obtained from different subpopulations, the standardization of attributes and factors within the subpopulations will give rise to systematic differences in common factor weights. This phenomenon argues further for the analysis of covariance matrices, since that would eliminate one of the sources of such differences. The issue of standardization of factors is more complex since, as noted above, if factors are not standardized some other approach must be taken to identify their scales. This issue will be considered for the case of subpopulations in Chapter 20.

6.4.2. Exploratory vs. Confirmatory Factor Analysis

Prior to conducting a factor analysis, the researcher must determine whether to use exploratory or confirmatory methods of analysis. The primary issue upon which this decision is based is the degree to which clear prior hypotheses are present regarding the factor structure underlying the battery of attributes. When such hypotheses are absent, or cannot be stated very explicitly, exploratory methods of factor analysis should be employed. This will generally be the case in early stages of research in a domain. Hypotheses will be at best loosely defined, and the general objective of the research will be to explore the factorial structure of the domain. The factor analysis methods employed in such studies involve estimation of all the parameters of the model (common factor weights and intercorrelations, and unique variances), and provide information to aid the researcher in determining the number of factors, interpreting the nature of those factors, and refining the battery of attributes for the purpose of further study of the domain. Exploratory factor analysis methods will be the subject of Chapters 7 through 13.

In later stages of research in a domain, the investigator is likely to have developed very specific hypotheses about the factorial structure of the battery of attributes. These hypotheses will concern the number of common factors, their pattern of intercorrelation, and the pattern of common factor weights. The presence of such explicit hypotheses can be taken into account via the use of confirmatory methods of factor analysis. Such methods allow the investigator to fit the common factor model to observed data under various types of constraints. For instance, the number of common factors would be defined to be a given number, and some of the parameters would be assigned fixed numerical values, rather than be estimated; e.g., certain common factor weights or intercorrelations could be fixed to zero. The remaining parameters of the model would then be estimated, and the goodness of fit of the solution to the data would be evaluated. The degree to which the solution fit the data would provide evidence for or against the prior hypotheses. A solution which fit well would lend support to the hypotheses, and would provide evidence for the construct validity of the attributes and the hypothesized factorial structure of the

domain as represented by the battery of attributes. A solution which fit poorly would indicate problems with the hypotheses and/or the data, and would call for further diagnosis and study. Methods of confirmatory factor analysis are presented in Chapters 14 through 17. A major issue to be resolved prior to conducting any factor analysis is to determine whether the study calls for exploratory or confirmatory analysis.

6.4.3. Lack of Fit of the Model to the Data

Regardless of what type of factor analytic methods are employed, researchers should not lose sight of the fact that all such methods involve fitting a model to data. As emphasized in numerous contexts previously in this book, the common factor model is a mathematical model developed from a theoretical view about the underlying structure of multiattribute data. Furthermore, it must be kept in mind that that model will virtually never represent the real world exactly. Thus, in factor analytic research, there will virtually always be some lack of fit of the model to the observed data. It is important that researchers understand the sources of this lack of fit, as well as how to attempt to reduce it.

As discussed in Chapter 4, there are two identifiable sources of lack of fit of model to the data. These are referred to as model error and sampling error. Model error can best be understood in the context of the population, where the common factor model simply will generally not precisely account for the variances and covariances of the surface attributes. The presence of model error is the reason for the distinction between surface attributes and modeled attributes used throughout this book. Sampling error is an additional source of error manifested by the fact that the random characteristics of a sample contribute additional error to the estimation of the model parameters. As discussed in detail in Chapter 4, this error arises from the fact that the simplifying assumption that unique factors will be uncorrelated with each other and with common factors will not hold exactly in a sample.

In empirical applications of factor analysis, the effects of model error and sampling error are combined and are not separable. Their presence is manifested by lack of perfect fit of model to the data, along with instability and error in the parameter estimates. Despite the fact that it is not possible to separate the effects of these two sources of error in practice, it is possible to take steps to attempt to reduce their impact. In particular, the impact of model error should be reduced as a battery of attributes is refined via successive studies in a domain. The gradual development of a battery of attributes in which a given set of common factors are strongly and clearly represented will have the effect of improving the correspondence between the theoretical model and the observed battery simply because this process eliminates problems causing a lack of such correspondence. The effect of sampling error, on the other hand, can be reduced by methods discussed earlier in this chapter; i.e., by employing a large sample, attempting to eliminate attributes with high unique variances, and analyzing covariance rather than correlation matrices.

A major concern of researchers in practice should be to make efforts to reduce the impact of both model error and sampling error. Regardless of the degree of success of these efforts, some lack of fit will almost always occur in fitting the model to real data, and the measurement and evaluation of this lack of fit will be an important part of any empirical factor analysis study.

CHAPTER 7
INTRODUCTION TO EXPLORATORY FACTOR ANALYSIS

From
Exploratory Factor Analysis
Ledyard R Tucker
and
Robert C. MacCallum

CHAPTER 7
INTRODUCTION TO EXPLORATORY FACTOR ANALYSIS

Factor analytic methodologies may be conceived on a continuum ranging from confirmatory techniques to pure exploratory procedures. Charles Spearman (1904 onward) was interested in confirming the idea of a general intelligence. With extended experimental evidence developed through years of studies involving larger test batteries given to larger samples of individuals, Spearman's theory of a single intellectual factor proved to be inadequate. Allowance had to be made for group factors. In the early 1930's, Thurstone broke with a common presumption based on prior assumptions as to the nature of factors and developed a general theory of multiple factor analysis. This is the theory presented in previous chapters. Thurstone's book "Vectors of Mind" (1935) presented the mathematical and logical basis for this theory. The only assumption made (and this is a question to be answered for each body of data) is that the interrelations among the measures of observed attributes may be represented by a smaller number of common factors. Needs for formal methods for checking the similarity of results from several studies have lead to development of matching procedures termed procrustes rotations. These procedures tend toward being confirmatory analyses. A return to a more general form of confirmatory factor analysis has occured with Jöreskog's developments in maximum likelihood factor analysis (1966, 1967, 1969). While Jöreskog's methods are derived in the statistics of confirming hypotheses, usage of alternate hypotheses brings this approach back toward exploratory methodology. Material in chapters 7 through 13 discusses methods from the exploratory side of the continuum.

A philosophic point presented in preceding paragraphs that almost all acceptable factor analytic models will not fit the real world needs to be emphasized for exploratory factor analytic studies. For experimenters and analysts to insist on a perfect fit in the population at all times is a route to chaos. For some factor methods a statistic has been developed to test whether analysis results fit the world in the population. Findings with large samples have been that quite reasonable analysis results have been rejected by these statistical tests. This point was first indicated by Tucker and Lewis (1973) who proposed a reliability coefficient for maximum likelihood factor analysis. Tucker and Lewis proposed this coefficient as a measure of goodness of fit. Many practitioners have come to realize the point that they should not expect acceptable models to fit perfectly in the population and have converted the chi square statistic to a measure of goodness of fit. A variety of other measures of goodness of fit have been proposed. One feature being investigated is the relation of these measures of goodness of fit to sample size, an ideal being that a measure of goodness of fit should not be affected by sample size on the average. An acceptable measure should estimate how well an analysis results fit in the population

and should not be biased by sample size. The chi square statistic is biased so as to favor small samples. To be sure, however, large enough batteries of attribute measures and samples of individuals should be used to yeild stable analysis results; but, there should not be an insistence that the factor analysis model fit precisely.

A related problem concerns relevant small effects which might be missed when results of an analysis are accepted with some extent of lack of fit. To be sure, we should avoid missing small factors which can be made large and important with special studies. All experimenters should be alert to the possibility that there are small, but important factors in their data. They should inspect residuals after extraction of a factor analytic model from a body of data for indications of possibly important factors that have been missed in the design of the study and which could be strengthend in future studies. This distinction between trivia and important small influences is a matter for experimenter insight and judgement. However, the complexities of factor analysis make it imperative that such a distinction be made. Inclusion of a number of very small factors in a study results in an unmanageably large dimensionality of the common factor space. Great care is required of an experimenter in making the decisions between possibly meaningful factors and trivia. There should be no doubt, however, that some trivia will exist so that an acceptable factor analytic model will not fit the data perfectly, nor would the model fit perfectly in the population. Experimenters should be prepared, on the other hand, for the possibility that no acceptable factor model can fit their data. That the factor analytic methodology does not fit all bodies of data is not a good reason to discard factor analytic methodology. Many factor analytic studies yield valuable results even though the fits to the data are imperfect.

7.1. Review of Basic Factor Analytic Formulas

Basic formulas presented in Chapters 3 and 4 are repeated here for convenience of reference. A minimum of description will be given here since more extensive discussions were given in the preceding chapters.

Relations among score vectors are given in equation (3.1).

$$\underline{y} = \underline{z} + \ddot{\underline{z}} \tag{3.1}$$

where $\underline{y}$ is the score vector on observed attributes for an individual, $\underline{z}$ is the modeled score vector for the individual from the factor analytic model, and $\ddot{\underline{z}}$ is the vector of discrepancies of fit of the factor model to the observed scores.

Covariance relations in the population are given by the following formulas.

$$\boldsymbol{\Sigma}_{yy} = \boldsymbol{\Sigma}_{zz} + \boldsymbol{\Sigma}_{z\ddot{z}} + \boldsymbol{\Sigma}_{\ddot{z}z} + \boldsymbol{\Sigma}_{\ddot{z}\ddot{z}} \tag{3.88}$$

$$\boldsymbol{\Delta_{\Sigma}} = \boldsymbol{\Sigma_{z\ddot{z}}} + \boldsymbol{\Sigma_{\ddot{z}z}} + \boldsymbol{\Sigma_{\ddot{z}\ddot{z}}} \tag{3.89}$$

$$\boldsymbol{\Sigma_{yy}} = \boldsymbol{\Sigma_{zz}} + \boldsymbol{\Delta_{\Sigma}} \tag{3.90}$$

$\boldsymbol{\Sigma_{yy}}$ is the covariance matrix among the observed attributes, $\boldsymbol{\Sigma_{zz}}$ is the covariance matrix among the modeled attributes, $\boldsymbol{\Delta_{\Sigma}}$ is the discrepancy of fit of the model covariance matrix to the covariance matrix among the observed attributes in the population. Properties of $\boldsymbol{\Delta_{\Sigma}}$ will be discussed in subsequent paragraphs.

The factor model for covariance relations in the population is given by the following formulas.

$$\boldsymbol{\Sigma_{zz}} = \boldsymbol{B\Phi B'} + \boldsymbol{U^2} \tag{3.35}$$

where $\boldsymbol{B}$ is the matrix of weights on the common factors, $\boldsymbol{\Phi}$ is the matrix of intercorrelations among the common factors, $\boldsymbol{U^2}$ is the diagonal matrix of unique variances. Score vectors on common factors and unique factors are designated by $\underline{\boldsymbol{x}}_{\boldsymbol{\beta}}$ and $\underline{\boldsymbol{x}}_{\boldsymbol{\mu}}$ with the modeled score vector being given by the following formula.

$$\underline{z} = \underline{\boldsymbol{x}}_{\boldsymbol{\beta}}\boldsymbol{B'} + \underline{\boldsymbol{x}}_{\boldsymbol{\mu}}\boldsymbol{U'} \tag{3.39}$$

A trait matrix, $\boldsymbol{T}$, is defined by the following formula.

$$\boldsymbol{\Phi} = \boldsymbol{TT'} \tag{3.42}$$

A matrix of factor weights, $\boldsymbol{A}$, for uncorrelated factors is defined by the following formula

$$\boldsymbol{A} = \boldsymbol{BT} \tag{3.44}$$

The factor model for uncorrelated common factors in the population is given below.

$$\boldsymbol{\Sigma_{zz}} = \boldsymbol{AA'} + \boldsymbol{U^2} \tag{3.45}$$

Covariance relations in a sample are given by the following formulas.

$$\boldsymbol{C_{yy}} = \boldsymbol{C_{zz}} + \boldsymbol{C_{z\ddot{z}}} + \boldsymbol{C_{\ddot{z}z}} + \boldsymbol{C_{\ddot{z}\ddot{z}}} \tag{4.22}$$

$$\boldsymbol{\Delta_c} = \boldsymbol{C_{z\ddot{z}}} + \boldsymbol{C_{\ddot{z}z}} + \boldsymbol{C_{\ddot{z}\ddot{z}}} \tag{4.23}$$

$$\boldsymbol{C_{yy}} = \boldsymbol{C_{zz}} + \boldsymbol{\Delta_c} \tag{4.24}$$

$\boldsymbol{C_{yy}}$ is the covariance matrix among the observed attributes, $\boldsymbol{C_{zz}}$ is the covariance matrix among the modeled attributes, $\boldsymbol{\Delta_c}$ is the discrepancy of fit of the model covariance matrix to the covariance matrix among the observed attributes in the sample. Properties of $\boldsymbol{\Delta_c}$ will be discussed in subsequent paragraphs.

In a sample, the factor model fitted to an observed covariance matrix, $\boldsymbol{C_{yy}}$, is given by equation (4.29). Note that the estimated factor weight matrix, $\boldsymbol{B}$, is for common factor scores standardized in the sample. Matrix $\boldsymbol{R_{bb}}$ (note change in subscripts from $\beta\beta$ to bb) contains estimated intercorrelations among common factors and matrix $\boldsymbol{U^2}$ contains estimated unique variances.

$$\boldsymbol{C_{zz}^{+} = BR_{bb}B' + U^2} \tag{4.29}$$

$$\boldsymbol{C_{yy} = C_{zz}^{+} + \Delta_{y}^{+}} \tag{4.30}$$

The following formulas for operations in a sample parallel formulas given for transformations to uncorrelated factors in the population.

$$\boldsymbol{R_{bb} = TT'} \qquad (7.1, \text{parallel to } 3.42)$$

$$\boldsymbol{A = BT} \qquad (7.2, \text{parallel to } 3.44)$$

$$\boldsymbol{C_{zz}^{+} = AA' + U^2} \qquad (7.3, \text{parallel to } 3.45)$$

Matrix $\boldsymbol{T}$ is the trait matrix with trait vectors as rows. Matrix $\boldsymbol{A}$ contains the factor loadings on uncorrelated factors.

A geometric representation in the common factor space was presented in Chapter 3. Parallel equations for a sample are given here. In addition to the trait matrix $\boldsymbol{T}$ and the uncorrelated factor matrix $\boldsymbol{A}$, four new matrices are defined. Matrix $\boldsymbol{Q}$ contains covariances of modeled attributes with traits; matrix $\boldsymbol{F}$ contains, as rows, normals to hyperplanes; matrix $\boldsymbol{G}$ contains projections of the modeled attributes on the normals, these being part correlations of the modeled attributes with the factors other factors being partialled out; and matrix $\boldsymbol{D}$ contains cosines of angles between the trait vectors and the normals to the hyperplanes.

$$\boldsymbol{B = AT^{-1}} \qquad (7.4, \text{parallel to } 3.48)$$

$$\boldsymbol{Q = AT'} \qquad (7.5, \text{parallel to } 3.55)$$

$$\boldsymbol{TF' = D} \qquad (7.6, \text{parallel to } 3.56)$$

$$\boldsymbol{F' = T^{-1}D} \qquad (7.7, \text{parallel to } 3.58)$$

$$\boldsymbol{D = (Diag(R_{bb}^{-1}))^{-\frac{1}{2}}} \qquad (7.8, \text{ parallel to } 3.60)$$

$$\boldsymbol{G = AF'} \qquad (7.9, \text{parallel to } 3.61)$$

$$\boldsymbol{G} = \boldsymbol{BD} \qquad (7.10, \text{parallel to } 3.63)$$

7.2. Comments on Fitting a Factor Model to a Covariance Matrix

Almost always the distinctive portion of a factor analytic study starts with a covariance matrix. It is important, however, to remember that a factor analytic study starts with the design of the study, continues with selection and construction of attribute measures, arrangements for a sample of individuals, administration of the attribute measures to the individuals in the sample, scoring of the attribute measures and computation of the covariance matrix. These matters have been discussed in preceding chapters. Note that the analytic portion distinctive to factor analysis does not start with the attribute measures vectors, $\boldsymbol{y}$, for individuals in the sample. The reason for this is a matter related to factor score indeterminacy, a topic to be discussed in score indeterminacy, a topic to be discussed in a later chapter. Only the nature of the source of the problems in this area will be indicated here.

As given in equation (3.39), modeled score vector, $\underline{z}$, can be expressed in terms of a contribution from the common factors and a contribution from the unique factors. Let $\underline{\boldsymbol{z}}_{\beta}$ and $\underline{\boldsymbol{z}}_{\mu}$ symbolize these two contribution vectors:

$$\underline{\boldsymbol{z}}_{\beta} = \underline{\boldsymbol{x}}_{\beta}\boldsymbol{B}' \qquad (7.11)$$

$$\underline{\boldsymbol{z}}_{\mu} = \underline{\boldsymbol{x}}_{\mu}\boldsymbol{U}' \qquad (7.12)$$

Then, from equation (3.39), the modeled score vector can be expressed as:

$$\underline{z} = \underline{\boldsymbol{z}}_{\beta} + \underline{\boldsymbol{z}}_{\mu} \qquad (7.13)$$

or

$$\underline{\boldsymbol{z}}_{\beta} = \underline{z} - \underline{\boldsymbol{z}}_{\mu} \qquad (7.14)$$

Combining these equations with equation (3.1) yields:

$$\underline{\boldsymbol{z}}_{\beta} = \boldsymbol{y} - \underline{\boldsymbol{z}}_{\mu} - \ddot{\underline{\boldsymbol{z}}} \qquad (7.15)$$

The crux of the problem is that the score vectors $\underline{\boldsymbol{z}}_{\mu}$ and $\ddot{\underline{\boldsymbol{z}}}$ are not defined and until they are defined, $\underline{\boldsymbol{z}}_{\beta}$ can not be determined. If $\underline{\boldsymbol{z}}_{\beta}$ were determined, the solution in the common factor space would result from a principal components analysis as per equation (7.11). However, vector $\boldsymbol{y}$ is in an *n* dimensional space (*n* being the number of measured attributes). The combined vector [$\underline{\boldsymbol{x}}_{\beta}$, $\underline{\boldsymbol{x}}_{\mu}$] is in an *r* + *n* dimensional space (*r* being the number of common factors). Vector $\underline{z}$ is in an *n* dimensional subspace of the combined vector [$\underline{\boldsymbol{x}}_{\beta}$, $\underline{\boldsymbol{x}}_{\mu}$]. Given a solution in the population to equations (3.59) and (3.90) with a satisfactory discrepancy matrix $\boldsymbol{\Delta}_{\Sigma}$,

knowledge of $\boldsymbol{B}$, $\boldsymbol{\Phi}$, $\boldsymbol{U^2}$ and the modeled score vector $\underline{z}$ yields not one but many solutions for score vectors $\underline{\boldsymbol{x}}_{\beta}$ and $\underline{\boldsymbol{x}}_{\mu}$ as shown by Guttman (1955). This is the factor score indeterminacy problem.

A further problem is introduced by a lack of knowledge of the discrepancies vectors $\ddot{\underline{z}}$. In fact, the nature of these vectors depends on the method of fitting the model to the observations of measured attributes. Each method for fitting a model to observations depends on a measure of goodness of fit, or of lack of fit; and, when different measures of lack of fit are tried, different solutions are obtained for parameters of the model. Consider a very simple example of this general principle: we have a collection of observations of the value of a variable and wish a single, representative value. The arithmetic mean minimizes the sum of squared deviations, this being the measure of lack of fit. In contrast, the median minimizes the sum of absolute values of deviations, this being the alternate measure of lack of fit. It is well known that the arithmetic mean and the median are not necessarily equal. One might consider any of a number of other possible lack of fit measures and arrive at a number of other representative values for our collection of observations. Fitting a factor model to observations both at the score level and at the covariance level possesses this same general problem, but in greater complexity. Most lack of fit measures, or loss functions used in factor analysis are stated in terms of covariance or correlation matrices. Some are more explicit than others, for example, the loss function for matrix factoring is more obscure than the loss function for Jöreskog's and Goldberger's generalized least squares factor analysis (1972). However, the application of these loss functions at the score level are not well known. Consequently, the score discrepancy vectors $\ddot{\underline{z}}$ are not well defined.

As a consequence of the indeterminacies of score vectors $\underline{z}_{\beta}$, $\underline{z}_{\mu}$ and $\ddot{\underline{z}}$, factor analysis theory has been extended analytically to derived statements for covariance and correlation matrices. While factor scores can not be determined uniquely, analysis of the covariance and correlation structures will yield important information about the structure of behavior. Estimates of the factor scores are available for use in particular situations; the estimation procedures and problems will be discussed in a subsequent chapter. Considerations of covariance matrices in the population led to equations (3.35) and (3.90) and in a sample to equations (4.29) and (4.30). Further considerations of the transformation of factors problem led to alternate equations for uncorrelated factors, equation (3.45) in the population and equation (7.3) in a sample. Substitution from equation (7.3) into (4.30) yields:

$$\boldsymbol{C_{yy}} = \boldsymbol{AA'} + \boldsymbol{U^2} + \boldsymbol{\Delta_y^+} \tag{7.16}$$

This is the formula most frequently used in practice in fitting a factor analytic model to observations. Matrix $\boldsymbol{\Delta_y^+}$ often is termed the matrix of residual covariances (or residual correlations in case $\boldsymbol{C_{yy}}$ is a correlation matrix) or, for short, the matrix of residuals. A number

of the methods for fitting a model to an observed covariance matrix involve a measure of lack of fit, or loss function derived from matrix Δ_y^+. Further discussion of problems in measures of lack of fit, or loss functions will be presented in chapter 9.

A serious problem not dealt with previously in this chapter is the number of common factors to use in the fitted model. This is the number of columns in matrix $\boldsymbol{A}$ and the dimensionality of the common factor space. Several suggestions have been made as a result of serious study and will be described in subsequent chapters. Further, several suggestions have been made for supplementary information relative to this problem. Frequently, for any given method of fitting the factor model, solutions are obtained for several successive numbers of factors and the resulting measures of goodness of fit are compared. A number of measures of lack of fit (for which smaller values are better) reduce as the number of factors is increased. The question becomes whether the reduction in lack of fit from one number of factors to a larger number is warranted considering the additional number of entries in matrix $\boldsymbol{A}$. Detailed consideration of the number of factors will be given in following chapters; it is sufficient, here, to recognize this problem and to keep it in mind during following discussions.

We turn now to several older factor extraction methods which were applied to correlation matrices. Equation (7.16) may be rewritten in terms of the correlation matrix, $\boldsymbol{R_{yy}}$, among measured attributes.

$$\boldsymbol{R_{yy} = AA' + U^2 + \Delta_R} \tag{7.17}$$

where Δ_R is the matrix of discrepancies in fitting the model for the correlation matrix. Equation (7.17) may be rearranged to:

$$\boldsymbol{(R_{yy} - U^2) = AA' + \Delta_R} \tag{7.18}$$

In this form, factor extraction involved establishing some guess at the unique variances in $\boldsymbol{U^2}$ so as to establish the matrix of correlations with communalities in the diagonal, this matrix being $\boldsymbol{(R_{yy} - U^2)}$. Having this matrix, the procedures turned to factor extraction. We, knowingly, use the terms guessed uniqueness and guessed communalities since precise values do not exist in the population as a result of the idea that they are not defined until a method of fitting the factor analytic model has been selected. Two methods for obtaining guessed communalities will be presented in the next chapter. Also to be presented is Guttman's general theory for matrix factoring (1944) which gives a mathematical basis of factor extraction. Mehtods of factor extraction to be considered include the centroid method developed by L. L. Thurstone (1935, 1947), the group centroid method developed in several forms by Thurstone (1945) and discussed by Guttman (1952). An initial discussion will be presented in chapter 8 of principal factors extraction with a more extended discussion given in chapter 9 which includes determination of

cormnunalities by a least squares fitting procedure. Principal factor extraction from a correlation matrix having guessed communalities is a commonly used and highly recommended method for exploratory factor analysis.

Detailed discussion of factor extraction by matrix factoring techniques appear in chapter 8. Chapter 9 will cover factor extraction by methods using loss functions. Most of these techniques result in a factor matrix for uncorrelated factors which require transformation.

7.3. Properties of Transformations of Factors

A most important problem for exploratory factor analysis follows establishment of a common factor matrix $\boldsymbol{A}$ with uncorrelated factors. As discussed in Chapter 3 in terms of a geometric representation there is almost complete freedom in establishing trait vectors in matrix $\boldsymbol{T}$ and corresponding loadings on the derived correlated factors. Basic equations for operations in a sample are given in equations (7.4) through (7.10). Only two restrictions exist for matrix $\boldsymbol{T}$ in maintaining the representation of the input covariance or correlation matrix.
First: matrix $\boldsymbol{T}$ must be square and nonsingular so that an inverse exists:

$$\left| \boldsymbol{T} \right| \neq \boldsymbol{0} \tag{7.19}$$

Second: each trait vector (row of matrix $\boldsymbol{T}$) must be of unit length so that:

$$Diag(\boldsymbol{T}\boldsymbol{T}') = Diag(R_{bb}) = \boldsymbol{I} \tag{7.20}$$

Many matrices $\boldsymbol{T}$ may be written which satisfy the preceding conditions. Consider matrix $\boldsymbol{C_c}$ to be the portion of $\boldsymbol{C_{zz}^{+}}$ from the common factors. $\boldsymbol{C_c}$ is the covariance matrix among the score vectors $\underline{\boldsymbol{z}}_{\beta}$. Then, from equation (4.29) and (7.3):

$$\boldsymbol{C_c} = \boldsymbol{B}\boldsymbol{R_{bb}}\boldsymbol{B}' = \boldsymbol{A}\boldsymbol{A}' \tag{7.21}$$

and

$$\boldsymbol{C_{zz}^{+}} = \boldsymbol{C_c} + \boldsymbol{U^2} \tag{7.22}$$

There is an urgent need for some additional principle for use in developing a useful transformation of factors.

Thurstone (1935, 1947) proposed the principle of simple structure to resolve the problem of definition of the transformation of factors for many areas of study. His reasoning was that, for a diverse battery of attributes, each factor should not influence the scores on all attributes. This line of reasoning leads to a conclusion that there should be a number of zero or trivial loadings on each transformed factor. Transformation of factors would be a search for a location of trait vectors so as to have a number of zero or trivial loadings. This principle is widely accepted and

has led to inclusion in most computer programs of analytic procedures which tend toward a simple structure solution. Maybe, this acceptance has been overly wide spread and has led in many cases to uncritical acceptance of computer output. Every analyst has a responsibility to inspect results from a study to judge whether or not the principle of simple structure applies to the data at hand. Other principles may exist for other classes of cases than that for which simple structure was suggested. Maybe, of course, a given body of data is inadequate to define a simple structure. A quite serious possibility is that the computerized solution does not yield satisfactory results even when an acceptable simple structure exists for a given factor matrix $\boldsymbol{A}$. However, simple structure appears to apply to many bodies of data so as to yield useful conjectures as to the nature of constructs and dynamics underlying the observed covariations among the surface attributes.

The word structure is used in terms of the geometric representation of the common factor space. It includes the configuration of attribute vectors, a collection of trait vectors and hyperplanes along with the normals to these hyperplanes. In the sense used here, a structure is analogous to a building which includes all of the material in place in the building including the structural steel. A structure includes all of the parts and their relations to each other in an object. In the case of the common factor space, the parts are elements of the geometric representation.

For an example of a common factor vector configuration consider the covariance matrix $\boldsymbol{C_c}$ given in Table 7.1. Diagonal entries are the squares of the lengths of the vectors, the vector lengths being labeled h_j (square roots of the diagonal entries). Off diagonal entries are scalar products between pairs of vectors. Let θ_{jk} designate the angle between vectors *j* and *k* ; then:

$$C_{cjk} = \mathrm{h}_j \mathrm{h}_k cos\theta_{jk} \tag{7.23}$$

Conversion of the covariance matrix $\boldsymbol{C_c}$ in Table 7.1 to a correlation matrix R_c of Table 7.2 involves division of each covariance by the product of the square roots of the corresponding diagonal entries. Thus:

$$r_{cjk} = C_{cjk}/(\mathrm{h}_j \mathrm{h}_k) \tag{7.24}$$

which, with equation (7.23), yields:

$$r_{cjk} = cos\theta_{jk} \tag{7.25}$$

Given the correlations in Table 7.2, the angles between the vectors may be obtained from a table of trigonometric functions, the angles for the artificial example being given in Table 7.3. Two views of a vector configuration for our example are given in Figure 7.1.

A configuration of attribute vectors may be constructed from a covariance matrix by computing the vector lengths and angles between pairs of vectors as described in the preceding

Table 7.1

Covariance Matrix from Common Factors among Attributes

	1	2	3	4	h_j
1	.25				.5000
2	.03	.34			.5831
3	.22	.27	.37		.6083
4	.07	.31	.28	.29	.5385

Table 7.2

Correlation Matrix from Common Factors among Attributes

	1	2	3	4
1	1.00			
2	.10	1.00		
3	.72	.76	1.00	
4	.26	.99	.85	1.00

Table 7.3

Angles (in Degree) between Attribute Vector

	1	2	3	4
1	--			
2	84.09	--		
3	43.67	40.43	--	
4	74.93	9.16	31.26	--

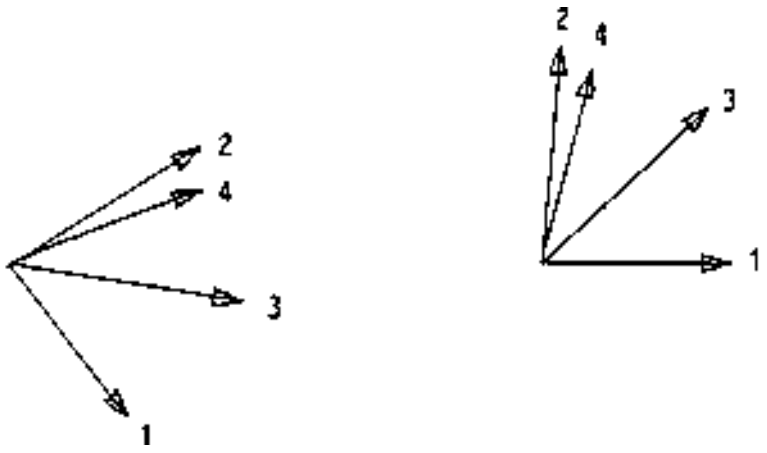

Figure 7.1: Two Views of a Geometric Vector Configuration
Representing Common Factor Covariance Matrix

paragraph and following the procedure to be described. This procedure does not depend upon obtaining coordinates of the terminals of the vectors on coordinate axes. Lay out vector 1 in any direction with length h_1. Lay out vector 2 at an angle θ_{12} with vector 1 and of length h_2. Best results are obtained by choosing vectors 1 and 2 to have an angle as near to 90° as possible. To achieve this result the attributes may have to be reordered. In the example no reordering was necessary. A third vector is laid out in a direction with angles θ_{13} and θ_{23} from the first two vectors and with length h_3. Some times a third dimension will be required; however, in no case will it be impossible to lay out this third vector. In some cases, as in the example, the third vector will lie in the plane defined by the first two vectors. Vector 4 is laid out in a similar fashion in terms of angles θ_{14}, θ_{24} and θ_{34} with length h_4. Additional attribute vectors could be added to the configuration in terms of the vectors' angles with preceding vectors and with the given lengths. With the addition of each successive vector a new dimension may have to be utilized; however, as remarked earlier, in no case will it be impossible to add a new attribute vector. This condition is normal for a covariance matrix so that when it is not possible to add a new vector there is a blunder in the covariance matrix. The number of dimensions required equals the rank of the covariance matrix. In our example the rank of the covariance matrix, $\boldsymbol{C_c}$ is 2 which equals the number of dimensions of the vector configuration.

Figure 7.2 presents a supplementary view of the vector configuration. All attribute vectors have been lengthened to unit length. Then, the relations among the vectors are determined by the correlation matrix, $\boldsymbol{R_c}$, by the angles between the pairs of vectors. Once the unit length vectors have been established, the original length vectors are established with the original vector lengths.

A most important point is that the vector configuration is determined solely by the covariance matrix $\boldsymbol{C_c}$. No axes have been defined. Given the covariance matrix $\boldsymbol{C_c}$, different methods of factor extraction will insert axes and yield matrices of coordinates on the axes. The locations of these axes will be different for different methods of factor extraction. Two pairs of orthogonal axes have been inserted in Figure 7.3. The first pair is labeled I and II while the second pair is labeled P and Q. Coordinates of the terminals of the attribute vectors on these pairs of axes are given in Table 7.4 as factor matrices $\boldsymbol{A_1}$ and $\boldsymbol{A_2}$. These factor matrices satisfy the condition that:

$$\boldsymbol{C_c = A_1 A_1' = A_2 A_2'} \tag{7.26}$$

Axis I was passed through the centroid of the attribute vectors, the centroid vector equaling the sum of the attribute vectors divided by the number of attributes. Axis II was set orthogonal to axis I. This is the result of the centroid method of factor extraction. For the second pair of axes, axis P was passed through the vector for attribute 1 and axis Q was set orthogonal. This is the result of the diagonal method of factor analysis which is related to the Cholesky decomposition

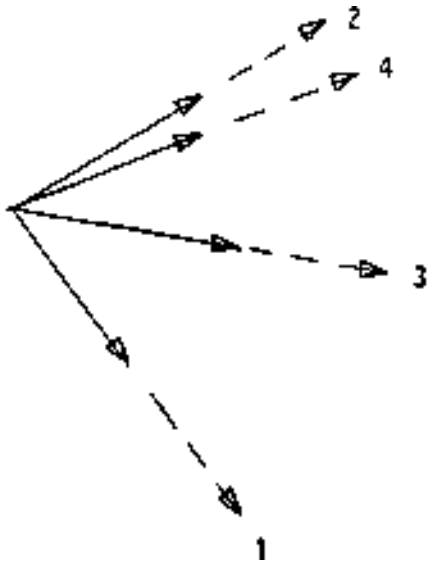

Figure 7.2: Construction of a Geometric Vector Configuration from Common Factor Covariance and Correlation Matrices

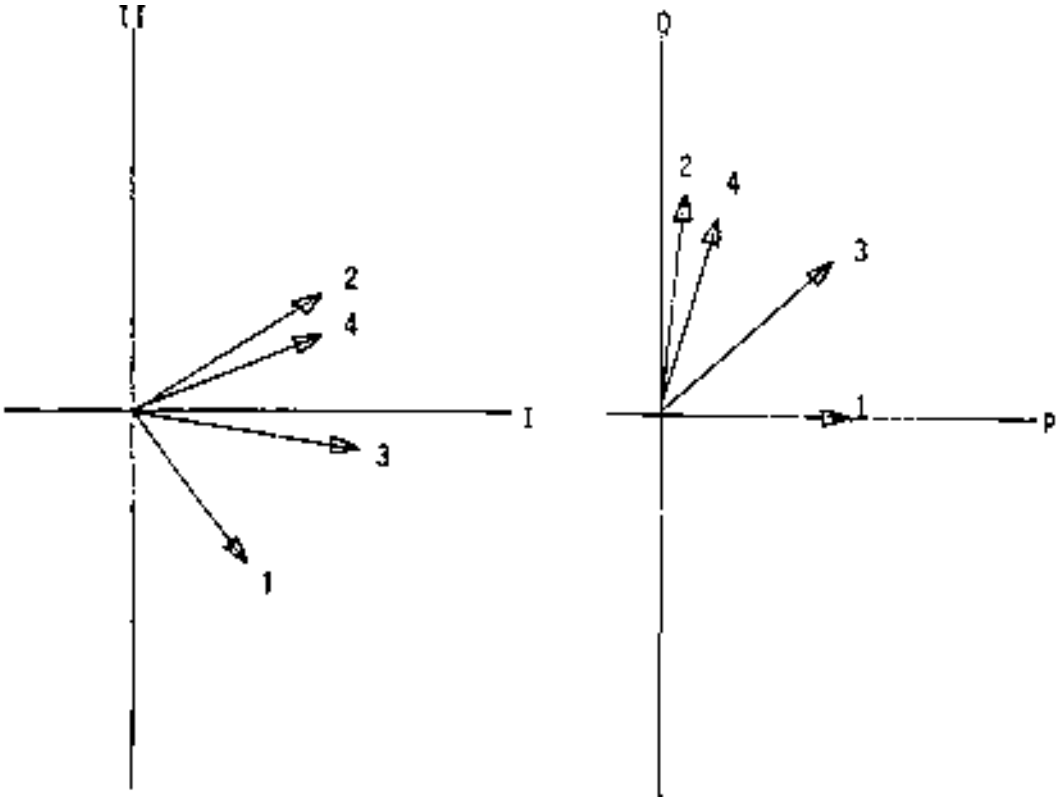

Figure 7.3: Geometric Vector Configuration with Two Sets of Coordinate Axes

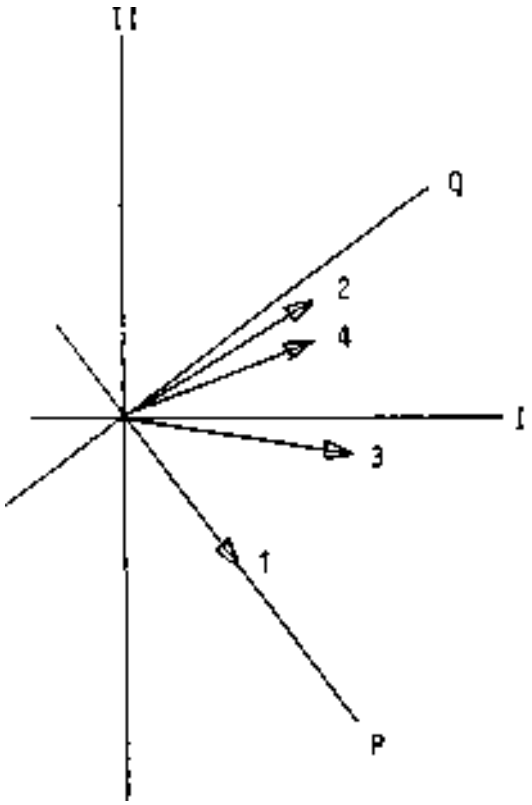

Figure 7.4: Transformation from One Set of Coordinate Axes to a Second Set

Table 7.4

Two Uncorrelated Factor Matrices from C_c

	A_1			A_2	
	I	II		P	Q
1	.30	-.40	1	.50	.00
2	.50	.30	2	.06	.58
3	.60	-.10	3	.44	.42
4	.50	.20	4	.14	.52

Table 7.5

Transformation Matrix from Factor Matrix A_1 to Factor Matrix A_2

	I	II
P	.60	-.80
Q	.80	.60

Table 7.6

Normalized Attribute Vectors from factor Matrix A_1

	A_1			Normalized A_1		
	I	II	h_j		P	Q
1	.30	-.40	.5000	1	.60	-.80
2	.50	.30	.5831	2	.86	.51
3	.60	-.10	.6083	3	.99	-.16
4	.50	.20	.5385	4	.93	.37

of a symmetric matrix. Both the centroid method and the diagonal method of factor extraction will be described in the next chapter.

Figure 7.4 shows the attribute vector configuration with both pairs of axes and Table 7.5 presents the relation between the two pairs of axes. Let $\boldsymbol{M_{12}}$ be a square, orthonormal matrix. Then:

$$\boldsymbol{M_{12}} = \boldsymbol{M_{12}^{-1}} \tag{7.27}$$

so that

$$\boldsymbol{M_{12}M_{12}'} = \boldsymbol{M_{12}M_{12}} = \boldsymbol{I} \tag{7.28}$$

Let $\boldsymbol{M_{12}}$ be determined so that:

$$\boldsymbol{A_1M_{12}'} = \boldsymbol{A_2} \tag{7.29}$$

This is always possible whenever $\boldsymbol{A_1}$ and $\boldsymbol{A_2}$ have the same number of columns which equals the rank of $\boldsymbol{C_c}$. The reverse transformation is:

$$\boldsymbol{A_2M_{12}'} = \boldsymbol{A_1} \tag{7.30}$$

Table 7.5 gives the matrix $\boldsymbol{M_{12}}$ for our example. Row P describes axis P in terms of axes I and II while row Q describes axis Q in terms of axes I and II . Entries in $\boldsymbol{M_{12}}$ are the cosines of the angles of axes P and Q with axes I and II ; they are termed direction cosines. Each row is a unit length vector in the direction of the transformed axes, P and Q .

Unit length attribute vectors may be obtained by "normalizing" the vectors in an orthogonal factor matrix such as $\boldsymbol{A_1}$. The term "normalize" refers to adjusting the length of a vector to a unit length by dividing each coordinate of the vector by the length of the vector. Consider Table 7.6. Matrix $\boldsymbol{A_1}$ is given at the left for our artificial example. Column h_j contains the lengths of the attribute vectors. Since matrix $\boldsymbol{A_1}$ is an orthogonal factor matrix, being for uncorrelated factors, the squares of the vector lengths are the sums of the squares of the coordinates, or loadings, with the vector lengths being the square roots of these squares of vector lengths. The matrix "NORMALIZED $\boldsymbol{A_1}$" is given on the right of Table 7.6 with each row being the row of $\boldsymbol{A_1}$ divided by the h_j for that row. These normalized attribute vectors are plotted in Figure 7.5 with axes I and II . Since these normalized vectors are of unit length their termini are on a unit radius circle which has been drawn in Figure 7.5. This property of normalized vectors will be generalized for three dimensions to the termini laying on the surface of a unit radius sphere. For cases with more than three dimensions, the termini will lie on the surface of a unit hypersphere.

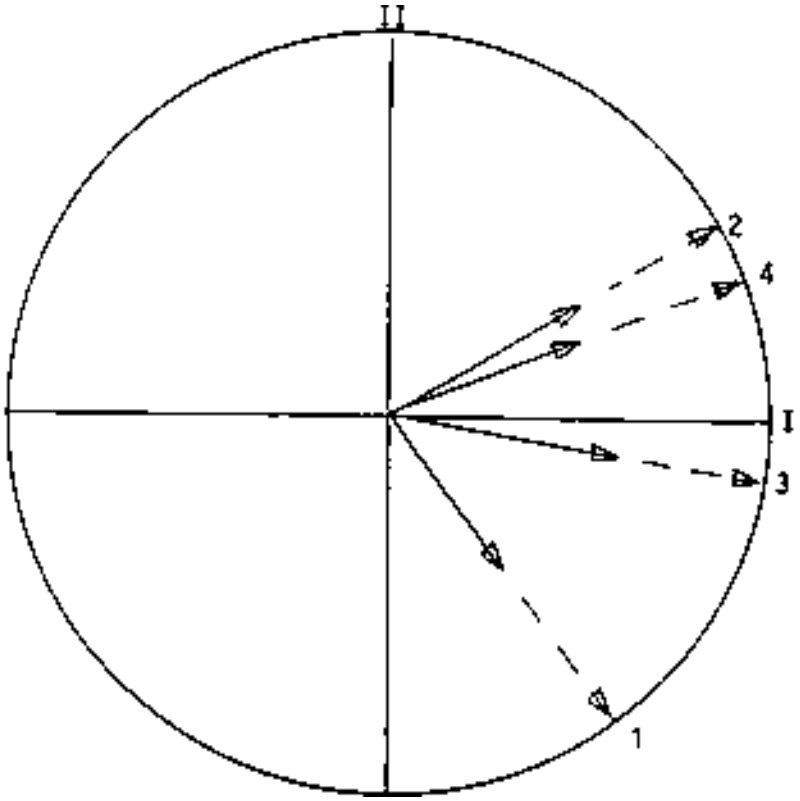

Figure 7.5: Normalized Attribute Vectors with a Set of Coordinate Axes

A drawing of a group of attribute vectors in three dimensions is given in Figure 7.6. This is a picture of a cardboard model on a stand with the vectors being drawn on three pieces of cardboard which form planes. These vectors are normalized, being of unit length, so that their termini would lie on the surface of a unit sphere. Figure 7.7 shows on the left three views of this model tipped forward by successive steps. The top view shows the outside of the front plane while the middle view is in line with this front plane and the bottom view shows the inside of the front plane. In the bottom view we are looking down into the model. On the right of Figure 7.7 are three corresponding drawings of a unit sphere with the termini of the attribute vectors indicated by closed circles. The planes of the model are shown also. Use of unit spheres is a convenient method for showing configurations of attribute vectors. Remember that the vectors emanate from the origin which is at the center of the sphere.

We return to the Nine Mental Tests example used in Chapter 1. Table 7.7 gives the factor matrix $\boldsymbol{A}$ for uncorrelated factors, the vector lengths in column h_j and the Normalized $\boldsymbol{A}$ computed by the procedure described in preceding paragraphs. Figure 7.8 shows the termini of the attribute vectors on the surface of a unit radius sphere. These termini are indicated by closed circles. A transformation of factors was computed by a method named DAPPFR which will be described in Chapter 11. Table 7.8 gives the resulting trait vectors and normal vectors. The trait vectors are rows of matrix $\boldsymbol{T}$ while the normal vectors are rows of matrix $\boldsymbol{F}$. Remember that the trait vectors and normal vectors are of unit length so that their termini would lie on the surface of the unit sphere. Figure 7.8 shows the trait vectors as open circles at the intersection of planes. The normal vectors are shown by crosses. The planes pass near to attribute vectors, this being a property of a successful transformation of factors to simple structure. Each normal vector is orthogonal to one of the planes and, thus, is orthogonal to every vector in the plane. Normal vector $\underline{\mathbf{f}}_1$ is orthogonal to the plane that passes through trait vectors $\underline{\mathbf{t}}_2$ and $\underline{\mathbf{t}}_3$. Consequently, normal vector $\underline{\mathbf{f}}_1$ is orthogonal to trait vectors $\underline{\mathbf{t}}_2$ and $\underline{\mathbf{t}}_2$. Note that normal vector $\underline{\mathbf{f}}_1$ is not orthogonal to trait vector $\underline{\mathbf{t}}_1$. In a similar fashion, each normal vector is orthogonal to all trait vectors except the corresponding trait vector. A result is that each trait vector is orthogonal to all normal vectors except the corresponding normal vector. The planes may be termed "base planes", or, in higher dimensions, "base hyperplanes". Each base plane may be considered as defined by the normal vector to that plane.

Three types of meaningful coefficients were described in Chapter 3 in terms of the geometric representation of the common factor model. These are structure coefficients relating the modeled attribute vectors to the trait vectors and normals. While they are computed from the common factor matrix $\boldsymbol{A}$ which contains vector coordinates on a set of orthogonal coordinate axes, these coefficients are properties of the common factor structure and are independent of the coordinate axes. They are not altered by an orthogonal rotation of the coordinate axes. Consider

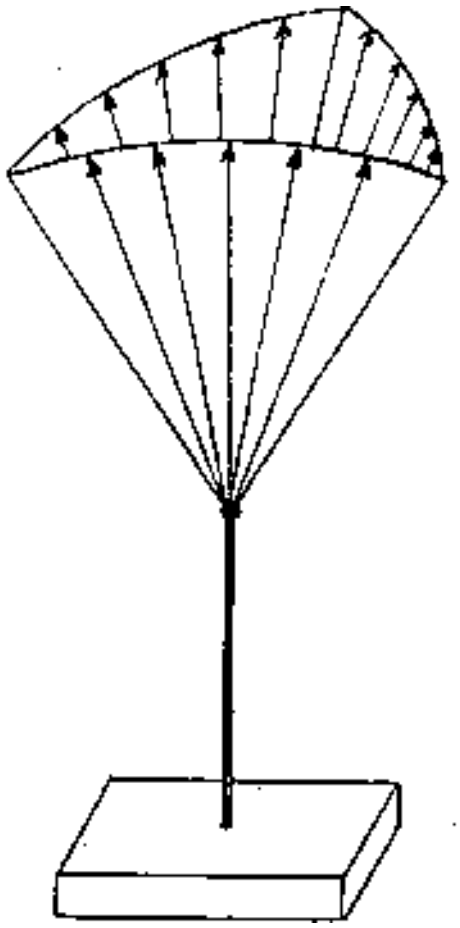

Figure 7.6: A Geometric Model of Normalized Vectors in Three Dimensions

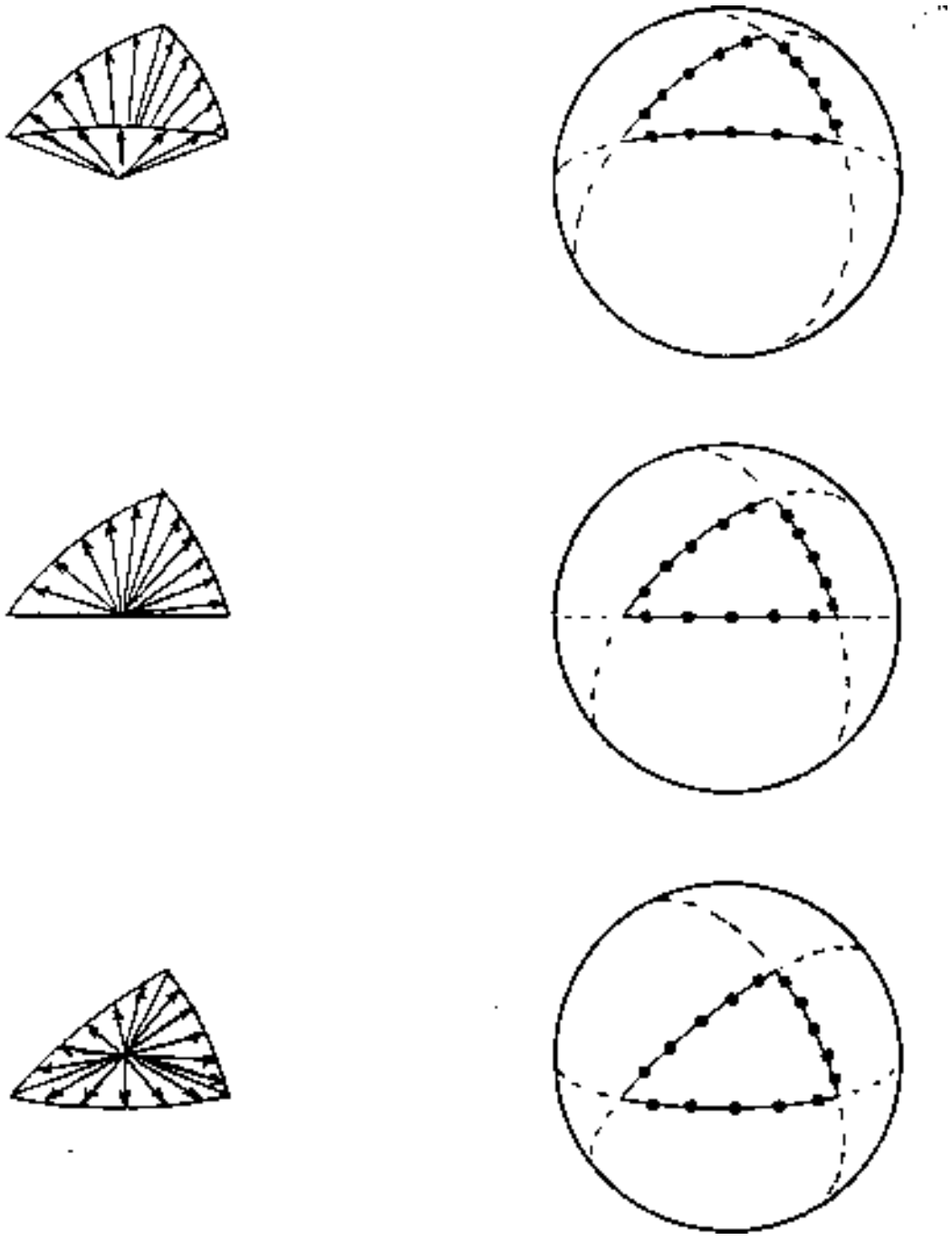

Figure 7.7: Three Views of the three Dimensional Vector Model with Rotation forward Accompanied by Spherical Representations of the Vector Terminals

Table 7.7

Uncorrelated Factor matrix for Nine Mental Tests

	Matrix A					Normalized A		
	1	2	3	h_j		1	2	3
1	.42	.36	.28	.6192	1	.67	.59	.45
2	.36	.54	.16	.7309	2	.64	.74	.21
3	.61	.16	.19	.6583	3	.92	.24	.30
4	.54	-.46	.09	.7126	4	.76	-.64	.13
5	.63	-.48	.05	.7960	5	.79	-.61	.06
6	.59	-.41	.06	.7161	6	.82	-.57	.09
7	.48	.40	-.21	.6618	7	.73	.61	-.31
8	.61	.00	-.29	.6738	8	.90	.00	-.44
9	.59	.16	-.25	.6603	9	.89	.24	-.38

Table 7.8

Factor Transformation Matrices for Nine Mental Tests

	Matrix T Trait Vectors				Matrix F Normal Vectors		
	1	2	3		1	2	3
1	.67	.67	.33	1	.31	.52	.80
2	.79	-.61	.09	2	.69	-.72	.08
3	.62	.53	-.58	3	.26	.20	-.94

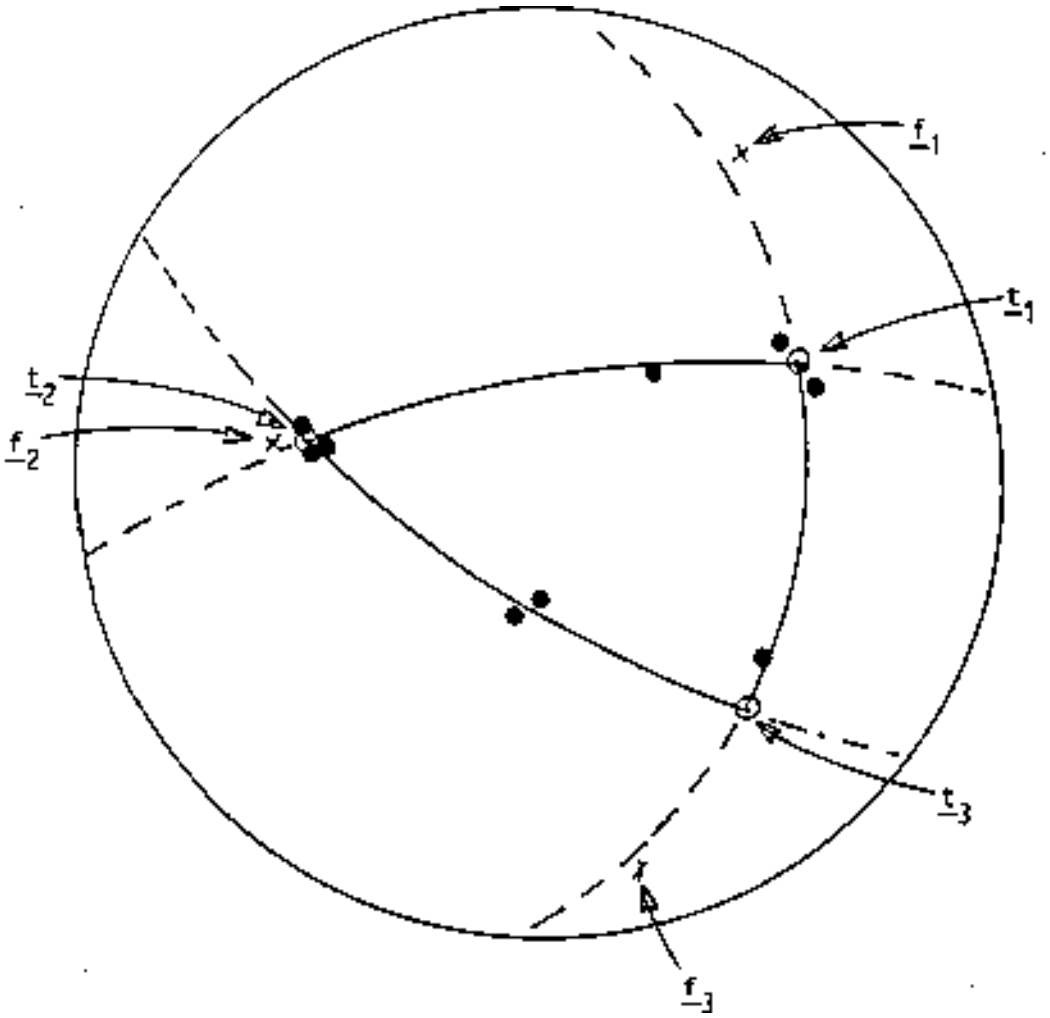

Figure 7.8: A Shperical Representation of the Nine Mental Tests Structure

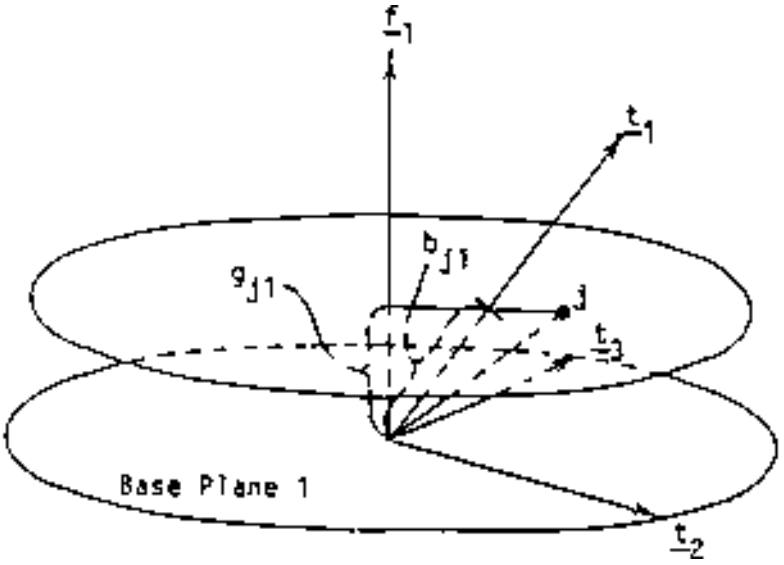

Figure 7.9: Illustration of Cartesian Coordinate and Projection on Normal in Three Dimensions

Table 7.9 which presents the matrices for the three types of structure coefficients for the nine mental test example. First is the matrix $\boldsymbol{B}$ of factor weights; second is the matrix $\boldsymbol{Q}$ of covariances of the modeled attributes with the factors; third is the matrix $\boldsymbol{G}$ of projections on the normals. Each of these will be discussed in turn.

As described in Chapter 3, the factor weights in matrix $\boldsymbol{B}$ are Cartesian coordinates of the attribute vectors on the trait vectors. In Figure 7.8 the base plane for factor 1 passes through trait vectors $\underline{\mathbf{t}}_2$ and $\underline{\mathbf{t}}_3$. This base plane is defined by normal $\underline{\mathbf{f}}_1$. Figure 7.9 gives an illustration of a Cartesian coordinate in three dimensions. Normal $\underline{\mathbf{f}}_1$ is vertical to base plane 1 horizontal. Trait vectors $\underline{\mathbf{t}}_2$ and $\underline{\mathbf{t}}_3$ are in this base plane. Attribute vector *j* is above the base plane and has had a projection plane passed through it parallel to the base plane. Trait vector $\underline{\mathbf{t}}_1$ passes through the projection plane at a distance b_{j1} from the origin. This is the Cartesian coordinate of attribute vector *j* with trait vector $\underline{\mathbf{t}}_1$. This is the geometric representation of a factor weight. It is the extent of the trait vector included in the attribute vector.

Structural coefficients in matrix $\boldsymbol{Q}$ are illustrated geometrically in Figure 7.10. These coefficients are the scalar products of the attribute vectors with the trait vectors. Since the trait vectors have unit length, the scalar products are the orthogonal projections of the attribute vectors on the trait vectors. Figure 7.10 is similar to Figure 7.9 involving the same normal $\underline{\mathbf{f}}_1$, base plane 1, trait vectors $\underline{\mathbf{t}}_1$, $\underline{\mathbf{t}}_2$ and $\underline{\mathbf{t}}_3$. The base plane 1 passes through trait vectors $\underline{\mathbf{t}}_2$ and $\underline{\mathbf{t}}_3$. Coefficient q_{jl} is the orthogonal projection of attribute vector *j* on trait vector $\underline{\mathbf{t}}_1$. This does not depend on the projection plane passed through the terminus of attribute vector *j* . As pictured, coefficient q_{jl} is larger than the factor weight b_{ji} shown in Figure 7.9. Note that the projection plane for an attribute vector in the base plane would have a zero factor weight while the orthogonal projection on the trait vector $\underline{\mathbf{t}}_1$ would not, necessarily, be zero. The coefficients in matrix $\boldsymbol{Q}$ are the covariances of the attributes with the traits.

Matrix $\boldsymbol{G}$ of projections on the normals is pictured in Figure 7.9 which shows the projection of attribute vector *j* on normal $\underline{\mathbf{f}}_1$. The projection plane parallel to the base plane and through the terminus of vector *j* cuts off a distance from the origin on normal $\underline{\mathbf{f}}_1$ equal to the projection g_{jl} . This projection is a measure of the contribution to attribute vector j which is independent of vectors in the base plane, including trait vectors $\underline{\mathbf{t}}_2$ and $\underline{\mathbf{t}}_3$. Statistically, these projections on the normals are semipartial covariances with trait vectors, having, for each trait vector, the effects of all other trait vectors partialled out.

As described in the preceding paragraphs, the structural coefficients in matrices $\boldsymbol{B}$, $\boldsymbol{Q}$ and $\boldsymbol{G}$ are geometrically independent of a reference set of axes. For mathematical and computational convenience the common factor structure is described in terms of an orthogonal axes frame starting with matrix $\boldsymbol{A}$. Matrices $\boldsymbol{T}$ of trait vectors and $\boldsymbol{F}$ or normals are referred to this same frame of orthogonal axes. The structural coefficients are independent of the location of

Table 7.9

Transformed Factor Matrices for Nine Mental Tests

(Structure Coefficient Matrices)

Matrix $B=AT^{-1}$

Factor Weights

		1	2	3
1	Addition	.66	.05	-.10
2	Multiplication	.67	-.05	.10
3	Three-Higher	.52	.32	.01
4	Figures	.00	.72	-.04
5	Cards	-.02	.79	.03
6	Flags	.02	.71	.02
7	Ident. Numbers	.23	.02	.49
8	Faces	-.06	.40	.54
9	Mirror Reading	.08	.28	.52

Matrix Q=AT'

Covariances with Factors

	1	2	3
1	.61	.13	.29
2	.72	.05	.48
3	.57	.40	.35
4	.08	.71	.04
5	.11	.80	.11
6	.14	.72	.11
7	.52	.11	.63
8	.31	.45	.55
9	.42	.34	.59

Matrix G=AF'

Projection on Normals

	1	2	3
1	.54	.05	-.08
2	.55	-.05	.08
3	.42	.32	.01
4	.00	.71	-.03
5	-.01	.79	.02
6	.02	.70	.02
7	.19	.02	.40
8	-.05	.39	.44
9	.06	.27	.42

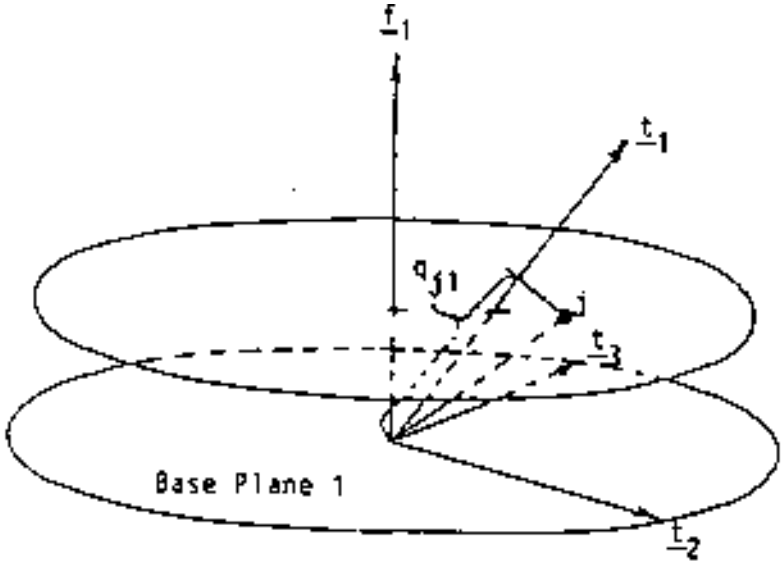

Figure 7.10: Illustration of Orthogonal Projection of an Attribute Vector on a Trait Vector

the orthogonal axes. They are not altered by an orthogonal rotation of axes. Consider that there is one frame of orthogonal axes on which the vectors have coordinates in matrices $\boldsymbol{A}_1$, $\boldsymbol{T}_1$ and $\boldsymbol{F}_1$. Consider an orthogonal rotation of axes by a matrix $\boldsymbol{M}_{12}$ as was used in equation (7.29) to effect such a rotation.

$$\boldsymbol{A}_1\boldsymbol{M}'_{12} = \boldsymbol{A}_2 \tag{7.29}$$

This rotation can be carried to matrices $\boldsymbol{T}_1$ and $\boldsymbol{F}_1$:

$$\boldsymbol{T}_1\boldsymbol{M}'_{12} = \boldsymbol{T}_2 \tag{7.30}$$

$$\boldsymbol{F}_1\boldsymbol{M}'_{12} = \boldsymbol{F}_2 \tag{7.31}$$

From the orthogonal properties of matrix $\boldsymbol{M}_{12}$:

$$\boldsymbol{T}_2^{-1} = \boldsymbol{M}_{12}\boldsymbol{T}_1^{-1} \tag{7.32}$$

Then,

$$\boldsymbol{A}_2\boldsymbol{T}_2^{-1} = \boldsymbol{A}_1\boldsymbol{M}'_{12}\boldsymbol{M}_{12}\boldsymbol{T}_1^{-1} = \boldsymbol{A}_1\boldsymbol{T}_1^{-1} = \boldsymbol{B} \tag{7.33}$$

A similar demonstration of invariance is possible for matrices $\boldsymbol{Q}$ and $\boldsymbol{G}$.

Table 7.10 presents an outline of terminology used for the structural coefficients for attribute measures in the common factor space. We have presented three types of coefficients with their interpretations. A terminology has come into common use following Harman (1976). He pointed out the existence of a fourth type of coefficient in matrix $\boldsymbol{A}\boldsymbol{F}^{-1}$. We do not have a ready interpretation for these coefficients. Computer packages frequently print matrices $\boldsymbol{B}$ and $\boldsymbol{G}$ with shortened titles of Pattern Loadings and Structure Loadings. We are using the term structure to refer to the collection of attribute vectors, trait vectors and normals to the hyperplanes along with the hyperplanes. Attachment of the term "structure loading" to one pair of types of coefficients appears to be a questionable interpretation of Thurstone's use of the word structure, as in simple structure. Thurstone used the projections on the normals as a measure of the independent contributions to the attributes.

Consider, now, the relations among and between the trait vectors and the normals. Table 7.11 presents matrices of these relations for the Nine Mental Tests example. Matrix $\boldsymbol{R}_{bb}$ has correlations among transformed factors, these correlations equaling the scalar products among the trait vectors. Since the trait vectors are unit vectors, these scalar products equal cosines of angles among the trait vectors. Matrix $\boldsymbol{F}\boldsymbol{F}'$ contains scalar products (equal to cosines of angles) among the normals. Due to an inverse relation between $\boldsymbol{R}_{bb}$ and $\boldsymbol{F}\boldsymbol{F}'$, the algebraic signs of the off-diagonal entries in these two matrices tend to be opposite. Matrix $\boldsymbol{D}$ contains scalar products between the trait vectors and the normals. Traits vectors and normals are paired,

Table 7.10
Outline of Structure Coefficients

Matrix	Terminology used here	Common terminology	
$AT^{-1} = B$	Factor Weights	Pattern loadings	on primary vectors
$AT' = Q$	Covariances with factors	Structure loadings	on primary vectors
AF^{-1}		Pattern loadings	on reference vectors
$AF' = G$	Projections on normals; semipartial covariances with factors	Structure loadings	on reference vectors

Table 7.11
Transformed Factors Relations Matrices for Nine Mental Tests

Matrix $R_{bb}=TT'$

	1	2	3
1	1.00	.14	.58
2	.14	1.00	.11
3	.58	.11	1.00

Matrix FF'

	1	2	3
1	1.00	-.10	-.57
2	-.10	1.00	-.03
3	-.57	-.03	1.00

Matrix D=TF'

	1	2	3
1	.81		
2		.99	
3			.82

one trait vector with one normal. The scalar product between the trait vector and the normal of a pair is a diagonal entry in $\boldsymbol{D}$. The off-diagonal entries in $\boldsymbol{D}$ are scalar products between non-paired trait vectors and normals. Each normal is orthogonal to all non-paired trait vectors. Likewise, each trait vector is orthogonal to all non-paired normals.

Various aspects of a structure in the common factor space have been discussed in the preceding paragraphs. Thurstone's conception of a simple structure is a particular form involving a theoretic postulate which may or may not be satisfied for any given body of data. As was presented earlier, the configuration of attribute vectors is defined by the covariance matrix $\boldsymbol{C}_c$ from the common factors. A structure is not complete until the trait vectors and normals have been defined. Thurstone's principal of simple structure involves a statement that the trait vectors could be located such that there would be many zero or near zero factor loadings. He argued that the nature of more basic factors should be such that each factor should influence scores on some attributes but not all attributes. If this be the case, trait vectors and normals could be located so that there were a number of trivial loadings on each factor. The hyperplanes could be defined by being near to a number of attribute vectors.

For examples of the concept of simple structure consider the nine mental tests example. In Figure 7.8 the base plane for factor 3 passes through trait vectors $\underline{\mathbf{t}}_1$ and $\underline{\mathbf{t}}_2$. This plane is moderately well defined by six points which are for the first six tests in the battery, see matrix $\boldsymbol{B}$ in Table 7.9, column 3 for factor 3. The last three tests have high loadings on this factor, they are the tests not near the base plane in Figure 7.8. These three tests involve highly speeded tasks of simple recognition of stimuli among distractors. This factor has been interpreted as a perceptual speed factor. While the first six tests were timed, performance depended more on conduct of mental tasks and to a trivial extent on speed of perceptual recognition of answers. By having only trivial dependence on perceptual speed, the first six tests determine the base plane for the third factor. The base plane for factor 1 passes through trait vectors $\underline{\mathbf{t}}_2$ and $\underline{\mathbf{t}}_3$. This plane is defined by five points in Figure 7.8 and is not as well determined as would be desirable. There is a point near trait vector $\underline{\mathbf{t}}_3$ near to the base plane but not "in" it. Consider the factor loadings for factor 1 in Table 7.9. The first three tests have high loadings; these tests involve numerical operations. Test 7, identical numbers, also involves simple numerical operations and has a lower loading on factor 1; it is the test with a vector near trait vector $\underline{\mathbf{t}}_3$. The base plane for factor 1 is defined by the vectors for the other five tests which do not involve numerical operations. The base plane for factor 2 is less well defined by only three points. These points are for tests 1, addition, 2, multiplication, and 7, identical numbers; all other tests appear to involve some form of spatial manipulation. However, each of the base planes in this example are defined by attribute vectors in the vector configuration. Thurstone's conception of simple structure provides for establishing

the trait vectors and normals so that the base planes, or hyperplanes, are defined by being near trait vectors.

A number of matters related to simple structure are illustrated in the following figures. Two major classes of simple structure exist: a simple structure with a positive manifold and a structure without a positive manifold. These two classes are illustrated in Figure 7.11. A positive manifold is generated in a domain for which only zero or positive effects exist. There appear to be very few inhibitive effects in mental abilities. A major observation is that ability measures correlate zero or positively which leads to a hypothesis that there will be no negative influences of factors on performance of ability tasks. Of course, there may be a rare exception so that investigators should be alert to this possibility and allow a negative factor loading in the transformation of factors. In contrast to the ability domain is the personality measures domain. Negative relations are commonly observed between personality measures which leads to the possibilities of negative factor loadings. The factor transformation problem is much more difficult when there is not a positive manifold.

Figure 7.12 presents relations between clusters of attributes and simple structure. The sphere on the left pictures an independent cluster configuration with the attribute vectors clustered at the corners of the configuration. Each attribute is dependent on only one factor. While this is an extreme of simple structure, it is not the only possibility. A common misconception of simple structure is that Thurstone's principle is satisfied only by an independent cluster configuration. The configuration on the right pictures a simple structure with no pure attributes at the corners. The planes are well defined by points between the corners. This configuration on the right has dependent clusters, one in the middle and one part way along the bottom plane. Some proposed methods of factor analysis have used clusters to define factors, both in the extraction of factors and in the transformation of factors from an original factor matrix. The existence of clusters of attributes may have very interesting scientific meaning; however, clusters are not reliable in defining simple structures. A well defined simple structure plane will have attribute vectors in the plane stretched out from one corner to the other. Being in a plane is to be interpreted as being only a trivial distance away from the plane. The definition of a simple structure plane, or hyperplane in higher dimensions, depends not only on the number of attribute vectors in the plane but also on the diversity of these attribute vectors.

The possibility of a general factor presents a problem for simple structure. If there is a common factor for a battery of attributes, there would be no attributes having zero loadings to define the hyperplane. The upper left configuration in Figure 7.13 illustrates the case of a common factor. There are no points on the lower plane. A possibility which has been followed in a few studies is to establish the planes for all factors except for the general factor and, then, to set the plane for the general factor orthogonal to the other planes.

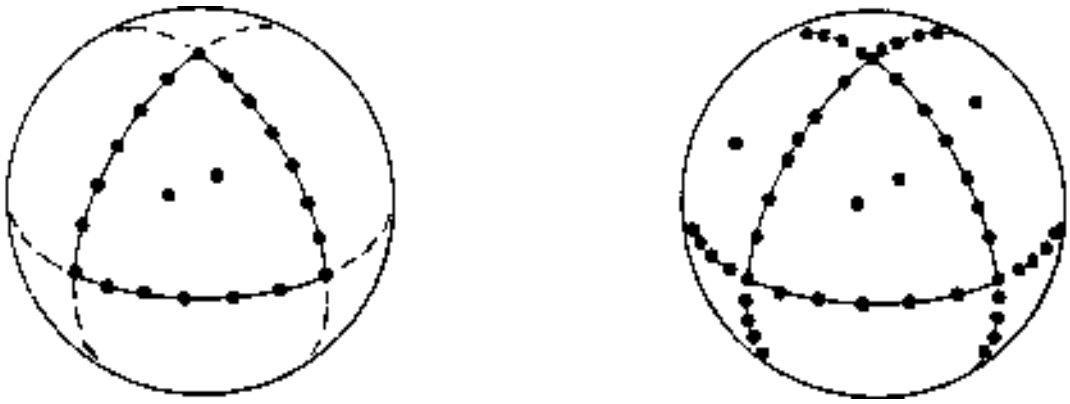

Figure 7.11: Shperical Representations of Two Classes of Simple Structure: 1. A Positive Manifold; 2. A Configuration without a Positive Manifold

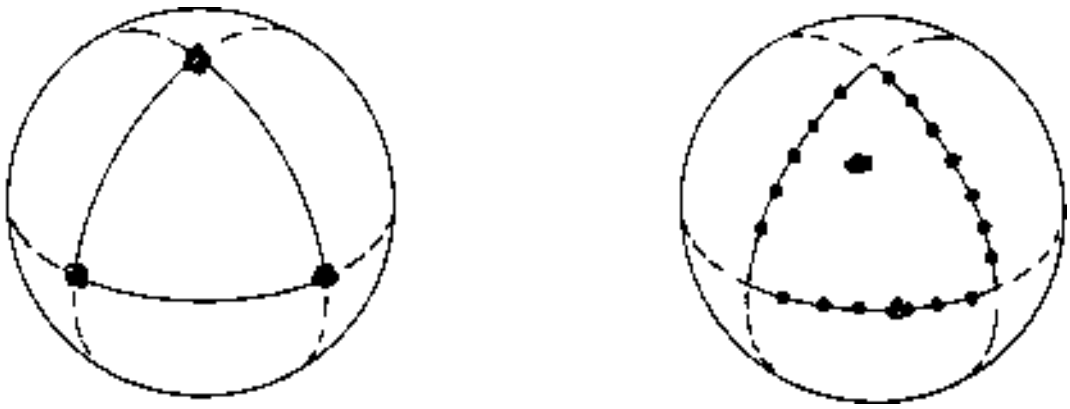

Figure 7.12: Simple Structures with Clusters of Attribute Vectors: 1. Independent Clusters; 2. Dependent Clusters

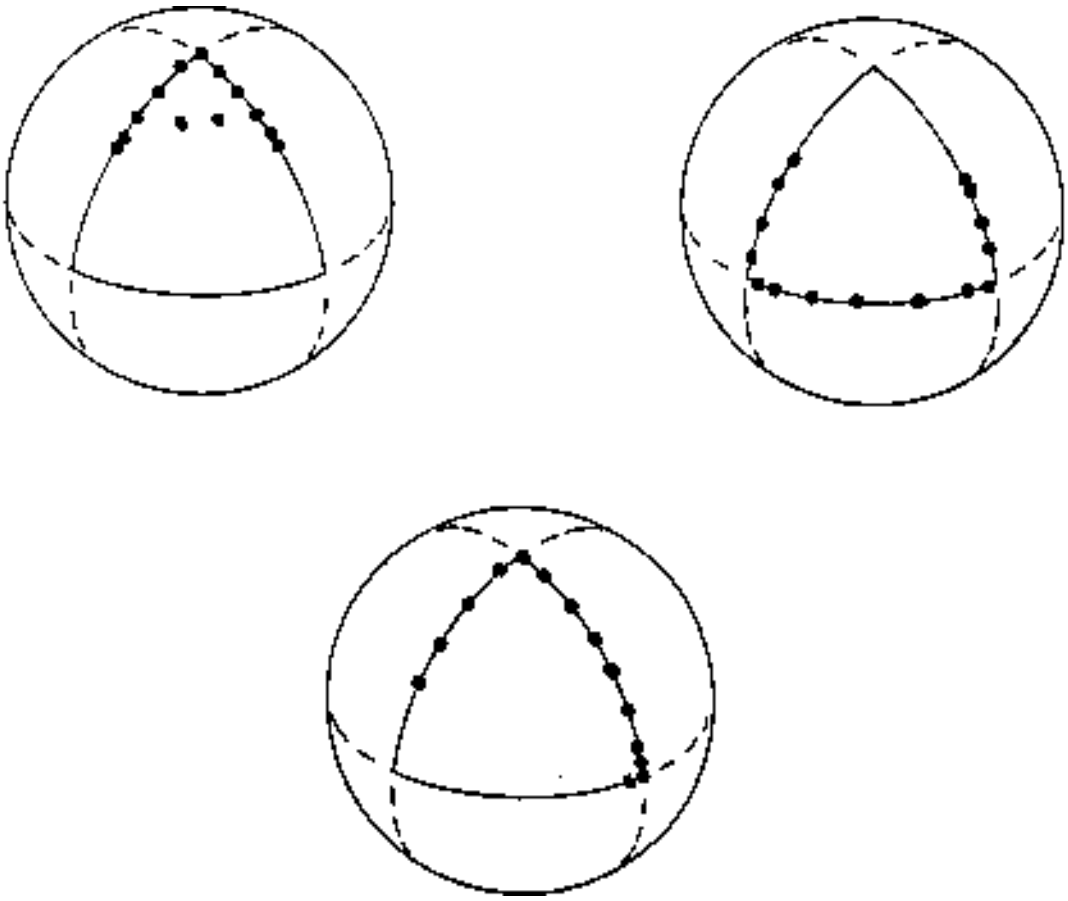

Figure 7.13: Incomplete Simple Structures: 1. A General Factor; 2. A Structure with a Corner Missing; 3. A structure with an Undefined Plane

The configuration at the upper right of Figure 7.13 presents another common type of problem: there are no points near the extreme of one factor. All attributes depend on other factors plus some dependence on the upper factor. For an example consider the case for a reasoning factor. The attributes that depend on this reasoning factor also depend upon a content type factor. There appears to be no pure reasoning tests. As a result, the hyperplanes for the other factors are less well defined by having points only part way toward the reasoning trait vector.

The bottom configuration of Figure 7.13 illustrates a difficulty derived from an incomplete battery of attributes. There are no attribute measures along the bottom plane in the direction of the left trait vector. This leaves this plane indeterminant. A solution may be available for future studies to include in the battery measures which would not depend on the factor for the upper trait but would depend upon the factor to the left side. A tentative solution for the present study would involve setting the left trait vector orthogonal to the upper trait vector.

In the preceding paragraphs we have reviewed some properties and problems of the transformation to simple structure. Many years of experience have indicated that Thursone's principal of simple structure is an extremely useful tool in the transformation of factors to meaningful solutions.

CHAPTER 8
FACTOR EXTRACTION BY MATRIX FACTORING TECHNIQUES

From
Exploratory Factor Analysis
Ledyard R Tucker
and
Robert C. MacCallum

CHAPTER 8
FACTOR EXTRACTION BY MATRIX FACTORING TECHNIQUES

In this chapter we delve into a number of intricacies of factor extraction by matrix factoring techniques. These methods are the older ones followed in many early studies when there were considerable computation limitations. These methods were developed prior to the advent of large scale computers. However, some of the techniques were not utilized due to the computing requirements but were considered as highly desirable. With modern computers these methods are quite feasible. For a general framework in considering the factor extraction techniques we refer to Guttman's (1944) general theory and methods for matrix factoring which provides a theoretic foundation for the matrix factoring techniques which had been in use or had been being considered for some time. The methods to be considered in this chapter maximize some function of the obtained factor loadings. Consideration of residuals is only tangential. Two areas of problems are closely related to the factor extraction procedures, these are: the problem of "guessed communalities" and the problem of number of factors to be extracted. After presentation of the general theory of matrix factoring, a section will discuss the subject of "guessed communalities". Determination of the number of factors appears to be closely related to the method of factor extraction and will be discussed with each such technique.

Before discussion of details of matrix factoring techniques there are several preliminary matters to be considered. These techniques are not scale free when applied to covariance matrices in general in that results vary with scaling of the attributes. The property of a factor extraction technique being scale free may be explained in reference to the following equations. Let $\boldsymbol{C}$ be an original covariance matrix to which a scaling diagonal matrix $\boldsymbol{D}$ is applied, $\boldsymbol{D}$ being finite, positive, non-singular. Let $\widetilde{\boldsymbol{C}}$ be the rescaled covariance matrix.

$$\widetilde{\boldsymbol{C}} = \boldsymbol{DCD}$$

A selected factor extraction technique is applied to $\widetilde{\boldsymbol{C}}$ to yield a common factor matrix $\widetilde{\boldsymbol{A}}$ which is scaled back to $\boldsymbol{A}_c$ for the original covariance matrix by:

$$\boldsymbol{A}_c = \boldsymbol{D}^{-1}\widetilde{\boldsymbol{A}}$$

For the factor extraction technique to be scale free, matrix $\boldsymbol{A}_c$ must be invariant with use of different scaling matrices $\boldsymbol{D}$. As is well known, a correlation matrix is independent of the scaling of a covariance matrix.

$$\boldsymbol{R} = \boldsymbol{S}^{-1}\boldsymbol{C}\boldsymbol{S}^{-1} = \widetilde{\boldsymbol{S}}^{-1}\widetilde{\boldsymbol{C}}\,\widetilde{\boldsymbol{S}}^{-1}$$

where

$$S^2 = Diag(C)$$

$$\widetilde{S}^2 = Diag(\widetilde{C}\,)$$

Note that:

$$\widetilde{S}\; = DS$$

A common factor matrix A obtained from R may be scaled back to $\widetilde{A}$ and A_c by:

$$\widetilde{A}\; = \widetilde{S}\; A$$

$$A_c = SA$$

In order to avoid scaling problems we follow the tradition of applying these techniques to correlation matrices. In a sense, this usage does make these techniques scale free when the obtained factor matrix may be scaled back to apply to the original attribute scales.

For notational convenience, the subscripts of the observed correlation matrix R_{yy} are dropped so that the observed correlation matrix is indicated by R . Also, matrix Δ_R will be indicated by A. Equation (7.17) becomes:

$$R = AA' + U^2 + \Delta \tag{8.1}$$

This is the basic equation considered in this chapter. An alternative equation is obtained by defining matrix R_h with adjusted diagonal entries with "guessed communalities" by:

$$R_h = R - U^2 \tag{8.2}$$

From equation (8.1):

$$R_h = AA' + \Delta \tag{8.3}$$

Matrix Δ contains residual correlations. Several letters used in transformations of factors will be used in the present context on a temporary basis to designate other matrices in this chapter.

8.1. General Theory of Matrix Factoring

Guttman (1944) developed a general theory of matrix factoring with which he described several existing methods of factor extraction. This theory applies, strictly in the present context, to Gramian matrices. These matrices need not be of full rank; however, for the present purposes they must not have imaginary dimensions (the least eigenvalue must be non-negative). A correlation matrix satisfies these conditions since it is the product of a score matrix times its transpose, an original definition of a Gramian matrix. Usually, however, matrix R_h with adjusted

diagonal entries containing "guessed communalities" is not Gramian. Nevertheless, the theory of matrix factoring is applied to matrix $\boldsymbol{R}_h$. In practice, with a few exceptions, this use appears to work satisfactorily.

The general procedure starts from a Gramian matrix $\mathbf{G_1}$ which is $n \times n$ and of rank r , greater than 0 and equal to or less than n. A matrix $\boldsymbol{W}$, $n \times m$ with m greater than 0 and equal to or less than n . $\boldsymbol{W}$ is to contain real numbers and satisfy a restriction stated later. Matrix $\boldsymbol{Q}$ is defined by:

$$\boldsymbol{Q} = \boldsymbol{G}_1\boldsymbol{W} \tag{8.4}$$

Note that $\boldsymbol{Q}$ is $n \times m$. Matrix $\boldsymbol{P}$, $m \times m$, is defined by:

$$\boldsymbol{P} = \boldsymbol{W}'\boldsymbol{G}_1\boldsymbol{W} = \boldsymbol{W}'\boldsymbol{Q} \tag{8.5}$$

Matrix $\boldsymbol{P}$ is Gramian and must be of full rank; this is the restriction on matrix $\boldsymbol{W}$. A square decomposition matrix $\boldsymbol{F}$ is determined such that:

$$\boldsymbol{F}\boldsymbol{F}' = \boldsymbol{P} \tag{8.6}$$

Any of several techniques may be used to determine $\boldsymbol{F}$. A section of a factor matrix, $n \times m$, on orthogonal axes is defined by:

$$\boldsymbol{A}_1 = \boldsymbol{Q}(\boldsymbol{F}^{-1})' \tag{8.7}$$

and a residual matrix $\boldsymbol{G_2}$ is defined by:

$$\boldsymbol{G_2} = \boldsymbol{G}_1 - \boldsymbol{A}_1\boldsymbol{A}_1' \tag{8.8}$$

There is a dual problem of proof. First, that the rank of $\boldsymbol{G_2}$ is (r - m) . Second, that $\boldsymbol{A_1}$ is a section with m columns of a complete, orthogonal factor matrix of $\boldsymbol{G_1}$.

The required proofs are expedited by considering a complete decomposition of $\boldsymbol{G_1}$ to a matrix $\boldsymbol{B}$, $n \times r$, such that:

$$\boldsymbol{G}_1 = \boldsymbol{B}\boldsymbol{B}' \tag{8.9}$$

$\boldsymbol{B}$ is a factor matrix on orthogonal axes and can be obtained by any of a number of procedures such as the procedure developed by Commandant A. L. Cholesky of the French Navy around 1915 and described by Dwyer (1944) as the square root method. There are many other possible procedures to obtain this decomposition. With equation (8.9), matrices $\boldsymbol{Q}$ and $\boldsymbol{P}$ become:

$$\boldsymbol{Q} = \boldsymbol{B}\boldsymbol{B}'\boldsymbol{W} \tag{8.10}$$

$$\boldsymbol{P} = \boldsymbol{W}'\boldsymbol{B}\boldsymbol{B}'\boldsymbol{W} \tag{8.11}$$

Then, matrix $\boldsymbol{A_1}$ becomes:.

$$\boldsymbol{A_1 = BB'W(F^{-1})'} \tag{8.12}$$

Define a matrix $\boldsymbol{T_1}$, $r \times m$:

$$\boldsymbol{T_1 = B'W(F^{-1})'} \tag{8.13}$$

Then, matrix $\boldsymbol{A_1}$ becomes:

$$\boldsymbol{A_1 = BT_1} \tag{8.14}$$

Matrix $\boldsymbol{T_1}$ is column wise orthonormal as shown by the following:

$$\begin{aligned} \boldsymbol{T_1T_1'} &= \boldsymbol{F^{-1}W'BB'(F^{-1})'} \\ &= \boldsymbol{F^{-1}P(F^{-1})'} \\ &= \boldsymbol{F^{-1}FF'(F^{-1})'} \\ &= \boldsymbol{I} \end{aligned} \tag{8.15}$$

Matrix $\boldsymbol{T}$, $r \times r$ orthonormal, is completed by adjoining section $\boldsymbol{T_2}$, $r \times (r - m)$ to $\boldsymbol{T_1}$.

$$\boldsymbol{T = [T_1, T_2]} \tag{8.16}$$

$\boldsymbol{T_2}$ is column wise orthonormal and orthogonal by columns to $\boldsymbol{T_1}$. Matrix $\boldsymbol{B}$ is rotated orthogonally by $\boldsymbol{T}$ to yield matrix $\boldsymbol{A}$ with sections $\boldsymbol{A_1}$ and $\boldsymbol{A_2}$ **.**

$$\boldsymbol{BT = B[T_1, T_2] = [BT_1, BT_2] = [A_1, A_2] = A} \tag{8.17}$$

Since $\boldsymbol{T}$ is an orthonormal rotation and from equation (8.9):

$$\boldsymbol{G_1 = AA' = A_1A_1' + A_2A_2'} \tag{8.18}$$

Then with equation (8.8):

$$\boldsymbol{G_2 = A_2A_2'} \tag{8.19}$$

Matrix $\boldsymbol{G_2}$ is of rank ($r - m$). The derivation of $\boldsymbol{A_1}$ has removed this section from the complete matrix $\boldsymbol{A}$. This completes the needed proof.

A point of interest is that $\boldsymbol{G_2}$ is orthogonal to $\boldsymbol{W}$ as shown as follows.

$$\boldsymbol{G_2W = (G_1 - A_1A_1')W = G_1W - A_1A_1'W} \tag{8.20}$$

Note, from equation (8.7), that:

$$\boldsymbol{A_1A_1' = Q(F^{-1})'F^{-1}Q' = QP^{-1}Q'} \tag{8.21}$$

With the results of equation (8.21), equation (8.20) becomes:

$$\begin{aligned} \boldsymbol{G_2 W} &= \boldsymbol{G_1 W} - \boldsymbol{Q P^{-1} Q' W} \\ &= \boldsymbol{Q} - \boldsymbol{Q P^{-1} P} \\ &= \boldsymbol{0} \end{aligned} \tag{8.22}$$

A second point of interest is the relation between matrix $\boldsymbol{W}$ and the obtained factor weight matrix $\boldsymbol{A_1}$. From equations (8.5), (8.6), and (8.7)

$$\begin{aligned} \boldsymbol{W' A_1} &= \boldsymbol{W' Q (F^{-1})'} \\ &= \boldsymbol{P (F^{-1})'} \\ &= \boldsymbol{F F' (F^{-1})'} \\ &= \boldsymbol{F} \end{aligned} \tag{8.23}$$

This result is important in the factor extraction methods.

Equations (8.4) through (8.8) are the bases of major steps in the factor extraction techniques to be considered in this chapter. These methods differ in the determination of matrix $\boldsymbol{W}$. Each of these techniques starts with a correlation matrix $\boldsymbol{R_{h1}}$, having adjusted diagonal entries, determines a weight matrix $\boldsymbol{W}$, computes a section of a factor matrix $\boldsymbol{A_1}$, then computes a residual matrix $\boldsymbol{R_{h2}}$ by equation (8.8). This process is repeated with the residual matrix to obtain the next section of the extracted factor matrix. This process is repeated with the succession of residual matrices until the full extracted factor matrix is obtained. As indicated earlier, the number of factors to be extracted is a decision frequently made from information obtained during the factor extraction process. The basis of this decision depends on the factor extraction technique employed.

8.2. The "Guessed Communalities" Problem

A preliminary operation in matrix factoring is to establish entries in the diagonals of the correlation matrix. The theory of common factor analysis presented in the preceding chapters establishes a basis for this operation. Note that the principal components procedure ignores the issue of common vs. unique variance and leaves unities in the diagonal of the correlation matrix. An early operational procedure using communality type values was the "Highest R" technique developed by Thurstone (1947) and was used in many of his studies as well as by many other individuals following Thurstone's lead. However, early in applications of digital computers there was a proposal to use principal components as an easy approximation to factor analysis. Kaiser (1960) described such a technique followed by VARIMAX transformation of the weight matrix. This procedure, which became known as the "Little Jiffy" after a suggestion by Chester Harris, is retained as an alternative in a number of computer packages. We can not recommend this

procedure and are concerned that many unwary individuals have been misled by the ease of operations. Serious problems exist.

A very simple example is presented in Table 8.1 with the constructed correlation matrix being given at the upper left. This correlation matrix was computed from a single common factor with uniqueness so that the theoretic communalities were known and have been inserted into the matrix on the upper right. There are neither sampling nor lack of fit discrepancies. A principal components analysis is given on the left and a principal factors analysis is given on the right. The principal factors procedure will be considered later in detail. For the principal components analysis an eigen solution (see Appendix A on matrix algebra for a discussion of the eigen problem) was obtained of the correlation matrix having unities in the diagonal. The series of eigenvalues are given on the left along with the component weights for one dimension. These weights are the entries in the first eigenvector times the square root of the first eigenvalue. The matrix of residuals after removing this first component is given at the bottom left. The principal factors analysis followed the same procedure as the principal components analysis but is applied to the correlation matrix having communalities in the diagonal. For the principal factors, both the weights for one factor and the uniqueness are given. These are the values used in constructing the correlation matrix.

There are a number of points to note. First, the eigenvalues for the principal components after the first value do not vanish as do the corresponding eigenvalues for the principal factors analysis. If one followed the procedure for the principal components of retaining only dimensions for which eigenvalues were greater than unity, only the first principal component would be used, this being the number of dimensions used in this example. For the principal factors analysis, only one factor existed since there was only one eigenvalue that did not vanish. A more important comparison concerns the obtained weights. All of the principal component weights are greater than the corresponding principal factor weights. This is especially true for the low to medium sized weights. Use of the principal components procedure exaggerates the values of the obtained weights. A further comparison is provided by the residual matrices. For the principal components analysis on the left, the diagonal values might be taken to be unique variances (a common interpretation). However, note the negative residuals. The component weights have removed too much from the correlations leaving a peculiar residual structure. For the principal factors analysis, all of the residual entries vanish in this example. The principal factors analysis yields a proper representation of the correlation matrix. A conclusion from this example is that leaving unities in the diagonals of a correlation matrix leads to a questionable representation of the structure of the correlation matrix.

Tables 8.2 and 8.3 provide a comparison when lack of fit is not included and is included in a constructed correlation matrix. The correlation matrix in Table 8.2 was computed from the

Table 8.1

Comparison of factor Extraction from a Correlation Matrix with Unities in the Diagonal versus Communalities in the Diagonal

Correlation Matrices

Unities in Diagonal (RW1)

	1	2	3	4
1	1.00	.35	.21	.07
2	.35	1.00	.15	.05
3	.21	.15	1.00	.03
4	.07	.05	.03	1.00

Communalities in Diagonal (RWH)

	1	2	3	4
1	.49	.35	.21	.07
2	.35	.25	.15	.05
3	.21	.15	.09	.03
4	.07	.05	.03	.01

Eigenvalues of (RW1)

1	1.50
2	.99
3	.87
4	.64

Eigenvalues of (RWH)

1	.84
2	.00
3	.00
4	.00

Principal Components

	Weights
1	.78
2	.73
3	.56
4	.22

Principal Factors

	Weights	Uniqueness
1	.70	.51
2	.50	.75
3	.30	.91
4	.10	.99

Residual Matrix from (RW1)

	1	2	3	4
1	.40	-.22	-.23	-.10
2	-.22	.46	-.26	-.11
3	-.23	-.26	.69	-.09
4	-.10	-.11	-.09	.95

Residual Matrix from (RWH)

	1	2	3	4
1	.00	.00	.00	.00
2	.00	.00	.00	.00
3	.00	.00	.00	.00
4	.00	.00	.00	.00

Table 8.2
Illustration of "Communality" Type Values
Simulated Correlation matrix without Discrepancies of Fit

	Major Domain Matrix		Variance Components		
	1	2	Major	Unique	Minor
1	.8	.0	.64	.36	.00
2	.6	.0	.36	.64	.00
3	.5	.1	.26	.74	.00
4	.2	.3	.13	.87	.00
5	.0	.2	.04	.96	.00
6	.0	.6	.36	.64	.00

Simulated Population Correlation Matrix

	1	2	3	4	5	6
1	1.00					
2	.48	1.00				
3	.40	.30	1.00			
4	.16	.12	.13	1.00		
5	.00	.00	.02	.06	1.00	
6	.00	.00	.06	.18	.12	1.00

"Communality" Type Values

	Attribute					
Method	1	2	3	4	5	6
Major Domain Variance	.64	.36	.26	.13	.04	.36
Highest R	.48	.48	.40	.18	.12	.18
Squared Multiple Correlation	.31	.25	.18	.07	.02	.05
Iterated Centroid Factors	.64	.36	.26	.13	.04	.36
Iterated Principal Factors	.64	.36	.26	.13	.04	.36
Alpha Factor Analysis	.64	.36	.26	.13	.04	.36
Unrestricted Maximum Likelihood	.64	.36	.26	.13	.04	.36

Table 8.3
Illustration of "Communality" Type Values
simulated Correlation matrix without Discrepancies of Fit

Major Domain Matrix			Variance Components		
	1	2	Major	Unique	Minor
1	.8	.0	.64	.36	.10
2	.6	.0	.36	.54	.10
3	.5	.1	.26	.64	.10
4	.2	.3	.13	.77	.10
5	.0	.2	.04	.86	.10
6	.0	.6	.36	.54	.10

Simulated Population Correlation Matrix

	1	2	3	4	5	6
1	1.00					
2	.49	1.00				
3	.37	.28	1.00			
4	.17	.16	.15	1.00		
5	-.04	-.02	.03	.02	1.00	
6	-.03	.02	.07	.21	.10	1.00

"Communality" Type Values

	Attribute					
Method	1	2	3	4	5	6
Major Domain Variance	.64	.36	.26	.13	.04	.36
Highest R	.49	.49	.37	.21	.10	.21
Squared Multiple Correlation	.31	.25	.16	.09	.01	.06
Iterated Centroid Factors	.67	.36	.23	.12	.01	.73
Iterated Principal Factors	.65	.37	.23	.15	.02	.50
Alpha Factor Analysis	.69	.37	.23	.11	.02	.63
Unrestricted Maximum Likelihood	.65	.37	.23	.16	.02	.45

major domain matrix and the listed uniqueness. This is a perfect population correlation matrix. In contrast, the technique described in Chapter 3, section 3.9, was used to add lack of fit discrepancies to the generated correlation matrix given in Table 8.3. The bottom section of each of these tables presents communality type values determined by a number of techniques. Discussion will compare results by these techniques between the cases when there are no discrepancies of fit and when there are discrepancies of fit.

Consider the case in Table 8.2 when no discrepancies of fit were included. The first row in the bottom section contains the major domain variances which may be considered as theoretic communalities. The second and third rows contain communality type values used in factor extraction techniques. These will be discussed in subsequent paragraphs. The last four rows give results from four methods of factor extraction which should result in ideal communalities in the population. These methods will be discussed in detail in subsequent sections. All four of these techniques utilize iterative computing methods to arrive at stable determinations of communalities. Note that all four techniques are successful in arriving at the theoretic values given in the row for the major domain variances. If a correlation matrix were determined from a sample of individuals from a population characterized by a given population matrix which does not include discrepancies of fit, an objective of factor extraction from this sample correlation matrix would be to arrive at estimates of the major domain variances.

Consider the case in Table 8.3 when discrepancies of fit were included. As before, the first row in the bottom section contains the major domain variances which, as will be demonstrated, no longer can be considered as theoretic communalities. As before, the second and third rows contain communality type values used in factor extraction techniques and will be discussed subsequently. The last four rows give results for this case from four methods of factor extraction which will be discussed in detail in subsequent sections. The values in these rows not only vary from the values in Table 8.2 but also differ between methods of factor extraction. Note, especially the values for attribute 6 for which the communality type values for the four methods of factor extraction vary from .73 to .45 and are all greater than the .36 in Table 8.2. Inclusion of the discrepancies of fit has had an effect on these values which is different for different methods of factor extraction. There no longer is a single ideal solution. The population communalities differ by method of factor extraction and provide different objectives to be estimated from sample correlation matrices. This conclusion poses a considerable problem for statistical modeling which ignores discrepancies of fit. We conclude that ignoring discrepancies of fit is quite unrealistic and raises questions concerning several factoring techniques. This is an illustration of the discussion in Chapter 3 of the effects of lack of fit on results obtained by different methods of fitting the factor model.

8.2.1. Highest R Procedure

Thurstone developed the highest R procedure during the analyses of large experimental test batteries such as in his study of Primary Mental Abilities (1938). Computations were performed using mechanical desk calculators so that a simple procedure was needed which would yield values in the neighborhood of what might be considered as the most desirable values. The highest R technique is based on intuitive logic; there is no overall developmental justification.

Application of the highest R technique involves, for each attribute, *j* , finding the other attribute, *i* , which correlates most highly in absolute value with *j*. This correlation, in absolute value, is taken as a communality type value for attribute *j.* This procedure is illustrated in Table 8.4. The correlation matrix in this table is the same as in Table 8.3 with reversals of directions of attributes 3 and 5. In making these reversals of directions, algebraic signs are reversed in rows and columns for these two attributes. Note that there are double reversals for the diagonal entries and the correlation between the two attributes which results in no sign changes for these entries. Consider the column for attribute 1. Ignore the diagonal entry and find the largest other entry in absolute value. This value is .49 in row 2 and is recorded in the "Highest R" row. Note that for column 3 the highest value in absolute value is .37, this is the -.37 in row 1. The foregoing procedure is followed for each of the attributes. Note that the sign changes in the correlation matrix between Tables 8.3 and 8.4 did not change the values of the highest R's.

Justification of the highest R technique rests mostly on the idea that attributes for which communality type values should be high should correlate more highly with other attributes in a battery than would be true for attributes for which the communality type values should be low. This relation should be more nearly true for larger sized batteries of attributes than for smaller sized batteries such as used in the illustration.

There are a few algebraic relations to be considered. Consider two attributes, *i* and *j* . From equation (8.1) when there are no discrepancies of fit the correlation between these two attributes is:

$$r_{ij} = \sum_k a_{ik} a_{jk} \tag{8.24}$$

This correlation can be expressed in trigonometric terms as the scalar product between two vectors.

$$r_{ij} = h_i h_j cos\theta_{ij} \tag{8.25}$$

where h_i and h_j are the lengths of these two vectors and $cos\theta_{ij}$ is the cosine of the angle between these two vectors. When the absolute value of the correlation is considered, as required in the highest R procedure:

Table 8.4
Illustration of "Highest R"

Correlation Matrix

	1	2	3	4	5	6
1	1.00					
2	.49	1.00				
3	-.37	-.28	1.00			
4	.17	.16	-.15	1.00		
5	.04	.02	.03	-.02	1.00	
6	-.03	.02	-.07	.21	-.10	1.00
Highest R	.49	.49	.37	.21	.10	.21

$$\left|r_{ij}\right| = h_i h_j \left|cos\theta_{ij}\right| \tag{8.26}$$

When the two attribute vectors are collinear

$$\left|cos\theta_{ij}\right| = 1$$

so that

$$\left|r_{ij}\right| = h_i h_j$$

If the two attribute vectors have equal length,

$$h_i = h_j$$

and

$$\left|r_{ij}\right| = h_i^2 = h_j^2$$

Thus, in this special case $\left|r_{ij}\right|$ would yield the desired communality type values. In case the two vectors are not of equal length, such as:

$$h_i > h_j$$

then:

$$h_i^2 > \left|r_{ij}\right| > h_j^2$$

so that $\left|r_{ij}\right|$ is too low for one of the h^2's and too high for the other. When the two attribute vectors are not collinear

$$\left|cos\theta_{ij}\right| < 1$$

and there is a tendency for $\left|r_{ij}\right|$ to be less the desired communality type values. The selection of attribute 1 to have a high correlation in absolute value will tend toward $\left|cos\theta_{ij}\right|$ having a value approaching unity.

8.2.2. Squared Multiple Correlations (SMC)'s

The "squared multiple correlation", or SMC, is the most commonly used communality type value used in factor extraction procedures. This coefficient is the squared multiple correlation for an attribute *j* in a multiple, linear regression of that attribute with all other attributes in a battery. Roff (1936) followed by Dwyer (1939) and Guttman (1940) showed that, in perfect cases, the communality of an attribute was equal to or greater than the SMC for that attribute. These developments presumed that the common factor model fit the correlation matrix precisely (there was no lack of fit) and that a population matrix was considered (there were no sampling discrepancies). Dwyer used a determinantal derivation while Guttman used super

matrices. We follow the Guttman form in our developments of this proposition. Guttman (1956) described the SMC's as the "best possible" systematic estimates of communalities; a conclusion justified in terms of a battery of attributes increasing indefinitely without increasing the number of common factors. Following this development, the SMC became widely adopted and incorporated into computer packages.

Standard procedures for computation of the SMC will be considered first. Initially, the case is to be considered when the correlation matrix is nonsingular. This case should cover the majority of factor analytic studies. Each attribute is considered in turn as a dependent variable with the remaining (*n* - 1) attributes being considered as a battery of independent attributes. A linear regression is considered relating attribute *j* to the battery of independent attributes with the smc_j being the squared multiple correlation in this regression. The variance of the discrepancies between the observed values of *j* and regressed values is designated by s_{ej}^2 . From regression theory:

$$smc_j + s_{ej}^2 = 1 \tag{8.27}$$

a super matrix is constructed as illustrated in the following equations.

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R_{jj}} & \boldsymbol{R_{jI}} \\ \boldsymbol{R_{Ij}} & \boldsymbol{R_{II}} \end{bmatrix} \quad \boldsymbol{R}^{-1} = \begin{bmatrix} \boldsymbol{R^{jj}} & \boldsymbol{R^{jI}} \\ \boldsymbol{R^{Ij}} & \boldsymbol{R^{II}} \end{bmatrix} \quad \boldsymbol{R}\boldsymbol{R}^{-1} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \tag{8.28}$$

The dependent attribute is indicated by the subscript *j* while the battery of independent attributes is indicated by the subscript I . Thus, $\boldsymbol{R_{jj}}$ contains unity for the variance of the dependent attribute, $\boldsymbol{R_{jI}}$ contains the correlations of the dependent attribute with the attributes in the independent battery. Similarly, $\boldsymbol{R_{Ij}}$ contains the correlations of the independent attributes with *j* and $\boldsymbol{R_{II}}$ contains the intercorrelations of the independent attributes. From regression theory:

$$smc_j = \boldsymbol{R_{jI}}\boldsymbol{R_{II}^{-1}}\boldsymbol{R_{Ij}} \tag{8.29}$$

The computational problem is to obtain smc_j.

The common computing procedure involves the inverse of the correlation matrix which is illustrated in equation (8.28) in super matrix form. Cells of the inverse matrix are indicated by superscripts. Only the $\boldsymbol{R^{jj}}$ cell is important in the present context. Note, in equation (8.28) that a super identity matrix is indicated as the product of $\boldsymbol{R}$ and its inverse. From this representation:

$$\boldsymbol{R_{jj}}\boldsymbol{R^{jj}} + \boldsymbol{R_{jI}}\boldsymbol{R^{Ij}} = \boldsymbol{I} \tag{8.30}$$

$$\boldsymbol{R_{Ij}}\boldsymbol{R^{jj}} + \boldsymbol{R_{II}}\boldsymbol{R^{Ij}} = \boldsymbol{0} \tag{8.31}$$

From equation (8.31):

$$\boldsymbol{R}^{Ij} = -\boldsymbol{R}_{II}^{-1}\boldsymbol{R}_{Ij}\boldsymbol{R}^{jj}$$

which yields $\boldsymbol{R}^{Ij}$ which may be substituted into equation (8.30) to obtain:

$$\boldsymbol{R}_{jj}\boldsymbol{R}^{jj} - \boldsymbol{R}_{Ij}\boldsymbol{R}_{II}^{-1}\boldsymbol{R}^{jj} = \boldsymbol{I}$$

This equation may be solved to yield the desired equation for $\boldsymbol{R}^{jj}$:

$$\boldsymbol{R}^{jj} = (\boldsymbol{R}_{jj} - \boldsymbol{R}_{jI}\boldsymbol{R}_{II}^{-1}\boldsymbol{R}_{Ij})^{-1} \tag{8.32}$$

Equations for other cells of the inverse matrix may be obtained by similar solutions when desired. With equation (8.29) equation (8.32) becomes:

$$\boldsymbol{R}^{jj} = (\boldsymbol{R}_{jj} - smc_j)^{-1}$$

so that:

$$smc_j = \boldsymbol{R}_{jj} - (\boldsymbol{R}^{jj})^{-1} \tag{8.33}$$

An alternative involves s_{ej}^2:

$$s_{ej}^2 = (\boldsymbol{R}_{jj})^{-1} \tag{8.34}$$

This equation is obtained from (8.27) and (8.34) with $\boldsymbol{R}_{jj}$ being substituted for its equivalent unity in (8.27).

Equation (8.34) is extended to involving all attributes in the battery by defining diagonal matrix $\boldsymbol{S}_e^2$ containing the s_{ej}^2 as the diagonal elements. Equation (8.34) may be expanded to:

$$\boldsymbol{S}_e^2 = [\boldsymbol{Diag}(\boldsymbol{R}^{-1})]^{-1} \tag{8.35}$$

To adjust the diagonal elements of $\boldsymbol{R}$ to having the SMC's , matrix $\boldsymbol{S}_e^2$ must be subtracted from $\boldsymbol{R}$ in accordance to equation (8.27). Let $\boldsymbol{R}_{smc}$ be the correlation matrix with the SMC's in the diagonal cells.

$$\boldsymbol{R}_{smc} = \boldsymbol{R} - \boldsymbol{S}_e^2 \tag{8.36}$$

The two preceding equation provide the basis for computations.

There is trouble in applying the preceding procedure when the correlation matrix is singular since, in this case, the inverse does not exist. A simple modification of this procedure was suggested by Finkbeiner and Tucker (1982). A small positive number, k , is to be added to each diagonal entry of $\boldsymbol{R}$ to yield a matrix $\widetilde{\boldsymbol{R}}$.

$$\widetilde{\boldsymbol{R}} = \boldsymbol{R} + \boldsymbol{kI} \tag{8.37}$$

Define:

$$\widetilde{S}_e^2 = [diag(\widetilde{R}^{-1})]^{-1} \tag{8.38}$$

and compute $\widetilde{R}_{smc}$ by:

$$\widetilde{R}_{smc} = \widetilde{R} - \widetilde{S}_e^2 \tag{8.39}$$

Two statements of approximation follow. As k approaches zero:

$(\widetilde{S}_e^2 - kI)$ approaches S_e^2;

$\widetilde{R}_{smc}$ approaches R_{smc}.

Finkbeiner and Tucker suggested a value of k = .000001 with which discrepancies in the approximations were in the fourth decimal place.

The advantage of the Finkbeiner and Tucker modification is that when R is a true Gramian matrix but is singular due to inclusion of dependent attributes in the battery, matrix $\widetilde{R}$ is not singular so that its inverse is possible. However, use of this modification does not remove the dependency; the procedure permits determination of which attributes form a dependent group. For example, if all scores of a multiple part test are included in the battery along with the total score, all part scores and the total score will be dependent so that their squared multiple correlations will be unity. Other attributes in the battery may not have unit multiple correlations. Use of the Finkbeiner and Tucker procedure will yield, within a close approximation, the unit squared multiple correlations and those that are not unity. Such dependencies should be eliminated from the battery by removing dependent measures such as the total score. These dependencies make what otherwise would be unique factors into common factors thus enlarging the common factor space as well as distorting this space.

There are several types of situations for which the correlation matrix may not be Gramian. One such type of situation is when the individual correlations are based on different samples. Missing scores can produce this situation. Another type situation is when tetrachoric correlations are used in the correlation matrix. The Finkbeiner and Tucker procedure usually will not correct for these situations. A possibility is to use the highest R technique.

The stated inequality between the communality and the squared multiple correlation is considered next. In the development here, the use of super matrices is continued. Assume that a population correlation matrix is being considered without discrepancies of fit so that the following equation holds:

$$R = AA' + U^2$$

where $\boldsymbol{A}$ is an $n \times r$ factor matrix on orthogonal axes. This equation may be expressed in super matrix form for the correlation matrix in equation (8.28):

$$\begin{bmatrix} \boldsymbol{R}_{jj} & \boldsymbol{R}_{jI} \\ \boldsymbol{R}_{Ij} & \boldsymbol{R}_{II} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_j \\ \boldsymbol{A}_I \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_j' & \boldsymbol{A}_I' \end{bmatrix} + \begin{bmatrix} \boldsymbol{u}_j^2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U}_I^2 \end{bmatrix} \tag{8.37}$$

A convenient transformation is considered next. For this transformation all uniquenesses for the attributes in battery I must be greater than zero. Let matrix $\boldsymbol{T}'$ contain eigenvectors and diagonal matrix $\boldsymbol{D}$ contain eigenvalues of the following matrix so that:

$$\boldsymbol{A}_I' \boldsymbol{U}_I^{-2} \boldsymbol{A}_I = \boldsymbol{T}' \boldsymbol{D} \tag{8.38}$$

Matrix $\boldsymbol{T}'$ is square, orthonormal. Following is the transformation.

$$\begin{bmatrix} \boldsymbol{A}_j \\ \boldsymbol{A}_I \end{bmatrix} = \begin{bmatrix} \boldsymbol{1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U}_I^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_j \\ \boldsymbol{A}_I \end{bmatrix} \boldsymbol{T}' \tag{8.39}$$

This transformation rescales the attributes in battery I to having unit uniquenesses and applies an orthogonal transformation on the common factor space. Then:

$$\boldsymbol{A}_I' \boldsymbol{A}_I = \boldsymbol{D} \tag{8.40}$$

and:

$$\begin{bmatrix} \boldsymbol{R}_{jj} & \boldsymbol{C}_{jI} \\ \boldsymbol{C}_{Ij} & \boldsymbol{C}_{II} \end{bmatrix} = \begin{bmatrix} \boldsymbol{1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U}_I^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{R}_{jj} & \boldsymbol{R}_{jI} \\ \boldsymbol{R}_{Ij} & \boldsymbol{R}_{II} \end{bmatrix} \begin{bmatrix} \boldsymbol{1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U}_I^{-1} \end{bmatrix} \tag{8.41}$$

Note that the resealing of attributes in battery I results in covariances for these attributes; however, attribute *j* is not rescaled. Also, the orthogonal transformation leaves the formula for the communality of attribute *j* at:

$$h_j^2 = \sum_k a_{jk}^2 \tag{8.42}$$

The transformation is applied to the uniquenesses by:

$$\begin{bmatrix} \boldsymbol{1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U}_I^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_j^2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U}_I^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U}_I^{-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_j^2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \tag{8.43}$$

The result of this transformation is that:

$$\begin{bmatrix} \boldsymbol{R_{jj}} & \boldsymbol{C_{jI}} \\ \boldsymbol{C_{Ij}} & \boldsymbol{C_{II}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A_j} \\ \boldsymbol{A_I} \end{bmatrix} \begin{bmatrix} \boldsymbol{A'_j} & \boldsymbol{A'_I} \end{bmatrix} + \begin{bmatrix} u_j^2 & 0 \\ 0 & \boldsymbol{U_I^2} \end{bmatrix} \tag{8.44}$$

This transformation leads to a simple form for the inverse of $\boldsymbol{C_{II}}$ which is used in the regression of attribute *j* on the battery I . Regression weights for the attributes in battery I have to be scaled accordingly; however, the squared multiple correlation is not altered by the scaling of battery I .

From multiple regression theory the normal equations relating attribute *j* to battery I are given by:

$$\boldsymbol{C_{II}B_{Ij}} = \boldsymbol{C_{Ij}} \tag{8.45}$$

where $\boldsymbol{B_{Ij}}$ is a column vector of regression weights. The inverse of $\boldsymbol{C_{II}}$ is given by the following equation:

$$\boldsymbol{C_{II}^{-1}} = \boldsymbol{I} - \boldsymbol{A_I}(\boldsymbol{D}+\boldsymbol{I})^{-1}\boldsymbol{A'_I} \tag{8.46}$$

so that the solution for $\boldsymbol{B_{Ij}}$ is:

$$\begin{aligned} \boldsymbol{B_{Ij}} &= \boldsymbol{C_{II}^{-1}C_{Ij}} = \boldsymbol{C_{Ij}} - \boldsymbol{A_I}(\boldsymbol{D}+\boldsymbol{I})^{-1}\boldsymbol{A'_I C_{Ij}} \\ &= \boldsymbol{A'_I A_j} - \boldsymbol{A_I}(\boldsymbol{D}+\boldsymbol{I})^{-1}\boldsymbol{A'_I A_I A'_j} \\ &= \boldsymbol{A_I}[\boldsymbol{I} - (\boldsymbol{D}+\boldsymbol{I})^{-1}\boldsymbol{D}]\boldsymbol{A'_j} \end{aligned} \tag{8.47}$$

The variance of the regressed values of *j* on battery I is designated by $v_{\hat{j}}$ and given by:

$$v_{\hat{j}} = \boldsymbol{B'_{Ij}C_{II}B_{Ij}} = \boldsymbol{B'_{Ij}C_{Ij}}$$

With equations (8.44) and (8.47):

$$\begin{aligned} v_{\hat{j}} &= \boldsymbol{A_j}[\boldsymbol{I} - \boldsymbol{D}(\boldsymbol{D}+\boldsymbol{I})^{-1}]\boldsymbol{A'_I A_I A'_j} \\ &= \boldsymbol{A_j}[\boldsymbol{I} - \boldsymbol{D}(\boldsymbol{D}+\boldsymbol{I})^{-1}]\boldsymbol{D A'_j} \\ &= \boldsymbol{A_j}[\boldsymbol{D} - \boldsymbol{D}(\boldsymbol{D}+\boldsymbol{I})^{-1}\boldsymbol{D}]\boldsymbol{A'_j} \end{aligned} \tag{8.48}$$

Let diagonal matrix $\boldsymbol{W}$ be defined in the present context by:

$$\boldsymbol{W} = \boldsymbol{D} - \boldsymbol{D}(\boldsymbol{D}+\boldsymbol{I})^{-1}\boldsymbol{D} \tag{8.49}$$

Equation (8.48) becomes:

$$v_{\hat{j}} = \boldsymbol{A_j W A'_j} = \sum_k a_{jk}^2 w_k \tag{8.50}$$

where w_k is the k'th diagonal entry in $\boldsymbol{W}$. The value of this entry can be expressed as:

$$w_k = d_k - d_k(d_k + 1)^{-1} d_k$$

which becomes with algebraic manipulation:

$$w_k = d_k/(d_k + 1) \tag{8.51}$$

With the variance of standardized measures of attribute *j* being unity, $\boldsymbol{R}_{jj}$ equaling unity, the variance, $v_{\hat{j}}$, of the regressed values of attribute *j* , equals the squared multiple correlation of attribute *j* on battery I .

$$smc_j = v_{\hat{j}} \tag{8.52}$$

The important relation with which this development is concerned compares the communality of attribute *j* with the squared multiple correlation of this attribute. For this comparison a difference is taken:

$$h_j^2 - smc_j = \sum_k a_{jk}^2 - \sum_k a_{jk}^2 w_k = \sum_k a_{jk}^2 (1 - w_k) \tag{8.53}$$

From equation (8.51):

$$(1 - w_k) = 1/(d_k + 1) \tag{8.54}$$

Note that the diagonal entries, d_k of matrix $\boldsymbol{D}$ are the sums of squares of the entries in columns of factor matrix $\boldsymbol{A}_\mathbf{I}$, as per equation (8.40), so that these d_k's must be positive. The possibility of a zero value of a d_k is discarded since this would imply a column of zero factor loadings which is ruled out by the definition that the factor matrix have a rank equal to its column order. Then, for all $k = 1, r$:

$$d_k > 0$$

and from equation (8.54)

$$0 < (1 - w_k) < 1 \tag{8.55}$$

The possibility of (1 - w_k) equaling 0 will be discussed later. Equations (8.53) and (8.55) lead to the following important inequality:

$$h_j^2 > smc_j \tag{8.56}$$

An illustration of the squared multiple correlations for the perfect case is given in Table 8.2 . The smc's for all six attributes are markedly less than the major domain variances which are the theoretic communalities for this case. The relation between the communalities and the squared multiple correlations is dependent on the values of the d_k's . An important effect is the relation of the d_k's to the battery size. Since each d_k is the sum of squares of scaled factor weights, as the

battery size increases without increasing the number of factors, each d_k will increase which will lead to a decrease in the value of (1 - w_k). This will lead to a reduction in the differences between the communalities and the squared multiple correlations so that the squared multiple correlations will become better estimates of the communalities as the battery size is increased. A limiting condition pointed out by Guttman (1956) is for the battery size to approach infinite; then the d_k's will approach infinity and (1 - w_k)'s will approach zero so that the difference between the communalities and the squared multiple correlations also will approach zero.

Application of the foregoing inequality in practice is accompanied by some uncertainties. First, inclusion of lack of fit of the model has unknown effects on squared multiple correlations. This is in addition to the idea that no longer are there fixed ideal communalities. Analyses of correlation matrices obtained from samples introduce further uncertainties. A well known effect is that the squared multiple correlation obtained from a sample of observations is biased upwards. This might lead to a violation of the inequality in case the communality is not biased similarly, the possible bias of the communality not being well known. However, use of squared multiple correlations has yielded satisfactory to good results in many practical applications.

8.2.3. Iterated Communalities

A procedure followed sometimes is to iterate the communality values. Such a procedure starts with trial communality values, extracts a factor matrix by one of the factor extraction techniques and computes output communalities from this matrix which are substituted into the diagonal entries of the correlation matrix as the next trial communalities. Each iteration starts with trial communalities and ends with output communalities which become next trial communalities. This procedure is continued until there are minimal changes from trial communalities to output communalities. Results of this type procedure for several methods of factor extraction are illustrated in Tables 8.2 and 8.3. The general idea is that these iterations lead to communality values with which the extracted factor matrix better fits the input correlation matrix.

As seen in Table 8.2 for the perfect case in the population, the iterated communalities settle to equaling the theoretic values of the major domain variances. The scheme of iterating the communalities works in this case. However, consider Table 8.3 which presents illustrations of iterated communalities when discrepancies of fit have been included in the correlation matrix. The output communalities do not equal the major domain variances nor do the communalities obtained by different methods of factor extraction equal each other. Introduction of sampling discrepancies will produce even more differences between obtained communality values and any theoretic values and among the obtained values from different methods of factor extraction. Considerable problems are raised for practical applications. Information concerning these problems might be obtained from extensive simulation, Monte Carlo type studies. Use of the

Tucker, Koopman, Linn simulation procedure (1966 and described in Chapter 3) is recommended so that obtained results may be compared with input major domain matrices. Preliminary results of such studies indicate that the iterated communalities procedure works better than more approximate procedures only for large samples and few factors compared with the battery sizes.

Convergence of the iteration procedure may be a problem. Often convergence is slow so that some techniques to speed convergence could be advisable. For methods of factor extraction, including the principal factors technique and the maximum likelihood factor analysis method, alternate computing routines have been developed for which the computing time is greatly reduced. Of the factor extraction methods illustrated in Tables 8.2 and 8.3, proof of convergence exists only for the principal factors technique and the maximum likelihood method. However, experience indicates that convergence does occur for the other techniques.

Another area of problems with iterated communalities consists of generalized Heywood cases. Initial work in this area was by H. B. Heywood (1931) who indicated that the rank of a correlation matrix using a limited number of common factors may imply either a communality greater than unity or a negative communality. The concern for iterated communalities is that one or more of the communalities becomes greater than unity which is not permissible since such a communality implies a negative uniqueness. Special procedures, to be discussed subsequently, are required to avoid this situation. Unfortunately, this case is ignored some times.

8.3. Centroid Method of Factor Extraction

The centroid method of factor extraction is presented partly for its historical value and partly for some very useful techniques used in special situations. Thurstone (1935, 1947) developed the centroid method of factor extraction in the 1930's for his major factor analytic studies. This time was prior to the age of electronic computers and used mechanical desk calculators to perform the needed computations. There was a need for simple procedures and the centroid method filled this need. Today, with the availability of electronic computers, much more computationally complex procedures are readily available. However, the sign change procedure and criterion L, to be described later, are very useful in obtaining a measure of complexity of relations in a covariance or correlation matrix.

Thurstone developed the centroid method using a geometric view involving a configuration of vectors to represent the attributes. The centroid vector is the mean vector through which a centroid axis is passed. This is a centroid factor with the orthogonal projections of attribute vectors on it being the factor weights. One factor is extracted at a time and a residual correlation matrix is computed. The next factor is extracted from the residual matrix and a further residual matrix is computed. Thus, there is a series of residual correlation matrices with a

centroid factor extracted from each residual matrix. One problem is that the configuration of vectors frequently has vectors splayed out in many directions. This is true of many correlation matrices among personality traits and is almost always true of residual correlation matrices. As a consequence there is a need to reverse some of vectors to obtain a more stable centroid. This is the sign change procedure. Factor extraction is continued until the residual correlations become small enough to be ignored and resulting factors have only quite small factor weights (in absolute value).

Rather than Thurstone's geometric approach, an algebraic approach is used here involving the general theory of matrix factoring presented earlier in this chapter. Let matrix $\boldsymbol{G}$ be any of the original correlation matrix and residual correlation matrices. One centroid factor is to be extracted from $\boldsymbol{G}$ and a new matrix of residual correlations is to be computed. A major restriction is that weights, w_j , in the single column of matrix $\boldsymbol{W}$ are to be either +1 or -1. The sign change procedure is used to reverse signs of weights for selected attributes. The signs are changed so as to maximize coefficient $\boldsymbol{P}$ defined in equation (8.5). After the sign change procedure, the sum of the absolute values of factor weights equals $\boldsymbol{F}$, the square root of $\boldsymbol{P}$. Thus, the centroid method combined with the sign change procedure tends to maximize the sum of absolute values of the factor weights. However, there may be several maxima and there is no guarantee that an absolute maximum is obtained. If an absolute maximum is not obtained in one factor extracted, a factor related to the absolute maximum is likely to be obtained from the next matrix of residual correlations.

There are several matters to be considered when the weights are restricted to +1 or -1 . First, it will be seen subsequently that the sign of the weight for an attribute has no effect on the contribution of the diagonal entry on coefficient $\boldsymbol{P}$. As a result, the diagonal entries in $\boldsymbol{G}$ are eliminated from the sign change computations. Table 8.5 presents a correlation matrix with zeros in the diagonal. Such a matrix may be symbolized by $\widetilde{\boldsymbol{G}}$ and defined by:

$$\widetilde{\boldsymbol{G}} = \boldsymbol{G} - \boldsymbol{Diag}(\boldsymbol{G}) \tag{8.57}$$

so that

$$\widetilde{\mathrm{g}}_{jk} = \mathrm{g}_{jk} \text{ for } j \neq k \text{ and}$$

$$\widetilde{\mathrm{g}}_{jj} = 0$$

Then:

$$\boldsymbol{G} = \widetilde{\boldsymbol{G}} + \boldsymbol{Diag}(\boldsymbol{G}) \tag{8.58}$$

Table 8.5
Illustration of Determination of Centroid Factor with Sign Change

Correlation Matrix with Zeros in Diagonal

	1	2	3	4	5	6	
1	--	.49	-.37	.17	.04	-.03	
2	.49	--	-.28	.16	.02	.02	
3	-.37	-.28	--	-.15	.03	-.07	
4	.17	.16	-.15	--	-.02	.21	
5	.04	.02	.03	-.02	--	-.10	
6	-.03	.02	-.07	.21	-.10	--	
D(R)	.49	.49	.37	.21	.10	.21	sum=1.87
W_1'	+1	+1	+1	+1	+1	+1	
$\widetilde{Q}_1$,	.30	.41	-.84	.37	-.03	.03	$\widetilde{p}_1$=.24
c_1			-2				
W_2'	+1	+1	-1	+1	+1	+1	
$\widetilde{Q}_2$,	1.04	.97	-.84	.67	-.09	.17	$\widetilde{P}_2$=3.60
c_2					-2		
W_3'	+1	+1	-1	+1	-1	+1	
$\widetilde{Q}_3$,	.96	.93	-.90	.71	-.09	.37	$\widetilde{p}_3$=3.96
Q'	1.45	1.42	-1.27	.92	-.19	.58	p=5.83
A'	.60	.59	-.53	.38	-.08	.24	

F=$\sqrt{P}$ =2.4145.

$\sum_K a_k w_k$ =2.42.

There will be several trials so that a trial number subscript will be used. Let $\boldsymbol{W_t}$ be the weight matrix for trial t with entries w_{tk} . Matrix $\boldsymbol{Q_t}$ is obtained from equation (8.4).

$$\boldsymbol{Q_t = GW_t}$$

and using equation (8.58):

$$\boldsymbol{Q_t = \widetilde{G}\, W_t + [Diag(G)]W_t}$$

Define $\widetilde{\boldsymbol{Q}}_t$ by:

$$\widetilde{\boldsymbol{Q}}_t = \widetilde{\boldsymbol{G}}\, \boldsymbol{W}_t \ . \tag{8.59}$$

Then:

$$\boldsymbol{Q_t = \widetilde{Q}_t + [Diag(G)]W_t} \ . \tag{8.60}$$

By equations (8.5) and (8.60):

$$\boldsymbol{P_t = W_t' Q_t = W_t' \widetilde{Q}_t + W_t'[Diag(G)]W_t} \ . \tag{8.61}$$

Define $\widetilde{\boldsymbol{P}}_t$ by:

$$\widetilde{\boldsymbol{P}}_t = \boldsymbol{W_t'}\widetilde{\boldsymbol{Q}}_t \ . \tag{8.62}$$

Equation (8.63) gives an interesting relation:

$$\boldsymbol{W_t'[Diag(G)]W_t = \sum_j w_{tj}^2 g_{jj} = \sum_j g_{jj}} \ . \tag{8.63}$$

Remember that the square of either +1 or -1 is +1. Then:

$$\boldsymbol{P_t = \widetilde{P}_t + \sum_j g_{jj}} \ . \tag{8.64}$$

This results supports the statement that the contribution of the diagonal entries of $\boldsymbol{G}$ is independent of the signs of the weights. From equation (8.6):

$$\boldsymbol{F_t = \sqrt{P_t}} \ . \tag{8.65}$$

From equation (8.7)

$$\boldsymbol{A_t = Q_t(1/F_t)} \ . \tag{8.66}$$

Then from equation (8.23)

$$\boldsymbol{F_t = W_t' A_t = \sum_k w_{tk} a_{tk}} \tag{8.67}$$

Changes from trial t to trial (t+1) are considered next. At each step, the sign of the weight for only one attribute, *i* , is to be changed. When $w_{ti} = +1,$ let $c_{ti} =$ -1, and when $w_{ti} =$ -1, let $c_{ti} = +1$; that is, the sign of c_{ti} is opposite to the sign of w_{ti}. Then:

$$w_{(t+1)i} = w_{ti} + c_{ti} \ ; \tag{8.68}$$

and, for $k \neq$ i:

$$w_{(t+1)k} = w_{tk} \ . \tag{8.69}$$

For all attributes, *j* :

$$\widetilde{q}_{(t+1)j} = \sum_h \widetilde{g}_{jh} w_{(t+1)h} = \sum_{k \neq i} \widetilde{g}_{jk} w_{(t+1)k} + \widetilde{g}_{ji} w_{(t+1)i} \ . \tag{8.70}$$

Substitution for equations (8.68) and (8.69) yields:

$$\begin{aligned} \widetilde{q}_{(t+1)j} &= \sum_{k \neq i} \widetilde{g}_{jk} w_{tk} + \widetilde{g}_{ji} w_{ti} + \widetilde{g}_{ji} c_{ti} \\ &= \sum_h \widetilde{g}_{jh} w_{th} + \widetilde{g}_{ji} c_{ti} \ . \end{aligned}$$

This yields the important result:

$$\widetilde{q}_{(t+1)j} = \widetilde{q}_{tj} + \widetilde{q}_{ji} c_{ti} \ ; \tag{8.71}$$

and, since $\widetilde{\mathbf{g}}_{ii}$ equals zero:

$$\widetilde{q}_{(t+1)i} = \widetilde{q}_{ti} \ . \tag{8.72}$$

Interpretation of equation (8.62) in terms of trial (t +1) yields:

$$\widetilde{\boldsymbol{P}}_{(t+1)} = \sum_h w_{(t+1)h} \widetilde{q}_{(t+1)h} = \sum_{k \neq i} w_{(t+1)k} \widetilde{q}_{(t+1)k} + w_{(t+1)i} \widetilde{q}_{(t+1)i} \ .$$

Substitution from equations (8.68) and (8.69) yields:

$$\widetilde{\boldsymbol{P}}_{(t+1)} = \sum_{k \neq i} w_{tk} \widetilde{q}_{tk} + c_{ti} \sum_{k \neq i} \widetilde{\mathbf{g}}_{ki} w_{tk} + w_{ti} \widetilde{q}_{ti} + c_{ti} \widetilde{q}_{ti}$$

which reduces to:

$$\widetilde{\boldsymbol{P}}_{(t+1)} = \widetilde{\boldsymbol{P}}_t + 2 c_{ti} \widetilde{q}_{ti} \tag{8.73}$$

The sign change procedure utilizes relations developed in the preceding paragraphs. This procedure is illustrated in Table 8.5 which gives a correlation matrix with zeros in the diagonal elements. This is matrix $\widetilde{\boldsymbol{G}}$. One general requirement is that the diagonal entries to be used subsequently must be all positive which is not necessarily true for residual correlation matrices when SMC's are used as communality like values. A procedure followed by Thurstone appears to

work very satisfactorily; that is, to use the highest R values for every correlation matrix and residual correlation matrix. Row D(R) of Table 8.5 contains the highest R (in absolute value) for the given correlation matrix. The sum of the entries in this row is given at the right. At trial 1, the weights in row W_1' are all taken to be +1 and the first trial $\widetilde{\boldsymbol{Q}}_1'$ contains the column sums of matrix $\widetilde{\boldsymbol{G}}$. Coefficient $\widetilde{\boldsymbol{P}}_1$ is the sum of the entries in row $\widetilde{\boldsymbol{Q}}_1'$. These are the preliminary steps before starting the sign change procedure.

Since, by equation (8.64), coefficient $\boldsymbol{P}_t$ is $\widetilde{\boldsymbol{P}}_t$ plus the sum of the entries to be inserted in the diagonal of $\boldsymbol{G}$ and this sum is necessarily positive, increasing $\widetilde{\boldsymbol{P}}_t$ necessarily increases $\boldsymbol{P}_t$. The objective of the sign change procedure is to increase $\widetilde{\boldsymbol{P}}_t$ as much as possible. In each step, or trial, the attribute is selected whereby $\widetilde{\boldsymbol{P}}_t$ is increased most to $\widetilde{\boldsymbol{P}}_{(t+1)}$. By equation (8.73), the change from $\widetilde{\boldsymbol{P}}_t$ to $\widetilde{\boldsymbol{P}}_{(t+1)}$ equals $2c_{ti}\widetilde{q}_{ti}$. For this change to be an increase, c_{ti} and $\widetilde{q}_{ti}$ must have the same algebraic sign. Note that c_{ti} is defined to have the opposite sign to w_{ti} . Therefore, w_{ti} and $\widetilde{q}_{ti}$ must have opposite signs. The strategy is to select that attribute for which w_{ti} and $\widetilde{q}_{ti}$ have opposite signs and $\widetilde{q}_{ti}$ is the largest in absolute value satisfying the signs condition. In the example in Table 8.5, since all weights in row W_1' are positive, the attribute with the most negative value in row $\widetilde{\boldsymbol{Q}}_1'$ is attribute 3 with an entry of -.84 . Consequently, attribute 3 was chosen and a change coefficient c_{13} was set at -2 , the sign being opposite to the sign of weight w_{13}.

After having established rows W_1', $\widetilde{Q}_1'$ and selected the attribute for the sign change, the next series of steps is to make the changes accompanying this first sign change. In row W_2' the signs of the weights for all attributes except attribute 3 remain unchanged at +1. The weight for selected attribute 3 is changed to -1 . These weights in row W_2' are the results of application of equations (8.68) and (8.69). Next, row $\widetilde{Q}_2'$ is established as per equations (8.71) and (8.72). For an example consider the first entry in row $\widetilde{Q}_2'$, $\widetilde{q}_{21}$. The value of 1.04 equals $\widetilde{q}_{21}$ plus c_{13} times $\widetilde{g}_{31}$; that is:

$$1.04 = .30 + (-2)(-.37)$$

These computations are continued for all attributes except the selected attribute 3 for which $\widetilde{q}_{23}$ equals $\widetilde{q}_{13}$, a value of -.84 . Coefficient $\widetilde{\boldsymbol{P}}_2$ can be computed two ways: one by obtaining the sum of products of entries in rows W_2' and $\widetilde{\boldsymbol{Q}}_2$, as per equation (8.62); and by equation (8.73). For the second method:

$$3.60 = .24 + 2(-2)(-.84)$$

When using hand computing with the aid of a desk calculator, this value should be computed both ways to provide a check.

A selection is made next of the second attribute to have the sign of its weight changed. Rows W_2' and $\widetilde{Q}_2'$ are inspected for those attributes having entries with opposite signs and that attribute is selected for which the absolute value of the entry in row $\widetilde{Q}_2'$ is largest. In the example, only for attribute 5 are the entries in rows W_2' and $\widetilde{Q}_2'$ opposite in sign; consequently, this attribute is selected to have its sign changed and a -2 is inserted into line c_2 for attribute 5.

Computations for trial 3 from the results in trial 2 are similar to those carried out in going from trial 1 to trial 2. The signs of weights in row W_3' are the same as those in row W_2' with exception of w_{35} , this being the weight having its sign changed. Entries in row $\widetilde{Q}_3'$ are obtained from row 5 of $\boldsymbol{G}$, the weights in row W_2' , and the entries in row $\widetilde{Q}_2'$. Coefficient $\widetilde{P}_3$ is obtained from $\widetilde{P}_2$, c_{25} , $\widetilde{q}_{25}$ as well as from the sum of products between entries in rows W_3' and $\widetilde{Q}_3'$.

Inspection of rows W_3' and $\widetilde{Q}_3'$ reveals that the signs of the entries in these two rows are in agreement. There are no more signs to be changed. Row W_3' is the final weight vector. A final row Q' is to be computed by adding in the diagonal entries of $\boldsymbol{G}$ with the proper signs, see equation (8.60). The entry in row $\boldsymbol{Q}'$ of Table 8.5 for the first attribute is:

$$1.45 = .96 + (.49)(+1) .$$

For attribute 3:

$$-1.27 = -.90 + (.37)(-1) .$$

For the final coefficient $\boldsymbol{P}$ see equation (8.64). The final $\boldsymbol{P}$ equals the final $\widetilde{P}_t$ plus the sum of the diagonal entries as well as equaling the sum of the products of the entries in the final $\boldsymbol{W}$ and $\boldsymbol{Q}$. The value for the example is:

$$5.85 = 3.96 + 1.87 .$$

Factor weights in row $\boldsymbol{A}'$ are obtained by dividing the entries in row $\boldsymbol{Q}'$ by F which equals the square root of $\boldsymbol{P}$. For the example the factor weight for attribute 1 is:

$$.60 = 1.45/2.4145 .$$

As indicated in equation (8.67), $\boldsymbol{F}$ equals the sum of products between the weights in the final row $\boldsymbol{W}'$ of the example and the factor weights in row $\boldsymbol{A}'$.

The example in Table 8.5 is too small to illustrate one kind of difficulty encountered with larger correlation matrices. Sometimes when the sign has been changed for one attribute and after a number of further changes the $\widetilde{q}$ for that attribute becomes positive which is, now, opposite to the sign of the weight. The sign of the weight has to be changed back to a +1 . In this case the change coefficient, c , is a +2 . Then equations (8.62), (8.71), (8.72), and (8.73) provide the means for making this reverse change.

It is of interest to note now that the use of vectors $\widetilde{\boldsymbol{Q}}$ without the diagonal entries provides a more effective sign change than would the use of vectors $\boldsymbol{Q}$ which includes the diagonal entries. For example, in trial 2 of the example in Table 8.5, adding in the diagonal entry of .10 to $\widetilde{q}_{25}$ of -.09 yields a value of .01 which agrees in sign with the weight w_{25}. If this value of .01 is compared with the weight of +1 , then the sign for attribute 5 would not be changed and the increase in $\boldsymbol{P}$ would have been missed.

We return to the maximization proposition for coefficient $\boldsymbol{P}$. With each trial resulting in an increase in $\boldsymbol{P}$, a maximum should be reached since the value can not exceed the sum of absolute values of entries in the correlation matrix. However, there is no guarantee that there is only one maximum nor that an obtained result yields the largest maximum. There is one statement possible using the vectors $\widetilde{\boldsymbol{Q}}$ without the diagonal entries: it is not possible to increase $\boldsymbol{P}$ further after reaching a solution by changing the sign of the weight for only one attribute. To go from one maximum to another must involve the changing of the weight signs for two or more attributes. Each solution, thus, involves at least a local maximum. We observe that when a major maximum has been missed, a solution involving the factor for this maximum is likely to appear in the results for the next factor extracted from the ensuing residual matrix.

Extraction of centroid factors from the correlation matrix among nine mental tests is given in Table 8.6 . This correlation matrix was given in Chapter 1, Table 1.1, which includes the names of tests. For the original correlation matrix given at the top of the first page of this table, row D(R) contains the highest correlation for each attribute. These values will be substituted for the unities in the diagonal. Since all correlations are positive, all sign change weights are +1 and the entries in row $\boldsymbol{Q}'$ are the column sums of the correlation matrix, the diagonal unity having been replaced by the entry in row D(R) . Coefficients $\boldsymbol{P}$ and $\boldsymbol{F}$ are given along with the factor weights in row $\mathbf{A}_1'$ for the first centroid factor. Coefficient L is a very useful criterion to measure the structure indicated in the correlation matrix. In general:

$$L = \widetilde{\boldsymbol{P}} \,/[\sum_{j} \sum_{k \neq j} |g_{jk}|] \tag{8.74}$$

This criterion may be used in decisions on the number of factors to be extracted from a correlation matrix, more about this later. Since all original correlations are positive in the example, L for this matrix is unity.

The first factor residual matrix is given at the bottom of the first page of Table 8.6. The diagonal entries are residuals from the substituted diagonals of the correlation matrix. The column sums including these residual diagonals are zero within rounding error. These sums provide an excellent check on hand computations. Revised diagonal entries are given in row D(R) . The signs were changed for attributes 4, 5 , 6, and 8 after which row $\boldsymbol{Q}'$ was obtained

Table 8.6
Extraction of Centroid Factors from
Correlation Matrix among Nine Mental Tests

Original Correlation Matrix

	1	2	3	4	5	6	7	8	9
1	1.000								
2	.499	1.000							
3	.394	.436	1.000						
4	.097	.007	.292	1.000					
5	.126	.023	.307	.621	1.000				
6	.085	.083	.328	.510	.623	1.000			
7	.284	.467	.291	.044	.114	.086	1.000		
8	.152	.235	.309	.319	.376	.337	.393	1.000	
9	.232	.307	.364	.213	.276	.271	.431	.489	1.000
D(R)*	.499	.499	.436	.621	.623	.623	.467	.489	.489
Q’	2.368	2.556	3.157	2.724	3.089	2.946	2.577	3.099	3.072
			P=25.588	F=5.058	L=1.000				
A_1’	.468	.505	.624	.539	.611	.582	.509	.613	.607

First Residual Correlation Matrix

	1	2	3	4	5	6	7	8	9
1	.290								
2	.262	.244							
3	.102	.121	.046						
4	-.155	-.265	-.044	.331					
5	-.160	-.286	-.074	.292	.250				
6	-.188	-.211	-.035	.196	.267	.284			
7	.046	.210	-.027	-.230	-.197	-.211	.207		
8	-.135	-.075	-.073	-.011	.002	-.020	.081	.114	
9	-.052	.000	-.015	-.114	-.095	-.083	.122	.117	.120
Sum	.000	.000	.001	.000	-.001	-.001	.001	.000	.000
D(R)*	.262	.286	.121	.292	.292	.267	.230	.135	.122
Q’	1.257	1.715	.528	-1.578	-1.665	-1.439	1.137	-.191	.351
			P=9.862	F=3.140	L=.859				
A_2’	.400	.546	.168	-.503	-.530	-.458	.362	-.061	.112

* Highest R.

Table 8.6 (Continued)
Extraction of Centroid Factors from
Correlation Matrix among Nine Mental Tests

Second Residual Correlation Matrix

	1	2	3	4	5	6	7	8	9
1	.102								
2	.044	-.013							
3	.034	.029	.092						
4	.046	.009	.040	.040					
5	.052	.004	.015	.026	.011				
6	-.004	.039	.042	-.034	.024	.057			
7	-.099	.012	-.088	-.048	-.005	-.045	.099		
8	-.110	-.041	-.063	-.041	-.030	-.048	.103	.131	
9	-.097	-.061	-.034	-.058	-.036	-.032	.081	.124	.109
Weighted Sum	.000	.000	-.001	-.002	-.001	.001	.000	.001	.000
D(R)*	.110	.061	.088	.058	.052	.048	.103	.124	.124
Q’	-.590	-.276	-.433	-.293	-.245	-.238	.561	.685	.645
			P=3.967	F=1.992	L=.941				
A_1’	-.296	-.139	-.217	-.147	-.123	-.120	.281	.344	.324

Third Residual Correlation Matrix

	1	2	3	4	5	6	7	8	9
1	.023								
2	.003	.042							
3	-.030	-.001	.041						
4	.003	-.011	.008	.036					
5	.016	-.013	-.012	.008	.037				
6	-.040	.022	.016	-.052	.010	.033			
7	-.016	.051	-.027	-.007	.030	-.011	.024		
8	-.009	.006	.012	.009	.012	-.006	.006	.006	
9	-.001	-.016	.037	-.010	.004	.007	-.010	.012	.019
Weighted Sum	-.001	-.001	.000	.000	.000	.001	.000	.000	.000
D(R)*	.040	.051	.037	.052	.030	.052	.051	.012	.037
Q’	-.119	.108	.082	-.109	.037	.180	.113	.054	.082
			P=.883	F=.940	L=.483				
A_1’	-.126	.115	.088	-.116	.039	.192	.120	.057	.088

* Highest R.

along with coefficients $\boldsymbol{P}$, $\boldsymbol{F}$, and L . For this matrix the sign change did not result in all positive contributions by the off-diagonal entries so that L is less than unity. Factor weights for the second factor in row A_2' are obtained from row $\boldsymbol{Q}'$ and coefficient $\boldsymbol{F}$.

The second factor residual correlation matrix is given at the top of the second page of Table 8.6. This matrix is obtained from the first factor residual correlation matrix with substituted diagonal entries and the second factor weights. The row of weighted sums uses the just preceding sign change weights as multipliers of the residual correlations. Again, these sums should equal zero within rounding error which provides a check on hand computations. See equation (8.22) for the basis for these zero weighted sums. Computation of the third factor weights progresses in the same manner as the computations for preceding factors.

The third factor residual correlation matrix is given at the bottom of the second page of Table 8.6. Computations for this matrix and the fourth factor weights are similar to the computations for preceding factors.

Decisions as to the number of factors to extract by the centroid method had only sketchy bases to support these decisions. Residual matrices were inspected for the magnitudes of the entries and factor extraction was stopped when these residuals were small so that they might be ignored. Table 8.7 gives three coefficients which might be used including the largest residual correlation in each residual matrix. For the example of nine mental tests the largest third factor residual was .052 and a decision might be made that this was small enough to be ignored. By this reason, the three factor solution would be accepted. Another coefficient which might be considered is the criterion L . Note in the example this coefficient is relatively high for the first three matrices and factors but drops substantially for the third factor residuals. In this example the low value of L for the third factor residuals could be taken as an indication to accept the three factor solution. A third criterion used by some analysts was the magnitude of the factor weights obtained. Frequently, when the largest factor weight was less than a value such as .2, a factor was not accepted. In the example this criterion would, again, indicate a three factor solution. Beyond such criteria as the foregoing, trial transformations of factors was considered and that number of factors accepted which led to the most meaningful solution. Some individuals advocated using an extra "residual" factor to help clean up transformed factors.

The centroid method of factor extraction has been presented partly for its historic value and partly to provide some useful techniques. For example, the sign change technique with criterion L has been found useful in testing the simplicity of special covariance matrices in research on the dimensionality of binary data such as item responses. Undoubtedly, there may be other cases of special covariance matrices for which a simple criterion related to complexity of the matrix would be helpful.

Table 8.7

Illustration of Indices used for Number of Factors

in Centroid Factor Extraction

	Largest Correlation	Criterion L	Largest Factor Loading*
Original Correlation matrix, Factor 1	.623	1.000	.624
First Residual Matrix, Factor 2	.292	.859	.546
Second Residual Matrix, Factor 3	.124	.941	.344
Third residual Matrix, Factor 4	.052	.483	.192

* In absolute value

8.4. Group Centroid Method of Factor Extraction

The group centroid method of factor extraction provides a simple technique which may be applied in special situations. In particular, a partitioning of the attributes into clusters with high within cluster correlations should be possible. The correlations between clusters should be low. For an example consider the correlation matrix in Table 8.8. Attributes 1, 2, and 4 intercorrelate relatively highly and have low correlations with the remaining attributes. These attributes are listed first in Table 8.9. Ignore the diagonal entries for the present. Attributes 5 and 6 have a high correlation while attribute 3 correlates negatively with them. A sign reversal of the correlations of attribute 3 produces moderately high, positive correlations with attributes 5 and 6. In making such a sign reversal, the signs of the correlations in both the rows and the columns are reversed for the attribute. Note that the diagonal entry remains positive since its sign is reversed twice. This operation yields the second cluster in Table 8.9. As seen in this table, there are two clusters of attributes with relatively high intercorrelations within the clusters and relatively low correlations between clusters. This is the type of situation for which the group centroid method of factor extraction could be appropriate.

For the operation of the group centroid method of factor extraction return to the original correlation matrix in Table 8.8; the clustered correlation matrix provided a guiding step but will not be used in the computations. A first consideration is the diagonal entries. The extracted factors will depend on the intercorrelations of the attributes in the clusters, these intercorrelations forming relatively small matrices. There is a problem in using SMC's as diagonal entries. As shown earlier, the SMC's tend to be smaller than desired for small matrices. For example, the SMC for attribute 1 in our example is .274 which appears small for the intercorrelations of the attributes in the example. In contrast, the "highest R" of .44 appears appropriate. Further, for computations using a desk calculator, the "highest R" technique is much more convenient. Thus, in general with the group centroid method of factor extraction, the "highest R" would be the preferred value to be inserted into the diagonal of the correlation matrix. This has been done for the middle matrix of Table 8.8 and were given in Table 8.9. Computations of the factor matrix will progress from the middle matrix of Table 8.8.

Equations used in the group centroid method of factor analysis method are repeated here for convenience. Matrix $\boldsymbol{G}_1$ is the correlation matrix with desired diagonal entries such as the middle matrix of Table 8.8.

$$\boldsymbol{Q} = \boldsymbol{G}_1\boldsymbol{W} \, . \tag{8.4}$$

$$\boldsymbol{P} = \boldsymbol{W}'\boldsymbol{Q} \, . \tag{8.5}$$

Table 8.8

Correlation matrices for

Illustration of Group Centroid Method of Factor Extraction

Correlation Matrix with Unities in Diagonal

	1	2	3	4	5	6
1	1.00	.44	-.04	.43	.04	.05
2	.44	1.00	-.06	.38	.07	.09
3	-.04	-.06	1.00	-.06	-.33	-.35
4	.43	.38	-.06	1.00	.09	.08
5	.04	.07	-.33	.09	1.00	.40
6	.05	.09	-.35	.08	.40	1.00

Correlation Matrix with Highest R's in Diagonal

	1	2	3	4	5	6
1	.44	.44	-.04	.43	.04	.05
2	.44	.44	-.06	.38	.07	.09
3	-.04	-.06	.35	-.06	-.33	-.35
4	.43	.38	-.06	.43	.09	.08
5	.04	.07	-.33	.09	.40	.40
6	.05	.09	-.35	.08	.40	.40

Residual Correlation Matrix

	1	2	3	4	5	6
1	-.01	.01	-.01	.01	-.01	.00
2	.01	.02	.00	-.03	-.01	.01
3	-.01	.00	.03	.01	.02	.01
4	.01	-.03	.01	.03	.01	-.01
5	-.01	-.01	.02	.01	.01	.01
6	.00	.01	.01	-.01	.01	.00

Table 8.9

Clustered Correlation Matrix for

Illustration of Group Centroid Method of Factor Extraction

Correlation Matrix with Highest R's in Diagonal

	1	2	4	5	6	-3
1	.44	.44	.43	.04	.05	.04
2	.44	.44	.38	.07	.09	.06
4	.43	.38	.43	.09	.08	.06
5	.04	.07	.09	.40	.40	.33
6	.05	.09	.08	.40	.40	.35
-3	.04	.06	.06	.33	.35	.35

$$\boldsymbol{F}\boldsymbol{F}' = \boldsymbol{P} \, . \tag{8.6}$$

$$\boldsymbol{A} = \boldsymbol{Q}(\boldsymbol{F}^{-1})' \, . \tag{8.7}$$

$$\boldsymbol{G_2} = \boldsymbol{G}_1 - \boldsymbol{A}\boldsymbol{A}' \tag{8.8}$$

where $\boldsymbol{G_2}$ is the matrix of residual correlations.

$$\boldsymbol{W}'\boldsymbol{A} = \boldsymbol{F} \, . \tag{8.23}$$

Reference will be made to these equations during the discussion of the computing procedures.

Computations for the example are given in Table 8.10. Weight matrix $\boldsymbol{W}$ is the starting point. This matrix reflects the clusters which have been determined during inspection of the correlation matrix. There is a column of $\boldsymbol{W}$ for each cluster, or group, and contains weights of +1 or -1 for attributes in the cluster and weights of 0 for attributes not in the cluster. Weights of +1 are assigned to attributes which are not reflected in sign and weights of -1 for attributes reflected in sign. In the example, the first cluster was composed of attributes 1, 2, and 4 without any reflections in sign. Consequently, the first column of weight matrix in Table 8.10 has +1's for these three attributes and 0's for the other attributes. The second column of the weight matrix is for the second cluster with +1's for attributes 5 and 6 and a -1 for attribute 3 since this attribute was reflected in sign to form the cluster. The weight of -1 performs this reflection. In this second column of weights, 0's are recorded for attributes 1, 2 , and 4 which are not in the second cluster. In general, the weight matrix reflects the clusters found during inspection of the correlation matrix.

Once the weight matrix has been established, computations follow the given equations. Matrix $\boldsymbol{Q}$ is computed by equation (8.4). Since the weight matrix contains only +1's , -l's , and 0's, this matrix multiplication involves only addition, or subtraction, of entries in the correlation matrix and may be accomplished quite readily with a desk calculator. Matrix $\boldsymbol{P}$ is obtained by equation (8.5) which, again, involves only addition or subtraction of entries in $\boldsymbol{Q}$. Matrix $\boldsymbol{F}$ is a decomposition of $\boldsymbol{P}$ as indicated by equation (8.6). A Cholesky decomposition of $\boldsymbol{P}$ to triangular matrix $\boldsymbol{F}$ is most convenient. At this point note a requirement that matrix $\boldsymbol{F}$ must be of full rank. This reflects on the composition of the weight matrix and the correlation matrix. Having matrix $\boldsymbol{F}$ its inverse is obtained, this being a simple solution when $\boldsymbol{F}$ is triangular. The matrix of factor weights is obtained by equation (8.7). Equation (8.23) gives an interesting relation.

Once the factor matrix has been determined, a matrix of residual correlations should be computed by equation (8.8). For the example, the matrix of residual correlations is given at the bottom of Table 8.8. For our example, these residuals are all quite tiny indicating that the

Table 8.10

Computation of Factor Matrix for

Illustration of Group Centroid Method of Factor Analysis

Weight Matrix

	1	2
1	+1	0
2	+1	0
3	0	-1
4	+1	0
5	0	+1
6	0	+1

Matrix Q

	1	2
1	1.31	.13
2	1.26	.22
3	-.16	-1.03
4	1.24	.23
5	.20	1.13
6	.22	1.15

Factor Matrix

	1	2
1	.67	-.04
2	.65	.02
3	-.08	-.56
4	.64	.02
5	.10	.61
6	.11	.62

Matrix P

	1	2
1	3.81	.58
2	.58	3.31

Factor Matrix

	1	2
1	1.952	.000
2	.000	1.795

Matrix $(F^{-1})'$

	1	2
5	.512	-.085
6	.000	.557

obtained factor matrix provides an excellent fit to the input correlation matrix. When the fit is not as good, further factors could be extracted from the residual matrix and added to the factor matrix by adjoining these new factors to those already obtained. New attribute clusters could be determined in the residual matrix and the group centroid method used to establish these new factors. An alternative is to apply the centroid method to the matrix of residual correlations.

The group centroid method of factor extraction appears to be a simple technique which could be useful in less formal analyses such as for pilot studies and analyses. For more formal studies more precise methods would be advisable.

8.5. Principal Factors Method of Factor Extraction

With the development of digital computers the principal factors method has become the most popular method for factor extraction. Prior to the large computer the calculation labor was prohibitively extensive to use this method on any but the most trivial sized matrices. The key to use of principal factors is the availability of solutions for eigenvalues and eigenvectors of real, symmetric matrices. Now, these solutions may be obtained quite readily for all but very large correlation and covariance matrices. The principal factors method has a number of desirable properties including a maximization of the sum of squares of factor weights on the extracted factors. Minimization of the sum of squares of residual correlations will be discussed in detail in the next chapter.

A numerical example is discussed before a presentation of mathematical properties of principal factors. Table 8.11 gives the correlation matrix for the nine mental tests example with SMC's in the diagonal. The eigenvalues and eigenvectors were computed for this matrix and are presented in Table 8.12. Note that all eigenvalues after the first three are negative. Use of SMC's in the diagonal of a correlation matrix must result in a number of negative eigenvalues. At the bottom of Table 8.12 is the principal factors matrix, each column of which being obtained by multiplying the entries in an eigenvector by the square root of corresponding eigenvalue. There are only three columns in the principal factors matrix since the square roots of the eigenvalues beyond the first three present problems with imaginary numbers. However, in some studies, the factor extraction may stop short of number of positive eigenvalues, this being a problem as to the number of factors which will be discussed subsequently.

Mathematical relations for principal factors are considered next. Let $\boldsymbol{C}$ be a square, symmetric matrix with real numbers. There is no restriction that $\boldsymbol{C}$ be Gramian as was stated in the general theory of matrix factoring; however, a number of the relations given in the general theory will be used. There must be considerable care in this usage to avoid violating several restrictions. For example, matrix $\boldsymbol{C}$ could be a correlation matrix with SMC's in the diagonal such as given for the nine mental tests in Table 8.11. As seen in Table 8.12, there are a number of

Table 8.11

Correlation and Residual Matrices for

Principal Factors for Nine Mental tests

Correlation Matrix with SMC's in Diagonal

	1	2	3	4	5	6	7	8	9
1	.297	.499	.394	.097	.126	.085	.284	.152	.232
2	.499	.424	.436	.007	.023	.083	.467	.235	.307
3	.394	.436	.356	.292	.307	.328	.291	.309	.364
4	.097	.007	.292	.428	.621	.510	.044	.319	.213
5	.126	.023	.307	.621	.535	.623	.114	.376	.276
6	.085	.083	.328	.510	.623	.440	.086	.337	.271
7	.284	.467	.291	.044	.114	.086	.350	.393	.431
8	.152	.235	.309	.319	.376	.337	.393	.361	.489
9	.232	.307	.364	.213	.276	.271	.431	.489	.349

First Residual Correlation Matrix

	1	2	3	4	5	6	7	8	9
1	.123	.303	.140	-.128	-.137	-.159	.083	-.101	-.014
2	.303	.204	.150	-.247	-.273	-.191	.241	-.050	.030
3	.140	.150	-.014	-.037	-.078	-.028	-.002	-.060	.005
4	-.128	-.247	-.037	.136	.280	.194	-.216	-.009	-.106
5	-.137	-.273	-.078	.280	.136	.254	-.190	-.007	-.097
6	-.159	-.191	-.028	.194	.254	.098	-.196	-.018	-.074
7	.083	.241	-.002	-.216	-.190	-.196	.118	.101	.147
8	-.101	-.050	-.060	-.009	-.007	-.018	.101	-.007	.131
9	-.014	.030	.005	-.106	-.097	-.074	.147	.131	.001

Second Residual Correlation Matrix

	1	2	3	4	5	6	7	8	9
1	-.009	.108	.083	.037	.038	-.011	-.064	-.102	-.072
2	.108	-.085	.066	-.001	-.014	.028	.024	-.051	-.055
3	.083	.066	-.039	.035	-.002	.036	-.066	-.061	-.020
4	.037	-.001	.035	-.071	.060	.008	-.032	-.008	-.034
5	.038	-.014	-.002	.060	-.096	.057	.004	-.006	-.020
6	-.011	.028	.036	.008	.057	-.069	-.031	-.017	-.010
7	-.064	.024	-.066	-.032	.004	-.031	-.045	.100	.083
8	-.102	-.051	-.061	-.008	-.006	-.017	.100	-.007	.131
9	-.072	-.055	-.020	-.034	-.020	-.010	.083	.131	-.024

Table 8.11(Continued)

Third Residual Correlation Matrix

	1	2	3	4	5	6	7	8	9
1	-.086	.064	.029	.012	.023	-.028	-.006	-.020	-.002
2	.064	-.110	.035	-.015	-.022	.019	.056	-.005	-.016
3	.029	.035	-.077	.017	-.012	.024	-.025	-.004	.029
4	.012	-.015	.017	-.079	.056	.002	-.013	.019	-.011
5	.023	-.022	-.012	.056	-.098	.054	.015	.009	-.007
6	-.028	.019	.024	.002	.054	-.073	-.018	.001	.006
7	-.006	.056	-.025	-.013	.015	-.018	-.088	.039	.031
8	-.020	-.005	-.004	.019	.009	.001	.039	-.093	.057
9	-.002	-.016	.029	-.011	-.007	.006	.031	.057	-.087

Table 8.12

Computation of Principal Factors Matrix for Nine Mental Tests Example from Correlation Matrix with SMC's in Diagonal

Eigenvalues

1	2	3	4	5	6	7	8	9
2.746	1.241	.346	-.048	-.064	-.125	-.153	-.175	-.255

Eigenvectors

	1	2	3	4	5	6	7	8	9
1	.252	.326	.472	.265	-.213	.511	.259	.052	.401
2	.283	.483	.266	.094	.305	-.267	.190	.228	-.595
3	.367	.141	.331	-.557	-.258	-.279	-.272	-.447	.084
4	.326	-.409	.154	.311	-.444	-.437	-.082	.456	.057
5	.381	-.433	.087	.349	.146	.349	-.213	-.440	-.394
6	.353	-.366	.106	-.355	.629	.035	.178	.263	.326
7	.291	.362	-.353	.393	.292	-.214	-.448	-.084	.407
8	.366	.002	-.499	.019	-.184	-.146	.684	-.305	.030
9	.356	.14	-.425	-.334	-.249	.456	-.265	.418	-.214

Principal Factor Matrix

	1	2	3
1	.417	.363	.278
2	.469	.538	.156
3	.609	.157	.195
4	.540	-.456	.090
5	.632	-.482	.051
6	.585	-.408	.062
7	.481	.404	-.208
8	.607	.003	-.293
9	.590	.159	-.250

negative eigenvalues for this matrix so that this matrix is not Gramian. Further, residual correlation matrices must have as many eigenvalues equal to zero as the number of factors that have been extracted. Allowance must be made for zero and negative eigenvalues in the development.

At the present, a single factor is considered so that there is a single column of factor weights. Also, matrix $\boldsymbol{W}$ has a single column which will be designated as the vector $\underline{\boldsymbol{w}}$. Analogous to equation (8.4), vector **q** is defined by:

$$\underline{\boldsymbol{q}} = \boldsymbol{C}\underline{\boldsymbol{w}} \, . \tag{8.24}$$

Matrix $\boldsymbol{P}$ of equation (8.5) is replaced by a scalar p :

$$p = \underline{\boldsymbol{w}}'\boldsymbol{C}\underline{\boldsymbol{w}} \, . \tag{8.25}$$

Since p is a scalar, f is a scalar also and equals the square root of p as from equation (8.6). The vector of factor weights, **a** , is obtained analogously to equation (8.7):

$$\underline{a} = \underline{q}[1/p^{\frac{1}{2}}] = C\underline{w}/[(\underline{w}'C\underline{w})^{\frac{1}{2}}] \, . \tag{8.26}$$

The major criterion for principal factors is that the sum of squares of the factor weights is to be a maximum. Let θ designate the sum of squares of the factor weights in vector **a** . Then:

$$\theta = \underline{a}'\underline{a} \tag{8.27}$$

with **a** determined so that θ is a maximum.

A major restriction on the solution for vector $\underline{\boldsymbol{w}}$ is that $\underline{\boldsymbol{w}}'\mathbf{C}\underline{\boldsymbol{w}}$ be greater than zero so that its square root is possible with a real value so as to satisfy equation (8.26). All possible solutions for which this is not true are to be rejected.

The solution for maximum θ can be simplified with a transformation using eigenvalues and eigenvectors of $\boldsymbol{C}$:

$$\boldsymbol{C} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}' \tag{8.28}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues in descending algebraic order and **V** is an orthonormal matrix containing the corresponding eigenvectors. Since **V** is orthonormal:

$$VV' = V'V = I \, . \tag{8.29}$$

Substitution from equation (8.28) into equation (8.26) yields:

$$\underline{\boldsymbol{a}} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}'\underline{\boldsymbol{w}}/[(\underline{w}'V\Lambda V'\underline{w})^{\frac{1}{2}}]$$

or

$$V'\underline{a} = \Lambda V'\underline{w}/[(\underline{w}'V\Lambda V'\underline{w})^{\frac{1}{2}}] \,. \tag{8.30}$$

Define vectors $\underline{\widetilde{\boldsymbol{a}}}$ and $\underline{\widetilde{\boldsymbol{w}}}$ by:

$$\underline{\widetilde{\boldsymbol{a}}} \;= \boldsymbol{V'}\underline{\boldsymbol{a}} \,; \tag{8.31}$$

$$\underline{\widetilde{\boldsymbol{w}}} \;= \boldsymbol{V'}\underline{\boldsymbol{w}} \,. \tag{8.32}$$

Then:

$$\underline{\widetilde{\boldsymbol{a}}} \;= \boldsymbol{\Lambda}\underline{\widetilde{\boldsymbol{w}}} \,/[(\underline{\widetilde{\boldsymbol{w}}}\,'\boldsymbol{\Lambda}\underline{\widetilde{\boldsymbol{w}}}\,)^{\frac{1}{2}}] \,, \tag{8.33}$$

and from equation (8.27)

$$\theta = \underline{\widetilde{\boldsymbol{a}}}\,'\boldsymbol{V'V}\underline{\widetilde{\boldsymbol{a}}} \;= \underline{\widetilde{\boldsymbol{a}}}\,'\underline{\widetilde{\boldsymbol{a}}} \;\,. \tag{8.34}$$

With equation (8.33)

$$\theta = \underline{\widetilde{\boldsymbol{w}}}\,'\boldsymbol{\Lambda^2}\underline{\widetilde{\boldsymbol{w}}} \,/(\underline{\widetilde{\boldsymbol{w}}}\,'\boldsymbol{\Lambda}\underline{\widetilde{\boldsymbol{w}}}\,) \,. \tag{8.35}$$

To obtain a maximum $\boldsymbol{\theta}$ a partial derivative with respect to the elements of vector $\underline{\widetilde{\boldsymbol{w}}}$ is set equal to zero. There may be several optima with the largest solution being chosen.

$$\frac{\partial \boldsymbol{\theta}}{\partial \underline{\widetilde{\boldsymbol{w}}}} = [2\boldsymbol{\Lambda^2}\underline{\widetilde{\boldsymbol{w}}}\,(\underline{\widetilde{\boldsymbol{w}}}\,'\boldsymbol{\Lambda}\underline{\widetilde{\boldsymbol{w}}}\,) - 2\boldsymbol{\Lambda}\underline{\widetilde{\boldsymbol{w}}}\,(\underline{\widetilde{\boldsymbol{w}}}\,'\boldsymbol{\Lambda^2}\underline{\widetilde{\boldsymbol{w}}}\,)]/(\underline{\widetilde{\boldsymbol{w}}}\,'\boldsymbol{\Lambda}\underline{\widetilde{\boldsymbol{w}}}\,)^{\boldsymbol{2}} = \boldsymbol{0} \,. \tag{8.36}$$

With the restriction that $(\underline{\widetilde{\boldsymbol{w}}}\,'\boldsymbol{\Lambda}\underline{\widetilde{\boldsymbol{w}}}\,)$ be finite and not equal to zero:

$$\boldsymbol{\Lambda^2}\underline{\widetilde{\boldsymbol{w}}} \;- [(\underline{\widetilde{\boldsymbol{w}}}\,'\boldsymbol{\Lambda}\underline{\widetilde{\boldsymbol{w}}}\,)/(\underline{\widetilde{\boldsymbol{w}}}\,'\boldsymbol{\Lambda}\underline{\widetilde{\boldsymbol{w}}}\,)]\boldsymbol{\Lambda}\underline{\widetilde{\boldsymbol{w}}} \;= \boldsymbol{0} \,. \tag{8.37}$$

Using equation (8.35) with algebraic manipulations yields:

$$(\boldsymbol{\Lambda} - \boldsymbol{\theta I})(\boldsymbol{\Lambda}\underline{\widetilde{\boldsymbol{w}}}\,) = \boldsymbol{0} \tag{8.38}$$

which is the equation for an eigen problem with eigenvalues θ and eigenvectors $(\boldsymbol{\Lambda}\underline{\widetilde{\boldsymbol{w}}}\,)$. Note that $(\boldsymbol{\Lambda}\underline{\widetilde{\boldsymbol{w}}}\,)$ is a vector with entries $(\boldsymbol{\lambda_i}\widetilde{w}_{\,i})$ for $i = 1, 2, \ldots, \mathrm{n}$.

Since $\boldsymbol{\Lambda}$ is a diagonal matrix, the eigenvalues θ of equation (8.38) equal the diagonal values of $\boldsymbol{\Lambda}$. Also, since the eigenvalues $\boldsymbol{\lambda}$ of $\boldsymbol{C}$ are in descending algebraic order, the maximum θ equals the first $\boldsymbol{\lambda}$. Thus:

$$\theta_1 = \boldsymbol{\lambda_1} \,. \tag{8.39}$$

The first eigenvector has an entry of unity for the first element and entries of zero for remaining entries.

$$(\boldsymbol{\lambda_1}\widetilde{w}_{\,1}) = 1 \tag{8.40}$$

$$(\boldsymbol{\lambda_i}\widetilde{w}_{\,i}) = 0 \;\text{ for } i = 2, \cdots, n\,. \tag{8.41}$$

From (8.40)

$$\widetilde{w}_{\,1} = 1/\boldsymbol{\lambda_1} \tag{8.42}$$

so that

$$(\widetilde{\underline{\boldsymbol{w}}}\,'\boldsymbol{\Lambda}\widetilde{\underline{\boldsymbol{w}}}\,) = \mathbf{1}/\boldsymbol{\lambda}_1 \tag{8.43}$$

Unless all eigenvalues of $\boldsymbol{C}$ are zero or negative, the obtained solution is acceptable.

The first factor weights are considered next. A subscript 1 is used with vectors $\widetilde{\underline{\boldsymbol{w}}}$, $\widetilde{\underline{\boldsymbol{a}}}$, and $\underline{\boldsymbol{a}}$ to designate this first factor. From equations (8.33) and (8.43):

$$\widetilde{\underline{\boldsymbol{a}}}_1 = (\boldsymbol{\Lambda}\widetilde{w}_{\,1})/(1/\boldsymbol{\lambda_1})^{\frac{1}{2}} = (\boldsymbol{\Lambda}\widetilde{w}_{\,1})\boldsymbol{\lambda}_1^{\frac{1}{2}}\,. \tag{8.44}$$

From equation (8.31):

$$\underline{\boldsymbol{a}}_1 = \boldsymbol{V}\widetilde{\underline{\boldsymbol{a}}}_1 = \boldsymbol{V}(\boldsymbol{\Lambda}\widetilde{\underline{\boldsymbol{w}}}_1)\boldsymbol{\lambda}_1^{\frac{1}{2}}$$

From equations (8.40) and (8.41) vector $(\boldsymbol{\Lambda}\widetilde{\underline{\boldsymbol{w}}}_1)$ has a first entry of unity with all other entries equal to zero. Let matrix $\boldsymbol{V}$ be partitioned as below:

$$\boldsymbol{V} = [\boldsymbol{V}_1, \boldsymbol{V}_2]$$

where $\boldsymbol{V}_1$ is $n \times 1$ containing the first column of $\boldsymbol{V}$ and $\boldsymbol{V}_2$ is $n \times (n\text{-}1)$ containing the remaining columns of $\boldsymbol{V}$. With this construction and nature of vector $(\boldsymbol{\Lambda}\widetilde{\underline{\boldsymbol{w}}}_1)$:

$$\underline{\boldsymbol{a}}_1 = \boldsymbol{V}_1\boldsymbol{\lambda}_1^{\frac{1}{2}} \tag{8.45}$$

which is the equation for the first factor weights.

The preceding paragraph concerned the first principal factor. Each of the eigenvalues $\boldsymbol{\lambda_j}$ greater than zero yields a principal factor. The sum of squares of the factor loadings on each of these factors is θ_j :

$$\theta_j = \boldsymbol{\lambda_j} \tag{8.46}$$

The *j*'th element of the eigenvector $(\boldsymbol{\Lambda}\widetilde{\underline{\boldsymbol{w}}}_j)$ equals unity while the remaining elements equal zero. Following similar steps which led to equation (8.45), the vector of factor weights is:

$$\underline{\boldsymbol{a}}_j = \boldsymbol{V}_j\boldsymbol{\lambda}_j^{\frac{1}{2}}\,. \tag{8.47}$$

As will be presented subsequently, each of these principal factors will be the first principal factor for a matrix of residual covariances.

The preceding results may be combined to yield a factor matrix for r principal factors. Let $\boldsymbol{A}_r$ be a factor matrix for the first r principal factors. Also, let $\boldsymbol{V}_r$ be an $n \times r$ matrix containing the first r eigenvectors and $\boldsymbol{\Lambda}_r$ be an $r \times r$ diagonal matrix containing the first r eigenvalues of $\boldsymbol{C}$. Then, equations (8.45) and (8.47) may be combined to:

$$\boldsymbol{A}_r = \boldsymbol{V}_r \boldsymbol{\Lambda}_r^{\frac{1}{2}} . \tag{8.48}$$

Since $\boldsymbol{V}_r$ is a vertical section of an orthonormal matrix:

$$\boldsymbol{V}_r' \boldsymbol{V}_r = \boldsymbol{I} .$$

Then:

$$\boldsymbol{A}_r' \boldsymbol{A}_r = \boldsymbol{\Lambda}_r . \tag{8.49}$$

The columns of $\boldsymbol{A}_\mathbf{r}$ are orthogonal and their sums of squares equal the corresponding eigenvalues of $\boldsymbol{C}$. For the nine mental tests example, three principal factors were extracted; thus, r = 3 . The principal factors matrix is given at the bottom of Table 8.12. Multipliers for the three columns of eigenvectors are the square roots of the first three eigenvalues, these square roots being: 1.657, 1.114, and .588. The three columns of principal factors matrix are obtained by multiplying the corresponding columns of eigenvectors by these multipliers.

Residual correlation matrices are considered next. When one principal factor is extracted at a time there is a sequence of residual matrices with one factor being obtained from one such matrix and a residual matrix being determined from the matrix used in determining the factor. For the nine mental tests example, this series of residual correlation matrices is given in Table 8.11. The operations will be similar for each of these steps so that only the first factor residual matrix will be considered explicitly. Let $\boldsymbol{C}_{\cdot 1}$ designate the first factor residual matrix. From equation (8.8):

$$\boldsymbol{C}_{\cdot 1} = \boldsymbol{C} - \underline{\boldsymbol{a}}_1 \underline{\boldsymbol{a}}_1' . \tag{8.50}$$

Equations are written in this section with expanded matrices involving partitions of 1 factor and (n-1) factors similar to the expansion used previously for matrix $\boldsymbol{V}$. The eigenvalues and eigenvectors of $\boldsymbol{C}$ in equation (8.28) are written as:

$$\boldsymbol{C} = [\boldsymbol{V}_1, \boldsymbol{V}_2] \begin{bmatrix} \lambda_1 , & \mathbf{0} \\ \mathbf{0} , & \boldsymbol{\Lambda}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1' \\ \boldsymbol{V}_2' \end{bmatrix} \tag{8.51}$$

where matrix $\boldsymbol{\Lambda}_\mathbf{2}$, in the present context, is an (n-1) $\times$ (n-1) diagonal matrix containing the eigenvalues after the first one. The matrix product $\underline{\boldsymbol{a}}_1 \underline{\boldsymbol{a}}_1'$ is expressed as:

$$\underline{a}_1 \underline{a}_1' = [\underline{a}_1, 0] \begin{bmatrix} \underline{a}_1' \\ 0 \end{bmatrix} .$$

with:

$$[\underline{a}_1, 0] = [V_1 , V_2] \begin{bmatrix} \lambda_1^{\frac{1}{2}}, & 0 \\ 0 & , \quad 0 \end{bmatrix}$$

$$\underline{a}_1 \underline{a}_1' = [V_1 , V_2] \begin{bmatrix} \lambda_1 , & 0 \\ 0 & , \quad 0 \end{bmatrix} \begin{bmatrix} V_1' \\ V_2' \end{bmatrix} \tag{8.52}$$

In the subtraction of $\underline{a}_1 \underline{a}_1'$ from C , as per equation (8.50), the eigenvectors on the left and right are factored out so that:

$$C_{\bullet 1} = [V_1 , V_2] \begin{bmatrix} 0, & 0 \\ 0, & \Lambda_2 \end{bmatrix} \begin{bmatrix} V_1' \\ V_2' \end{bmatrix} \tag{8.53}$$

Thus, the first eigenvalue of C has been replaced by a zero for $C_{\bullet 1}$ so that the largest eigenvalue of $C_{\bullet 1}$ is the second largest eigenvalue of C . Note that the eigenvectors have not been changed. As a consequence of the foregoing, the first principal factor of $C_{\bullet 1}$ is the second principal factor of C . In a similar manner going from the first factor residual matrix to the second factor residual matrix sets the second eigenvalue of C to zero and the third principal factor of C is the first principal factor of the second residual matrix. These relations continue through as many factors as are extracted.

Several types of information are used in judging the number of factors to be extracted. However, not one criterion can be trusted completely so that an analyst must consider several indications available before making a judgment as to the number of factors. This is in contrast to a common procedure in computer packages which use a single criterion to automate this judgment so that each analysis can be completed automatically in a single run. Several of the types of information for number of factors will be discussed in following paragraphs.

An important point is that factor extraction is only the first part of a complete analysis. After an original factor matrix has been established, there is factor transformation. The transformed factors do not correspond directly, one to one, to the extracted factors but are mixtures of the extracted factors. A final criterion for number of factors extracted is the validity and interpretability of the transformed factor structure.

Major indicators for the number of factors are derived from the series of eigenvalues of the correlation matrix with unities in the diagonal cells and with SMC's in the diagonal cells. A procedure, called by some individuals as "root staring", involves inspection of the series of

eigenvalues, especially of the correlation matrix with SMC's in the diagonal cells. Table 8.13 lists the eigenvalues of the correlation matrix with unities in the diagonal cells and with SMC's in the diagonal cells for the nine mental tests example. Figure 8.1 presents a graph of the eigenvalues of the correlation matrix with SMC's in the diagonal. In this graph, the eigenvalues are plotted against the number of the eigenvalue. This graph illustrates results frequently obtained for well edited test batteries, the series of eigenvalues after a few large ones form an almost straight line. This phenomena may be interpreted as indicating that there are two influences in the formation of the data: first, a relevant factor structure and a second influence of random noise. Cattell (1966) described his "Scree Test" for the number of common factors based on the foregoing observation. The points in an eigenvalue graph are not to be interpreted as goodness of fit measures. If they were so interpreted, factor extraction would continue until a satisfactory goodness of fit is obtained. In contrast, factor extraction should be continued as long as eigenvalues are above the random noise line. Thus, for the nine mental tests example a three factor solution would be accepted. However, Thurstone as well as Cattell advocated extracting one or more extra factors which might be used in the factor transformation process to "clean up" the meaningful transformed factors. Such an operation should be followed only with great care.

An alternative to making an eigenvalue graph is illustrated in Table 8.13. On the right of the section for the eigenvalues of the correlation matrix with SMC's in the diagonal is a column of differences. These values are the differences between consecutive eigenvalues. Geometrically, for a straight line of points, such differences would be equal. Since the eigenvalues are ordered in descending algebraic value, all of the differences between consecutive eigenvalues must be equal to or greater than zero. Then, a series of points which approximate a straight line would have almost equal, positive differences. For the nine mental tests series of differences in Table 8.13, starting with the difference between the fourth and fifth eigenvalue, the values of the differences are quite small with little variation when compared with preceding differences. The last large difference is between the third and fourth eigenvalue indicating that the third eigenvalue is the last one above the random noise line.

Guttman (1964) developed three lower bounds for the number of common factors for a correlation matrix. In this development he considered only cases for population correlation matrices for which the common factor model fitted exactly. Guttman's strongest lower bound for the number of common factors is the number of non-negative eigenvalues of the correlation matrix with SMC's in the diagonal cells. That is, the number of common factors is equal to or greater than the number of these eigenvalues which are positive, including those equal to zero. As can be seen from Table 8.13 for the nine mental tests example, by this criterion there are at least 3 common factors for this matrix. However, as shown by Kaiser and Hunka (1973) from analyses of 64 correlation matrices found in the literature, this criterion leads to the extraction of

Table 8.13

Information for Number of Factors from

Eigenvalues of Correlation Matrix for Nine Mental Tests Example

Correlation Matrix with

Unities in Diagonal		SMC's in Diagonal		
	Eigenvalues		Eigenvalues	Differences
1	3.347	1	2.746	
				1.505
2	1.820	2	1.241	
				.894
3	.997	3	.346	
				.395
4	.580	4	-.048	
				.016
5	.549	5	-.064	
				.061
6	.497	6	-.125	
				.028
7	.476	7	-.153	
				.022
8	.412	8	-.175	
				.050
9	.322	9	-.225	

Parallel Analysis for Eigenvalues

	Real Data	Approximate Random Data	Differences
1	2.746	.189	2.557
2	1.241	.127	1.114
3	.346	.087	.259
4	-.048	.049	-.097

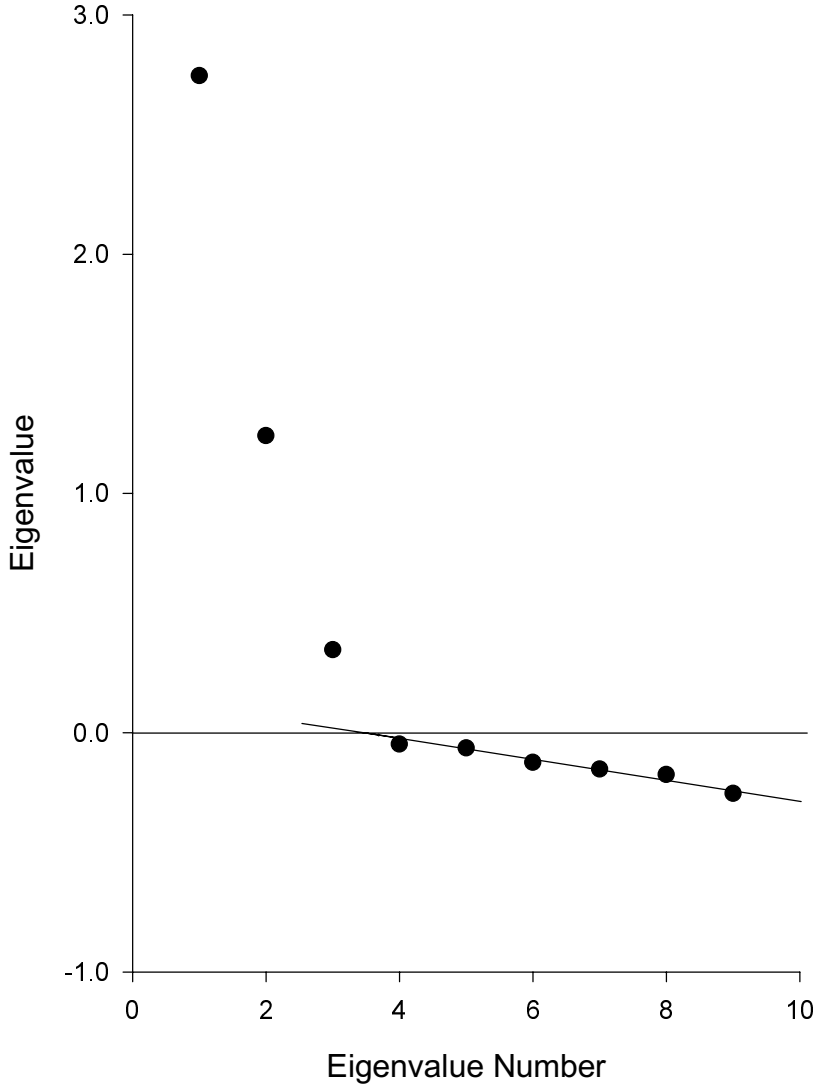

Figure 8.1: Eigenvalue graph for correlation matrix with SMC's in the diagonal, nine mental tests example

a large number of factors. They conclude that this "lower bound is not of practical use in determining the effective number of common factors." An illustration of this stronger lower bound leading to an undesirably large number of extracted factors is given in a second example presented subsequently.

Guttman's weaker lower bound for the number of common factors is the number of eigenvalues of the correlation matrix with unities in the diagonal cells equal to or greater than one. This is a very commonly used value in computer packages for the number of factors to be extracted. Following considerable experience in analyzing a variety of correlation matrices, Kaiser (1960) suggested a simple, approximate procedure using principal components analysis and extracting the number of components equal to Guttman's weaker lower bound. Later, this procedure became called "Little Jiffy" after a remark by Chester Harris. Use of this weaker lower bound has been carried over to computer packages. Analysts, however, must remember that this is a lower bound and may lead to extracting too few factors. See the first column of values in Table 8.13 for the nine mental tests example, these are the eigenvalues of the correlation matrix with unities in the diagonal cells. There are two eigenvalues greater than 1.000 so that Guttman's weaker lower bound for the number of factors would indicate that there are at least two common factors. The third eigenvalue is just less than 1.000; however, a computer output from a number of computer packages would extract only two factors.

Table 8.14 presents results from transformations of the two factor solution and the three factor solution for the nine mental tests example. The two factor solution was indicated by a blind following of the procedure based on Guttman's weaker lower bound. The three factor solution was indicated by the series of eigenvalues of the correlation matrix with SMC's in the diagonal. These results illustrate the types of difficulties which may be encountered when too few factors are extracted. For the three factor solution the three transformed factors are the previously identified numerical operations factor, the spatial factor, and the perceptual speed factor. For the two factor solution, the perceptual speed factor has been collapsed into the numerical operations factor. The spatial factor representation in the two factor solution was OK. Limiting the common factor space by extracting too few factors causes a loss of some weaker factors with the attributes then having improper weights on other factors. As a result, the transformed factor solution is defective. As Kaiser has put it (personal communication) "it's a sin to extract too few factors." Remember that eigenvalues (roots) greater than one of the correlation matrix with unities in the diagonal is a lower bound to the number of factors. Analysts should inspect computer outputs to see if too few factors were extracted. If more factors are indicated, a computer parameter should be set for a rerun to use a larger number of factors.

Humphreys with Ilgen and Montanelli (see Humphreys and Ilgen (1959), Humphreys and Montanelli (1975), Montanelli and Humphreys (1976)) developed a different type of information

Table 8.14
Transformed Factor Weight Matrices from

Principal Factors for Nine Mental Tests Example

Factor Weights

	Two Factor Solution			Three Factor Solution			
Test		1	2		1	2	3
Addition	1	.55	.02	1	.66	.05	-.09
Multiplication	2	.72	-.07	2	.67	-.05	.11
Three-Higher	3	.50	.31	3	.52	.32	.01
Figures	4	-.03	.71	4	.00	.72	-.04
Cards	5	.00	.79	5	-.02	.80	.03
Flags	6	.03	.71	6	.02	.71	.02
Identical Numbers	7	.62	.04	7	.24	.03	.49
Faces	8	.38	.42	8	-.06	.40	.53
Mirror Reading	9	.49	.30	9	.08	.28	.52

Factor Intercorrelations

Two Factor Solution			Three Factor Solution			
	1	2		1	2	3
1	1.00	.16	1	1.00	.15	.57
2	.16	1.00	2	.15	1.00	.11
			3	.57	.11	1.00

relevant to the number of factors to be extracted. Their suggestion was to compare results obtained from the real data with results obtained from random data. Paralleling the real data score matrix, they drew a matrix of random normal deviates (mean = 0, SD = 1) having the same number of rows and columns as the real data matrix. They, then obtained a correlation matrix for the random data and found the eigenvalues of this matrix with SMC's in the diagonal. Their idea was to continue factor extraction until the eigenvalues for the real data were not larger than the eigenvalues for the random data. See the bottom section of Table 8.13 for an example. The first column for "Real Data" is a copy of the first four eigenvalues given above for the correlation matrix having SMC's in the diagonal. The middle column for "Approximate Random Data" was computed by a procedure to be given later. The third column gives the differences between the real data eigenvalues and the approximate random data eigenvalues. Note that the first three real data eigenvalues are materially greater than the approximate random data eigenvalues while there is a switch at the fourth pair of eigenvalues. The parallel analysis criterion indicates that three factors should be extracted from this correlation matrix. In general, Humphreys and associates suggest that when the real data eigenvalues are not greater than the random data eigenvalues the real data eigenvectors and factors contain no more real information than exists for the random data. Consequently, factor extraction can be stopped.

To implement the parallel analysis procedure Montanelli and Humphreys (1976) ran a large Monte Carlo study involving replications of random data analyses for a selection of matrix sizes. The numbers of statistical individuals were 25 , 96 , 384 . or 1533 . They used a total of 21 battery sizes ranging from 6 to 90 . The number of replications per cell varied from 10 to 40 with the cells having fewer replications being those for the larger sample sizes. For each matrix order, N $\times$ *n* , a series of mean eigenvalues across replications was computed, each of these mean values being for the i'th eigenvalue. Montanelli and Humphreys provided a system to approximate such series of eigenvalues using tabled weights. The data of means was provided to Tucker who developed a system for use with a computer to approximate the mean eigenvalues. A first point is that the eigenvalues for random data become negative after *m* eigenvalues where *m*=$\frac{1}{2}n$ for *n* even and $\frac{1}{2}$(*n* - 1) for *n* odd. As a consequence, the series of eigenvalues is truncated after the m'th eigenvalue. Following are the equations for Tucker's system.

$$n' = n(N+1)/N \tag{8.54}$$

$$x_{Ni} = ln\{N - 1 + 100(3N - 4n)/N(N + 4n)\} \tag{8.55}$$

$$x_{ni} = ln\{\frac{1}{2}n'(n'+1) - ni - 10n/\{1 + 2[ln(50i/n)]^2\}\} \tag{8.56}$$

$$a_i = .81848 - .39125 \bullet ln(i - .54151) \tag{8.57}$$

$$b_{Ni} = -.53365 - .03711 \bullet ln(i + 4.66402) \tag{8.58}$$

$$b_{ni} = .40202 + .03169 \bullet ln(i - .71657) \tag{8.59}$$

where N is the sample size, *n* is the battery size, and *i* is the eigenvalue number. Then, the approximation to the random data eigenvalue is $\widehat{\rho}_{Nni}$:

$$\widehat{\rho}_{Nni} = exp(a_i + b_{Ni}x_{Ni} + b_{ni}x_{ni}) \tag{8.60}$$

The approximate random data eigenvalues in Table 8.13 were computed by this system. Tucker used an empirical, trial and error procedure. Measures of goodness of fit to the Montanelli and Humphreys data indicated a satisfactory level for practical use.

A second example of principal factor extraction is presented in Tables 8.15 through 8.17 and Figure 8.2. This example uses 18 verbal and numerical tests selected from a 66 test battery by Thurstone and Thurstone (1941). Table 8.15 gives the correlation matrix with the test names being given in Table 8.17. First consideration is given to the number of factors to be extracted. For the roots greater than one from the correlation matrix with unities in the diagonal, the fourth eigenvalue is 1.122 while the fifth eigenvalue is .741. By this criterion there are four common factors. Figure 8.2 presents the eigenvalue plot for the correlation matrix with SMC's in the diagonal. There are four eigenvalues above a well defined line for random noise eigenvalues. By this criterion there appear to be four common factors to be extracted. The parallel analysis is given in Table 8.16. The first four real data eigenvalues are distinctly greater than the corresponding approximate random data eigenvalues while there is a switch at eigenvalue 5 for which the real data eigenvalue is less than the approximate random data eigenvalue. By this criterion, also, there appears to be four common factors. From the convergence of these three criteria for the number of factors to be extracted, a decision to consider a four factor solution is well justified.

A second look at the Guttman stronger lower bound for the number of factors is provided in Table 8.16. The first 9 real data eigenvalues are listed for the correlation matrix with SMC's in the diagonal. Guttman's stronger lower bound states that there are at least as many common factors as the number of these eigenvalues that are non-negative. For the 18 verbal and numerical test example, this criterion indicates that there are at least 8 common factors. This is an undesirable answer when compared with the number of factors indicated by the criteria discussed in the preceding paragraph. The Kaiser and Hunka (1973) conclusion appears to be upheld that the Guttman stronger lower bound is not usable for real world data. Table 8.17 contains the principal factors matrix for the 18 verbal and numerical tests example.

Table 8.15

Correlation Matrix among 18 Verbal and Numerical Tests Example

	1	2	3	4	5	6	7	8	9
1	1.000								
2	.264	1.000							
3	.232	.231	1.000						
4	.276	.465	.311	1.000					
5	.221	.248	.378	.460	1.000				
6	.284	.034	.431	.221	.237	1.000			
7	.234	.243	.377	.367	.357	.286	1.000		
8	.207	.395	.378	.439	.400	.154	.437	1.000	
9	.499	.208	.307	.328	.281	.467	.295	.203	1.000
10	.160	.092	.294	.236	.254	.299	.294	.321	.242
11	.195	.444	.220	.476	.384	.054	.237	.308	.183
12	.126	.403	.307	.420	.349	.116	.341	.473	.126
13	.159	.430	.209	.497	.454	-.008	.226	.335	.133
14	.184	.420	.206	.511	.455	.026	.187	.342	.172
15	.168	.365	.240	.486	.497	.126	.258	.326	.210
16	.165	.329	.331	.342	.256	.238	.396	.473	.163
17	.394	.454	.364	.411	.341	.291	.334	.376	.436
18	.217	.230	.385	.490	.541	.407	.291	.303	.356

	10	11	12	13	14	15	16	17	18
1									
2									
3									
4									
5									
6									
7									
8									
9									
10	1.000								
11	.112	1.000							
12	.218	.385	1.000						
13	.048	.624	.396	1.000					
14	.090	.623	.418	.769	1.000				
15	.112	.550	.396	.730	.661	1.000			
16	.296	.239	.459	.212	.241	.223	1.000		
17	.274	.324	.313	.281	.331	.310	.330	1.000	
18	.224	.418	.387	.439	.479	.500	.277	.364	1.000

* Selected from 66 test study by Thurstone and Thurstone (1941), Factorial studies of intelligence; samole size=710.

Table 8.16

Eigenvalues and Parallel Analysis for 18 Verbal and Numerical Tests Example

	Real Data	Approximate Random Data	Differences
1	6.031	.316	5.715
2	1.649	.254	1.395
3	.764	.209	.554
4	.532	.175	.357
5	.112	.146	-.035
6	.054	.120	-.066
7	.020	.094	-.074
8	.011	.065	-.054
9	-.013	.017	-.030

Table 8.17

Principal Factors Matrix for 18 Verbal and Numerical Tests Example

Test		1	2	3	4
Addition	1	.39	.28	.22	.31
Arithemetic (Word Problems)	2	.57	-.12	-.17	.36
Mirror Reading	3	.52	.33	-.03	-.14
Directions	4	.71	-.04	-.01	.06
Disarranged Sentences	5	.65	.01	.09	-.22
Identical Numbers	6	.35	.53	.24	-.16
Letter Grouping	7	.52	.27	-.18	-.06
Letter Series	8	.61	.12	-.36	.00
Multiplication	9	.45	.41	.34	.19
Number Patterns	10	.34	.34	-.12	-.11
Paragraph Recall	11	.64	-.32	.05	.07
Pedigrees	12	.60	-.05	-.31	-.06
Vocabulary	13	.70	-.52	.10	-.02
Sentences (Completion)	14	.70	-.45	.11	.00
Same or Opposite	15	.69	-.35	.16	-.11
Secret Writing	16	.51	.20	-.37	-.03
Three-Higher	17	.60	.23	.03	.26
Verbal Enumeration	18	.67	.05	.23	-.26

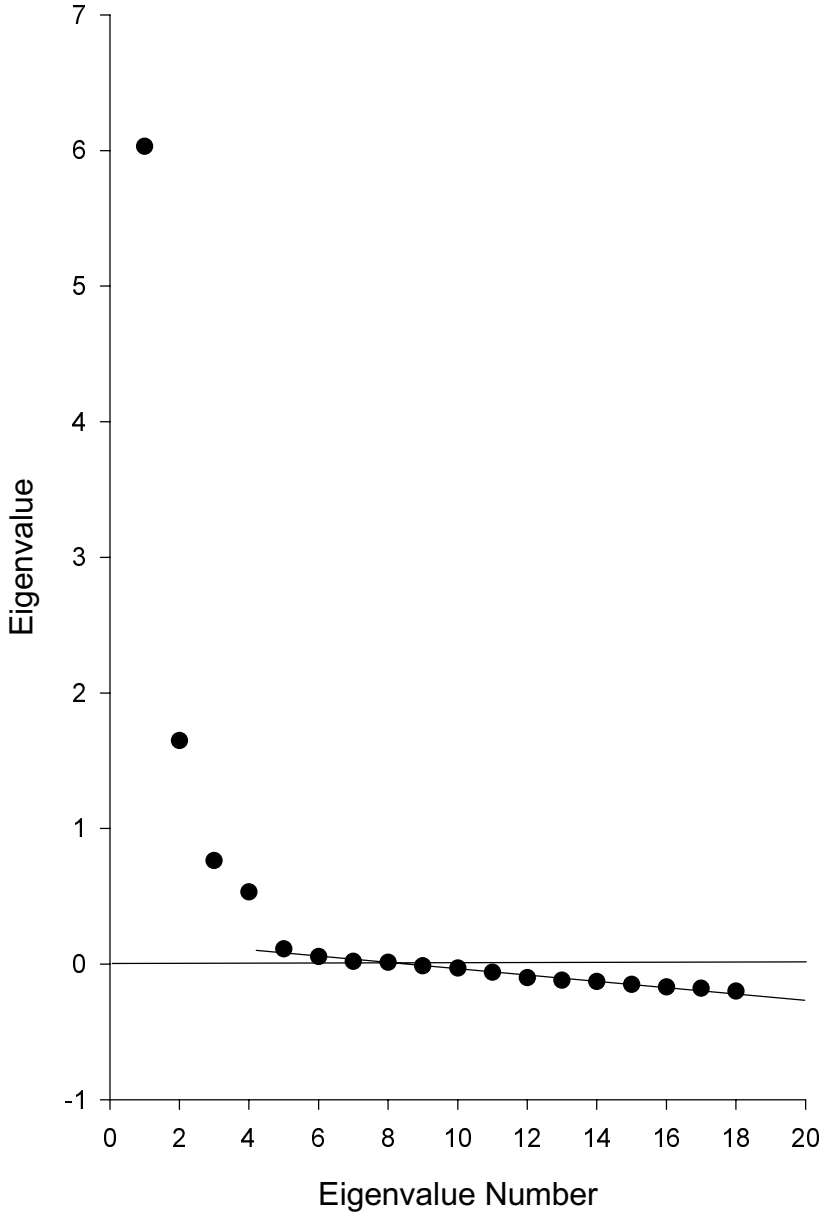

Figure 8.2: Eigenvalue graph for correlation matrix with SMC's in diagonal, 18 verbal & numerical test example.

8.6. Alpha Factor Analysis

Kaiser and Caffrey (1965) presented alpha factor analysis as a psychometric approach to factor analysis in contrast to what they termed statistical factor analysis. They emphasized, in a population of individuals, the generalization of results to a universe of content from observations on a battery of attributes which they considered to be a sample (usually nonrandom) from the universe of content. They contrasted this conception with statistical analysis generalization to a population of individuals from an observed sample (usually random) of individuals. Alpha factor analysis considers relations in a population of individuals and does not consider the sampling of individuals problem. The argument is that this sampling of attributes and generalization to a universe of content is an extremely important psychometric problem not considered in statistical factor analysis.

Alpha factor analysis considers the common parts of observed attribute measures. In accordance with the factor analysis model presented in Chapter 3 and outlined in Chapter 7, the vector of common parts scores is related to a vector of common factor scores for uncorrelated factors by:

$$\underline{\boldsymbol{z}}_{\boldsymbol{\alpha}} = \underline{\boldsymbol{x}}_{\boldsymbol{\alpha}}\mathring{\mathbf{A}}' \tag{8.61}$$

This equation is an interpretation of equation (7.11) in terms of uncorrelated factors (Remember that score vectors are taken as row vectors.). In the present context, reference to uncorrelated common factors is a matter of convenience. Problems with transformations to correlated common factors are to be discussed subsequently. A basic relation in alpha factor analysis involves an inverse from equation (8.61) which expresses common factor scores as weighted linear combinations of the common part scores. For a given battery, the common factor score for an individual on factor *k* is expressed as:

$$x_{\alpha k} = \sum_{j=1}^{n} w_{kj} z_{\alpha j} \tag{8.62}$$

where *j* refers to attributes in the battery. This relation pertains to a given battery since the common factor scores and common part scores may, and probably will change with a change in the battery. In comparison, the common factor score for the individual as related to the universe of content is related to the common part scores by:

$$\xi_{\alpha k} = \sum_{p=1}^{\infty} w_{kp} z_{\alpha p} \tag{8.63}$$

where p refers to attributes in the universe of content. The correlation between $x_{\alpha k}$ and $\xi_{\alpha k}$ is used as a measure of generalizability from the given battery to the universe of content. Following

the work by Kuder and Richardson (1937) on test reliability which was extended by Cronbach (1951) and with Rajaratnam and Glesser (1963), an adaptation of coefficient alpha was developed for generalizability in the present situation.

$$\alpha = \left[\frac{n}{n-1}\right]\left[1 - \frac{w'H^2w}{w'(R - \boldsymbol{U^2})\boldsymbol{w}}\right] \tag{8.64}$$

where **w** is a column vector of weights for a given factor. In factor extraction this measure is to be maximized for successive factors. Solution for the maximum is facilitated by defining:

$$\boldsymbol{\lambda = \frac{w'(R - U^2)w}{w'H^2w}} \tag{8.65}$$

Setting the partial derivative of $\boldsymbol{\lambda}$ with respect to **w** equal to zero leads, with algebraic operations, to:

$$[(R - \boldsymbol{U^2}) - \boldsymbol{\lambda H^2}]\boldsymbol{w} = \boldsymbol{0} \tag{8.66}$$

With column vector **v** defined by:

$$v = Hw \tag{8.67}$$

Equation (8.66) can be written as follows.

$$[H^{-1}(R - \boldsymbol{U^2})\boldsymbol{H^{-1}} - \boldsymbol{\lambda I}]\boldsymbol{v} = \boldsymbol{0} \tag{8.68}$$

$\boldsymbol{\lambda}$ is an eigenvalue of $H^{-1}(R - \boldsymbol{U^2})\boldsymbol{H^{-1}}$ and **v** is the corresponding unit eigenvector. From equations (8.64) and (8.65) the k'th value of α is:

$$\alpha_k = \left[\frac{n}{n-1}\right]\left[1 - \frac{1}{\boldsymbol{\lambda_k}}\right] \tag{8.69}$$

Note that the maximum α corresponds to the maximum λ.

Matrix ($R - \boldsymbol{U^2}$) has the communalities, diagonal entries of $\boldsymbol{H^2}$, in its diagonal and, thus, is the correlation matrix of the common parts of the observed attributes. Premultiplication and postmultiplication of this matrix by H^{-1} scales the attribute measures to unit communalities, thus, yielding the correlation matrix among the common parts. The eigensolution of equation (8.68) may be used to yield principal factors of this correlation matrix. For r factors (the number of factors to be extracted will be discussed in subsequent paragraphs.) the principal factors matrix of $H^{-1}(R - \boldsymbol{U^2})\boldsymbol{H^{-1}}$ is:

$$\widetilde{\mathring{\mathbf{A}}} = \boldsymbol{V_r\Lambda_r^{\frac{1}{2}}} \tag{8.70}$$

To obtain the factor matrix for the attributes in terms of the original scale, it is necessary to perform an inverse scaling:

$$\mathring{\mathbf{A}} = H\widetilde{\mathring{\mathbf{A}}} \ . \tag{8.71}$$

Combining equations (8.70) and (8.71):

$$\mathring{\mathbf{A}} = \boldsymbol{H V_r \Lambda_r^{\frac{1}{2}}} \ . \tag{8.72))}$$

To this point the communalties in $\boldsymbol{H^2}$ have been taken to be known. However, these values are not known so that a solution for them is necessary. For convenience, repeating equation (3.49):

$$\boldsymbol{H^2 = Diag(\mathring{A}\mathring{A}')} \tag{3.49}$$

Also, from equation (3.34):

$$\boldsymbol{U^2 = I - H^2} \tag{8.73}$$

An iterative solution involves starting with an initial $\boldsymbol{H^2}$ and performing the following steps.

1. Obtain $\boldsymbol{U^2}$ as per equation (8.73).
2. Form matrix $H^{-1}(R - \boldsymbol{U^2})\boldsymbol{H^{-1}}$.
3. Obtain Λ_r and $\boldsymbol{V_r}$ from an eigensolution of $H^{-1}(R - \boldsymbol{U^2})\boldsymbol{H^{-1}}$.
4. Obtain matrix $\mathring{\mathbf{A}}$ as per equation (8.72).
5. Obtain $\boldsymbol{H^2}$ as per equation (3.49).
6. Return to step 1 until there is a minimal change of $\boldsymbol{H^2}$.

Kaiser and Caffrey (1965) outlined an efficient program to implement this solution incorporating some short cuts to speed up the computations. While there is no proof that this system will converge, experience indicates that a converged solution will be obtained in almost all cases.

The ALPHA factor analysis of the correlation matrix in Table 1.1 for the nine mental tests example is given in Table 8.18 for two factors. The eigenvalues of $H^{-1}(R - \boldsymbol{U^2})\boldsymbol{H^{-1}}$ are given in the first row and the corresponding values of α are given in the second row. Only the first three α's are given for reasons to be described later. The iterated factor matrix and communalities are presented in the lower section of the table.

Kaiser and Caffrey (1965) suggest that the number, r , of factors to be extracted from a correlation matrix equals the number of positive α's ; that is, all factors, and only those factors, are to be extracted for which the generalizability is positive. Thus, the number of ALPHA factors to be extracted for the nine mental tests example is two, this being the number of positive α's . Note from equation (8.69) that for a positive α_k the eigenvalue $\boldsymbol{\lambda_k}$ must be greater than one. This relation supports a commonly used principal that the number of factors to be extracted from a

Table 8.18

Alpha Factor Analysis for Nine Mental Tests Example

Two Factors Extracted

	Dimension								
	1	2	3	4	5	6	7	8	9
Eigenvalue	6.38	2.62	.99	.06	.03	-.07	-.27	-.34	-.41
Alpha	.95	.70	-.01						

Factor Matrix			
	1	2	Communality
1	.44	-.29	.28
2	.58	-.55	.64
3	.63	-.09	.40
4	.48	.53	.51
5	.58	.60	.70
6	.53	.49	.52
7	.54	-.32	.39
8	.60	.09	.37
9	.62	-.07	.38

correlation matrix equals the number of eigenvalues of the correlation matrix which are greater than one. To support the transition from eigenvalues of $H^{-1}(R - \boldsymbol{U^2})\boldsymbol{H^{-1}}$ to eigenvalues of the correlation matrix $\boldsymbol{R}$ consider the following development. From equation (8.73):

$$H^{-1}(R - \boldsymbol{U^2})\boldsymbol{H^{-1}} = H^{-1}(R - \boldsymbol{I} + \boldsymbol{H^2})\boldsymbol{H^{-1}} = H^{-1}(R - \boldsymbol{I})\boldsymbol{H^{-1}} + \boldsymbol{I} \, .$$

Equation (8.68) becomes:

$$[H^{-1}(\boldsymbol{R} - \boldsymbol{I})\boldsymbol{H^{-1}} - (\boldsymbol{\lambda} - \mathbf{1})\boldsymbol{I}]\boldsymbol{V} = \mathbf{0} \qquad (8.74)$$

Several logical steps follow.

1. For any α to be positive, the corresponding ($\boldsymbol{\lambda}$ - 1) must be positive.
2. Since **($\boldsymbol{\lambda}$-1)** is an eigenvalue of $H^{-1}(\boldsymbol{R} - \boldsymbol{I})\boldsymbol{H^{-1}}$, the number of positive α's equals the number of positive eigenvalues of $H^{-1}(\boldsymbol{R} - \boldsymbol{I})\boldsymbol{H^{-1}}$.
3. By Sylvester's Law of Inertia (see: Guttman, 1954 and Bôcher, 1907) the number of positive eigenvalues of value of $H^{-1}(\boldsymbol{R} - \boldsymbol{I})\boldsymbol{H^{-1}}$ is independent of the value of $\boldsymbol{H}$.
4. A possible H to be considered is an identity matrix so that the number of positive eigenvalues of $(\boldsymbol{R} - \mathbf{I})$ equals the number of positive eigenvalues of $H^{-1}(\boldsymbol{R} - \boldsymbol{I})\boldsymbol{H^{-1}}$ for any other value of H .
5. Each eigenvalue of $(\boldsymbol{R} - \mathbf{I})$ equals the corresponding eigenvalue of $\boldsymbol{R}$ decreased by 1.
6. The number of positive eigenvalues of $(\boldsymbol{R} - \mathbf{I})$ equals the number of eigenvalues of $\boldsymbol{R}$ which are greater than one.
7. The number of positive α's equals the number of eigenvalues of $\boldsymbol{R}$ greater than one.

From the preceding, the number of factors, r , may be set from the eigenvalues of the correlation matrix and not changed with different approximations of the communalties.

Kaiser and Caffrey (1965) did not consider the topic of transformations of factors from the obtained ALPHA factors. For the nine mental tests example, Table 8.19 presents a transformation to correlated factors of the two factor ALPHA factor matrix. The Alpha coefficients for the transformed factors were obtained by entering the transformed factor weights into equation (8.64). The results in Table 8.19 differ from the solution given in Table 1.2 which included three factors. The first transformed ALPHA factor in Table 8.19 appears to be a combination of a numbers factor and a perceptual speed factor, that is, a combination of factors 1 and 3 of Table 1.2. This inspection suggests that too few factors had been extracted using the principle of positive generalizability.

Consider the series of eigenvalues in Table 8.18, the third eigenvalue is only slightly less than one and the corresponding α is only slightly negative. There is a distinct break in the eigenvalue series between the third and fourth eigenvalue which suggests that a three factor

Table 8.19

Transformed Alpha factors for Nine Mental Tests Example

Two Factor Solution

Factor Weights		
Attribute	1	2
1. Addition	.52	.04
2. Multiplication	.80	-.08
3. Three-Higher	.51	.32
4. Figures	-.03	.71
5. Cards	.00	.84
6. Flags	.03	.72
7. Ident. Numbers	.61	.07
8. Faces	.37	.44
9. Mirror Reading	.49	.33

Factor Correlations		
1	1	2
2	.52	.04

Alpha for Transformed Factors		
	1	2
	.80	.86

solution might be appropriate. Table 8.20 presents a three factor ALPHA solution and Table 8.21 presents the transformed solution. Note that the transformed factors correspond to the three factors in Table 1.2 . A point of interest is that, while the third ALPHA factor had slightly negative generalizability, all three transformed factors had positive α's .

The preceding material suggests a distinction between two views: that of factors being determined from the observed scores and the view of factor analytic studies being conducted to obtain indications of major internal attributes, or latent variables which are characteristics in a domain of mental behavior. Use of the generalizability of factors provides no mechanism to distinguish between major dimensions and possible trivial dimensions due to lack of fit. As noted by Kaiser and Caffrey, enlarging a battery of measures will lead to increasing numbers of factors. A small battery such as the nine mental tests example may lead to extraction of too few factors some of which are combinations of major factors which might be obtained with a larger battery. An argument may be made that this small battery is not adequate to determine the common factor space. This is a question of battery adequacy. However, extraction of three factors, ignoring the negative generalizability of the third factor, does lead to a quite interpretable transformed solution. This battery appears to be adequate to provide indications of three major internal attributes. There appears to be a contrast between two opinions as to the purpose of factor analytic studies. One opinion is that factor analytic studies are conducted to provide information about the structure of the dynamics of mental behavior by identifying internal attributes. The other opinion appears to be that factor analytic studies are conducted to determine factors, including factor scores, from the observed attributes. ALPHA factor analysis appears to be a representative of a procedure to determine factors.

8.7. Image Factor Analysis

Jöreskog (1963) described a most interesting but little used factor analysis model and analytic procedure which, subsequently, became known as Image Factor Analysis due to mathematical relations to Guttman's (1953) image theory. However, Jöreskog (1969) commented that Image Factor Analysis was a model in its own right. In Jöreskog's notation, his model for uncorrelated factors is:

$$\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi} \tag{8.75}$$

where $\mathbf{\Sigma}$ is the population dispersion (covariance) matrix among the observed attributes, $\mathbf{\Lambda}$ is the factor weights matrix, and $\mathbf{\Psi}$ is the diagonal matrix of unique variances. Note that Jöreskog does not incorporate a term for lack of fit of the model. For Image Factor Analysis, Jöreskog specializes this model by assuming that:

Table 8.20

Alpha Factor Analysis for Nine Mental Tests Example

Three Factors Extracted

	Dimension								
	1	2	3	4	5	6	7	8	9
Eigenvalue	5.53	2.48	.99	.14	.10	.03	-.03	-.10	-.14
Alpha	.92	.67	-.01						

	Factor Matrix			
	1	2	3	Communality
1	.46	-.35	.32	.43
2	.54	-.56	.20	.64
3	.63	-.11	.21	.45
4	.51	.51	.12	.53
5	.61	.57	.09	.71
6	.55	.46	.08	.52
7	.52	-.37	-.25	.47
8	.63	.06	-.36	.53
9	.62	-.11	-.29	.47

Table 8.21

Transformed Alpha factors for Nine Mental Tests Example

Three Factor Solution

Factor Weights			
Attribute	1	2	3
1. Addition	.70	.06	-.12
2. Multiplication	.76	-.06	.08
3. Three-Higher	.53	.34	.02
4. Figures	.01	.73	-.04
5. Cards	-.01	.84	.02
6. Flags	.02	.72	.03
7. Ident. Numbers	.24	.02	.53
8. Faces	-.09	.41	.62
9. Mirror Reading	.08	.28	.56

Factor Correlations			
	1	2	3
1	1.00	.12	.52
2	.12	1.00	.08
3	.52	.08	1.00

Alpha for Transformed Factors		
1	2	3
.24	.85	.72

$$\boldsymbol{\Psi} = \boldsymbol{\theta}[diag(\boldsymbol{\Sigma}^{-1})]^{-1} \tag{8.76}$$

where θ is a constant parameter of the model.

A translation from Jöreskog's notation to that used in this book is needed. Our matrix $\mathbf{Å}$ replaces Jöreskog's matrix $\boldsymbol{\Lambda}$ and $\boldsymbol{U^2}$ replaces $\boldsymbol{\Psi}$. Our $\boldsymbol{\Delta_\Sigma}$ is taken to be zero so that:

$$\Sigma = \boldsymbol{\Sigma_{yy} = \Sigma_{zz} = ÅÅ' + U^2.} \tag{8.77}$$

Equation (8.35) gives the squares of the standard errors of estimating the attributes from the remaining attributes in a battery in a sample. The equivalent relation for the population and involving the covariance matrix Σ is:

$$\Sigma_e = [diag(\Sigma^{-1})]^{-1} \tag{8.78}$$

with diagonal entries σ_{ej}^2 which are the error variances in estimating scores on attribute *j* from the remaining attributes in the battery. With these translations of notation and equations (8.76) and (8.78), Jöreskog's model becomes:

$$\Sigma = \mathbf{ÅÅ}' + \theta\Sigma_e \, . \tag{8.79}$$

Understanding of the Image Factor Analysis model is facilitated by consideration of the inequality of equation (8.56). This equation may be revised to apply to covariance matrices instead of correlation matrices so as to yield with algebraic operations an inequality between uniqueness and error of estimate variances:

$$u_j^2 < \boldsymbol{\sigma_{ej}^2} \tag{8.80}$$

Usually this inequality includes possible equality; however, this equality occurs only for very special situations and is ignored here. With θ a positive constant between zero and one, u_j^2 might be approximated by $\theta\sigma_{ej}^2$. This would be especially true when both u_j^2 and $\theta\sigma_{ej}^2$ are nearly constant over the attributes in the battery. With this approximation, the Image Factor Analysis model may be written as:

$$(\Sigma - \theta\Sigma_e) = \mathbf{ÅÅ}' \tag{8.81}$$

In this form, use of $\theta\sigma_{ej}^2$ appears to provide a solution to the communality problem. There remains to be the solution for an appropriate value for θ . This will be discussed in terms of analysis for a sample. The factor matrix could be determined by the principal factors procedure.

In a sample: Σ is replaced by $\boldsymbol{C}$, $\mathbf{Å}$ by $\boldsymbol{A}$ and Σ_e by S_e^2 with, per equation (8.35),

$$S_e^2 = [diag(\boldsymbol{C}^{-1})]^{-1}$$

With the addition of a residual term, the model of equation (8.81) may be written for the sample as:

$$(\boldsymbol{C} - \boldsymbol{\theta S_e^2}) = \boldsymbol{AA'} + \boldsymbol{\Delta} \tag{8.82}$$

Jöreskog considers $(\boldsymbol{\theta S_e^2} + \boldsymbol{\Delta})$ to be the residual in representing $\boldsymbol{C}$ by $\boldsymbol{AA'}$ **.** In the computing procedure Jöreskog described in 1963 he performed a scaling transformation instead of directly factoring $(\boldsymbol{C} - \boldsymbol{\theta S_e^2})$ to obtain:

$$S_e^{-1}(\boldsymbol{C} - \boldsymbol{\theta S_e^2})S_e^{-1} = S_e^{-1}AA'S_e^{-1} + S_e^{-1}\boldsymbol{\Delta}S_e^{-1}$$

so that

$$(S_e^{-1}\boldsymbol{C}S_e^{-1} - \theta I) = S_e^{-1}AA'S_e^{-1} + S_e^{-1}\boldsymbol{\Delta}S_e^{-1} \tag{8.83}$$

With the following definitions:

$$\widetilde{\boldsymbol{C}} = S_e^{-1}\boldsymbol{C}S_e^{-1} \tag{8.84}$$

$$\widetilde{\boldsymbol{A}} = S_e^{-1}A \tag{8.85}$$

$$\widetilde{\boldsymbol{\Delta}} = S_e^{-1}\boldsymbol{\Delta}S_e^{-1} \tag{8.86}$$

equation (8.83) yields:

$$(\widetilde{\boldsymbol{C}} - \boldsymbol{\theta I}) = \widetilde{\boldsymbol{A}}\,\widetilde{\boldsymbol{A}}' + \widetilde{\boldsymbol{\Delta}} \tag{8.87}$$

An important point is that $\widetilde{\boldsymbol{C}}$ can be shown to be invariant within reflection of the attributes for any rescaling of the attribute scores. The form of equation (8.87) is very convenient: the eigenvectors of $(\widetilde{\boldsymbol{C}} - \boldsymbol{\theta I})$ are invariant with changes in θ and the eigenvalues change by an additive procedure. The eigensolution for $\boldsymbol{C}$ is (the eigenvalues are in descending algebraic order):

$$\widetilde{\boldsymbol{C}} = \boldsymbol{V\Lambda V'} \tag{8.88}$$

so that the eigensolution for $(\widetilde{\boldsymbol{C}} - \boldsymbol{\theta I})$ is:

$$(\widetilde{\boldsymbol{C}} - \boldsymbol{\theta I}) = \boldsymbol{V}(\boldsymbol{\Lambda} - \boldsymbol{\theta I})\boldsymbol{V'}\,. \tag{8.89}$$

This is in accord with the general theory of eigensolutions. A principal factors solution to r factors of $(\widetilde{\boldsymbol{C}} - \boldsymbol{\theta I})$ yields:

$$\widetilde{\boldsymbol{A}} = \boldsymbol{V_r}(\boldsymbol{\Lambda_r} - \boldsymbol{\theta I_r})^{\frac{1}{2}} \tag{8.90}$$

where $\mathbf{V}_r$ contains the first *r* eigenvectors, $\mathbf{\Lambda}_r$ is a diagonal matrix containing the first *r* eigenvalues, and I_r is an $r \times r$ identity matrix. Then the factor matrix $\boldsymbol{A}$ in terms of the original scaling may be obtained from equation (8.85) as:

$$\boldsymbol{A} = \boldsymbol{S}_e \widetilde{\boldsymbol{A}} \, . \tag{8.91}$$

Jöreskog's suggested solution for θ was to set it equal to the mean of the (n - r) discarded eigenvalues of $\widetilde{\boldsymbol{C}}$

$$\boldsymbol{\theta} = \sum_{k=r+1}^{n} \boldsymbol{\lambda}_k / (\boldsymbol{n} - \boldsymbol{r}) \tag{8.92}$$

With this value of θ the sum of the discarded eigenvalues of $(\widetilde{\boldsymbol{C}} - \boldsymbol{I})$ equals zero, that is:

$$\sum_{k=r+1}^{n} (\boldsymbol{\lambda}_k - \boldsymbol{\theta}) / (\boldsymbol{n} - \boldsymbol{r}) = \boldsymbol{0} \tag{8.93}$$

A major problem is the choice of the number of factors to be extracted. Fortunately, the solution for $\widetilde{\boldsymbol{C}}$ and the eigensolution do not depend upon the choice of r so that several choices may be considered. Jöreskog suggested a statistical coefficient for testing for significant departure of a chosen model (number of factors) from the number for the population. Subsequently, he found that this coefficient did not follow an expected chi-square distribution so that he advised (personal communication) that this coefficient would not be useful. Consequently, this coefficient will not be presented here. Choice of the number of factors remains a matter of judgment. Information from the series of eigenvalues might be useful; also, the size of residuals could be used.

Jöreskog (1969) described a maximum likelihood method for estimating the parameters in the Image Factor Analysis model and presented a measure of goodness of fit. The method described in the preceding paragraphs sacrifices some efficiency in the parameter estimates for the sake of speed. The maximum likelihood estimates are fully efficient in large samples. However, this procedure is relatively slow involving an iterative solution. In this respect it is very similar to the maximum likelihood factor analysis to be discussed in Chapter 9. While the maximum likelihood solution for image factor analysis is of theoretic interest it is very seldom used.

Reference notes for Sylvester's "Law of Inertia":

Guttman, L. Some necessary conditions for common-factor analysis. Psychometrika. 1954, 19, 149-161.

Bôcher, Maxime. Introduction to higher algebra. New York, Macmillan, 1907. (Sixteenth printing, 1952).

Kaiser, Henry F. & Caffrey. Alpha factor analysis. Psychometrika. 1965, 30, 1-14.

References for Image Factor Analysis

Guttman, L. Image theory for the structure of quantitative variates. Psychometrika, 1953, 18, 277-296.

Jöreskog, K. G. Statistical estimation in factor analysis. Stockholm: Almqvist & Wiksell, 1963.

Jöreskog, K. G. Efficient estimation in image factor analysis. Psychometrika, 1969, 34, 51-75.

CHAPTER 9
FACTOR FITTING BY STATISTICAL FUNCTIONS

From
Exploratory Factor Analysis
Ledyard R Tucker
and
Robert C. MacCallum

CHAPTER 9
FACTOR FITTING BY STATISTICAL FUNCTIONS

Determination of factor matrices from covariance matrices according to several statistical functions is discussed in this chapter. In Chapter 8 we concentrated on factor extraction by matrix factoring techniques. A contrast may be considered for an example between the principal factors method of factor extraction described in Chapter 8 and the minimum squared residuals method (MINRES) to be described in this chapter. The principal factors method obtains factors to maximize the sum of squares of the factors determined from ($\boldsymbol{R_{yy}} - \boldsymbol{U^2}$) where $\boldsymbol{R_{yy}}$ is the observed correlation matrix. The MINRES method minimizes the sum of squares of the residuals in matrix $\boldsymbol{\Delta}$ where:

$$\boldsymbol{\Delta} = \boldsymbol{R_{yy}} - (\boldsymbol{AA'} + \boldsymbol{U^2})$$

This sum of squares of the residuals will be expressed as a statistical function which is minimized by the determination of the common factor matrix $\boldsymbol{A}$ and uniqueness $\boldsymbol{U^2}$. Two other statistical functions will be considered: maximum likelihood factor analysis and factor analysis by generalized least squares. However, there are several preliminary, general topics to be discussed before getting to the particular factoring methods.

Bentler & Bonnett (1980) provided a general least squares function as an extension of the function used by Jöreskog & Goldberger (1972) for their generalized least squares factor analysis. The Bentler & Bonett function may be written as:

$$G = \frac{1}{2} tr[\boldsymbol{W}(\boldsymbol{C_{yy}} - \boldsymbol{C_{zz}})]^2 \tag{9.1}$$

where $\boldsymbol{C_{yy}}$ is the observed covariance matrix and $\boldsymbol{C_{zz}}$ is the covariance matrix obtained from the factor model:

$$\boldsymbol{C_{zz}} = \boldsymbol{AA'} + \boldsymbol{U^2} \tag{9.2}$$

and $\boldsymbol{W}$ is a weight matrix. There is a problem, however, the matrix product $[\boldsymbol{W}(\boldsymbol{C_{yy}} - \boldsymbol{C_{zz}})]$ is not symmetric so that the trace of its square is not a sum of squares. This problem may be resolved when $\boldsymbol{W}$ is symmetric, positive definite so that it may be decomposed into the product of a matrix $\boldsymbol{T}$ times its transpose:

$$\boldsymbol{W} = \boldsymbol{TT'} \tag{9.3}$$

with $\boldsymbol{T}$ square and non-singular so that its inverse exists. Then:

$$
\begin{aligned}
tr[\boldsymbol{W}(\boldsymbol{C_{yy}}-\boldsymbol{C_{zz}})]^2 &= tr[\boldsymbol{W}(\boldsymbol{C_{yy}}-\boldsymbol{C_{zz}})\boldsymbol{W}(\boldsymbol{C_{yy}}-\boldsymbol{C_{zz}})] \\
&= tr[\boldsymbol{TT'}(\boldsymbol{C_{yy}}-\boldsymbol{C_{zz}})\boldsymbol{TT'}(\boldsymbol{C_{yy}}-\boldsymbol{C_{zz}})] \\
&= tr[\boldsymbol{T'}(\boldsymbol{C_{yy}}-\boldsymbol{C_{zz}})\boldsymbol{TT'}(\boldsymbol{C_{yy}}-\boldsymbol{C_{zz}})\boldsymbol{T}]
\end{aligned}
$$

so that:

$$
tr[\boldsymbol{W}(\boldsymbol{C_{yy}}-\boldsymbol{C_{zz}})]^2 = tr[\boldsymbol{T'}(\boldsymbol{C_{yy}}-\boldsymbol{C_{zz}})\boldsymbol{T}]^2 \tag{9.4}
$$

With matrices $\boldsymbol{C_{yy}}$ and $\boldsymbol{C_{zz}}$ being symmetric, matrix $[\boldsymbol{T'}(\boldsymbol{C_{yy}}-\boldsymbol{C_{zz}})\boldsymbol{T}]$ is symmetric so that the trace of its square is a sum of squares.

Scale freeness of a factoring method is considered to be very important by theoretically oriented individuals. This topic, which was discussed in Chapter 8, applies to all three methods of factor determination as described in this chapter. The idea is that when the attribute measures are rescaled the factor matrix will be unchanged except for a rescaling of the rows of the factor matrix in exactly the same fashion as the attributes were rescaled. The uniquenesses also will be rescaled in the same fashion. Let $\boldsymbol{D}$ be a rescaling diagonal matrix so that:

$$
\widetilde{\boldsymbol{C}}_{\boldsymbol{yy}} = \boldsymbol{DC_{yy}D}\,. \tag{9.5}
$$

With

$$
\widetilde{\boldsymbol{A}} = \boldsymbol{DA} \tag{9.6}
$$

$$
\widetilde{\boldsymbol{U}}^{\boldsymbol{2}} = \boldsymbol{DU^2D} \tag{9.7}
$$

so that:

$$
\widetilde{\boldsymbol{C}}_{\boldsymbol{zz}} = \widetilde{\boldsymbol{A}}\,\widetilde{\boldsymbol{A}}' + \widetilde{\boldsymbol{U}}^{\boldsymbol{2}} \tag{9.8}
$$

$$
\widetilde{\boldsymbol{C}}_{\boldsymbol{zz}} = \boldsymbol{DC_{zz}D} \tag{9.9}
$$

The modeled covariance matrix is rescaled in the same fashion as the observed covariance matrix. Jöreskog & Goldberger (1972) gave a general condition for scale freeness. Let ϕ be a general function yielding a numerical value. Their condition may be written as:

$$
\phi(\boldsymbol{C_{yy}},\boldsymbol{C_{zz}}) = \phi(\boldsymbol{DC_{yy}D},\boldsymbol{DC_{zz}D}) \tag{9.10}
$$

or

$$
\phi(\boldsymbol{C_{yy}},\boldsymbol{C_{zz}}) = \phi(\widetilde{\boldsymbol{C}}_{\boldsymbol{yy}},\widetilde{\boldsymbol{C}}_{\boldsymbol{zz}})\ . \tag{9.11}
$$

The application of this condition will be presented for all three of the methods as described in this chapter.

A large sample statistical function has been derived for each of the three methods to be described in this chapter. These functions result in tests whether a factor model with a specified number of attributes and number of factors exactly represent the population covariance matrix. That is, the lack of fit in a sample may be allocated to sampling fluctuations with no model error. While these statistical tests may be useful in decisions as to the numbers of factors to be established, these tests must be used with some care. A problem, which was discussed in Chapter 3, is generated by the proposition that almost all theoretic models are not correct representations of the real world. Then, there is an expectation that a very useful factor model will not exactly represent the population covariance matrix in any particular study. As indicated in Chapter 4, the discrepancy matrix, $\boldsymbol{\Delta}_y$, between an observed covariance matrix and a modeled covariance matrix is a sum of model error and sampling error. With a small sample many factor models might not be rejected due to the lack of power of the statistical tests while all reasonable factor models would be rejected with large sample data. The power of these statistical tests is a direct function of sample size. As a general proposition, statistical stability of results of an analysis is highly desirable. This does not indicate, however, that trivial, but stable, dimensions should be retained in a factor analysis. An analyst, nevertheless, must be alert to the possible small factor which may be made more important with a redesigned battery of attributes. There is a necessary judgment as to the number of scientifically meaningful factors involved in the observed covariance matrix. The statistical tests will help on the issue of statistical stability.

Some of the issues indicated in the preceding paragraph may be clarified by consideration of classes of the factor analysis model in the population which were discussed in Chapter 4. Equation (3.35) expresses the common factor model in the population with correlated factors

$$\boldsymbol{\Sigma}_{zz} = \boldsymbol{B}\boldsymbol{\Phi}\boldsymbol{B}' + \boldsymbol{U}^2 \,. \tag{3.35}$$

The corresponding formula for uncorrelated factors is obtained from equation (3.45)

$$\boldsymbol{\Sigma}_{zz} = \boldsymbol{A}\boldsymbol{A}' + \boldsymbol{U}^2 \tag{9.12}$$

Factor transformations among and between uncorrelated factors and correlated factors was discussed in Chapter 3. These equations provide the model in abstract terms. The general model may be termed the abstract common factor model. In contrast there are many concrete models for which the numerical values of the model are specified. These specifications include not only the number of attributes and the number of common factors but also the values of the uniqueness and numerical values in the common factor structure. This common factor structure includes all matrices $\boldsymbol{B}$, $\boldsymbol{\Phi}$, and $\boldsymbol{A}$ which are interrelated by factor transformations. A conditional abstract

model may be considered which involves specification of the number of attributes and number of common factors with the specification that any matrix $\boldsymbol{A}$ in the factor structure is of full column rank. The abstract model contains an infinite set of concrete models and a conditional abstract model contains a subset of concrete models. It is well known that any population covariance matrix, $\boldsymbol{\Sigma}_{yy}$, may be satisfied by equations (3.35) and (9.12) provided that a large enough number of common factors is employed. Thus, a population covariance matrix among surface attributes may be represented by a concrete model for n attributes and r_y common factors. However, when a common factor model is determined for r_z common factors with r_z less than r_y by any of the factoring methods, there is bound to be a discrepancy which may be termed model error as in Chapter 3 and defined, from equation (3.90) as:

$$\boldsymbol{\Delta}_{\Sigma} = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{zz} \quad . \tag{9.13}$$

A crucial point is that a specification of r_z along with the number of attributes defines a conditional abstract model which contains many concrete models.

Each of the factor fitting methods by a statistical function is defined in terms of an observed covariance matrix. However, each of these methods has a population counterpart obtained by substituting the population covariance matrix among the surface attributes for an observed covariance matrix. Since observed covariance matrices approach population covariance matrices with increase in sample size, each of these factor fitting techniques is statistically consistent so that the results using one of these methods approaches results that would be obtained from the population covariance matrix. Consider that the number of common factors defined for the fitted factor matrix is less than the number of common factors associated with the population covariance matrix among the observed attributes; that is: r_z is less than r_y . Each of the factor fitting methods will result in a solution in the conditional abstract model subset with the given number of attributes and number of common factors. There will be a model error, as previously described, since the population covariance matrix among the observed attributes can not be represented in this conditional abstract model subset. In a small sample the model error will be incorporated with considerable sampling error in the discrepancy between the observed covariance matrix and the model covariance matrix determined in the sample. The sampling error may hide the model error so that the model will not be rejected by the statistical test function. In comparison, there will be smaller sampling error in a larger sample so that the model error will not be hidden with a consequent that the model will be rejected by the statistical test function.

The concept of concrete factor analysis models is illustrated numerically in Tables 9.1, 9.2, and 9.3. A simulated population covariance matrix was generated by the Tucker-Koopman-Linn technique described in Chapter 3. Table 9.1 presents the data generating parameters used in

Table 9.1

Data Generating Parameters for

Creation of Simulated Population Correlation Matrix

	Major Domain Factors			Unique Variance	Unique Variance
	1	2	3		
1	.95	.00	.00	.0000	.0975
2	.95	.00	.00	.0000	.0975
3	.95	.00	.00	.0000	.0975
4	.95	.00	.00	.0000	.0975
5	.00	.70	.00	.4125	.0975
6	.00	.70	.00	.4125	.0975
7	.00	.70	.00	.4125	.0975
8	.00	.00	.45	.7000	.0975
9	.00	.00	.45	.7000	.0975
10	.00	.00	.45	.7000	.0975

$\varepsilon = .20$; Number of Minor Factors = 50

Table 9.2

Simulated Population Correlation Matrix

Variables	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	.87	1.00								
3	.98	.85	1.00							
4	.98	.89	.97	1.00						
5	.00	.02	-.02	-.02	1.00					
6	-.07	.04	-.07	-.06	.51	1.00				
7	.01	.00	.01	.02	.50	.54	1.00			
8	-.07	.08	-.09	-.06	.01	.07	-.01	1.00		
9	.03	-.04	.05	.04	.03	-.05	-.02	.15	1.00	
10	.03	-.05	.04	-.01	-.03	-.07	-.07	.16	.22	1.00

Table 9.3

Concrete Models Obtained from Analysis of

Simulated Population Correlation Matrix*

	MINRES Solution					Maximum Likelihood Solution				
		Factor		Unique-			Factor		Unique-
Variable	1	2	3	ness	Variable	1	2	3	ness
1	.99	.00	.01	.01	1	.99	.00	.12	.01
2	.88	.06	-.03	.23	2	.93	.07	-.37	.00
3	.98	-.01	.02	.04	3	.98	-.01	.14	.03
4	1.00	.01	-.01	.01	4	.99	.00	.06	.03
5	-.02	.68	.08	.53	5	-.01	.68	.05	.54
6	-.06	.75	.01	.43	6	-.05	.76	-.10	.41
7	.00	.72	.00	.48	7	.00	.73	.14	.45
8	-.04	.03	.32	.90	8	-.04	.04	-.30	.90
9	.02	-.04	.46	.79	9	.02	-.03	.15	.98
10	.01	-.10	.49	.75	10	.01	-.09	.14	.97

*Factor matrices were orthogonally transformed to principal axes orientation.

this example. The first four attributes had very high loadings on the first major domain factors, the next three attributes had moderately high loadings on the second major domain factor while the last three attributes had only moderate loadings on the third major domain factor. The minor variances were set equal for all ten attributes to a small value of .0975. A considerable extent of lack of fit was introduced by use of a relatively large value of ϵ and small number of minor factors. Table 9.2 presents a simulated population correlation matrix (covariance matrix for this example). A different matrix could be produced using the same generating parameters with different random numbers in the generation of the minor factors. Note the larger correlations in the boxes corresponding to the attributes with high loadings on the three major domain factors as compared to the correlations not in the boxes. Concrete models were computed by two of the factor fitting methods to be described in this Chapter: MINRES and Maximum Likelihood Factoring methods.

Table 9.3 presents two concrete factor analysis models produced by the two methods of factor fitting using three common factors. MINRES results include all three major domain factors while the Maximum Likelihood results include the first two major domain factors but yielded a peculiar third factor which, undoubtedly, is associated with the lack of fit introduced in the data generation procedure. Our major point is that these two methods of factor fitting yielded different concrete models from the same population correlation matrix. When larger and larger, multidimensional normal samples are drawn from the population characterized by the simulated correlation matrix, application of the two methods of factor fitting will yield results approaching the concrete models produced in the population. The MINRES results will approach the MINRES solution in the population while the Maximum Likelihood results will approach the Maximum Likelihood solution in the population. In Monte Carlo studies with similar simulated covariance matrices and using samples as small as 100 , results indicate that MINRES results from the samples reflect the MINRES solution in the population and not the Maximum Likelihood solution in the population. Likewise, the Maximum Likelihood results in the samples reflect the Maximum Likelihood solution in the population and not the MINRES solution in the population. Thus, for these two methods of factor fitting, results for each method from samples estimate the parameters in the concrete model established in the population by that method.

In contrast to the statistical test functions discussed in preceding paragraphs is a topic of goodness of fit of a concrete model to the real world population covariance matrix. There have been several measures developed for structural models analysis, confirmatory factor analysis being a form of structural model analysis. These will be discussed in later Chapters. The first contribution to a coefficient for indicating goodness of fit was by Tucker & Lewis (1973) with a proposed reliability coefficient for Maximum Likelihood Factor Analysis. This coefficient is generalized in this chapter to Generalized Least Squares Factor Analysis and MINRES factor

fitting. Consider the discrepancy matrix, $\mathbf{\Delta}_{\boldsymbol{y}}$, between the observed covariance matrix and the modeled covariance matrix. In the Maximum Likelihood system the uniqueness is added into the diagonals and the resulting matrix standardized by pre and post multiplication by a diagonal matrix to having unit diagonal entries. The resulting matrix contains partial correlations after the common factors have been removed. Jöreskog's function F_r after determination of r common factors approximates the sum of squares of the entries on one side of the diagonal of this partial correlation matrix. A mean square, M_r , may be obtained by dividing F_r by the remaining degrees of freedom, df_r .

$$\mathrm{M}_r = \mathrm{F}_r/\mathrm{df}_r \ . \tag{9.14}$$

Having a mean square suggest a components of variation analogy with an intraclass correlation as a reliability coefficient. Let M_0 be the value of M for no common factors. For the variance components model let α_r be a variance associated with a model having r common factors, δ_r be a variance representing the deviation of the model from actuality, and ϵ_r be a variance associated with sampling. Using expected values, consider the components of variance as:

$$Exp(M_0) = \alpha_r + \delta_r + \epsilon_r \ , \tag{9.15}$$

$$Exp(M_r) = \delta_r + \epsilon_r \ . \tag{9.16}$$

The intraclass correlation which is used as a reliability coefficient is defined by:

$$\rho_r = \frac{\alpha_r}{\alpha_r + \delta_r} \ . \tag{9.17}$$

To obtain a solution in terms of observed coefficients, Tucker & Lewis substitute observed values of the M coefficients for the expected values. A value of ϵ_r is obtained by assuming that, if the factor model fit exactly in r common factors, δ_r would equal zero and the chi square would equal unity. In this case, $\epsilon_r$would equal M_r which, in turn, would equal $1/\mathrm{N}_r^{\bigstar}$ with:

$$\mathrm{N}_r^{\bigstar} = \mathrm{N} - 1 - (2n + 5)/6 - 2r/3 \tag{9.18}$$

Solution for the reliability coefficient yields:

$$\rho_r \doteq \frac{\mathrm{M}_0 - \mathrm{M}_r}{\mathrm{M}_0 - 1/\mathrm{N}_r^{\bigstar}} \ . \tag{9.19}$$

Tucker & Lewis (1973, page 5) indicate that: "This reliability coefficient may be interpreted as indicating how well a factor model with r common factors represents the covariances among the attributes for a population of objects". This coefficient has been adapted from Maximum Likelihood Factor Analysis to General Least Squares Factor Analysis and MINRES factor fitting.

9.1. Maximum Likelihood Factor Analysis.

Lawley (1940) developed the method of maximum likelihood factor analysis at a time when the field had considerable need for a sound theoretic statistical foundation. The centroid method was in use at the Thurstone laboratory for analyses of major factor analytic studies. The number of factors was a major problem. In that the maximum likelihood method provided a statistical test for significance of factors this was a most promising development. However, at that time, computing facilities were very limited so that the extensive computations required by the maximum likelihood method made this method unavailable. Application remained limited until developments of modern digital computers made them sufficiently powerful to accomplish the required computations. Lord (1956) used Whirlwind I computer in the analysis of a covariance matrix among 39 attributes. He used an iterative technique suggested by Lawley (1942). Convergence was very slow. This method had several other problems such as converging on secondary maximum. Jöreskog (1967) contributed advanced procedures which could use more powerful analysis methods such as the Newton-Raphson iterations. With the computer developments and Jöreskog's contributions, maximum likelihood factor analysis is quite feasible. This method has many desirable statistical properties such as consistency, normality, efficiency.

Theoretic developments of maximum likelihood factor analysis starts with a population covariance matrix, $\boldsymbol{\Sigma}$, and sample score vectors, $\boldsymbol{y}$, which are multivariate normal drawn from the population characterized by the population covariance matrix. Note the normality assumption which is added to the assumptions of the factor analysis model which have been given previously. The number of attributes is taken to be n and the sample size is N . Consider all possible observed covariance matrices, $\boldsymbol{C}$, for samples of size N among n attributes when the score vectors are drawn multivariate normal from a population having a covariance matrix $\boldsymbol{\Sigma}$. Wishart (1928) gave the density function for these observed covariance matrices.

$$\phi(\boldsymbol{C}\big|\boldsymbol{\Sigma},\mathrm{N}) = \mathrm{K}\big|\boldsymbol{\Sigma}\big|^{-\frac{1}{2}(\mathrm{N}-1)}\big|\boldsymbol{C}\big|^{\frac{1}{2}(\mathrm{N}-n-2)}\mathrm{EXP}[-\frac{1}{2}(\mathrm{N}-1)tr(\boldsymbol{C}\boldsymbol{\Sigma}^{-1})] \qquad (9.20)$$

where K is a constant involving only N and n .

The likelihood of a given observed covariance matrix, $\boldsymbol{C}_{yy}$, is taken equal to the value of the density function for that observed covariance matrix. It is illogical to consider that the probability of this covariance matrix is other than unity since this covariance matrix already has occurred. Thus, the likelihood is a measure of typicality versus rarity of an observation. In the present case, the likelihood of an observed $\boldsymbol{C}_{yy}$ is given by:

$$\mathrm{L}(\boldsymbol{C}_{yy}\big|\boldsymbol{\Sigma},\mathrm{N}) = \mathrm{K}\big|\boldsymbol{\Sigma}\big|^{-\frac{1}{2}(\mathrm{N}-1)}\big|\boldsymbol{C}_{yy}\big|^{\frac{1}{2}(\mathrm{N}-n-2)}\mathrm{EXP}[-\frac{1}{2}(\mathrm{N}-1)tr(\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{yy})] \qquad (9.21)$$

Subsequent developments and discussions are expedited by a logarithmic conversion to the log-likelihood function:

$$\Lambda = ln[\mathrm{L}(\boldsymbol{C_{yy}}\big|\boldsymbol{\Sigma}\,,\mathrm{N})]\ . \tag{9.22}$$

Substitution from equation (9.21) yields:

$$\begin{aligned}\Lambda = ln(\mathrm{K}) - \frac{1}{2}(\mathrm{N}-1)[\ ln\big|\boldsymbol{\Sigma}\big|] + \frac{1}{2}(\mathrm{N}-\mathrm{n}-2)[\ ln\ \big|\boldsymbol{C_{yy}}\big|] \\ - \frac{1}{2}(\mathrm{N}-1)[tr(\boldsymbol{\Sigma}^{-1}\boldsymbol{C_{yy}})]\ .\end{aligned} \tag{9.23}$$

A statistical function is developed using the likelihood ratio which compares a theoretical model which is to be fit to the data with a more general comparison model. The comparison model is to include the theoretical model as a special case but has a larger number of parameters than does the theoretical model. For present considerations, the comparison model is to have a sufficient number of parameters to completely account for the observed covariance matrix, $\boldsymbol{C_{yy}}$. For each of the theoretic model and the comparison model an estimate of the population covariance matrix is developed along with the likelihood of the solution. For the theoretic model, hereafter designated model m , the estimate is $\widehat{\Sigma}_{\mathbf{m}}$. The comparison model is designated model M and yields estimate $\widehat{\boldsymbol{\Sigma}}_{\mathbf{M}}$. The likelihood functions for these two models are designated $\mathrm{L_m}$ and $\mathrm{L_M}$ while the log-likelihood functions are designated Λ_{m} and Λ_{M}. The likelihood ratio, λ **,** is defined by:

$$\lambda = \mathrm{L_m}/\mathrm{L_M}\ . \tag{9.24}$$

A coefficient ξ based on the likelihood ratio provides an approximate chi square for a statistical test on whether or not the theoretic model is a complete representation of the population covariance matrix.

$$\xi = -2[\ ln\,(\lambda)]\ . \tag{9.25}$$

Substitution from equation (9.24) yields:

$$\begin{aligned}\xi &= -2[ln(L_m) - ln(L_M)] \\ &= -2[\Lambda_m - \Lambda_M] \\ &= (\mathrm{N}-1)\{ln\big|\Sigma_m\big| + tr(\boldsymbol{\Sigma}_m^{-1}\boldsymbol{C_{yy}}) - ln\big|\boldsymbol{\Sigma_M}\big| - tr(\boldsymbol{\Sigma_M^{-1}}\boldsymbol{C_{yy}})\}\ .\end{aligned} \tag{9.26}$$

Let η(m) and η(M) be the number of parameters involved in models m and M . The number of degrees of freedom for ξ is $\nu(\xi)$ and equals the difference between the number of parameters used in the two models:

$$\nu(\xi) = \eta(\mathrm{M}) - \eta(\mathrm{m})\ . \tag{9.27}$$

At this point it is necessary to trace through the maxima - minima relations among the likelihood, L($\boldsymbol{C_{yy}}\big|\boldsymbol{\Sigma}$,**N**) , the log-likelihood, Λ , likelihood ratio, $\boldsymbol{\lambda}$, and coefficient ξ . From equation (9.22), Λ is an increasing monotonic function of L($\boldsymbol{C_{yy}}\big|\boldsymbol{\Sigma}$,N) so that maximum Λ occurs with the maximum L($\boldsymbol{C_{yy}}\big|\boldsymbol{\Sigma}$, N) ; that is, maximizing the log-likelihood maximizes the likelihood. Consider that the comparison model is fixed so that L_M and Λ_M can be considered to be constants. Then, the likelihood ratio, λ , is a monotonic increasing function of L_m so that a maximum likelihood corresponds to a maximum likelihood ratio. However, from equation (9.25), the coefficient ξ is a decreasing monotonic function of λ so that the maximum likelihood is associated with a minimum coefficient ξ . Analysis procedures are required to minimize ξ .

Comparison model M is considered first. The number of parameters in this model equals the number of variances and independent covariances in $\boldsymbol{C_{yy}}$ (since matrix $\boldsymbol{C_{yy}}$ is symmetric, the number of independent covariances equals the number of entries on one side of the diagonal). Then:

$$\eta(M) = \frac{1}{2}n(n+1) \ . \tag{9.28}$$

With this number of parameters an exact solution may be obtained so that:

$$\boldsymbol{\widehat{\Sigma}_M = C_{yy}} \ . \tag{9.29}$$

Then:

$$\boldsymbol{\widehat{\Sigma}_M^{-1} C_{yy} = I} \tag{9.30}$$

and

$$tr(\boldsymbol{\widehat{\Sigma}_M^{-1} C_{yy}}) = n \ . \tag{9.31}$$

Taking the estimated theoretic covariance matrix to be $\boldsymbol{\widehat{\Sigma}_m}$ and substituting from equations (9.29) and (9.31) into equation (9.26) yields:

$$\xi = (N-1)\{ln\big|\boldsymbol{\widehat{\Sigma}_m}\big| + tr(\boldsymbol{\widehat{\Sigma}_m^{-1} C_{yy}}) - ln\big|\boldsymbol{C_{yy}}\big| - n\} \ . \tag{9.32}$$

Jöreskog (1967) suggested a convenient function free from the effects of sample size as:

$$F = \xi/(N-1) \ . \tag{9.33}$$

$$F = ln\big|\boldsymbol{\widehat{\Sigma}_m}\big| + tr(\boldsymbol{\widehat{\Sigma}_m^{-1} C_{yy}}) - ln\big|\boldsymbol{C_{yy}}\big| - n \ . \tag{9.34}$$

The estimated theoretic covariance matrix is to be established so as to minimize this function F with the number of free parameters used in this estimate being η(m) . This function is the basis

not only for the maximum likelihood procedures in exploratory factor analysis but also in confirmatory factor analysis and the more general covariance structure analyses. Only the solution for exploratory factor analysis will be considered in this chapter.

The estimated theoretic covariance matrix may be taken as our matrix $\boldsymbol{C_{zz}}$ for a sample. This usage along with the factor analysis model for uncorrelated factors may be expressed as:

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{m}} = \boldsymbol{C_{zz}} = \boldsymbol{AA'} + \boldsymbol{U^2} \tag{9.35}$$

where $\boldsymbol{A}$ is the estimated factor matrix and $\boldsymbol{U^2}$ is the estimated uniqueness matrix. Note that this usage does not involve the discrepancy matrix $\boldsymbol{\Delta_y}$ which we saw in Chapter 4 included both sampling error and model error. The maximum likelihood procedure ignores the model error substituting, instead, the statistical test to indicate a solution for which there is no model error. In practice, the statistical test is taken as advisory to indicate small model error when the chi square indicates that the estimated model should be rejected with only a small probability associated with the chi square. With a very tiny probability the model should be rejected.

Early computing procedures suggested by Lawley (1940, 1941, 1949) involved alternating iterations between solution for matrix $\boldsymbol{A}$ conditioned on a trial $\boldsymbol{U^2}$ and solution for $\boldsymbol{U^2}$ conditioned on a trial matrix $\boldsymbol{A}$. Since more effective methods have been developed subsequently by Jöreskog (1967, 1969) and by Jennrick and Robinson (1969) only a summary of these early methods will be presented here. The conditional solution for $\boldsymbol{A}$ given a trial $\boldsymbol{U^2}$ is obtained by setting the derivative of F with respect to $\boldsymbol{A}$ equal to zero with the results that follow. Obtain the eigenvalues, q_k , and unit length eigenvectors $\underline{\mathbf{v}}_{\boldsymbol{k}}$, of the matrix $\boldsymbol{U^{-1}C_{yy}U^{-1}}$, with matrix $\boldsymbol{Q}$ diagonal and matrix $\boldsymbol{V}$ containing the eigenvectors as columns. The eigenvalues are arranged in descending algebraic order. With $\boldsymbol{C_{yy}}$ symmetric this solution may be expressed by:

$$\boldsymbol{U^{-1}C_{yy}U^{-1}} = \boldsymbol{VQV'} \tag{9.36}$$

For r factors, let $\boldsymbol{V_r}$ contain the first r eigenvectors and $\boldsymbol{Q_r}$ be an $r \times r$ diagonal matrix containing the first r eigenvalues. Matrix $\boldsymbol{A}$ is:

$$\boldsymbol{A} = \boldsymbol{UV_r}(\boldsymbol{Q_r} - \boldsymbol{I})^{\frac{1}{2}} \tag{9.37}$$

Solution for $\boldsymbol{U^2}$ is obtained by setting the derivative of F with respect to $\boldsymbol{U^2}$ for a given trial matrix $\boldsymbol{A}$ equal to zero. This results in:

$$\boldsymbol{U^2} = diag(\boldsymbol{C_{yy}}) - diag(\boldsymbol{AA'}) \ . \tag{9.38}$$

The procedure was to start with a trial $\boldsymbol{U^2}$, compute a factor matrix with equations (9.36) and (9.37), then compute a new $\boldsymbol{U^2}$ with equation (9.38) and return to the solution for a new factor

matrix. These cycles were repeated until there were minimal changes in values from one cycle to the next. A point to note is that the $\boldsymbol{U^2}$ is set such that the residual diagonal is set to zero. This results in the communality of each attribute plus the uniqueness for that attribute summing to the observed variance in the diagonal of $\boldsymbol{C_{yy}}$. In addition to this procedure converging slowly it stopped for some data before true convergence was obtained. A further major problem occurred when, for some data on a cycle, the iterated values of $\boldsymbol{U^2}$ contained one or more negative entries. In this case $\boldsymbol{U^{-1}}$ did not exist as a real number and the solution stopped. This is known as a generalized Heywood (1931) case. With these problems a more effective solution was very desirable.

Two iterative methods are possible for solution for the $\boldsymbol{U^2}$ values: a gradient method and a Newton-Raphson method. The gradient method is to be considered first. Let f be a function of a vector $\underline{\boldsymbol{x}}$. The first derivative of f with respect to the entries in $\underline{\boldsymbol{x}}$ form a vector $\mathbf{g}$. This vector is termed the gradient vector. Element i of $\mathbf{g}$ is:

$$\mathrm{g}_i = \partial f / \partial x_i \quad . \tag{9.39}$$

Let $\underline{\boldsymbol{x}}_t$ be a trial vector which is to be improved by the gradient method. The gradient vector corresponding to $\underline{\boldsymbol{x}}_{(t+1)}$ is $\mathbf{g}_t$. The next trial vector, $\underline{\boldsymbol{x}}_{(t+1)}$is obtained by:

$$\underline{\boldsymbol{x}}_{(t+1)} = \underline{\boldsymbol{x}}_t + m\mathbf{g}_t \tag{9.40}$$

where m is a step size coefficient. To iterate to a maximum, m is to be positive; to iterate to a minimum, m is to be negative. Various techniques are available to indicate an advantageous value of step size to be used. With the absolute value of m infinitesimal, the iteration will follow what is called the orthogonal trajectory of the function f and eventually end at a maximum for a positive m or at a minimum for a negative m . The number of iterations may be reduced materially by using larger absolute values of m . Using quite large absolute values of m can lead to a change in $\underline{\boldsymbol{x}}$ which leads to either a decreased value of f during iterations to a maximum or an increase in f during iterations to a minimum. When this occurs, a trial could be repeated with a smaller step size. An advantageous step size might be determined by repeating a trial with a number of step sizes and choosing the step size which yields the best results. When a satisfactory step size has been established it could be carried over to subsequent trials. There are analytic methods for establishing approximations to the step size which yields the most desirable results for each trial. These methods, however, are beyond the scope for this book.

The Newton-Raphson iterative method uses both the first and second derivatives of function f with respect to elements of vector $\underline{\boldsymbol{x}}$. The first derivative vector $\mathbf{g}$ was discussed in the preceding paragraph. The second derivatives of f with respect to pairs of entries in vector $\underline{\boldsymbol{x}}$ are entries h_{ij} in matrix $\boldsymbol{H}$.

$$h_{ij} = \partial^2 f / \partial x_i \partial x_j \ . \tag{9.41}$$

For a trial vector $\underline{\boldsymbol{x_t}}$ the first and second derivatives are $\mathbf{g_t}$ and $\boldsymbol{H_t}$. Then, the next trial vector, $\underline{\boldsymbol{x}}_{(t+1)}$ is given by:

$$\underline{\boldsymbol{x}}_{(t+1)} = \underline{\boldsymbol{x_t}} - \boldsymbol{H_t^{-1}}\mathbf{g}_t \ . \tag{9.42}$$

This would be an exact solution for a quadratic function f and works well when the actual function f is approximately quadratic. This condition appears to be quite good in many cases near an optimum, either a maximum or a minimum. At times the iteration from one trial to the next the value of the function will go the wrong way (decrease for iteration to a maximum or increase for an iteration to a minimum). In such cases, the iteration should revert to the gradient method for one or more trials before attempting the Newton-Raphson procedure. The determinant of the second derivative matrix should be negative for iterations to a maximum and negative for iterations to a minimum.

Both the gradient and Newton-Raphson methods work best starting from good first estimates of the solution vector $\underline{\boldsymbol{x}}$.

In the application of the preceding iterative methods to the maximum likelihood solution, function F of equation (9.34) is transformed from a function of the estimated factor matrix $\boldsymbol{A}$ and estimated uniqueness $\boldsymbol{U^2}$ to a function of the eigenvalues of matrix $\boldsymbol{U^{-1}C_{yy}U^{-1}}$ eliminating estimated factor matrix $\boldsymbol{A}$ from the solution. When a solution for $\boldsymbol{U^2}$ is obtained the estimate of factor matrix $\boldsymbol{A}$ is obtained from equation (9.37). In order to avoid negative estimates of uniqueness a transformation of variables is taken from each u_i^2 to a ς_i .

$$\varsigma_i = ln(u_i^2)$$

$$u_i^2 = [EXP(\varsigma_i)] \tag{9.43}$$

With iteration on real valued ς_i , no uniqueness, u_i^2 , can become negative. A limit can be placed on ς_i not to be less than some negative number such as -10. The first and second derivatives of function F with respect to elements ς_i are obtained and used in a Newton-Raphson iteration.

Steps in the developments of the maximum likelihood solution will be illustrated with a small simulated sample covariance matrix. This sample covariance matrix was generated using the Tucker-Koopman-Linn procedure discussed in Chapter 3 with the generating parameters given in Table 9.4. The simulated covariance matrix for a sample of 200 is given in Table 9.5. This is considered as matrix $\boldsymbol{C_{yy}}$. First trial u_i^2 are given in Table 9.4 along with the corresponding ς_i . These initial u_i^2 are given by:

Table 9.4

Generating Parameters for Illustrative
Sample Covariance Matrix

	Major Domain Factor Weight	Unique Variance	Minor Domain Variance
1	15	30	45
2	20	45	30
3	10	15	15
4	25	70	65
5	12	15	10

$\varepsilon = .05$ Number of Minor Factors = 135
Sample Size = 200

Table 9.5

Illustrative Sample Covariance Matrix, C_{yy}
N = 200

	1	2	3	4	5
1	277.6				
2	307.8	496.1			
3	148.0	215.4	131.4		
4	381.3	522.5	266.2	772.0	
5	178.0	264.8	122.1	314.6	177.8

$$u_i^2 = 1/c^{ii} \tag{9.44}$$

where c^{ii} is the i'th diagonal entry of $\boldsymbol{C}_{yy}^{-1}$.

The first step in the development is to obtain the solution for the factor matrix conditional on a matrix $\boldsymbol{U^2}$. Partial derivatives of function F of equation (9.34) with respect to the entries in $\boldsymbol{A}$. Reference will be made to $\widehat{\boldsymbol{\Sigma}}_m$ as given in equation (9.35).

$$\begin{aligned} \partial F/\partial a_{\text{jk}} &= 2(\widehat{\boldsymbol{\Sigma}}_m^{-1})_j' \boldsymbol{A_k} - 2(\widehat{\boldsymbol{\Sigma}}_M^{-1})_j' \boldsymbol{C}_{yy}(\widehat{\boldsymbol{\Sigma}}_m^{-1})\boldsymbol{A_k} \\ &= 2(\widehat{\boldsymbol{\Sigma}}_m^{-1})_j' (\boldsymbol{I} - \boldsymbol{C}_{yy}\widehat{\boldsymbol{\Sigma}}_m^{-1})\boldsymbol{A_k} \end{aligned} \tag{9.45}$$

$(\widehat{\boldsymbol{\Sigma}}_m^{-1})_j$ is the j'th column of $(\widehat{\boldsymbol{\Sigma}}_m^{-1})$; $\boldsymbol{A_k}$ is the k'th column of $\boldsymbol{A}$. When the derivatives for all entries in $\boldsymbol{A}$ are set to zero the following result is obtained:

$$2(\widehat{\boldsymbol{\Sigma}}_m^{-1})(\boldsymbol{I} - \boldsymbol{C}_{yy}\widehat{\boldsymbol{\Sigma}}_m^{-1})\boldsymbol{A} = \boldsymbol{0}\,.$$

Canceling out the first $2(\widehat{\boldsymbol{\Sigma}}_m^{-1})$ yields:

$$(\boldsymbol{I} - \boldsymbol{C}_{yy}\widehat{\boldsymbol{\Sigma}}_m^{-1})\boldsymbol{A} = \boldsymbol{0}\ . \tag{9.46}$$

Following is a most interesting relation which is used in equation (9.45).

$$\widehat{\boldsymbol{\Sigma}}_m^{-1} = \boldsymbol{U^{-2}} - \boldsymbol{U^{-2}A}(\boldsymbol{I} + \boldsymbol{A'U^{-2}A})^{-1}\boldsymbol{A'U^{-2}} \tag{9.47}$$

This relation substituted into equation (9.46) yields:

$$\begin{aligned} (\boldsymbol{I} - \boldsymbol{C}_{yy}\widehat{\boldsymbol{\Sigma}}_m^{-1})\boldsymbol{A} &= [\boldsymbol{I} - \boldsymbol{C}_{yy}\boldsymbol{U^{-2}} + \boldsymbol{C}_{yy}\boldsymbol{U^{-2}A}(\boldsymbol{I} + \boldsymbol{A'U^{-2}A})^{-1}\boldsymbol{A'U^{-2}}]\boldsymbol{A} \\ &= \boldsymbol{0}\,. \end{aligned} \tag{9.48}$$

Algebraic operations yield

$$\boldsymbol{C}_{yy}\boldsymbol{U^{-2}A}(\boldsymbol{I} + \boldsymbol{A'U^{-2}A})^{-1} = \boldsymbol{A}$$

or:

$$\boldsymbol{C}_{yy}\boldsymbol{U^{-2}A} = \boldsymbol{A}(\boldsymbol{I} + \boldsymbol{A'U^{-2}A})$$

and:

$$(\boldsymbol{U^{-1}C}_{yy}\boldsymbol{U^{-1}})(\boldsymbol{U^{-1}A}) = (\boldsymbol{U^{-1}A})(\boldsymbol{I} + \boldsymbol{A'U^{-2}A}) \tag{9.49}$$

Given any situation, matrix $\boldsymbol{A}$ may be orthogonally transformed by columns such that $\boldsymbol{A'U^{-2}A}$ is an $r \times r$ diagonal matrix as is, also, $(\boldsymbol{I} + \boldsymbol{A'U^{-2}A})$. With:

$$(\boldsymbol{I} + \boldsymbol{A'U^{-2}A}) = \boldsymbol{Q}_r \tag{9.50}$$

equation (9.49) becomes:

$$(\boldsymbol{U}^{-1}\boldsymbol{C}_{yy}\boldsymbol{U}^{-1})(\boldsymbol{U}^{-1}\boldsymbol{A}) = (\boldsymbol{U}^{-1}\boldsymbol{A})\boldsymbol{Q}_r \ . \tag{9.51}$$

This equation is in the form for an eigen solution with eigenvalues $\boldsymbol{Q}_r$ and eigenvectors $(\boldsymbol{U}^{-1}\boldsymbol{A})$. These eigenvectors are not of unit length but have a length to satisfy equation (9.50). The solution for matrix $\boldsymbol{A}$ is given in equations (9.36) and (9.37).

Instead of using the conditional solution for $\boldsymbol{U}^2$ dependent on the factor matrix as was done for the alternating solution outlined earlier, and expression for function F is derived dependent upon $\boldsymbol{U}^2$ and the eigenvalues of $\boldsymbol{U}^{-1}\boldsymbol{C}_{yy}\boldsymbol{U}^{-1}$. Then, the derivatives of F with respect to elements of $\boldsymbol{U}^2$ are obtained so as to utilize the gradient and Newton-Raphson iterative procedures. The first step is to substitute the results in equation (9.37) into equation (9.35).

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}}_m &= \boldsymbol{U}\boldsymbol{V}_r(\boldsymbol{Q}_r - \boldsymbol{I})\boldsymbol{V}_r'\boldsymbol{U} + \boldsymbol{U}^2 \\ &= \boldsymbol{U}[\boldsymbol{V}_r(\boldsymbol{Q}_r - \boldsymbol{I})\boldsymbol{V}_r' + \boldsymbol{I}]]\boldsymbol{U}\end{aligned} \tag{9.52}$$

It is convenient to define an $n \times n$ diagonal matrix $\boldsymbol{Q}_{rI}$.

$$\boldsymbol{Q}_{rI} = \begin{bmatrix} \boldsymbol{Q}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \tag{9.53}$$

The full, $n \times n$, matrix of eigenvalues is partitioned into matrix $\boldsymbol{V}_r$, $n \times r$, and a second section $\boldsymbol{V}_2$, $n \times (n - r)$.

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{V}_r & \boldsymbol{V}_2 \end{bmatrix} \tag{9.54}$$

The matrix difference $(\boldsymbol{Q}_{\mathbf{r}\mathbf{I}} - \boldsymbol{I})$ is shown below.

$$\begin{bmatrix} \boldsymbol{Q}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} - \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} = \begin{bmatrix} (\boldsymbol{Q}_r - \boldsymbol{I}) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \tag{9.55}$$

With this result:

$$\boldsymbol{V}_r(\boldsymbol{Q}_r - \boldsymbol{I})\boldsymbol{V}_r' = \boldsymbol{V}(\boldsymbol{Q}_{rI} - \boldsymbol{I})\boldsymbol{V}' \tag{9.56}$$

which is important since it involves the complete matrix $\boldsymbol{V}$ of eigenvectors. Then from equation (9.52):

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}}_m &= \boldsymbol{U}[\boldsymbol{V}(\boldsymbol{Q}_{rI} - \boldsymbol{I})\boldsymbol{V}' + \boldsymbol{I}]\boldsymbol{U} \\ &= \boldsymbol{U}[\boldsymbol{V}\boldsymbol{Q}_{rI}\boldsymbol{V}']\boldsymbol{U} \ . \end{aligned} \tag{9.57}$$

Functions of $\widehat{\boldsymbol{\Sigma}}_m$ are to be obtained to be entered into the equation for function F in equation (9.34). First is the determinant of $\widehat{\boldsymbol{\Sigma}}_m$.

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}}_m &= |\boldsymbol{U}|\bullet|\boldsymbol{V}\boldsymbol{Q}_{rI}\boldsymbol{V}'|\bullet|\boldsymbol{U}| = |\boldsymbol{U}|^2\bullet|\boldsymbol{V}|\bullet|\boldsymbol{Q}_{rI}|\bullet|\boldsymbol{V}'| \\ &= |\boldsymbol{U}^2|\bullet|\boldsymbol{Q}_{rI}|\end{aligned}$$

So that:

$$ln|\widehat{\boldsymbol{\Sigma}}_m| = ln|\boldsymbol{U}^2| + \sum_{k=1}^{r} ln(q_k) \ . \tag{9.58}$$

Next is the trace of $\widehat{\boldsymbol{\Sigma}}_m^{-1}\boldsymbol{C}_{yy}$. Note that:

$$\widehat{\boldsymbol{\Sigma}}_m^{-1} = \boldsymbol{U}^{-1}[\boldsymbol{V}\boldsymbol{Q}_{rI}^{-1}\boldsymbol{V}']\boldsymbol{U}^{-1}$$

so that:

$$\widehat{\boldsymbol{\Sigma}}_m^{-1}\boldsymbol{C}_{yy} = \boldsymbol{U}^{-1}[\boldsymbol{V}\boldsymbol{Q}_{rI}^{-1}\boldsymbol{V}']\boldsymbol{U}^{-1}\boldsymbol{C}_{yy} \ . \tag{9.59}$$

From equation (9.36):

$$\boldsymbol{U}^{-1}\boldsymbol{C}_{yy} = \boldsymbol{V}\boldsymbol{Q}\boldsymbol{V}'\boldsymbol{U}$$

so that:

$$\begin{aligned}\widehat{\boldsymbol{\Sigma}}_m^{-1}\boldsymbol{C}_{yy} &= \boldsymbol{U}^{-1}[\boldsymbol{V}\boldsymbol{Q}_{rI}^{-1}\boldsymbol{V}']\boldsymbol{V}\boldsymbol{Q}\boldsymbol{V}'\boldsymbol{U} \\ &= \boldsymbol{U}^{-1}\boldsymbol{V}\boldsymbol{Q}_{rI}^{-1}\boldsymbol{Q}\boldsymbol{V}'\boldsymbol{U} \ .\end{aligned}$$

From the above with algebraic operations:

$$\begin{aligned}tr(\widehat{\boldsymbol{\Sigma}}_m^{-1}\boldsymbol{C}_{yy}) &= tr(\boldsymbol{U}^{-1}\boldsymbol{V}\boldsymbol{Q}_{rI}^{-1}\boldsymbol{Q}\boldsymbol{V}'\boldsymbol{U}) \\ &= tr(\boldsymbol{Q}_{rI}^{-1}\boldsymbol{Q}) \ . \end{aligned} \tag{9.60}$$

Following is an illustration of the product $\boldsymbol{Q}_{rI}^{-1}\boldsymbol{Q}$.

$$\begin{bmatrix} \boldsymbol{Q}_r^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{Q}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_2 \end{bmatrix}$$

Diagonal matrix $\boldsymbol{Q}_2$ contains the last $(n - r)$ eigenvalues. Then:

$$tr(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{m}}^{-1}\boldsymbol{C_{yy}}) = r + \sum_{k=r+1}^{n} q_k \ . \tag{9.61}$$

An expression for $ln|\boldsymbol{C_{yy}}|$ is needed. From equation (9.36):

$$\boldsymbol{C_{yy}} = \boldsymbol{UVQV'U}$$

so that:

$$|\boldsymbol{C_{yy}}| = |\boldsymbol{U}|\bullet|\boldsymbol{V}|\bullet|\boldsymbol{Q}|\bullet|\boldsymbol{V'}|\bullet|\boldsymbol{U}|$$

Then:

$$\begin{aligned} ln|\boldsymbol{C_{yy}}| &= ln|\boldsymbol{U^2}| + ln|\boldsymbol{Q}| \\ &= ln|\boldsymbol{U^2}| + \sum_{k=1}^{n} ln(q_k) \ . \end{aligned} \tag{9.62}$$

A reduced expression for function F is obtained by substitution into equation (9.34) the results given in equations (9.58), (9.61) and (9.62):

$$F = ln|\boldsymbol{U^2}| + \sum_{k=1}^{r} ln(q_k) + r + \sum_{k=r+1}^{n} q_k - ln|U^2| - \sum_{k=1}^{n} ln(q_k) - n$$

Which reduces to:

$$F = \sum_{k=r+1}^{n} [q_k - ln(q_k)] - (n-r) \ . \tag{9.63}$$

This is a most interesting and important representation of function F. It provides the basis for further mathematical developments of the derivatives of F with respect to the uniqueness u_i^2.

Operations to this point in the developments are illustrated for the example in Table 9.6 and 9.7 for the first trial of the iterative procedure. As discussed earlier, the initial trial uniqueness are given in the first column of Table 9.6. Matrix $\boldsymbol{U^{-1}C_{yy}U^{-1}}$ is given in Table 9.7 along with its eigenvectors and eigenvalues (these being the q_k). The illustrative solution is for a single factor so that $r = 1$. Substitution of the last four eigenvalues into equation (9.63) yields the value of function F given at the bottom of Table 9.6.

To obtain the derivatives of F with respect to the u_i, the derivatives of the eigenvalues, q_k, and elements, v_{ik} , of the eigenvectors are needed. Jöreskog and Goldberger (1972, see

Appendix A2) developed the derivatives for matrix $\boldsymbol{U}\boldsymbol{C}_{\boldsymbol{yy}}^{-1}\boldsymbol{U}$. An analogous development for matrix $\boldsymbol{U}^{-1}\boldsymbol{C}_{\boldsymbol{yy}}\boldsymbol{U}^{-1}$ yields:

$$(\partial q_k/\partial u_i) = (-2q_k/u_i)v_{ik}^2 \tag{9.64}$$

$$(\partial v_{ik}/\partial u_i) = (-1/u_i)v_{ik}\sum_{m\neq k}^{n}[(q_k+q_m)/(q_k-q_m)]v_{im}^2 \ . \tag{9.65}$$

The partial derivative of F with respect to u_i follows.

$$(\partial F/\partial u_i) = \sum_{k=r+1}^{n}\left[(\partial q_k/\partial u_i) - (1/q_k)(\partial q_k/\partial u_i)\right] \ .$$

This equation with equation (9.64) yields:

$$(\partial F/\partial u_i) = (-2/u_i)\{\sum_{k=r+1}^{n} q_k v_{ik}^2 - \sum_{k=r+1}^{n} v_{ik}^2\} \ . \tag{9.66}$$

For subsequent steps equation (9.66) is converted to involving only the first r eigenvalues and eigenvectors. Note that:

$$\sum_{k=1}^{r} q_k v_{ik}^2 + \sum_{k=r+1}^{n} q_k v_{ik}^2 = \sum_{k=1}^{n} q_k v_{ik}^2 = c_{ii}u_i^{-2}$$

so that:

$$\sum_{k=r+1}^{n} q_k v_{ik}^2 = c_{ii}u_i^{-2} - \sum_{k=1}^{r} q_k v_{ik}^2 \ .$$

Also:

$$\sum_{k=1}^{r} v_{ik}^2 + \sum_{k=r+1}^{n} v_{ik}^2 = \sum_{k=1}^{n} v_{ik}^2 = 1$$

so that:

$$\sum_{k=r+1}^{n} v_{ik}^2 = 1 - \sum_{k=1}^{r} v_{ik}^2 \ .$$

With the preceding results equation (9.66) becomes:

$$(\partial F/\partial u_i) = (-2/u_i)\{c_{ii}u_i^{-2} - \sum_{k=1}^{r}(q_r - 1)v_{ik}^2 - 1\} \ . \tag{9.67}$$

Table 9.6
Trial 1 Uniqueness and Zeta

	Uniqueness	Zeta (ζ)
1	70.704	4.259
2	70.642	4.258
3	30.639	3.422
4	143.727	4.968
5	30.416	3.415

Function F = .15477

Table 9.7

Trial 1 Matrix ($U^{-1}C_{yy}U^{-1}$) with Eigen Solution

$U^{-1}C_{yy}U^{-1}$

	1	2	3	4	5
1	3.926	4.355	3.180	3.782	3.838
2	4.355	7.023	4.630	5.185	5.713
3	3.180	4.630	4.289	4.011	4.000
4	3.782	5.185	4.011	5.371	4.758
5	3.838	5.713	4.000	4.758	5.846

Eigenvectors

	1	2	3	4	5
1	.371	.254	.675	.527	.256
2	.529	-.424	-.233	.405	-.567
3	.391	.467	-.665	.157	.402
4	.451	.473	.197	-.588	-.434
5	.475	-.580	.094	-.434	.514

Eigenvalues

1	2	3	4	5
23.037	1.078	.929	.845	.565

The diagonal cells of the second derivatives matrix are handled separately from the off-diagonal cells.

$$\partial^2 F/\partial u_i^2 = (2/u_i^2)\{c_{ii}u_i^{-2} - \sum_{k=1}^{r}(q_k - 1)v_{ik}^2 - 1\}$$
$$+ (2/u_i)\{2c_{ii}u_i^{-3} + \sum_{k=1}^{r} v_{ik}^2(\partial q_k/\partial u_i)$$
$$+ \sum_{k=1}^{r}(q_k - 1)v_{ik}^2(\partial v_{ik}/\partial u_i)\} \qquad (9.68)$$

Substitution from equations (9.64) and (9.65) into equation (9.68) yields with algebraic operations:

$$\partial^2 F/\partial u_i^2 = (2/u_i^2)\{c_{ii}u_i^{-2} - \sum_{k=1}^{r}(q_k - 1)v_{ik}^2 - 1\}$$
$$+ (4/u_i^2)\{c_{ii}u_i^{-2} - \sum_{k=1}^{r} q_k v_{ik}^2$$
$$- \sum_{k=1}^{r}(q_k - 1)v_{ik}^2 \sum_{m\neq k}^{n}[(q_k + q_m)/(q_k - q_m)]v_{im}^2\} \ . \qquad (9.69)$$

For the off-diagonal entrees in the second derivative matrix, i ≠ j:

$$\partial^2 F/\partial u_i \partial u_j = (-2/u_i)\{-\sum_{k=1}^{r} v_{ik}^2(\partial q_k/\partial u_j)$$
$$- 2\sum_{k=1}^{r}(q_k - 1)v_{ik}(\partial v_{ik}/\partial u_j)\} \qquad (9.70)$$

Substitution from equations (9.64) and (9.65) into equation (9.70) yields with algebraic operations:

$$\partial^2 F/\partial u_i \partial u_j = (-4/u_i u_j)\{\sum_{k=1}^{r} q_k v_{ik}^2 v_{jk}^2$$
$$+ \sum_{k=1}^{r}(q_k - 1)v_{ik}v_{jk}\sum_{m\neq k}^{n}[(q_k + q_m)/(q_k - q_m)]v_{im}v_{jm}\} \qquad (9.71)$$

In order to avoid problems involving negative u_i^2 values a transformation of variables from u_i^2 to ς_i indicated in equation (9.43) is taken. The needed derivatives to accomplish this transformation follow.

$$\partial F/\partial \varsigma_i = (u_i/2)(\partial F/\partial u_i) \ ; \qquad (9.72)$$

$$\partial^2 F/\partial \varsigma_i^2 = (u_i/4)(\partial F/\partial u_i) + (u_i^2/4)(\partial^2 F/\partial u_i^2) \; ; \tag{9.73}$$

$$\partial^2 F/\partial \varsigma_i \partial \varsigma_j = (u_i u_j/4)(\partial^2 F/\partial u_i \partial u_j) \qquad \text{for } i \neq j \; . \tag{9.74}$$

For the first derivative of F with respect to ς_i , equation (9.72) is used with equations (9.67) to obtain:

$$\partial F/\partial \varsigma_i = 1 + \sum_{k=1}^{r} (q_i - 1) v_{ik}^2 - c_{ii}^{-2} u_i \; . \tag{9.75}$$

For the diagonal entries in the second derivative matrix, equation (9.73) is used with equation (9.67) and (9.69) to obtain:

$$\begin{aligned} \partial^2 F/\partial \varsigma_i^2 = & -\sum_{k=1}^{r} q_k v_{ik}^4 - \sum_{k=1}^{r} (q_k - 1) v_{ik}^2 \sum_{m \neq k}^{n} [(q_k + q_m)/(q_k - q_m)] v_{im}^2 \\ & + c_{ii} u_i^2 \; . \end{aligned} \tag{9.76}$$

For the off-diagonal entries in the second derivative matrix, equation (9.74) is used with equation (9.71) to obtain, i ≠ j :

$$\begin{aligned} \partial^2 F/\partial \varsigma_i \partial \varsigma_j = & -\sum_{k=1}^{r} q_k v_{ik}^2 v_{jk}^2 \\ & - \sum_{k=1}^{r} (q_k - 1) v_{ik} v_{jk} \sum_{m \neq k}^{n} [(q_k + q_m)/(q_k - q_m)] v_{im} v_{jm} \; . \end{aligned} \tag{9.77}$$

Equation (9.75) gives the gradient for the gradient method of iteration while equations (9.76) and (9.77) give the second derivative matrix used in Newton-Raphson method of iteration. For the example, Table 9.8 presents these derivatives for the first trial for one factor. Note that the determinant of the second derivative matrix is positive which is necessary for a minimum solution.

Operation of the Newton-Raphson procedure involves solution of the linear equations indicated in equation (9.42). Table 9.9 gives the inverse of the second derivative matrix and the change vector, $-\boldsymbol{H}_1^{-1}\mathbf{g}_1$, for the first trial of the example. The Newton-Raphson iteration took six trials to converge for the one factor for the example. Progress over these trials is outlined in Table 9.10 which gives the value of the function F for each of these trials and the first derivative vector for each trial. The change in function F from trial to trial diminishes as the trials progress and, finally, shows no change to five decimals from trial 5 to trial 6. Entries in the first derivative vectors diminish, also, from trial to trial with all entries becoming zero to five decimal places at trial 6. The convergence test involves the entries in the first derivative vector. Convergence may

Table 9.8

Trial 1 Derivatives of Function F

	First	Second Derivative Matrix*				
	Derivative	1	2	3	4	5
1	.099	.674	.026	.026	-.005	.044
2	.155	.026	.452	.017	.097	-.027
3	.088	.026	.017	.657	-.011	.071
4	.121	-.005	.097	-.011	.564	.031
5	.120	.044	-.027	.071	.031	.535

Determinant of second derivative matrix = .056 .

Table 9.9

Trial 1 Derivatives Matrix Inverse
with change Vector for Zeta

	Change	Second Derivative Matrix Inverse				
	Vector	1	2	3	4	5
1	-.055	1.498	-.099	-.042	.037	-.126
2	-.102	-.099	2.319	-.079	-.412	.159
3	-.049	-.042	-.079	1.550	.056	-.209
4	-.070	.037	-.412	.056	1.854	-.138
5	-.075	-.126	.159	-.209	-.138	1.922

Table 9.10

Outline of Progress over Trials

Function F over Trials

1	2	3	4	5	6
.09872	.05343	.01014	.00038	.00000	.00000

First Derivative Vectors Over Trials

	1	2	3	4	5	6
1	.09872	.05343	.01014	.00038	.00000	.00000
2	.15481	.10005	.03060	.00353	.00015	.00000
3	.08828	.04741	.00880	.00029	-.00001	.00000
4	.12070	.06777	.01248	.00018	-.00003	.00000
5	.12004	.073.4	.01887	.00150	.00004	.00000

Table 9.11

Maximum Likelihood Results for One Factor

	Factor Matrix 1	Communality	Uniqueness
1	14.6	214.2	63.4
2	21.1	443.5	52.6
3	10.2	103.5	27.9
4	25.4	646.3	125.7
5	12.4	152.6	25.2

$\xi\ (\chi^2) = 19.9$ df = 5 p = .001

Tucker-Lewis Reliability = .974

be declared when every entry in the first derivative vector is less in absolute value than some preset small value.

The value of function F should be checked for each Newton-Raphson trial. In case the value of F does not decrease, the trail should be discarded and one or more gradient method trials attempted. In case the value of F does not decrease on a gradient trial, this trail should be repeated with a smaller step size. When a decrease in F has been obtained the Newton-Raphson iteration may resumed.

Table 9.11 presents the solution for one factor for the example. For each attribute, the communality plus the uniqueness equals the variance of that attribute in Table 9.5. Coefficient ξ uses a revised value for the number of cases, this revised number of cases which was suggested by Bartlett (1950) is given in equation (9.18) and termed $N_r^{\bigstar}$. From equation (9.33) using $N_r^{\bigstar}$

$$\xi = N_r^{\bigstar} F \tag{9.78}$$

The degree of freedom, $\nu(\xi)$, for ξ interpreted as a chi square involves the number of parameters used in the theoretic model, this number being the number of uniqueness plus the number of entries in the factor matrix less the number of entries which may be set to zero by an orthogonal transformation of the factors.

$$\eta(m) = n + nr - r(r-1)/2\,. \tag{9.79}$$

Substitution of this value along with the value of η(M) from equation (9.28) into equation (9.27) yields:

$$\nu(\xi) = n(n+1)/2 - n - nr + r(r-1)/2\ . \tag{9.80}$$

The value of ξ and $\nu(\xi)$ for the example for one factor is given in Table 9.11 along with the corresponding p for ξ as a chi square with the given number of degrees of freedom. With p equaling .001 the statistical hypothesis that a one factor solution would fit in the population should be rejected. This is not an unexpected result since the generation of the population correlation matrix included a number of minor factors. However, there is a question whether the one factor solution should be considered as satisfactory since only one major domain factor was used in the generation of the simulated covariance matrix. In many applications of the maximum likelihood factor method to real data there is a problem concerning acceptance of a solution when the statistical test indicates that the solution should be rejected. Any decision on the number of common factors to use should consider a number of items of information, not just the statistical test.

The Tucker-Lewis reliability coefficient was discussed earlier. For the present application, the value of M_0 for no common factors is required. From equation (9.38)

$$\boldsymbol{U}_0^2 = diag(\boldsymbol{C_{yy}})$$

so that from equation (9.35)

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{m0}} = \boldsymbol{U}_0^2 \ .$$

With this result substitution into equation (9.32) yields with algebraic operations

$$F_0 = ln\left|\boldsymbol{U_0^2}\right| - ln\left|\boldsymbol{C_{yy}}\right| = \sum_{i=1}^{n} ln(c_{ii}) - ln\left|\boldsymbol{C_{yy}}\right| \tag{9.81}$$

where c_{ii} are the diagonal entries in $\boldsymbol{C_{yy}}$. The degrees of freedom for zero common factors is:

$$\nu(\xi)_0 = n(n-1)/2 \ . \tag{9.82}$$

By equation (9.14):

$$M_0 = F_0/\nu(\xi)_0 \ . \tag{9.83}$$

Also, from equation (9.14), for r common factors:

$$M_r = F_r/\nu(\xi)_r \ . \tag{9.84}$$

The Tucker-Lewis reliability coefficient is given by equation (9.19), repeated here for convenience.

$$\rho_r \doteq \frac{M_0 - M_r}{M_0 - 1/N_r^{\bigstar}} \tag{9.19}$$

The Tucker-Lewis reliability for the example for one factor is given at the bottom of Table 9.11. This value of .974 indicates a quite good fit of the factor results. Experience has indicated that this reliability coefficient is a very useful measure of goodness of fit but is not of much use in deciding how many factors to be extracted. The number of factors should be decided on other grounds and, then, the reliability coefficient used to indicate the quality of the results.

A two factor solution is given in Table 9.12. While the first factor is quite close to the major domain loadings in Table 9.4 the second factor loadings are smaller and may have been introduced by some combination of the minor domain factors. Note that the uniqueness of attribute 2 is very small. This illustrates a frequent result obtained with maximum likelihood factor analysis: one or more of the uniqueness iterates to very small values and, in fact, for solutions not using the transformation from u_i^2 to ς_i the iterations tend to obtaining negative values of u^2 . During the iterations, on trial 3, the value of the function increased so that one gradient method trial was required. After this, the movement was slow so that a total of 26 trials was required before convergence was obtained. The chi square statistic yielded a p of .422 so that

Table 9.12

Maximum Likelihood Results for Two Factors

	Factor Matrix		Communality	Uniqueness
	1	2		
1	14.5	.3	211.3	66.3
2	21.3	-6.5	496.1	.0
3	10.1	.1	102.7	28.7
4	26.2	5.5	715.3	56.7
5	12.2	-.8	149.0	28.8

$\xi\ (\chi^2) = .64$ df = 1 p = .422

Tucker-Lewis Reliability = 1.003

the two factor solution would not be rejected. There is a problem, however, concerning subsequent factor transformations for the two factor matrix. Maybe, the second factor should be ignored as a small 'residual' dimension. This would result in accepting the first factor as it stands. Note that the first factor is very similar to the single factor for the one factor solution given in Table 9.11 so that, maybe, the one factor solution should be accepted even though the statistical test indicated that it should be rejected. The Tucker-Lewis reliability coefficient for the two factor solution is slightly greater than unity, a result which can be obtained due to the approximations involved with this coefficient.

9.2. Generalized Least Squares Factor Analysis

Jöreskog and Goldberger (1972) introduced generalized least squares factor analysis as a second statistical method for factor analysis which they based on Aitken's (1934-35) generalized least squares principle. Application of this principle chooses parameter estimates to minimize the function

$$G_p = \frac{1}{2} tr\{[\boldsymbol{\Sigma}_{yy}^{-1}(\boldsymbol{C}_{yy} - \boldsymbol{\Sigma}_{zz})]^2\} \tag{9.85}$$

where, as previously, $\boldsymbol{\Sigma}_{yy}$ is the population covariance matrix among the observed attributes, $\boldsymbol{C}_{yy}$ is the observed covariance matrix in a sample and $\boldsymbol{\Sigma}_{zz}$ is the modeled population covariance matrix. However, $\boldsymbol{\Sigma}_{yy}$ is not available so that the unbiased estimate $\boldsymbol{C}_{yy}$ is substituted to yield

$$\begin{aligned} G &= \frac{1}{2} tr\{[\boldsymbol{C}_{yy}^{-1}(\boldsymbol{C}_{yy} - \boldsymbol{\Sigma}_{zz})]^2\} \\ &= \frac{1}{2} tr\{(I - \boldsymbol{C}_{yy}^{-1}\boldsymbol{\Sigma}_{zz})^2\} \end{aligned} \tag{9.86}$$

Note that the function of (9.85) is an application of the general function given in equation (9.1) with $\boldsymbol{\Sigma}_{yy}^{-1}$ being the weight matrix $\boldsymbol{W}$. Since $\boldsymbol{\Sigma}_{yy}^{-1}$ and $\boldsymbol{C}_{yy}^{-1}$ are symmetric, function G represents a true least squares.

Application of the Jöreskog & Goldberger (1972) general condition for scale freeness given in equation (9.10) demonstrates that the generalized least squares criterion is scale free. The previous discussion of the scale free condition involved matrices $\boldsymbol{C}_{yy}$ and $\boldsymbol{C}_{zz}$. For the present case matrix $\boldsymbol{\Sigma}_{zz}$ is substituted for matrix $\boldsymbol{C}_{zz}$. Then, steps similar to those in equations (9.5) through (9.11) are followed to demonstrate the scale freeness of function G .

A number of relations exist between maximum likelihood factor analysis and generalized least squares factor analysis. These relations will be pointed out during discussion of the derivations that follow. A first relation is an approximate relation derived by Jöreskog &

Goldberger between the maximum likelihood function F and the generalized least squares function G . They show that:

$$F \approx G \tag{9.87}$$

which is important in the statistical distribution associated with G . Thus, coefficient may be defined for generalized least squares factor analysis as:

$$\xi = N_r^{\bigstar} G \tag{9.88}$$

where the value of $\mathrm{N}_r^{\bigstar}$ is given in equation (9.18). This definition parallels equation (9.78) for maximum likelihood factor analysis. With normality assumptions and a minimum G , ξ is distributed approximately as χ^2 with $\nu(\xi)$ degrees of freedom, see equation (9.80) for $\nu(\xi)$.

Solution for minimum G involves two steps, first is a conditional solution for the estimated factor matrix, $\boldsymbol{A}$, dependent on values of the uniqueness, second is an overall solution for the estimated uniqueness matrix $\boldsymbol{U}$. In the present a switch is being made from the population factor matrix and uniqueness to estimates derived from the sample data. The sample factor model is given in equation (9.2). With this switch the function G is written as:

$$G = \frac{1}{2} tr\{(\boldsymbol{I} - \boldsymbol{C}_{yy}^{-1}\boldsymbol{C}_{zz})^2\} \ . \tag{9.89}$$

To obtain the conditional minimum for G dependent upon values of the uniqueness a partial derivative of G with respect to the factor matrix $\boldsymbol{A}$ is obtained:

$$\frac{\partial G}{\partial \boldsymbol{A}} = 2\boldsymbol{C}_{yy}^{-1}(\boldsymbol{C}_{zz} - \boldsymbol{C}_{yy})\boldsymbol{C}_{yy}^{-1}\boldsymbol{A} \ . \tag{9.90}$$

This derivative was developed in their Appendix A1. Setting $\partial \mathrm{G}/\partial \boldsymbol{A}$ equal to zero with algebraic steps:

$$\boldsymbol{C}_{yy}^{-1}\boldsymbol{A} = \boldsymbol{C}_{zz}^{-1}\boldsymbol{A} \ . \tag{9.91}$$

Using equation (9.47) with algebraic steps:

$$(\boldsymbol{U}\boldsymbol{C}_{yy}^{-1}\boldsymbol{U})(\boldsymbol{U}^{-1}\boldsymbol{A}) = (\boldsymbol{U}^{-1}\boldsymbol{A})(\boldsymbol{I} + \boldsymbol{A}'\boldsymbol{U}^{-2}\boldsymbol{A})^{-1} \tag{9.92}$$

The similarity of this equation to equation (9.49) for maximum likelihood factor analysis is to be noted. With an orthogonal transformation of the factor matrix $\boldsymbol{A}$

$$(\boldsymbol{I} + \boldsymbol{A}'\boldsymbol{U}^{-2}\boldsymbol{A})^{-1} = \boldsymbol{\theta}_1 \tag{9.93}$$

where $\boldsymbol{\theta}_1$ is an $r \times r$ diagonal matrix. Comparison of this definition with equation (9.50) for maximum likelihood indicates that $\boldsymbol{\theta}_1$ is the inverse of $\mathbf{Q}_r$. The form of equation (9.92) indicates

that $\boldsymbol{\theta}_1$ contains eigenvalues of $\boldsymbol{U}\boldsymbol{C}_{yy}^{-1}\boldsymbol{U}$ and that $(\boldsymbol{U}^{-1}\boldsymbol{A})$ contains the corresponding eigenvectors scaled according to equation (9.93). A complete eigen solution of $\boldsymbol{U}\boldsymbol{C}_{yy}^{-1}\boldsymbol{U}$ yields:

$$(\boldsymbol{U}\boldsymbol{C}_{yy}^{-1}\boldsymbol{U}) = \begin{bmatrix} \boldsymbol{V}_1 & \boldsymbol{V}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\theta}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1' \\ \boldsymbol{V}_2' \end{bmatrix} \tag{9.94}$$

with $\begin{bmatrix} \boldsymbol{V}_1 & \boldsymbol{V}_2 \end{bmatrix}$ containing all n eigenvectors as unit vectors and is $n \times n$ orthonormal so that:

$$\begin{bmatrix} \boldsymbol{V}_1' \\ \boldsymbol{V}_2' \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1 & \boldsymbol{V}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I} \end{bmatrix} \tag{9.95}$$

and

$$\begin{bmatrix} \boldsymbol{V}_1 & \boldsymbol{V}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1' \\ \boldsymbol{V}_2' \end{bmatrix} = \boldsymbol{V}_1\boldsymbol{V}_1' + \boldsymbol{V}_2\boldsymbol{V}_2' = \boldsymbol{I} \tag{9.96}$$

$\boldsymbol{\theta}_2$ is $(n - r) \times (n - r)$ diagonal so that $\begin{bmatrix} \boldsymbol{\theta}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\theta}_2 \end{bmatrix}$ contains the n eigenvalues of $\boldsymbol{U}\boldsymbol{C}_{yy}^{-1}\boldsymbol{U}$.
Dimensions of the eigen solution are to be in ascending order of the eigenvalues.

The five attribute example used with the maximum likelihood discussion will be used to illustrate points for the generalized least squares development. The covariance matrix is given in Table 9.5. Table 9.13 gives the inverse of the covariance matrix. Jöreskog & Goldberger recommend refined starting values for the uniqueness. Since the standard errors of estimate, as developed in Chapter 8, are overstatements of the uniqueness, multiplication by a fraction appears to be desirable. The recommended stating values of u_j^2 are:

$$u_j^2 = (1 - r/2n)(1/c^{jj}) \ . \tag{9.97}$$

where c^{jj} is the j'th diagonal entry in $\boldsymbol{C}_{yy}^{-1}$. When comparing this formula with equation (9.44), note that a multiplying fraction of (1 − r/2n) is used. The initial uniqueness and values of ς are given in Table 9.14. Note that these uniqueness are .9 times the values in Table 9.6. Table 9.15 gives matrix $\boldsymbol{U}\boldsymbol{C}_{yy}^{-1}\boldsymbol{U}$.

There are close relations of generalized least square factor analysis to maximum likelihood factor analysis. For an area of relations note, for a given matrix $\boldsymbol{U}$, that:

$$(\boldsymbol{U}\boldsymbol{C}_{yy}^{-1}\boldsymbol{U})^{-1} = \boldsymbol{U}^{-1}\boldsymbol{C}_{yy}\boldsymbol{U}^{-1} \ . \tag{9.98}$$

Table 9.13

Inverse of Illustrative Sample Covariance Matrix, C^{-1}_{yy}

	1	2	3	4	5
1	.01414	-.00373	-.00188	-.00298	-.00203
2	-.00373	.01416	-.00721	-.00072	-.01112
3	-.00188	-.00721	.03264	-.00522	-.00056
4	-.00298	-.00072	-.00522	.000696	-.00467
5	-.00203	-.01112	-.00056	-.00467	.03288

Table 9.14

GLS Trial 1 Uniqueness and Zeta

	Uniqueness	Zeta (ζ)
1	63.333	4.153
2	63.585	4.152
3	27.573	3.317
4	129.345	4.863
5	27.374	3.310

Table 9.15

Trial 1 Matrix ($U^{-1}C_{yy}U^{-1}$) with Eigen Solution

$U^{-1}C_{yy}U^{-1}$

	1	2	3	4	5
1	.900	-.237	-.079	-.271	-.085
2	-.237	.900	-.302	-.065	-.464
3	-.079	-.302	.900	-.312	-.016
4	-.271	-.065	-.312	.900	-.278
5	-.085	-.464	-.016	-.278	.900

Eigenvectors

	1	2	3	4	5
1	.371	.254	.675	.527	.256
2	.529	-.424	-.233	.405	-.567
3	.391	.467	-.665	.157	.402
4	.451	.473	.197	-.588	-.434
5	.475	-.560	.094	-.434	.514

Eigenvalues

1	2	3	4	5
.039	.835	.969	1.065	1.592

The eigen solution of $\boldsymbol{U}^{-1}\boldsymbol{C}_{yy}\boldsymbol{U}^{-1}$ given in equation (9.36) may be written as:

$$\boldsymbol{U}^{-1}\boldsymbol{C}_{yy}\boldsymbol{U}^{-1} = \begin{bmatrix} \boldsymbol{V}_r & \boldsymbol{V}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{Q}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_r' \\ \boldsymbol{V}_2' \end{bmatrix} \tag{9.99}$$

Since the inverse of a matrix has identical eigenvectors as the matrix and the eigenvalues are inverted:

$$\boldsymbol{V}_1 = \boldsymbol{V}_r \tag{9.100}$$

and

$$\begin{bmatrix} \boldsymbol{\theta}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\theta}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{Q}_r^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_2^{-1} \end{bmatrix} \tag{9.101}$$

Note that the eigenvalues of $\boldsymbol{U}\boldsymbol{C}_{yy}^{-1}\boldsymbol{U}$ are in ascending order when the eigenvalues of $\boldsymbol{U}^{-1}\boldsymbol{C}_{yy}\boldsymbol{U}^{-1}$ are in descending order. Table 9.15 gives the eigenvalues and eigenvectors for trial 1 of the GLS factoring for the illustrative example. The eigenvectors given in Table 9.15 are identical to the eigenvectors for $\boldsymbol{U}^{-1}\boldsymbol{C}_{yy}\boldsymbol{U}^{-1}$ given in Table 9.7 for maximum likelihood factoring. However, the eigenvalues in Table 9.15 are .9 times the reciprocals of the eigenvalues in Table 9.7, the multiplier of .9 being due to the proportionality of the uniqueness used in the two analyses.

The conditional solution for the factor matrix $\boldsymbol{A}$ is similar to the conditional solution for maximum likelihood factor analysis. To satisfy equation (9.93)

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{V}_1(\boldsymbol{\theta}_1^{-1} - \boldsymbol{I})^{\frac{1}{2}} \tag{9.102}$$

which is to be compared with equation (9.37).

The next step is to substitute the conditional solution for factor matrix $\boldsymbol{A}$ into the expression for G in equation (9.86) to obtain a relation for G dependent only on $\boldsymbol{U}$. From equation (9.2) and (9.102):

$$\begin{aligned} \boldsymbol{C}_{zz} &= \boldsymbol{U}\boldsymbol{V}_1(\boldsymbol{\theta}_1^{-1} - \boldsymbol{I})^{\frac{1}{2}}(\boldsymbol{\theta}_1^{-1} - \boldsymbol{I})^{\frac{1}{2}}\boldsymbol{V}_1'\boldsymbol{U} + \boldsymbol{U}^2 \\ &= \boldsymbol{U}[\boldsymbol{V}_1(\boldsymbol{\theta}_1^{-1} - \boldsymbol{I})\boldsymbol{V}_1' + \boldsymbol{I}]\boldsymbol{U} \\ &= \boldsymbol{U}[\boldsymbol{V}_1\boldsymbol{\theta}_1^{-1}\boldsymbol{V}_1' - \boldsymbol{V}_1\boldsymbol{V}_1' + \boldsymbol{I}]\boldsymbol{U} \ . \end{aligned}$$

With equation (9.96):

$$\boldsymbol{C_{zz} = U[V_1\theta_1^{-1}V_1' + V_2V_2']U} \quad . \tag{9.103}$$

An expression is required for $\boldsymbol{C_{yy}^{-1}C_{zz}}$ which appears in equation (9.86).

$$\boldsymbol{C_{yy}^{-1}C_{zz} = C_{yy}^{-1}U[v_1\theta_1^{-1}V_1' + V_2V_2']U} \quad .$$

With equation (9.94)

$$\boldsymbol{C_{yy}^{-1}C_{zz} = U^{-1}\boxed{\begin{matrix} V_1 & V_2 \end{matrix}}\boxed{\begin{matrix} \theta_1 & 0 \\ 0 & \theta_2 \end{matrix}}\boxed{\begin{matrix} V_1' \\ V_2' \end{matrix}}\boxed{V_1\theta_1^{-1}V_1' + V_2V_2'}U} \quad .$$

Algebraic manipulation involving equations (9.95) and (9.96) yields:

$$\boldsymbol{C_{yy}^{-1}C_{zz} = U^{-1}[I - V_2(I - \theta_2)V_2']}$$

$$\boldsymbol{(I - C_{yy}^{-1}C_{zz}) = U^{-1}V_2(I - \theta_2)V_2'U}$$

$$\boldsymbol{(I - C_{yy}^{-1}C_{zz})^2 = U^{-1}V_2(I - \theta_2)^2V_2'U} \tag{9.104}$$

Then:

$$\begin{aligned} tr(\boldsymbol{I - C_{yy}^{-1}C_{zz}})^2 &= tr[\boldsymbol{U^{-1}V_2}(\boldsymbol{I - \theta_2})^2\boldsymbol{V_2'U} \\ &= tr(\boldsymbol{I - \theta_2})^2 \\ &= \sum_{m=k+1}^{n}(1 - \theta_m)^2 \\ &= \sum_{m=r+1}^{n}(\theta_m - 1)^2 \end{aligned} \tag{9.105}$$

Substitution into equation (9.89) yields:

$$G = (1/2)\sum_{m=r+1}^{n}(\theta_m - 1)^2 \tag{9.106}$$

This is the reduced function which does not involve the estimated factor matrix. When a solution for $\boldsymbol{U}$ for a minimum G is obtained, the solution for the factor matrix $\boldsymbol{A}$ may be obtained by equation (9.102).

An efficient solution for $\boldsymbol{U}$ uses a Newton-Raphson iterative procedure. To obtain the required derivatives of G with respect to entries in $\boldsymbol{U}$, the derivatives of eigenvalues θ_m and

entries v_{im} in the eigenvectors of $\boldsymbol{U}\boldsymbol{C}_{\boldsymbol{yy}}^{-1}\boldsymbol{U}$ are required. Jöreskog & Goldberger (1972) provided these derivatives in their Appendix A_2 .

$$\frac{\partial \theta_m}{\partial u_i} = \frac{2}{u_i}\theta_m v_{im}^2 \tag{9.107}$$

$$\frac{\partial v_{im}}{\partial u_j} = \frac{1}{u_j} v_{jm} \sum_{\substack{k=1 \\ k \neq m}}^{n} \frac{\theta_m + \theta_k}{\theta_m - \theta_k} v_{ik} v_{jk} \tag{9.108}$$

Using these equations:

$$\frac{\partial G}{\partial u_i} = \frac{2}{u_i} \sum_{m=r+1}^{n} (\theta_m^2 - \theta_m) v_{im}^2 \ ; \tag{9.109}$$

$$\begin{aligned} \frac{\partial^2 G}{\partial u_i \partial u_j} = \frac{4}{u_i u_j} \sum_{m=r+1}^{n} \{ & (2\theta_m^2 - \theta_m) v_{im}^2 v_{jm}^2 \\ & + (\theta_m^2 - \theta_m) v_{im} v_{jm} \sum_{\substack{k=1 \\ k \neq m}}^{n} \frac{\theta_m + \theta_k}{\theta_m - \theta_k} v_{ik} v_{jk} \\ & - \frac{1}{2} \delta_{ij} (\theta_m^2 - \theta_m) v_{im} v_{jm} \} \end{aligned} \tag{9.110}$$

With $\delta_{ij} = 1$ for $i = j$ and $\delta_{ij} = 0$ for $i \neq j$.

As with maximum likelihood factor analysis there are possible troubles with negative uniqueness; consequently, the same transformation is used as for maximum likelihood factor analysis from u_i^2 to ς_i . This transformation, defined in equation (9.43), has:

$$\varsigma_i = ln(u_i^2) \ .$$

Needed derivatives associated with this transformation are:

$$\frac{\partial G}{\partial \varsigma_i} = \frac{u_i}{2} \frac{\partial G}{\partial u_i} \tag{9.111}$$

$$\frac{\partial^2 G}{\partial \varsigma_i \partial \varsigma_j} = \frac{u_i u_j}{4} \frac{\partial^2 G}{\partial u_i \partial u_j} + \delta_{ij} \frac{u_i}{4} \frac{\partial G}{\partial u_i} \tag{9.112}$$

Using these derivatives, equation (9.109) becomes:

$$\frac{\partial G}{\partial \varsigma_i} = \sum_{m=r+1}^{n} (\theta_m^2 - \theta_m) v_{im}^2 \ . \tag{9.113}$$

Also, equation (9.110) becomes:

$$\frac{\partial^2 G}{\partial \varsigma_i \partial \varsigma_j} = \sum_{m=r+1}^{n} (2\theta_m^2 - \theta_m) v_{im}^2 v_{jm}^2 + \sum_{m=r+1}^{n} (\theta_m^2 - \theta_m) v_{im} v_{jm} \sum_{\substack{k=1 \\ k \neq m}}^{n} \frac{\theta_m + \theta_k}{\theta_m - \theta_k} v_{ik} v_{jk} \ . \qquad (9.114)$$

Equation (9.113) gives the gradient vector and equation (9.114) gives the second derivative matrix for use in a Newton-Raphson iteration for the vector of ς's for a minimum value of G . For a given trial vector of ς's , values of trial uniqueness are obtained by the second form of equation (9.43):

$$u_i^2 = EXP(\varsigma_i) \ .$$

Then matrix $\boldsymbol{U}\boldsymbol{C}_{\boldsymbol{yy}}^{-1}\boldsymbol{U}$ is computed and the eigen solution is obtained. Pertinent values for trial 1 for the example are given in Tables 9.14 and 9.15. The corresponding gradient vector and second derivative matrix are computed by equations (9.113) and (9.114). For trial 1 of iterations for one factor, Table 9.16 gives these values. The Newton-Raphson iteration follows the scheme outlined for maximum likelihood method in equation (9.42). The value of the criterion G for the trial may be obtained by equation (9.106). For the example trial 1 this value is .19167. Table 9.17 gives a summary from the iterations over trials for the example for one factor. The criterion stopping the iterations is the value of the derivative vector. This vector should approach zero as a minimum value of the criterion is approached. A very tight termination criterion was employed in the program producing the example so that the iteration was carried to 6 trials when a convergence was apparent at trial 5.

Table 9.18 gives the solution for one factor for the example. The model variances, which are the sums of the communalities and uniqueness, are less than the input variances given in Table 9.5. The generalized least squares solution does not restrict these model variances to equaling the input variances. The value of ξ was computed by equation (9.88) and the degrees of freedom was obtained by equation (9.80). When ξ is interpreted as a χ^2, the value of p is equal to .001 which indicates that the one factor model does not, statistically, fit the data. However, this is no surprise since lack of fit variation was included in the generation of the simulated data. There remains a question of goodness of fit.

As discussed in the introduction to this chapter, the Tucker & Lewis (1973) coefficient of reliability may be adapted to use with generalized least squares factor analysis. Function G is substituted in equation (9.14) for function F. This yields:

$$M_0 = G_0 / d\mathrm{f}_0$$

for no common factors and:

Table 9.16

GLS Trial 1 Derivatives of Function G for One Factor

	First Derivative	Second Derivative Matrix*				
		1	2	3	4	5
1	.058	.859	.059	.008	.075	.010
2	.288	.059	1.802	.094	.009	.220
3	.110	.008	.094	.910	.009	.003
4	.170	.075	.009	.099	.967	.081
5	.219	.010	.220	.003	.081	1.015

Table 9.17

Outline of Progress over GLS Trials for One Factor

Function G over Trials

1	2	3	4	5	6
.19167	.11483	.10917	.10904	.10904	.10904

First Derivative Vectors Over Trials

	1	2	3	4	5	6
1	.05821	.00479	.00032	.00001	.00000	.00000
2	.28764	.06899	.00970	.00036	.00000	.00000
3	.11035	.01355	.00053	.00002	.00000	.00000
4	.16953	.03031	.00187	.00001	.00000	.00000
5	.21854	.04173	.00268	.00004	.00000	.00000

Table 9.18

Generalized Least Squares Results for One Factor

	Factor Matrix 1	Communality	Uniqueness	Model Variance
1	14.6	213.7	61.7	275.4
2	21.1	446.6	44.3	491.0
3	10.2	103.8	25.4	129.2
4	25.4	650.1	106.2	756.3
5	12.4	153.2	22.7	175.9

ξ (χ^2) = 21.4 df = 5 p = .001
Tucker-Lewis Reliability = .683

$$M_r = G_r / d\mathrm{f}_r$$

for r common factors. Coefficient G_r and degrees of freedom df_r may be obtained from the solution for r common factors. However, coefficient G_0 and degrees of freedom df_0 are required for 0 common factors. Derivation of a solution follows.

For 0 common factors the factor matrix $\boldsymbol{A}$ does not exist so that, from equation (9.2):

$$\boldsymbol{C_{zz0}} = \boldsymbol{U_0^2} \ . \tag{9.115}$$

Then from equation (9.89):

$$G_0 = \frac{1}{2} tr(\boldsymbol{I} - \boldsymbol{C_{yy}^{-1}} \boldsymbol{U_0^2})^2 \ . \tag{9.116}$$

Some algebraic operations follow.

$$G_0 = \frac{1}{2} tr(\boldsymbol{I} - 2\boldsymbol{C_{yy}^{-1}} \boldsymbol{U_0^2} + \boldsymbol{C_{yy}^{-1}} \boldsymbol{U_0^2} \boldsymbol{C_{yy}^{-1}} \boldsymbol{U_0^2}) \, .$$

Elements of $\boldsymbol{C_{yy}^{-1}}$ are designated with superscripts as c^{ij} .

$$G_0 = \frac{1}{2} n - 2 \sum_{i=1}^{n} c^{ii} u_{0i}^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} c^{ij} u_{0j}^2 c^{ji} u_{0i}^2 \ .$$

Due to the symmetry of $\boldsymbol{C_{yy}}$

$$c^{ij} = c^{ji} \, .$$

Then:

$$G_0 = \frac{1}{2} \{ n - 2 \sum_{i=1}^{n} c^{ii} u_{0i}^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} (c^{ij})^2 u_{0i}^2 u_{0j}^2 \} \ .$$

To obtain a minimum G_0 the derivatives with respect to the u_{0k}^2 (k=1, n) are obtained and set equal to zero.

$$\begin{aligned} \frac{\partial G_0}{\partial u_{0k}^2} &= \frac{1}{2} \{ -2c^{kk} + \sum_{j=1}^{n} (c^{kj})^2 u_{0j}^2 + \sum_{i=1}^{n} (c^{ik})^2 u_{0i}^2 \} \\ &= \frac{1}{2} \{ -2c^{kk} + 2 \sum_{i=1}^{n} (c^{ik})^2 u_{0i}^2 \} = 0 \, . \end{aligned}$$

Then:

$$\sum_{i=1}^{n}(c^{ik})^2 u_{0i}^2 = c^{kk} \ . \qquad (9.117)$$

Equation (9.117) is in linear form with the u_{0i}^2 to be determined. The solution is simplified by some temporary definitions.

$\boldsymbol{M}$ is a matrix containing elements $(c^{ik})^2$;

$\underline{\boldsymbol{x}}$ is a column vector of u_{0i}^2 ;

$\underline{\mathbf{c}}$ is a column vector of c^{kk} .

Then equation (9.117) may be written:

$$\boldsymbol{M}\underline{\boldsymbol{x}} = \underline{\boldsymbol{c}}$$

With a solution:

$$\underline{\boldsymbol{x}} = \boldsymbol{M}^{-1}\underline{\boldsymbol{c}} \ .$$

The entries in $\underline{\boldsymbol{x}}$ may be inserted into the diagonal of matrix $\boldsymbol{U_0^2}$. Then, the value of G_0 may be obtained by substitution into equation (9.116).

The degrees of freedom, df_0 , for no common factors equals the number of covariances on one side of the diagonal of $\boldsymbol{C_{yy}}$:

$$df_0 = \frac{1}{2}n(n-1) \ .$$

Coefficients obtained for the illustrative example are given below.

$G_0 = .57763$	$df_0 = 10$	$M_0 = .05776$
$G_1 = .10904$	$df_1 = \ 5$	$M_1 = .02181$

With N = 200 , n = 5, r = 1, $N_1^{\bigstar}$ = 195.833 by equation (9.18). Then the Tucker & Lewis type reliability given in equation (9.19) is .683 which is given at the bottom of Table 9.18. This result appears to indicate a poor fit of the one factor model to the data. Also, the statistical test which resulted in a p = .001 indicates a lack of fit. However, the construction of this simulated data involved only one major domain dimension. This indicates a problem in deciding on whether to accept this one factor model or not to accept it. Maybe these results have been produced by the small number of attributes, 5, and the large amount of lack of fit variance introduced by the minor domain factors. Experimental trials with larger number of attributes for real data indicate more acceptable results.

Table 9.19 presents the results for a two factor model for the example. The largest changes induced by the addition of the second factor are the increases in communality for attributes 2 and 4 with accompanying decreases in the uniqueness for these two attributes. The

Table 9.19

Generalized Least Squares Results for Two Factors

	Factor Matrix 1	Factor Matrix 2	Communality	Uniqueness	Model Variance
1	13.8	4.5	211.4	66.0	277.4
2	22.3	.0	496.1	.0	496.1
3	9.7	3.0	102.8	28.5	131.3
4	23.5	12.8	714.9	57.1	772.0
5	11.9	2.8	149.0	28.6	177.6

ξ (χ^2) = .66 df = 1 p = .416

Tucker-Lewis Reliability = 1.033

statistical p of .416 indicates that the two factor model should not be rejected. Also, the Tucker & Lewis type reliability has risen to 1.033.

9.3. Minimum Squared Scaled Residual (MINSSR) Factor Fitting

Minimum squared scaled residuals is the third method of factor fitting considered in this chapter. It is closely related to the method of principal factors discussed in Chapter 8; in a sense, MINSSR is the complement of principal factors which was developed in terms of maximizing the sum of squares of the factor loadings. In contrast MINSSR is developed to minimize residual covariances after the extraction of factors. The scaling of the residuals is an important aspect of this method. Two cases are to be considered: determination of the factors in terms of the original scales for the attributes and determination of the factors in terms of attribute scores scaled to standard scores. The principal factors technique considered in Chapter 8 was restricted to consideration of the attribute scores scaled to standard scores. This is implied by application of the principal factors technique to correlation matrices as was the case in Chapter 8.

Equation (9.1) with the weight matrix $\boldsymbol{W}$ restricted to a diagonal scaling matrix is the basis for the MINSSR methods. Two special cases are considered: raw MINSSR , for which $\boldsymbol{W}$ is an identity matrix; normal MINSSR , for which $\boldsymbol{W}$ contains the reciprocals of the observed variances of the attribute scores, these variances being the diagonal entries in $\boldsymbol{C_{yy}}$. Criterion G_1 is used for the raw MINSSR:

$$G_1 = \frac{1}{2} tr[\boldsymbol{I}(\boldsymbol{C_{yy}} - \boldsymbol{C_{zz}})]^2 = \frac{1}{2} tr(\boldsymbol{C_{yy}} - \boldsymbol{C_{zz}})^2 \ . \tag{9.118}$$

Criterion G_2 is used for the normal MINSSR:

$$G_2 = \frac{1}{2} tr[\boldsymbol{S^{-1}}(\boldsymbol{C_{yy}} - \boldsymbol{C_{zz}})]^2 \tag{9.119}$$

Where

$$\boldsymbol{S^2} = diag(\boldsymbol{C_{yy}}) \ . \tag{9.120}$$

An algebraic transformation of criterion G_2 is useful. From equation (9.119):

$$G_2 = \frac{1}{2} tr\{(\boldsymbol{S^{-1}C_{yy}S^{-1}} - \boldsymbol{S^{-1}C_{zz}S^{-1}})\}^2 . \tag{9.121}$$

Note that the sample correlation matrix $\boldsymbol{R}$ is:

$$\boldsymbol{R} = \boldsymbol{S^{-1}C_{yy}S^{-1}} \ . \tag{9.122}$$

Let:

$$\boldsymbol{C}_{zzR} = \boldsymbol{S}^{-1}\boldsymbol{C}_{zz}\boldsymbol{S}^{-1} \ . \tag{9.123}$$

Then:

$$G_2 = \frac{1}{2}tr[\boldsymbol{R} - \boldsymbol{C}_{zzR}]^2 \ . \tag{9.124}$$

The foregoing transformation may be carried on to the factor analysis model given in equation (9.2) by defining:

$$\boldsymbol{A}_R = \boldsymbol{S}^{-1}\boldsymbol{A} \tag{9.125}$$

and

$$\boldsymbol{U}_R^2 = \boldsymbol{S}^{-1}\boldsymbol{U}^2\boldsymbol{S}^{-1} \ . \tag{9.126}$$

The factor model for the correlation matrix becomes:

$$\boldsymbol{C}_{zzR} = \boldsymbol{A}_R\boldsymbol{A}_R' + \boldsymbol{U}_R^2 \ . \tag{9.127}$$

At a solution the matrices $\boldsymbol{A}$ and $\boldsymbol{U}^2$ for the attribute scores in their original scales can be obtained by solving equations (9.125) and (9.126):

$$\boldsymbol{A} = \boldsymbol{S}\boldsymbol{A}_R \ ; \tag{9.128}$$

$$\boldsymbol{U}^2 = \boldsymbol{S}\boldsymbol{U}_R^2\boldsymbol{S} \ . \tag{9.129}$$

There are some concerns about the scale freeness of the MINSSR solutions. The raw MINSSR is not scale free. However application of the material in equations (9.5) through (9.11) indicates that the normal MINSSR is scale free. This result for normal MINSSR supports the long standing application of principal factors and MINSSR to correlation matrices.

Two computational procedures are to be discussed: principal factors with iterated uniqueness; the Harman and Jones (1966) MINRES procedure. For the raw MINSSR, the computing procedure is applied to the observed covariance matrix; for normal MINSSR, the computing procedure is applied to the observed correlation matrix. (Remember that the correlation matrix can be considered to be a covariance matrix among standardized scores.)

The principal factors with iterated uniqueness involves a series of trials with each trial reducing the sum of squared residuals. Each trial has two phases:

1) Given a trial matrix $\boldsymbol{U}^2$, notated $\boldsymbol{U}_t^2$, the factor matrix $\boldsymbol{A}_t$ is determined to minimize the sum of squared residuals.
2) Given the matrix $\boldsymbol{A}_t$, the entries in $\boldsymbol{U}_t^2$ are adjusted to set the diagonal residuals to zero. Since this operation does not affect the off-diagonal residuals, the sum of

squares of the residuals as a whole is reduced by the sum of squares of the diagonal entries before this adjustment.

Convergence of this procedure occurs when there are minimal changes in the trial $\boldsymbol{U^2}$ from trial to trial.

The procedure to minimize the sum of squared residuals will be discussed in terms of a given covariance matrix $\boldsymbol{C_{yy}}$. For analysis of a correlation matrix, the correlation matrix is substituted for the covariance matrix. For a given trial $\boldsymbol{U_t^2}$, a matrix $\widetilde{\boldsymbol{C}}_{\boldsymbol{yyt}}$ is defined by:

$$\widetilde{\boldsymbol{C}}_{\boldsymbol{yyt}} = \boldsymbol{C_{yy}} - \boldsymbol{U_t^2} \tag{9.130}$$

Cell entries in $\widetilde{\boldsymbol{C}}_{\boldsymbol{yyt}}$ are designated by $\widetilde{\mathrm{c}}_{tij}$. Let the matrix of residuals be:

$$\boldsymbol{E_t} = \widetilde{\boldsymbol{C}}_{\boldsymbol{yyt}} - \boldsymbol{A_t A_t'} \ . \tag{9.131}$$

For the ij cell of E_t :

$$e_{tij} = \widetilde{\mathrm{c}}_{tij} - \sum_{k=1}^{r} a_{tik} a_{tjk} \ . \tag{9.132}$$

where r is the number of factors. The sum of squared residuals over the entire matrix $\boldsymbol{E_t}$ is represented by the function g_t .

$$\mathrm{g}_t = \sum_{i=1}^{n} \sum_{j=1}^{n} e_{tij}^2 \ . \tag{9.133}$$

To minimize this function, the partial derivatives of g_t with respect to the factor weights a_{thk} are obtained:

$$\frac{\partial \mathrm{g}_t}{\partial a_{thk}} = 2 \sum_{j=1}^{n} e_{thj}(-a_{tjk}) + 2 \sum_{i=1}^{n} e_{tih}(-a_{tik}) \ . \tag{9.134}$$

These partial derivatives are set equal to zero. With the fact that the matrix $\boldsymbol{E_t}$ is symmetric, algebraic operations yield:

$$\sum_{i=1}^{n} e_{tih} a_{tik} = 0 \tag{9.135}$$

With equation (9.132):

$$\sum_{i=1}^{n} (\widetilde{\mathrm{c}}_{tih} - \sum_{m=1}^{r} a_{tim} a_{thm}) a_{tik} = 0 \ . \tag{9.136}$$

Then:

$$\sum_{i=1}^{n} \widetilde{c}_{tih} a_{tik} = \sum_{m=1}^{r} a_{thm} \sum_{i=1}^{n} a_{tim} a_{tik} \ . \tag{9.137}$$

Consider that the factor matrix has been transformed to principal axes so that:

$$\boldsymbol{A}_t' \boldsymbol{A}_t = \boldsymbol{\Lambda}_t \ ,$$

where $\boldsymbol{\Lambda}_t$ is a diagonal matrix so that:

$$\sum_{i=1}^{n} a_{tim} a_{tik} = \sum_{i=1}^{n} a_{tik}^2 = \boldsymbol{\lambda}_{tk} \qquad \text{for } m = k \tag{9.138a}$$

$$\sum_{i=1}^{n} a_{tim} a_{tik} = 0 \qquad\qquad \text{for } m \neq k \ . \tag{9.138b}$$

Equation (9.137) becomes:

$$\sum_{i=1}^{n} \widetilde{c}_{tih} a_{tik} = a_{thk} \lambda_{tk} \ . \tag{9.139}$$

In matrix form with $\underline{\mathbf{a}}_{tk}$ being a column vector of a_{tik} or a_{thk} , equation (9.139) yields:

$$\widetilde{\boldsymbol{C}}_{\boldsymbol{yyt}} \, \underline{\boldsymbol{a}}_{\boldsymbol{tk}} = \underline{\boldsymbol{a}}_{\boldsymbol{tk}} \lambda_{tk}$$

or

$$(\widetilde{\boldsymbol{C}}_{\boldsymbol{yyt}} - \lambda_{tk} \boldsymbol{I}) \underline{\boldsymbol{a}}_{\boldsymbol{tk}} = \mathbf{0} \ . \tag{9.140}$$

The $\boldsymbol{\lambda}_{tk}$ are eigenvalues of $\mathbf{C}_{\boldsymbol{yyt}}$ and the $\underline{\mathbf{a}}_{\boldsymbol{tk}}$ are corresponding eigenvectors which are scaled according to equation (9.138a). Note that the $\underline{\mathbf{a}}_{\boldsymbol{tk}}$ are principal factors of $\widetilde{\boldsymbol{C}}_{\boldsymbol{yyt}}$.

There remain questions both as to the value of the criterion g_t and as to the dimensions of the eigen solution to be used. From equations (9.132) and (9.133):

$$g_t = \sum_{i=1}^{n} \sum_{j=1}^{n} (\widetilde{c}_{tij} - \sum_{k=1}^{r} a_{tik} a_{tjk})^2$$

which may be written as:

$$g_t = \sum_{i=1}^{n} \sum_{j=1}^{n} (\widetilde{c}_{tij} - \sum_{k=1}^{r} a_{tik} a_{tjk}) (\widetilde{c}_{tij} - \sum_{m=1}^{r} a_{tim} a_{tjm}) \ .$$

Algebraic operations yield:

$$g_t = \sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{c}_{tij}^2 - \sum_{k=1}^{r}\sum_{m=1}^{r}\sum_{i=1}^{n} a_{tik}a_{tim}\sum_{j=1}^{n} a_{tjk}a_{tjm} \ .$$

Using equations (9.138a) and (9.138b):

$$g_t = \sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{c}_{tij}^2 - \sum_{k=1}^{r}\lambda_k^2 \ . \tag{9.141}$$

From equation (9.138a) only positive eigenvalues are permissible. Function g_t is minimized by selecting the r eigen dimensions having the largest, positive eigenvalues λ_k . In case there are not r positive λ_k , no solution is permissible.

The second phase of a trial involves adjustment of the trial uniqueness to:

$$\boldsymbol{U}_{t+}^2 = \boldsymbol{U}_t^2 + Diag(\boldsymbol{E}_t) \ . \tag{9.142}$$

The matrix of discrepancies, $\boldsymbol{E}_{t+}$, uses the $\boldsymbol{A}_t$ computed in the first phase and is:

$$\begin{aligned} \boldsymbol{E}_{t+} &= \boldsymbol{C}_{yy} - \boldsymbol{U}_{t+}^2 - \boldsymbol{A}_t\boldsymbol{A}_t' \\ &= \boldsymbol{C}_{yy} - \boldsymbol{U}_t^2 - Diag(\boldsymbol{E}_t) - \boldsymbol{A}_t\boldsymbol{A}_t' \\ &= \boldsymbol{E}_t - Diag(\boldsymbol{E}_t) \ . \end{aligned} \tag{9.143}$$

Then:

$$Diag(\boldsymbol{E}_{t+}) = \boldsymbol{0} \ . \tag{9.144}$$

An operator SSQ[X] is used to designate the sum of squares of the entries in matrix $\boldsymbol{X}$. From equation (9.143) and the fact that the diagonal entries in E_{t+} equal 0 :

$$SSQ[\boldsymbol{E}_t] = SSQ[\boldsymbol{E}_{t+}] + SSQ[Diag(\boldsymbol{E}_t) \ .] \tag{9.145}$$

The following logic will hold when there are no negative entries in the adjusted uniqueness matrix $\boldsymbol{U}_{t+}^2$. When any diagonal entry, e_{tii} , in $\boldsymbol{E}_t$ does not equal zero:

$$SSQ[Diag(\boldsymbol{E}_t)] > 0$$

and

$$SSQ[\boldsymbol{E}_{t+}] < SSQ[\boldsymbol{E}_t] \ . \tag{9.146}$$

Thus, the adjustment of the uniqueness reduces the sum of squares of the residuals. For the case when there are no negative entries in $\boldsymbol{U}_{t+}^2$, this matrix becomes the $\boldsymbol{U}_t^2$ for the next trial. Convergence occurs when there are minimal changes from one trial to the next.

Sometimes one of the uniqueness in $\boldsymbol{U}_{t+}^2$ becomes negative, this being termed a generalized Haywood case. Since negative uniqueness are not permissible, a common practice is

to employ a 'fix up' of replacing the negative uniqueness to zero in the transition from $\boldsymbol{U}_{t+}^{2}$ to the next trial $\boldsymbol{U}_{t}^{2}$. This violates the logic discussed in the preceding paragraph. However, the solution appears to converge. Note that the obtained factor matrix still implies a negative uniqueness after the solution of the first phase of the last trial.

A remaining problem is the initial values of uniqueness. Common usage involves the error variances in estimating the attributes from the other attributes in the battery. These values were given for the maximum likelihood method in equation (9.44). Then, trial 1, phase 1 is identical with principal factors with squared multiple correlations in the diagonal method for analysis of correlation matrices method described in Chapter 8.

While the principal factors with iterated uniqueness is easily programmed, rate of convergence sometimes is a problem. In studies having a relatively large number of factors the iterations seem just to crawl along with no evidence of immediate convergence. A large number of trials become necessary. This slowness of convergence along with the problems of handling the generalized Haywood cases lessen the attractiveness of this approach.

The MINRES procedure developed by Harman and Jones (1966) is much more effective. A true solution for the generalized Haywood case was developed by Harman and Fukuda (1966). This method starts with a trial factor matrix instead of trial uniqueness. With the given trial factor matrix the method cycles through the attributes one at a time replacing the row of factor loadings so as to minimize the sum of squares of the residuals of each attribute in turn with the other attributes. Since the trial factor matrix has been changed by the solutions for the various rows of the matrix, repeated solutions have to be made for the sequence of rows of the factor matrix. These cycles of solutions are continued until no changes occur between cycles.

Let $\boldsymbol{A}$, $n \times r$, be a trial factor matrix for the n attributes on r factors. Note that the solution assumes a given number of factors so that separate solutions must be made for different numbers of factors. Let $\boldsymbol{A}_{-i}$, $(n-1) \times r$, be the trial factor matrix with row i deleted where i is a pivot attribute. A solution is desired for row vector $\underline{\mathbf{a}}_{i}$ such as to minimize the sum of squared residuals, e_{ij}, $j \neq i$, with all other attributes. Note that the diagonal residual is excluded. These residuals are given by:

$$e_{ij} = c_{ij} - \sum_{k=1}^{r} a_{ik} a_{jk} \quad \text{with } j \neq i \tag{9.147}$$

where c_{ij} is the ij entry in the given covariance matrix $\boldsymbol{C}_{yy}$. In matrix form:

$$\underline{\boldsymbol{e}}_{i,-i} = \underline{\boldsymbol{c}}_{i,-i} - \underline{\boldsymbol{a}}_{i} \boldsymbol{A}'_{-i} \tag{9.148}$$

where $\underline{\boldsymbol{e}}_{\boldsymbol{i}},_{-\boldsymbol{i}}$ is the row vector of residuals for attribute i excluding attribute i and $\underline{\mathbf{c}}_{\boldsymbol{i}},_{-\boldsymbol{i}}$ is the corresponding row vector for attribute i from matrix $\boldsymbol{C}_{\boldsymbol{yy}}$ with attribute i excluded. A criterion g_i can be defined as:

$$g_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} e_{ij}^2 = \underline{\boldsymbol{e}}_{\boldsymbol{i}},_{-\boldsymbol{i}} \underline{\boldsymbol{e}}'_{\boldsymbol{i}},_{-\boldsymbol{i}} \ . \tag{9.149}$$

At a solution the uniqueness for attribute i is given by:

$$u_i^2 = c_{ii} - \sum_{k=1}^{r} a_{ik}^2 = c_{ii} - \underline{\boldsymbol{a}}_{\boldsymbol{i}} \underline{\boldsymbol{a}}'_{\boldsymbol{i}} \ . \tag{9.150}$$

Two solutions are considered: one an unrestricted solution and second a restricted solution which is used when the unrestricted solution yields a negative uniqueness, that is, a generalized Heywood case.

Solution for the unrestricted case is quite simple. The problem from equations (9.148) and (9.149) is of the form of a general least squares. Minimization of g_i involves a set of simultaneous linear equations with the entries a_{ik} in vector $\underline{\mathbf{a}}_{\boldsymbol{i}}$ as unknowns. These equations in matrix form are:

$$\underline{\boldsymbol{a}}_{\boldsymbol{i}} (\boldsymbol{A}'_{-\boldsymbol{i}} \boldsymbol{A}_{-\boldsymbol{i}}) = (\underline{\boldsymbol{c}}_{\boldsymbol{i}},_{-\boldsymbol{i}} \boldsymbol{A}_{-\boldsymbol{i}}) \ . \tag{9.151}$$

For any given pivot attribute i on any cycle of solutions when the uniqueness computed by equation (9.150) is equal to or greater than zero, the unrestricted solution is acceptable. When the uniqueness is negative, the unrestricted solution is to be discarded and a restricted solution obtained.

For the restricted solution the uniqueness is constrained to equaling zero which is expressed in the following equation.

$$c_{ii} - \sum_{k=1}^{r} a_{ik}^2 = c_{ii} - \underline{\boldsymbol{a}}_{\boldsymbol{i}} \underline{\boldsymbol{a}}'_{\boldsymbol{i}} = 0 \ . \tag{9.152}$$

The revised criterion, g_{ri} , is:

$$g_{ri} = \sum_{\substack{j=1 \\ j \neq i}}^{n} e_{ij}^2 + \beta (c_{ii} - \sum_{k=1}^{r} a_{ik}^2) \tag{9.153}$$

with β being an unknown LaGrange multiplier. A solution for optimum values of g_{ri} involves the partial derivatives of g_{ri} with respect to the a_{ik} .

$$\frac{\partial \mathbf{g}_{ri}}{\partial a_{ik}} = 2\sum_{\substack{j=1 \\ j\neq i}}^{n} e_{ij}(-a_{jk}) - 2\beta a_{ik} \ . \tag{9.154}$$

Setting these derivatives to zero yields:

$$\sum_{\substack{j=1 \\ j\neq i}}^{n} e_{ij}a_{jk} + \beta a_{ik} = 0 \ . \tag{9.155}$$

Substitution from equation (9.147) yields:

$$\sum_{\substack{j=1 \\ j\neq i}}^{n} c_{ij}a_{jk} - \sum_{\substack{j=1 \\ j\neq i}}^{n} a_{jk}\sum_{m=1}^{r} a_{im}a_{jm} + \beta a_{ik} = 0 \ . \tag{9.156}$$

With algebraic operations equation (9.156) may be written in matrix form as:

$$\underline{\boldsymbol{c}}_{\boldsymbol{i},-\boldsymbol{i}}\boldsymbol{A}_{-\boldsymbol{i}} - \underline{\boldsymbol{a}}_{\boldsymbol{i}}(\boldsymbol{A}'_{-\boldsymbol{i}}\boldsymbol{A}_{-\boldsymbol{i}} - \beta\boldsymbol{I}) = \boldsymbol{0} \ . \tag{9.157}$$

Note that when β equals zero this equation reduces to equation (9.151) for the unrestricted case. However, for the restricted case a solution for β is needed so as to satisfy equation (9.157) and the constraining equation (9.152). There appear to be several possible solutions similar to the case for an eigen problem. However, there appears to be no simple solution leading to a standard form. Assume that the matrix formed by the product $\boldsymbol{A}'_{-\boldsymbol{i}}\boldsymbol{A}_{-\boldsymbol{i}}$ is positive, definite, thus being nonsingular. Then a solution for $\underline{\mathbf{a}}_{\boldsymbol{i}}$ is:

$$\underline{\boldsymbol{a}}_{\boldsymbol{i}} = \underline{\boldsymbol{c}}_{\boldsymbol{i},-\boldsymbol{i}}\boldsymbol{A}_{-\boldsymbol{i}}(\boldsymbol{A}'_{-\boldsymbol{i}}\boldsymbol{A}_{-\boldsymbol{i}} - \boldsymbol{\beta I})^{-1} \ . \tag{9.158}$$

The communality for attribute i from Chapter 3, is:

$$h_i^2 = \underline{\boldsymbol{a}}_{\boldsymbol{i}}\underline{\boldsymbol{a}}'_{\boldsymbol{i}} \ .$$

From equation (9.158):

$$h_i^2 = \underline{\boldsymbol{c}}_{\boldsymbol{i},-\boldsymbol{i}}\boldsymbol{A}_{-\boldsymbol{i}}(\boldsymbol{A}'_{-\boldsymbol{i}}\boldsymbol{A}_{-\boldsymbol{i}} - \beta\boldsymbol{I})^{-2}\boldsymbol{A}'_{-\boldsymbol{i}}\underline{\boldsymbol{c}}'_{\boldsymbol{i},-\boldsymbol{i}} \ . \tag{9.159}$$

The constraint of equation (9.152) may be written as:

$$h_i^2 = c'_{ii} \ . \tag{9.160}$$

In order for the matrix $(\boldsymbol{A}'_{-\boldsymbol{i}}\boldsymbol{A}_{-\boldsymbol{i}} - \beta\boldsymbol{I})$ to remain positive, definite, the value of β must be less than the least eigenvalue of $\boldsymbol{A}'_{-\boldsymbol{i}}\boldsymbol{A}_{-\boldsymbol{i}}$. When h_i^2 is considered as a function of the slope of this function is positive in the range from minus infinity to the least eigenvalue of $\boldsymbol{A}'_{-\boldsymbol{i}}\boldsymbol{A}_{-\boldsymbol{i}}$. Since this restricted case is used only when the unrestricted case yields a value of h_i^2 greater than c_{ii} , that is, when β equals zero, the range of interest for β is from minus infinity to zero. This range

is included in the range when the slope of the function for h_i^2 is positive. Thus, the desired value of β is negative. Further, there is only one value of β for which h_i^2 equals c_{ii} . These observations lead to a possible solution involving interpolation between values of β which yield communalities less than and greater than c_{ii} . Such a procedure can be made quite efficient.

Table 9.20 presents a MINRES solution for the nine mental tests example for three common factors. The upper section gives the MINRES output factor matrix along with the communality and uniqueness for each attribute. Since none of the uniqueness are negative, the restrictive solution was not used. A transformation of factors to correlated factors was performed with the transformed factor matrix being given in the middle section of the table. The factor correlations matrix is given in the bottom section of the table. These results may be compared with those given in Table 1.2 which were obtained using principle factors with SMC's in the diagonal.

Table 9.21 gives some statistics for the MINRES factor extraction for the nine mental tests example. These statistics are given for zero through four factor solutions. For each of these solutions the root mean square off-diagonal residual is given as RMS Residuals, There are considerable decreases in these values from zero factors through three factors with a much lesser decrease from three factors to four factors. The RMS Residual is relatively small for the three factor solution, this supporting the idea of accepting the three factor solution. Rippe (1953) proposed an "Application of a large sampling criterion to some sampling problems in factor analysis" which has been suggested for use with MINRES factor analysis (see: Harman (1976), page 184). This statistic is identical to the maximum likelihood statistic, see equation (9.34). Let F_R be:

$$F_R = ln\left|\boldsymbol{C_{zz}}\right| - ln\left|\boldsymbol{C_{yy}}\right| + tr(\boldsymbol{C_{zz}^{-1}C_{yy}}) - n \ . \tag{9.161}$$

Rippe's chi-square statistic is given by:

$$\xi_R = N^{\bigstar} F_R \ . \tag{9.162}$$

Rippe gives $N^{\bigstar} = N - 1$; however, a better value from Bartlett (1950) is given in equation (9.18) and repeated here for convenience.

$$N^{\bigstar} = N - 1 - (2n + 5)/6 - 2r/3 \ . \tag{9.18}$$

Rippe gave a value for the degrees of freedom larger than the one for maximum likelihood factor analysis. Study of his values raised some questions so that a small Monte Carlo study was performed for MINSSR factoring. Results indicated that the degrees of freedom computed from the maximum likelihood factor analysis formula was more appropriate than the Rippe values for

Table 9.20

Factor Matrices from MINRES Factor Extraction
Nine Mental Tests Example
Three Common Factors

MINRES Output Factor Matrix

	1	2	3	Communality	Uniqueness
1	.331	.252	.499	.422	.578
2	.391	.501	.494	.648	.352
3	.555	.098	.353	.442	.558
4	.573	-.451	.030	.533	.467
5	.690	-.484	.003	.710	.290
6	.613	-.386	.028	.525	.475
7	.473	.497	.017	.472	.528
8	.673	.163	-.224	.529	.471
9	.618	.284	-.103	.473	.527

Transformed Factor Matrix

	1	2	3
1	.692	.060	-.115
2	.770	-.062	.074
3	.522	.337	.021
4	.004	.732	-.039
5	-.015	.843	.019
6	.028	.718	.023
7	.249	.022	.522
8	-.083	.405	.614
9	.078	.278	.560

Factor Correlation Matrix

	1	2	3
1	1.000	.121	.514
2	.121	1.000	.086
3	.514	.086	1.000

Table 9.21

Statistics for MINRES Factor Extraction
Nine Mental Tests Example
(N = 710)

Number of Factors	RMS Residuals	Significance Statistics			T – L Type Reliability
		Chi-Square	DF	P	
0	.331	1975.6	36	.000	.000
1	.151	833.1	27	.000	.445
2	.053	169.4	19	.000	.853
3	.015	28.6	12	.005	.974
4	.010	12.1	6	.059	.981

MINSSR factoring. These degrees of freedom are given in equation (9.80), repeated here for convenience.

$$\nu(\xi) = n(n+1)/2 - n - nr + r(r-1)/2 \ . \tag{9.80}$$

The Rippe statistics are given in Table 9.21. For the three factor solution the value of p equal to .005 indicates that this model should be rejected. However, as has been indicated several times previously, there is no surprise in this observation in that a worth while factor model is quite likely not to fit the real world even for samples approaching the population of individuals. Table 9.21 also gives the Tucker-Lewis type reliability values for the several factor solutions. These values were computed according to the material presented earlier in this chapter. See the material associated with equations (9.14) and (9.19). The value of .974 appears quite adequate to accept the three factor solution.

There is some interest in the scale freeness of factor solutions. The Rippe function is not scale free when used with the raw MINSSR solution since the raw MINSSR solution is not scale free. However, the Rippe function is scale free when used with the normal MINSSR solution.

The Rippe function is different from the function minimized in MINSSR. The maximum likelihood solution would minimize Rippe's function but, probably, would not minimize the MINSSR function. Conversely, considering the normal MINSSR case, the normal MINSSR solution, probably, would not minimize the maximum likelihood and the Rippe function. Only when the MINSSR solution is a good approximation to the maximum likelihood solution would the Rippe function be appropriate. For the nine mental tests example, the MINSSR solution appears to be a good approximation. Table 9.22 gives the maximum likelihood statistics for this example. Note that the maximum likelihood chi-squares are slightly lower than the Rippe chi-squares in Table 9.21. For other, larger examples the maximum likelihood chi-squares are markedly less than the Rippe chi-squares. A conclusion would be that the Rippe function should be used with considerable caution.

9.4. Remarks on the Number of Factors

Considerations as to the number of factors should take into account a variety of matters arising from the differences between factor analytic models and the real world. As discussed in Chapters 3 and 4 there are model errors in both the population and the sample. Further, these model errors may be different for different methods of factor extraction as indicated in Table 9.3 and associated text earlier in this Chapter. These matters of model error affect the statistical functions associated with the factor methods discussed in this chapter. One should expect that a good model determined for a given number of factors would be rejected by the chi-square

Table 9.22

Maximum Likelihood Factor Statistics
Nine Mental Tests Example
(N = 710)

Number of Factors	Significance Statistics			T – L Type
	Chi-Square	DF	P	Reliability
0	1975.6	36	.000	.000
1	824.9	27	.000	.451
2	167.6	19	.000	.855
3	28.1	12	.005	.975
4	11.4	6	.077	.983

statistics. However, the chi-square should be low enough to indicate statistical stability for an accepted solution.

In addition to chi-square statistics generated in the factor extraction, consideration should be given to other items of information. For example, the Guttman weaker lower bound on the number of factors (the number of eigenvalues of the correlation matrix with unities in the diagonal equal to or greater than unity) could be considered. Further, as described in Chapter 8 for principal factors extraction, the series of eigenvalues of the correlation matrix with SMC's in the diagonal should be inspected for a break between larger eigenvalues and a straight line of decreasing eigenvalues. Also, the parallel analysis described in Chapter 8 should be considered. For the nine mental tests example, material in Chapter 8 indicates that the number of eigenvalues equal to or greater than unity of the correlation matrix is two. However, there appears to be a break in the eigenvalue series for the correlation matrix with SMC's in the diagonal between three and four factors. This break along with the parallel analysis indicates a three factor solution. However, as given in Table 9.22, the chi-square for three factors yields a p of .005 which, statistically, indicates that a three factor solution should be rejected. A decision to use a three factor solution could be defended in that the chi-square is low compared with the preceding values for fewer factors. Also, the Tucker-Lewis reliability is acceptably high.

Table 9.23 gives the maximum likelihood statistics for the 18 verbal and numerical tests selected from the Thurstone and Thurstone (1941) battery. The correlation matrix is given in Table 8.15 with results from principal factors analysis given in Tables 8.16 and 8.17. A graph of the series of eigenvalues of the correlation matrix with SMC's in the diagonal is given in Figure 8.2. From this graph a conclusion of four factors appears justified. The parallel analysis given in Table 8.16 agrees with a decision of four factors. However, from Table 9.23, the maximum likelihood chi-square for four factors yields a p of .000 indicating that a four factor solution should be rejected. The Tucker-Lewis reliability of .974 for four factors is quite acceptable. Maybe, a four factor solution is OK for the principal factors analysis while the maximum likelihood factor analysis requires more than four factors. This difference may reflect a difference in the representation of the correlation matrix. A five factor maximum likelihood solution is problematic with the p of .010. Some study is indicated after factor transformations to more meaningful factors before deciding which factor solution to accept.

A final point is that the factor matrices obtained by any of the factor extraction methods establishes a factor space in which factor transformations are to be performed. The original factors obtained are possible mixtures of more meaningful factors. Extraction of too few factors yields confused results for the transformed factors. As Kaiser has said (informally) "it is a sin to extract too few factors." Extraction of too many factors will result in an extra, less meaningful factor which may result in some problems in the transformation of factors part of the analysis.

Table 9.23

Maximum Likelihood Factor Statistics
Thurston & Thurston 18 V & N Tests
(N = 710)

Number of Factors	Significance Statistics			T – L Type
	Chi-Square	DF	P	Reliability
0	5429.4	153	.000	.000
1	1784.4	135	.000	.645
2	732.2	118	.000	.849
3	396.9	102	.000	.916
4	165.4	87	.000	.974
5	103.8	73	.001	.988

When possible with good data, the extraction of a number of factors which differentiates between factors with considerable good contributions to the variance of the attributes and factors having "junk" contributions is desirable.

CHAPTER 10
FACTOR TRANSFORMATIONS: GRAPHICAL ROTATION

From
Exploratory Factor Analysis
Ledyard R Tucker
and
Robert C. MacCallum

CHAPTER 10
FACTOR TRANSFORMATIONS: GRAPHICAL ROTATION

After a factor matrix has been extracted from a covariance matrix by one of the methods discussed in Chapters 8 and 9, the next part of exploratory factor analysis is the transformation of the factors. The basic theory for transformation of factors was discussed in Chapter 7 along with a geometric model. In this chapter procedures will be discussed for transformation of factors by graphical methods which utilize the geometric model. A simple situation involves a two factor solution. When there are more than two factors the procedures are more complex.

The procedures to be discussed represent attempts to establish transformed factors in terms of Thurstone's principle of simple structure. Such a transformed solution is not always possible for all bodies of data; there is a matter of judgment in deciding the extent to which a simple structure has been achieved. It is not a matter of a simplest structure as interpreted by some individuals but is a matter of absolute judgment whether or not a simple structure was achieved. The strength of a simple structure may be judged in terms of the number and variety of attributes that have trivial loadings on each transformed factor. As discussed in Chapter 7 there are a variety of simple structures some of which are not complete. The completeness of a structure is a function of the attributes in a battery for which an analysis is being conducted. Each of the incomplete structures poses problems in the transformation of factors and in the interpretation of the factorial results. A major goal of factor transformations is to provide bases for conjectures as to the structure of the phenomena underlying the measured attributes and the relations among these attributes.

The graphical rotations of factors operate in terms of an original factor matrix $\boldsymbol{A}$. Usually, the attributes are assumed to be scaled in terms of standard scores. In case the original factor matrix was derived from analysis of a covariance matrix, the obtained factor matrix should be scaled so as to refer to standardized scores on the attributes. Let $\widetilde{\boldsymbol{A}}$ be the original factor matrix derived from the analysis of a covariance matrix and $\boldsymbol{S}$ be the diagonal matrix of standard deviations of the attributes. Then, $\boldsymbol{A}$, the factor matrix for standard scores on the attributes is obtained by:

$$\boldsymbol{A} = \boldsymbol{S}^{-1}\widetilde{\boldsymbol{A}} \quad . \tag{10.1}$$

The discussions in this chapter will consider only factor matrices for standard scores on the attributes.

The coordinate system for transformed factors is defined by base entities, base lines for two factor studies, base planes for three factor studies, and base hyperplanes for studies having more than three factors. These base entities may be defined by normals orthogonal to the base

entities. Matrix $\boldsymbol{F}$ contains as row vectors the coordinates of the normals. Since the normals are unit length vectors:

$$Diag(\boldsymbol{F}\boldsymbol{F}') = \boldsymbol{I} \tag{10.2}$$

Once having the normals matrix $\boldsymbol{F}$, the trait matrix, with trait vectors as rows, may be obtained by a solution from equation (7.6) as:

$$\boldsymbol{T} = \boldsymbol{D}(\boldsymbol{F}')^{-1} \ . \tag{10.3}$$

By equation (7.20), the trait vectors are of unit length so that:

$$Diag(\boldsymbol{T}\boldsymbol{T}') = \boldsymbol{I} \ . \tag{10.4}$$

Diagonal matrix $\boldsymbol{D}$ may be obtained by:

$$\boldsymbol{D} = [Diag(\boldsymbol{F}\boldsymbol{F}')^{-1}]^{-\frac{1}{2}} \ . \tag{10.5}$$

The factor correlation matrix is given by equation (7.1) as:

$$\boldsymbol{R}_{bb} = \boldsymbol{T}\boldsymbol{T}' \ . \tag{7.1}$$

An alternative formula for the factor correlation matrix is:

$$\boldsymbol{R}_{bb} = \boldsymbol{D}(\boldsymbol{F}\boldsymbol{F}')^{-1}\boldsymbol{D} \ . \tag{10.6}$$

There are four types of coefficients for the attribute vectors in the transformed factor structure. Only three of these will concern us here: matrix $\boldsymbol{G}$ contains the orthogonal projections of the attribute vectors on the normals; matrix $\boldsymbol{B}$ of factor weights of the modeled attributes; matrix $\boldsymbol{Q}$ of covariances of modeled attributes with the traits. Matrix $\boldsymbol{G}$ is given by equation (7.9)

$$\boldsymbol{G} = \boldsymbol{A}\boldsymbol{F}' \ . \tag{7.9}$$

Matrix $\boldsymbol{B}$ is given by equation (7.4)

$$\boldsymbol{B} = \boldsymbol{A}\boldsymbol{T}^{-1} \ . \tag{7.4}$$

Matrix $\boldsymbol{Q}$ is given by equation (7.5)

$$\boldsymbol{Q} = \boldsymbol{A}\boldsymbol{T}' \ . \tag{7.5}$$

Graphical transformation of factors uses the projections of the attribute vectors on the normals as measures of importance of the transformed factors to the modeled attributes. Note that these coefficients are proportional, factor by factor, to the factor weights in matrix $\boldsymbol{B}$. By equation (7.10):

$$\boldsymbol{G} = \boldsymbol{BD} \ . \tag{7.10}$$

An important point is that each column of $\boldsymbol{G}$ depends only upon the corresponding column of $\boldsymbol{F}$ while each column of factor weights in $\boldsymbol{B}$ depends upon the location of all the normals in matrix $\boldsymbol{F}$. This independence permits the shifting of each normal independently from the location of the other normals.

10.1. Graphical Rotation for Two Factor Studies

For a two factor study a single graph provides a view of the configuration of attribute vectors and transformed coordinate vectors. This single graph provides a convenient view of the entire transformed factorial structure. For a three factor study a unit sphere may be used to provide a view of the configuration of attribute vectors and transformed coordinate vectors as illustrated in Chapter 7, Figure 7.8; however, this is quite inconvenient and is not used in practice. The two dimensional representation in a graph provides a convenient opportunity to view the configuration of attribute vectors and the transformed coordinate vectors. The combination of attribute vectors and transformed coordinate vectors is termed a structure. Thurstone's simple structure is a structure well defined by many attribute vectors near the base entities, base lines for two dimensional studies.

Graphical rotation of factors for a two factor study will be illustrated by an analysis of a selected battery of verbal comprehension and word fluency tests taken from the Thurstone & Thurstone (1941) factorial studies of intelligence. The correlation matrix for the selected nine tests is given in Table 10.1 and the principal factors matrix is given in Table 10.2. This principal factors matrix is the given matrix $\boldsymbol{A}$ for which the factors are to be transformed. Figure 10.1 provides the graph of the attribute vector configuration. Each row of matrix $\boldsymbol{A}$ contains the coordinates of the terminus of an attribute vector on the principal factor axes. In the graph of Figure 10.1, the horizontal axis represents coordinates on principal factor 1 while the vertical axis represents coordinates on principal axis 2. There is a point on this graph for the terminus of each attribute vector. For clarity of the graph, the vectors are not drawn. Each vector emanates from the origin to its terminus represented by a point. In this illustration the points form two clusters. Such a configuration is not required for a simple structure. It would be acceptable to have a few points in the middle between two clusters. Also, the points do not have to form a cluster; rather, a rough radial line of points could replace a cluster.

In Figure 10.2 the lines of the transformed factor system have been added to the configuration of attribute points shown in Figure 10.1. This system can be thought of as defined by base lines, there being one such base line for each transformed factor. Consider Base Line 1 and Base Line 2 drawn on Figure 10.2. Each base line is restricted to passing through the origin.

Table 10.1

Correlation Matrix for V & W Tests Example

Variable	1	2	3	4	5	6	7	8	9
1	1.000								
2	.769	1.000							
3	.718	.681	1.000						
4	.730	.661	.672	1.000					
5	.227	.189	.280	.241	1.000				
6	.296	.219	.311	.311	.554	1.000			
7	.237	.212	.313	.245	.461	.479	1.000		
8	.243	.226	.348	.290	.506	.530	.520	1.000	
9	.304	.291	.374	.306	.408	.425	.514	.473	1.000

*Selected from 66 test study by Thurstone and Thurstone (1941). Factorial studies of intelligence; sample size = 710.

Table 10.2

Principal Factors Matrix for V & W Tests Example

Variable		1	2
Chicago Reading Tests, Vocabulary	1	.765	-.435
Chicago Reading Tests, Sentences	2	.707	-.437
Completion	3	.766	-.281
Same or Opposite	4	.727	-.337
Prefixes	5	.530	.425
Suffixes	6	.583	.401
First and Last Letters	7	.552	.415
First Letters	8	.583	.420
Four-Letter Words	9	.575	.294

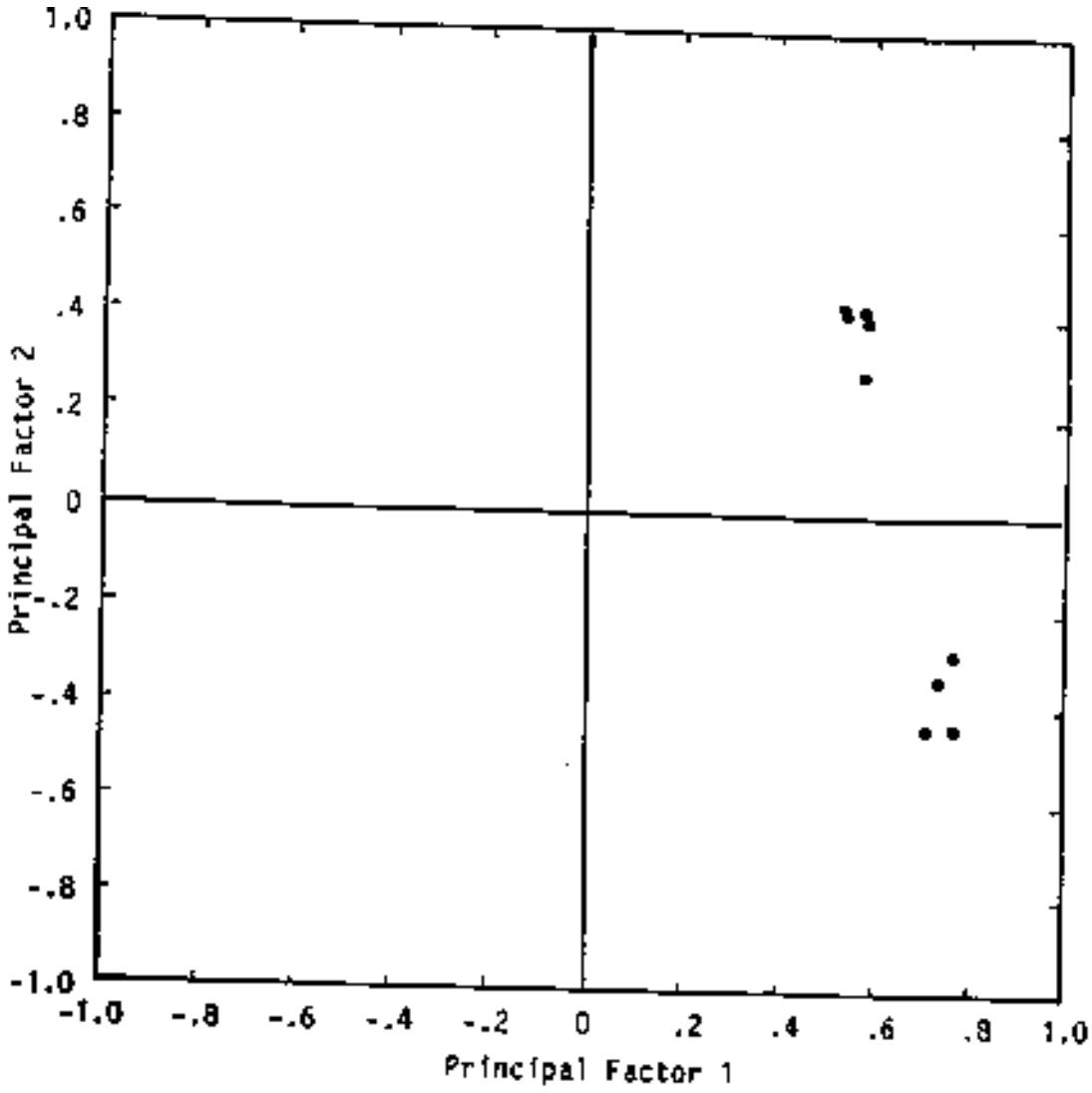

Figure 10.1 Principal Factors Configuration for V & W Tests Example.

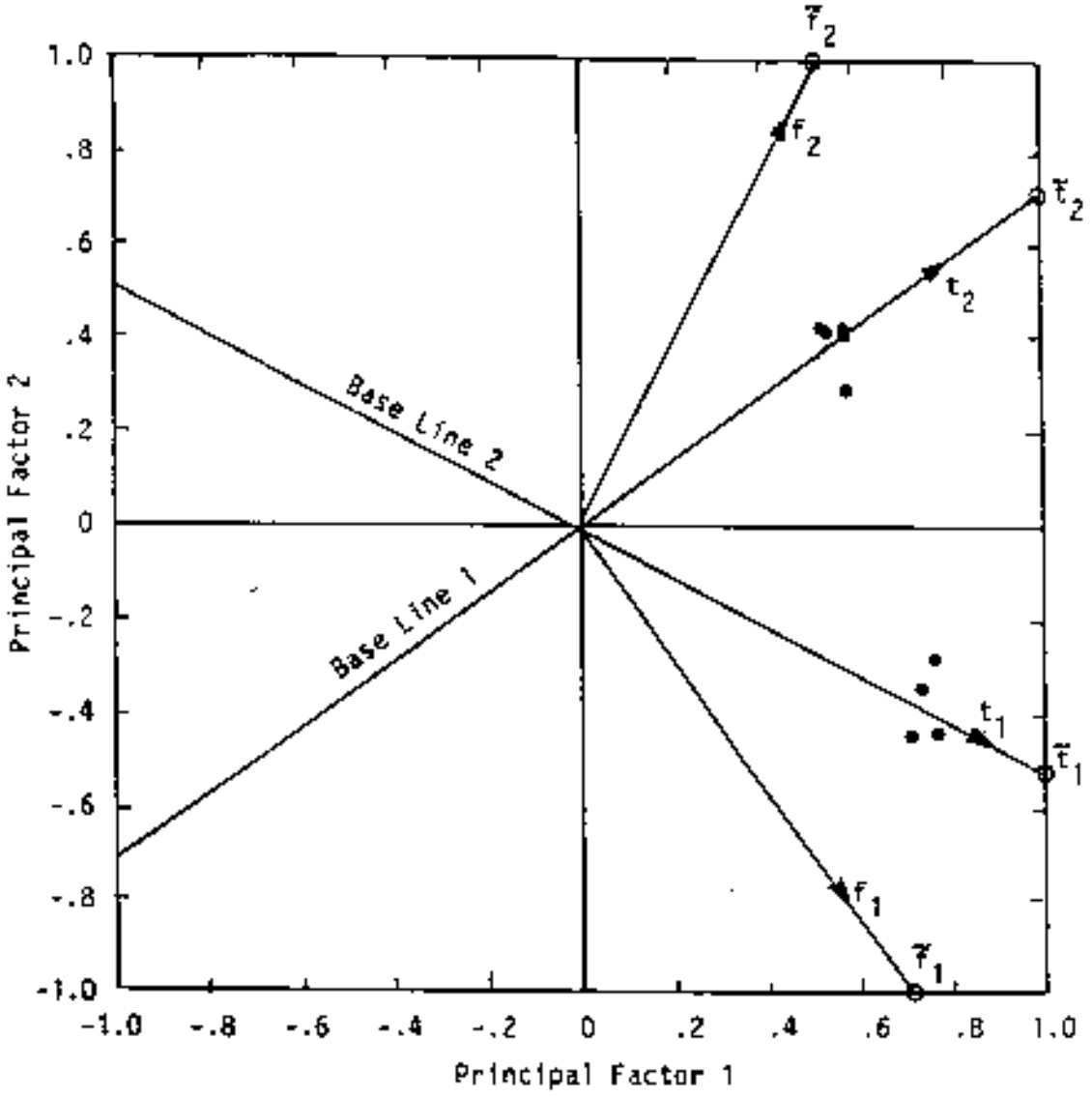

Figure 10.2 Factor Structure for V & W Tests Example.

In graphical rotation of factors subjective judgments are involved in drawing the base lines. For a simple structure each base line should pass near to a number of attribute points which may be in a rough radial line, there being no restriction that the points be in a cluster as in the illustration. Each base line is important in that it forms the zero for measuring influence of the transformed factor on each attribute. For each attribute having a point at a trivial distance from the base line the factor has a trivial influence. As the distances to points increase, influences of the factor on the attributes increase. While theoretic statements emphasize points at zero distances from a base line, in practice points are considered as being "in a base line" when these points are within a small interval from the base line. Frequently, this interval is considered to have a width of $\pm.10$. Subjective judgment is involved in seeing a radial line of points in the configuration which lie within such an interval from a possible base line and in drawing the base line.

Base Line 1 is drawn on Figure 10.2 from the lower left through the origin and the cluster of points in the upper right quadrant to a point $\tilde{t}_2$ which was chosen to have coordinates easily read from the graph. It is not necessary that such points be at a unit distance from the origin; when necessary, unit length vectors can be computed. The coordinates of point $\tilde{t}_2$ are (1.00, .71) which are recorded in Table 10.3 as line 2 of the Raw Traits Matrix $\tilde{\boldsymbol{T}}$. The normal to Base Line 1 is orthogonal to this base line and emanates from the origin to point $\tilde{f}_1$ at the lower right. This normal could have been directed toward the upper left but was chosen to be directed to the lower right so that the attribute points distant from the base line would be on the positive side of the base line. Coordinates of $\tilde{f}_1$ are (.71, -1.00) which are obtained from the coordinates of point $\tilde{t}_2$ by interchanging the coordinates of this point and changing the sign of one of the coordinates. Choice of which coordinate for which the sign is to be changed is according to the direction the normal is to take. This interchange of coordinates and change in the sign of one of the coordinates makes the scalar product equal to zero of vectors emanating from the origin to points $\tilde{t}_2$ and $\tilde{f}_1$. The coordinates of point $\tilde{f}_1$ are recorded in Table 10.3 in line 1 of the Raw Normals Matrix $\tilde{\boldsymbol{F}}$.

Base Line 2 is drawn on Figure 10.2 from the upper left through the origin and the cluster of attribute points at the lower right to point $\tilde{t}_1$ with coordinates (1.00, -.51) which are recorded in line 1 of the Raw Traits Matrix $\tilde{\boldsymbol{T}}$ in Table 10.3. The normal to Base Line 2 is orthogonal to this base line, emanates from the origin to point $\tilde{f}_2$ with coordinates (.51, 1.00) which are recorded in line 2 of the Raw Normals Matrix $\tilde{\boldsymbol{F}}$ in Table 10.3. The coordinates for $\tilde{f}_2$ were obtained by interchanging the coordinates of point $\tilde{t}_1$ and changing the sign of one of these coordinates. Choice of which coordinate to have its sign changed is made so that the points distant from the base line will be on the positive side of the base line.

Once the raw vectors have been obtained in the Raw Traits Matrix $\tilde{\boldsymbol{T}}$ and the Raw Normals Matrix $\tilde{\boldsymbol{F}}$ the unit length vectors are to be obtained in Traits Matrix $\boldsymbol{T}$ and Normals

Table 10.3

Graphical Transformation Matrices for V & W Tests Example

Raw Traits Matrix $\widetilde{T}$

	1	2	L^2	L
1	1.00	-.51	1.2601	1.1225
2	1.00	.71	1.5401	1.2264

Traits Matrix T

	1	2
1	.891	-.454
2	.815	.579

Raw Traits Matrix $\widetilde{F}$

	1	2	L^2	L
1	.71	-1.00	1.5041	1.2264
2	.51	1.00	1.2601	1.1225

Normals Matrix F

	1	2
1	.579	-.815
2	.454	.891

Cosines of Angles Between Normals FF'

	1	2
1	1.000	-.463
2	-.463	1.000

Factor Correlation Matrix R_{bb} = TT'

	1	2
1	1.000	.463
2	.463	1.000

Cosines of Angles Between Normals and Trait Vectors FT'

	1	2
1	.886	.000
2	.000	.886

Matrix $\boldsymbol{F}$. Let $\widetilde{\mathbf{x}}_{\boldsymbol{k}}$ with entries $\widetilde{x}_{jk}$ represent each of the raw vectors. The normalization process can be symbolized by computing the squared length, $\widetilde{\mathrm{L}}_{k}^{2}$, of the vector by:

$$\widetilde{\mathrm{L}}_{k}^{2} = \sum_{j=1}^{r} \widetilde{x}_{jk}^{2} \quad . \tag{10.7}$$

Each element in the vector is divided by the length of the vector to produce the unit length, or normalized vector:

$$\boldsymbol{x}_{\boldsymbol{jk}} = \widetilde{\boldsymbol{x}}_{\boldsymbol{jk}} / \widetilde{\mathrm{L}}_{k} \quad . \tag{10.8}$$

This procedure is applied to each of the raw vectors in matrices $\widetilde{\boldsymbol{T}}$ and $\widetilde{\boldsymbol{F}}$ to yield the Traits Matrix $\boldsymbol{T}$ and $\boldsymbol{F}$. Normals Matrix $\boldsymbol{F}$ in Table 10.3. These unit length vectors are shown by arrows on Figure 10.2.

The remaining matrices in Table 10.3 contain scalar products among the vectors in matrices $\boldsymbol{T}$ and $\boldsymbol{F}$. Since the vectors in matrices $\boldsymbol{T}$ and $\boldsymbol{F}$ have been normalized to unit length vectors, these scalar products between these vectors are cosines of the angles between these vectors. The first of these scalar product matrices is the product $\boldsymbol{FF'}$ and contains the cosines of the angles between the normals. The diagonal entries of unity are the lengths of the normal vectors and must be unity since these vectors were normalized. The next matrix is the scalar product matrix among the trait vectors and is obtained by the product $\boldsymbol{TT'}$. By equation (7.1) this is the factor correlation matrix $\boldsymbol{R_{bb}}$. In the two dimensional factor rotation case the off-diagonal entries in $\boldsymbol{R_{bb}}$ must be equal to the negative of the off-diagonal entries in $\boldsymbol{FF'}$. The bottom matrix contains the scalar products of the vectors in $\boldsymbol{F}$ and $\boldsymbol{T}$. This is a diagonal matrix due to the construction of the normals vectors being orthogonal to the base planes which contain the trait vectors. This is the diagonal matrix $\boldsymbol{D}$ as used in equations (7.6) and (10.3).

Table 10.4 gives the matrices for coefficients of the modeled attributes on the transformed factors. Matrix $\boldsymbol{G}$ contains the projections of the modeled attributes on the normals. These are the perpendicular distances of the attribute points from the base lines and are computed by equation (7.9); they are the scalar products of the attribute vectors with the normals. A convenient interpretation for the projection of an attribute vector on a normal is that it represents the contribution of a transformed factor to the attribute which is independent of the other transformed factors. It depends only on the location of the particular base line and not on the location of the other base line. In matrix $\boldsymbol{G}$ for the illustration the first four attributes have high coefficients on the first transformed factor. These are attributes for which their points were some distance from Base Line 1 which was drawn to be near the points for the last five attributes. Base Line 1 was drawn to be near these last five points. Note that the projections of these last five points are in the interval of $\pm.10$. Factor 1 is a good representation of a simple structure factor.

Table 10.4

Graphical Transformed Factor Matrices for V & W Test Example

Matrix G of Projections on Normals

	1	2
1	.798	-.040
2	.766	-.068
3	.673	.098
4	.696	.030
5	-.040	.619
6	.011	.622
7	-.019	.620
8	-.005	.639
9	.093	.523

Factor Weight Matrix B

	1	2
1	.900	-.045
2	.864	-.077
3	.759	.110
4	.785	.034
5	-.045	.699
6	.012	.702
7	-.021	.700
8	-.006	.721
9	.105	.590

Matrix Q of Covariances of Modeled Attributes with Traits

	1	2
1	.879	.372
2	.828	.323
3	.810	.482
4	.801	.398
5	.279	.678
6	.337	.708
7	.303	.690
8	.329	.719
9	.379	.639

For factor 2, the first four attributes have projections in the ±.10 interval which provides a nice representation of a simple structure factor. The locations of the base lines is dependent on the points having trivial projections on the normals.

Factor Weight Matrix $\boldsymbol{B}$ is the second matrix of Table 10.4 and was computed by equation (7.4). Its relation to the matrix of projections is given by equation (7.10) which can be rewritten as:

$$\boldsymbol{B} = \boldsymbol{G}\boldsymbol{D}^{-1} \ . \tag{10.9}$$

The factor weights are proportional to the projections on the normals, the constants of proportionality being the reciprocals of the diagonal entries in matrix $\boldsymbol{D}$. For uncorrelated transformed factors matrix $\boldsymbol{D}$ is an identity matrix and matrices $\boldsymbol{G}$ and $\boldsymbol{B}$ are equal. As the factors become more correlated, the entries in $\boldsymbol{D}$ become smaller so that the entries in $\boldsymbol{B}$ become larger for the same sized entries in $\boldsymbol{G}$. As with multiple regression, when the independent variables are highly correlated the regression weights are large and do not represent contributions to the dependent variable. This is true for the factor weights.

Matrix $\boldsymbol{Q}$ of Table 10.4 contains the covariances of Modeled Attributes with the traits. These are scalar products of the attribute vectors in matrix $\boldsymbol{A}$ with the trait vectors as given in equation (7.5). Equations (7.1), (7.4), and (7.5) may be combined to produce:

$$\boldsymbol{Q} = \boldsymbol{B}\boldsymbol{R}_{bb} \ . \tag{10.10}$$

It is apparent from this equation that the coefficients in $\boldsymbol{Q}$ combine both the factor weights and the factor correlations. Therefore, these coefficients do not represent the independent contributions of the factors to the modeled attributes. Note that: for uncorrelated factors matrix $\boldsymbol{D}$ is an identity matrix so that matrices $\boldsymbol{G}$, $\boldsymbol{B}$, and $\boldsymbol{Q}$ are equal. Some investigators prefer this type transformation; however, when base lines are well defined by radial lines of points, the correlated factors transformation would provide a preferred simple structure solution.

10.2. Graphical transformation of Factors for Three and More Factor Studies.

For a three factor study, a vector model of the configuration of points can be constructed and a transformed coordinate system of planes, normals and trait vectors inserted. As an alternate, Thurstone (1947) described a procedure to plot the termini of attribute vectors extended to unit length on the surface of blackboard sphere and to draw in the coordinate planes as great circles. The coordinates of the trait vectors and normals could be read on this sphere. Such a procedure was quite unwieldy and not used except as a demonstration. Further, such a procedure was not available for studies involving four or more factors. A usable procedure was to make a set of two dimensional graphs plotting projections on normals for all pairs of factors. This

procedure involved a series of rotations with each rotation made from a set of graphs for tentative transformed normals. Projections on a new tentative normals were computed and a new set of graphs were made. This procedure was continued until a satisfactory set of graphs were obtained. Details for this procedure are described in the following paragraphs with an illustration using the nine mental tests example.

Graphical transformations start from some initial transformation. In the case of the illustration the procedure was started from the principal factors for the nine mental tests example. This implied an initial normals matrix equal to an identity matrix. Table 10.5 gives the initial normals matrix and projections on these normals (which equal the loadings on the principal axes). Frequently some one of the analytical transformation methods (to be described in Chapter 11) is used to provide the initial normals matrix and projections on the normals. A normal VARIMAX transformation provides a good initial normals matrix and projections on the normals. Sometimes the computer output does not give the normals matrix, only the factor weights, matrix $\boldsymbol{B}$, and/or the projections on the normals, matrix $\boldsymbol{G}$, being given along with the correlations among the factors, matrix $\boldsymbol{R_{bb}}$. In this case the initial normals matrix, $\boldsymbol{F_0}$, must be computed by the appropriate one of the following formulas.

$$\boldsymbol{F_0} = \boldsymbol{G'A}(\boldsymbol{A'A})^{-1} \quad . \tag{10.11}$$

$$\boldsymbol{F_0} = \boldsymbol{DB'A}(\boldsymbol{A'A})^{-1} \tag{10.12}$$

where

$$\boldsymbol{D} = [Diag(\boldsymbol{R_{bb}^{-1}})]^{-\frac{1}{2}} \quad . \tag{7.8}$$

Matrix $\boldsymbol{G_0}$ of projections on the normals is given by:

$$\boldsymbol{G_0} = \boldsymbol{AF_0'} \quad . \tag{7.9}$$

An initial set of factor plots is made using projections on the normals in matrix $\boldsymbol{G_0}$. Each pair of factors is used in making one of the plots in the set; thus, there are r(r $-$ 1)/2 plots in the set. For the three factor, nine mental tests example there are three plots in Figures 10.3a, 10.3b, 10.3c. The attribute points are labeled with the attribute numbers for convenience in cross referencing between the several plots in the set. A base line is drawn using subjective judgment for each of the transformed factors on one or another of the plots in the set. Each base line should pass near a number of the attribute points. With aptitude and intelligence measures each base line should be near one side of the configuration of points in accordance with an assumption that all factor weights in these cases should be positive. This is a very powerful principal termed a positive manifold.

Table 10.5

Initial Graphical Transformation
Nine mental Tests Example

Normals Matrix F_0

	1	2	3
1	1.000	.000	.000
2	.000	1.000	.000
3	.000	.000	1.000

Projections on Normals

Matrix G^*_0

Variable		1	2	3
Addition	1	.42	.36	.28
Multiplication	2	.47	.54	.16
Three-Higher	3	.61	.16	.19
Figures	4	.54	-.46	.09
Cards	5	.63	-.48	.05
Flags	6	.59	-.41	.06
Identical Numbers	7	.48	.40	-.21
Faces	8	.61	.00	-.29
Mirror Reading	9	.59	.16	-.25

*Principal factor of R with SMC's in diagonal.

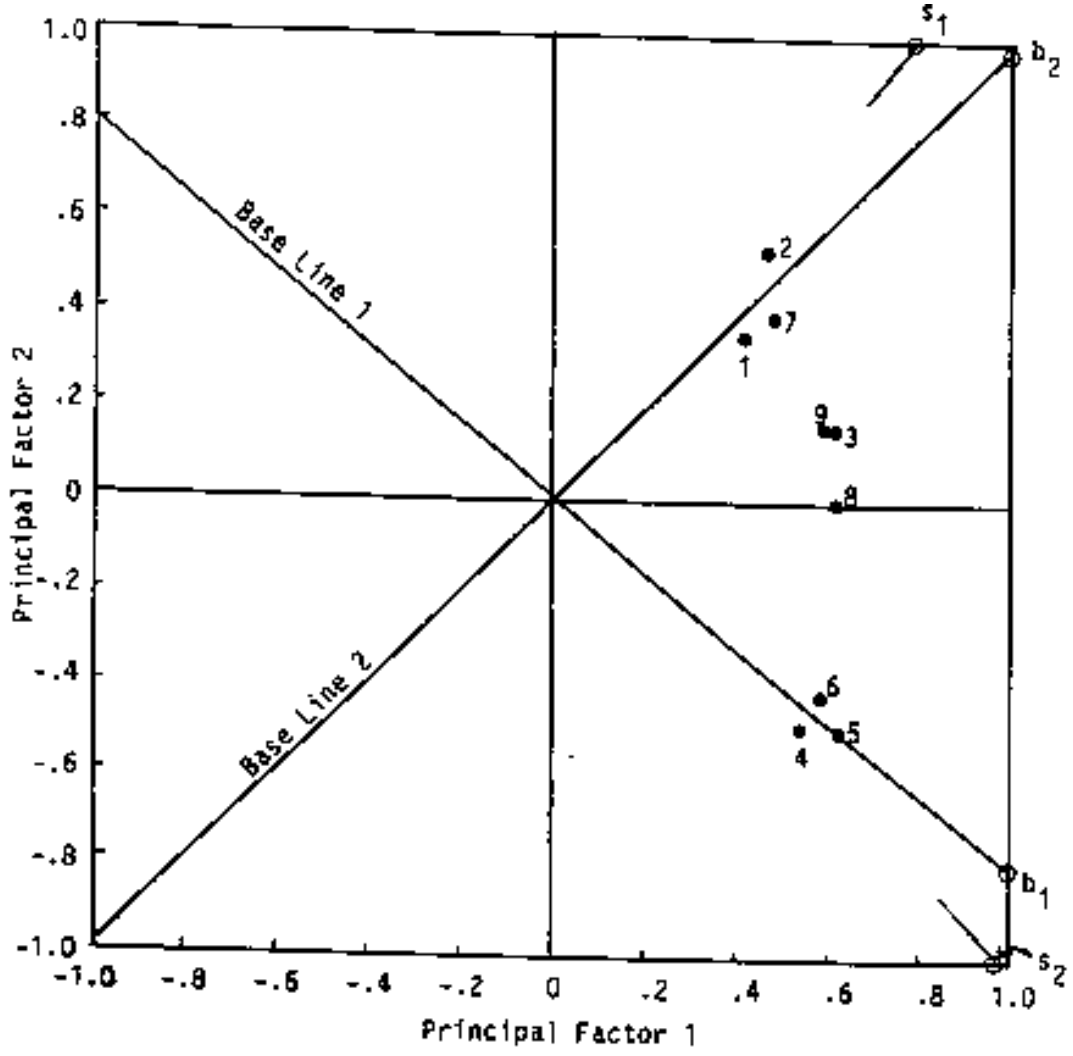

Figure 10.3a Nine Mental Tests: Factor Plot for Principal Factors 1 and 2.

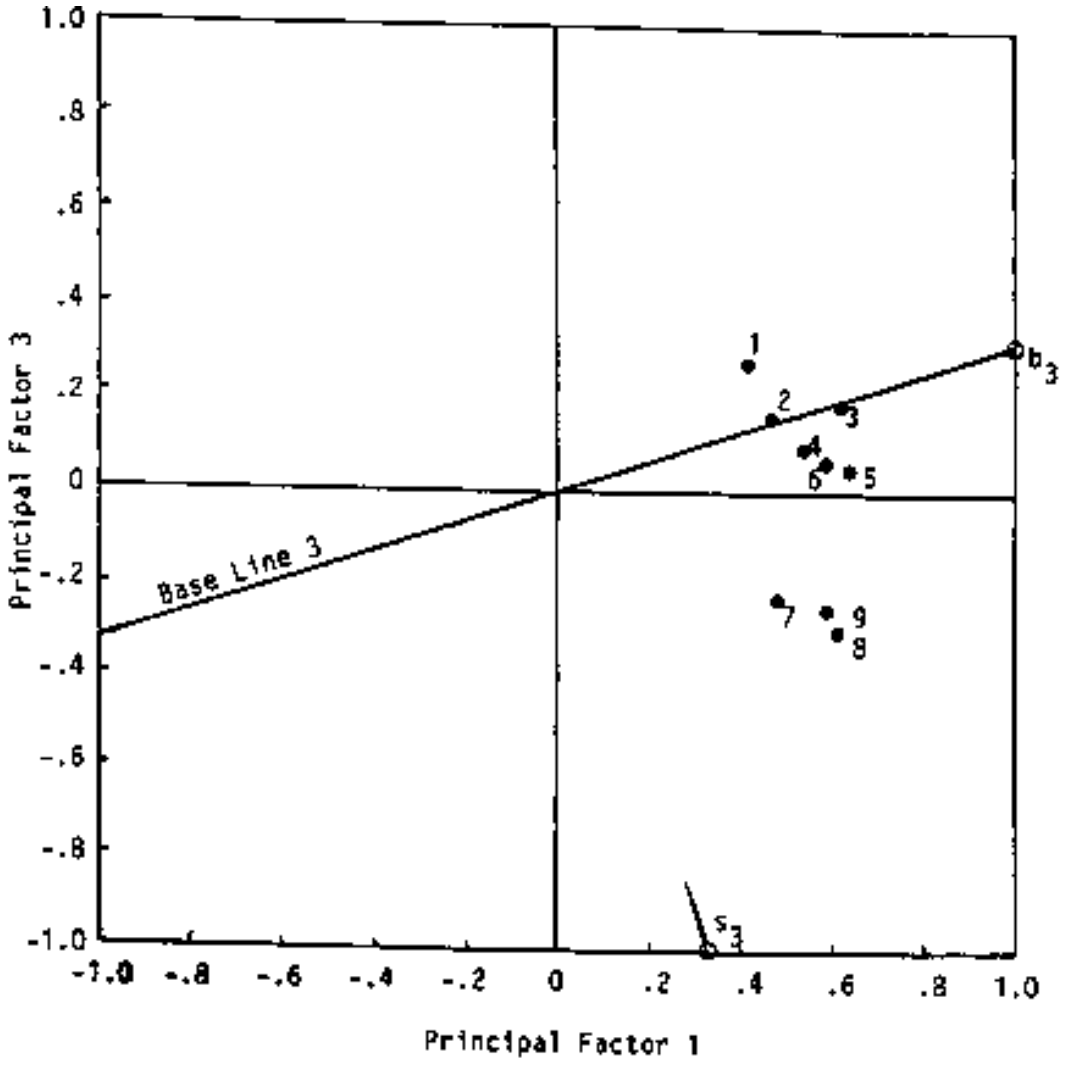

Figure 10.3b Nine Mental Tests: Factor Plot for Principal Factors 1 and 3.

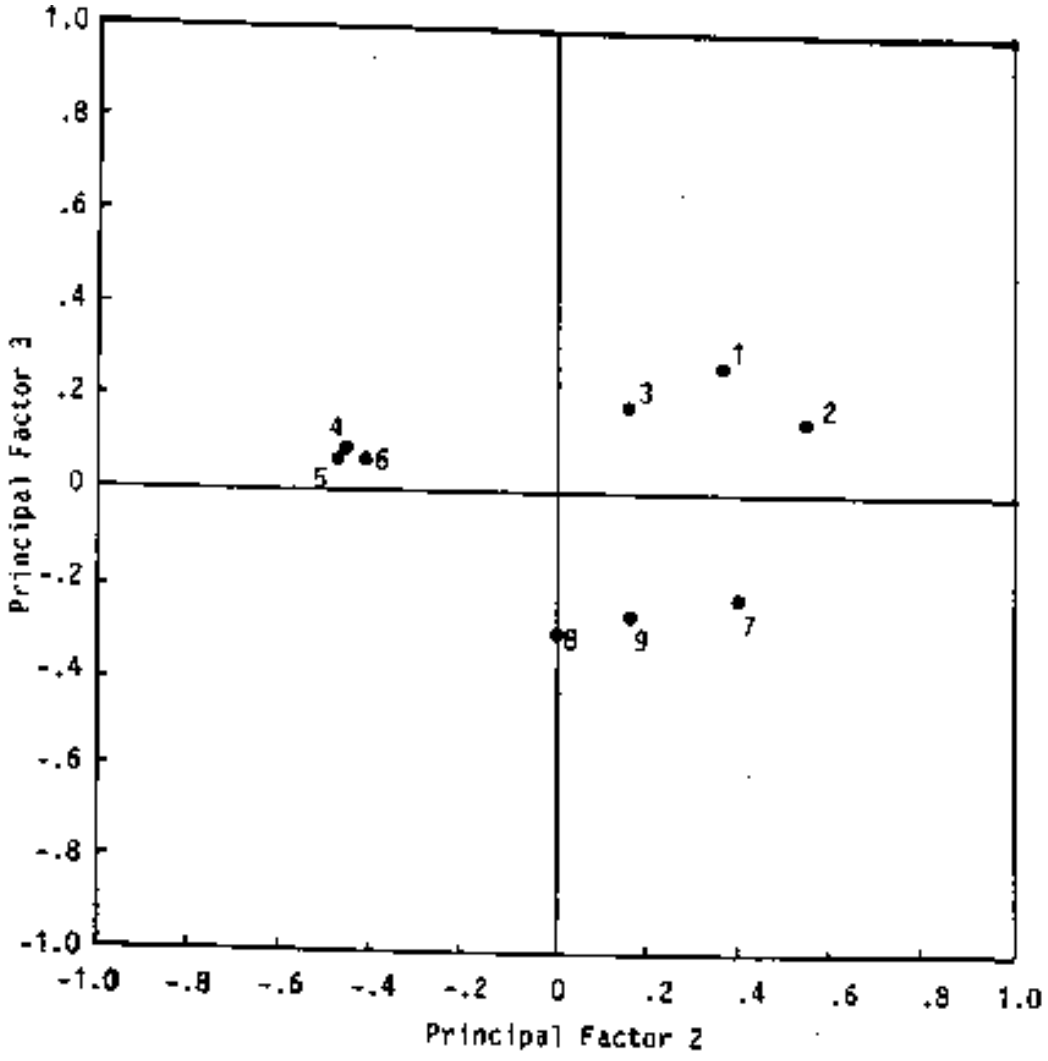

Figure 10.3c Nine Mental Tests: Factor Plot for Principal Factors 2 and 3.

Base Line 1 for transformed factor 1 is drawn on the plot for factors 1 and 2 on Figure 10.3a to be near points 4, 5, 6. Point b_1 is chosen on Base Line 1 to have conveniently read coordinates of (1.00, -.80) and point s_1 is indicated to be on a line from the origin orthogonal to Base Line 1. The coordinates of s_1 , (.80, 1.00), have the coordinates of b_1 reversed with one of them reversed in sign. Choice as to which coordinate to reverse in sign is dictated by the location desired for point s_1. The desired point may be plotted on the graph as shown on Figure 10.3a. This direction of the point was chosen so that the majority of the attribute points would be on the positive side of Base Line 1.

Base Line 2 was drawn on the plot between factors 1 and 2 on Figure 10.3a from the lower left through the origin and near points 1, 2, 7 to point b_2 with coordinates (1.00, .98). Point s_2 is at the lower right on a line from the origin orthogonal to Base Line 2. The coordinates of s_2 are (.98, -1.00) which are the coordinates of b_2 interchanged and the sign of one coordinate reversed in sign. The direction to the lower right was chosen such that the attribute points not near Base Line 2 would be on the positive side of Base Line 2. Note that the cluster of points 3, 8, 9 was ignored in drawing base lines since this cluster is in the middle of the configuration so that some attribute points would be on the negative side of such a base line while other attribute points would be on the positive side. To choose such a base line would violate the principal of a positive manifold.

Base Line 3 was drawn on the plot for factors 1 and 3 on Figure 10.3b near a loose cluster of all attribute points except for points 7, 8, 9. Point b_3 has coordinates (1.00, .32). Point s_3 has coordinates (.32, -1.00).

The raw trait vectors, $\widetilde{\mathbf{t}_k}$, of the plot for a two factor study illustrated on Figure 10.2 have been replaced by base line marked points b_k . Likewise, the raw normals vectors, $\widetilde{f_k}$, have been replaced by shift vectors $\mathbf{s}_k$ which operate to shift the normals vectors to new raw normals vectors. These shift vectors are recorded as rows in a shift matrix $\boldsymbol{S}$. Consider Table 10.6; the three shift vectors are defined by points s_1 , s_2 , s_3 from the factor plots in Figures 10.3a, 10.3b. The coordinates of these shift vectors are recorded in Shift Matrix $\boldsymbol{S_1}$. The new raw normals matrix, $\widetilde{F_1}$, is obtained by:

$$\widetilde{F_1} = \boldsymbol{S_1 F_0} \tag{10.13}$$

The vectors in raw normals matrix $\widetilde{\boldsymbol{F_1}}$ are normalized to unit length vectors in the normals matrix $\boldsymbol{F_1}$. Matrix product $\boldsymbol{F_1 F_1'}$ containing the cosines of the angles between the normals appears at the lower left of Table 10.6. These cosines of angles will be important to consider in the next set of graphs. The projections on the normals, matrix $\boldsymbol{G_1}$, is computed by:

$$\boldsymbol{G_1 = A F_1'} \ . \tag{7.9}$$

Table 10.6

Graphical Transformation 1
Nine Mental tests Example

Shift Matrix S_1

	1	2	3
1	.80	1.00	.00
2	.98	-1.00	.00
3	.32	.00	-1.00

Raw Normals Matrix $\widetilde{F}_1$

	1	2	3	L^2	L
1	.800	1.000	.000	1.6400	1.2806
2	.980	-1.000	.000	1.9604	1.4001
3	.320	.000	-1.000	1.1024	1.0500

Normal Matrix F_1

	1	2	3
1	.625	.781	.000
2	.700	-.714	.000
3	.305	.000	-.952

Cosines of Angles between Normals F_1F_1'

	1	2	3
1	1.00	-.12	.19
2	-.12	1.00	.21
3	.19	.21	1.00

Projections on Normals Matrix G_1

	1	2	3
1	.54	.04	-.14
2	.72	-.06	-.01
3	.51	.31	.01
4	-.02	.71	.08
5	.02	.78	.14
6	.05	.71	.12
7	.61	.05	.35
8	.38	.43	.46
9	.49	.30	.42

After the initial graphical transformation is accomplished new factor plots are made using the projections on the normals in matrix $\boldsymbol{G}_1$. The normals are no longer orthogonal as shown in matrix $\boldsymbol{F}_1\boldsymbol{F}_1'$ which contains the cosines of the angles between the normals. A true representation of the configuration would require plotting on oblique axes. Consider Figure 10.4 for an illustration of the procedure to make these plots. Two normals, $\mathbf{f}_1$ and $\mathbf{f}_2$, are shown at an angle of 120° (cosine = -.50) from each other. Zero Line 1 is drawn at a right angle to normal vector $\mathbf{f}_1$. This is the locus of all points which have a zero projection on normal $\mathbf{f}_1$. Similarly, Zero Line 2 is drawn at a right angle to the vector for normal $\mathbf{f}_2$. Point j has projections on these normals of .3 and .5 . The projection of .3 of j on normal $\mathbf{f}_1$ is measured off on this normal and a projection line is drawn at a right angle to the normal. This projection line is parallel to Zero Line 1 and is the locus of all points which have a projection of .3 on normal $\mathbf{f}_1$. Similarly, the projection of .50 of point j on normal $\mathbf{f}_2$ is measured off on this normal and a projection line is drawn at a right angle to this normal. Point j is located at the intersection of the two projection lines. However, this procedure is quite tedious and need not be resorted to except when two normals are very oblique. The factor plots may be made on orthogonal axes with very little distortion except when two normals are very oblique (cosine of the angle between them more discrepant from zero than some value of $\pm.70$). Figure 10.5 provides an illustration of the distortion when the cosine of the angle between the normals is -.50. Eight points are on the sides of a square when an oblique plot is made as in the upper plot of Figure 10.5. When these points are plotted using orthogonal axes as in the lower figure the square is skewed to a diamond shape. Distances from upper left to lower right are increased while distances from lower left to upper right are decreased. However, all straight lines remain straight lines. Zero Line 1 has been rotated to coincide with normal vector $\mathbf{f}_2$; also, Zero Line 2 has been rotated to coincide with normal vector $\mathbf{f}_1$. Distances of points from these zero lines have remained unchanged in the transformation from oblique axes to orthogonal axes. While normal vectors appear to move around with reference to the configuration of points, the zero lines appear to have a constant relation to the configuration of points. Consequently, there is little effect on the distances of points from new base lines when these base lines are near zero lines. Experience has shown that the distortion from using orthogonal axes has little effect on subjective judgments in driving new base lines near lines of points.

The factor plots for rotation 1 of the nine mental tests example were made from the projections on the normals in matrix $\boldsymbol{G}_1$ of Table 10.6. These plots were studied for improvements in points near each of the zero lines. These improvements result in shifts of the normals. In the plot between factors 1 and 3 there are no points distance from the origin near the zero line for factor 1. This lead to a Base Line 1 being drawn among points 7, 8, 9. Then coordinates of base point b_1 were read from the graph and the coordinates of shift point s_1

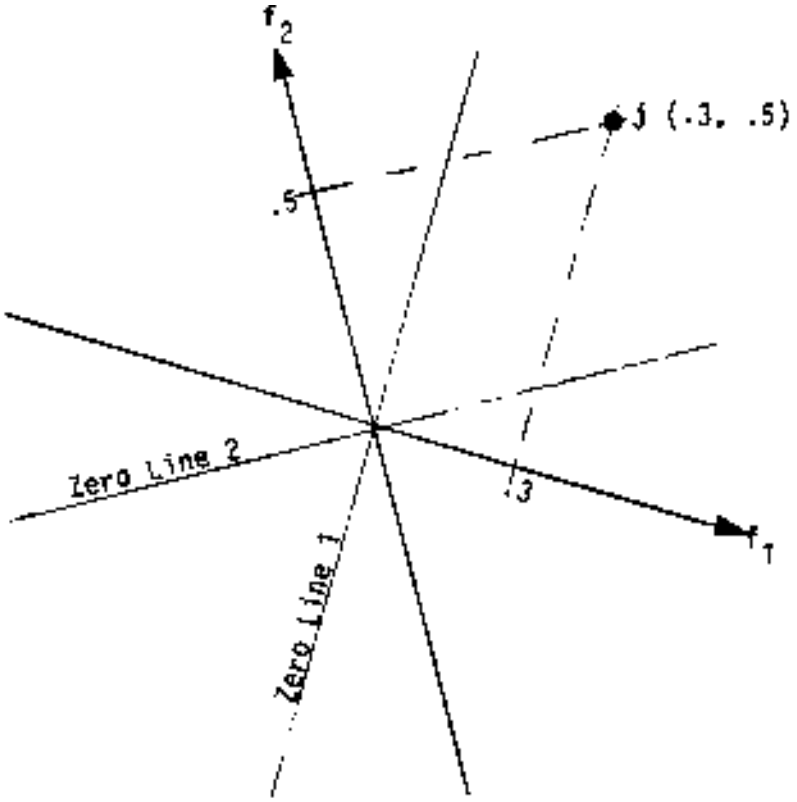

Figure 10.4 Illustration of plotting a point on oblique axes; angle between axes = 120° , cosine = -.5.

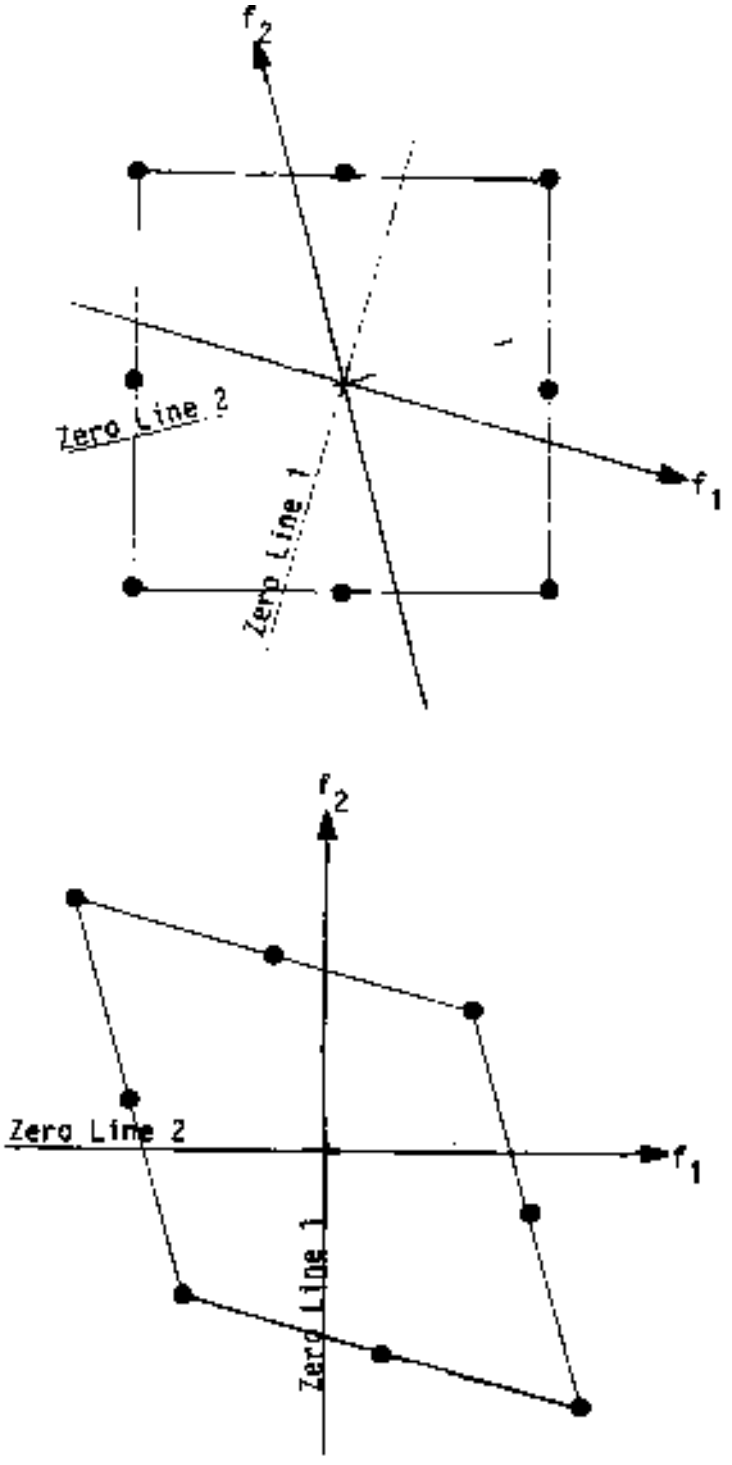

Figure 10.5 Distortion of a configuration plotted on orthogonal axes when the normals are oblique.

obtained as previously described. The coordinates of s_1 are recorded in the first row of matrix $\boldsymbol{S_2}$ in Table 10.7. A more problematic situation occurs on Figure l0.6c for the zero line for factor 2. Point 7 is near the zero line for factor 2. However, points 7, 8, 9 might form a rather wide cluster. A decision was made to try having such a cluster centered at zero for factor 2. Base Line 2 was drawn through the approximate center of this cluster. For factor 3 on the plot for factors 2 and 3 of Figure 10.6c there is a very distinct improvement resulting drawing Base Line 3 through the cluster of points 4, 5, 6 . Shift points s_2 and s_3 were developed from base points b_2 and b_3. The coordinates of s_2 and s_3 are recorded in rows of shift matrix $\boldsymbol{S_2}$ in Table 10.7.

The raw normals matrix, $\widetilde{\boldsymbol{F_2}}$ for rotation 2 is given in Table 10.7 and was computed by:

$$\widetilde{\boldsymbol{F_2}} = \boldsymbol{S_2 F_1} \ .$$

Rows of $\widetilde{\boldsymbol{F_2}}$ were normalized to unit length in normals matrix $\boldsymbol{F_2}$ and the matrix, $\boldsymbol{F_2 F_2'}$, was computed. The matrix of projections on the revised normals, $\boldsymbol{G_2}$ was computed by:

$$\boldsymbol{G_2 = A F_2'} \ .$$

Factor plots for rotation 2 are given in Figures 10.7a, 10.7b, 10.7c and were inspected for further adjustments of the zero lines to be near radial lines of points. A small adjustment for factor 1 appeared desirable on the plot for factors 1 and 2 in Figure 10.7a. Also, a small adjustment for factor 3 appeared desirable on the plot for factors 1 and 3 in Figure 10.7b. A larger adjustment for factor 2 was decided upon, see Figure 10.7c for the plot between factors 2 and 3. This adjustment approximately reverses a decision made for rotation 1 on Figure 10.6c. At that time it was noted that points 7, 8, 9 might form a loose cluster. The further view given in Figure 10.7c lead to a judgment that these points were too widely separated to form such a cluster. Two other matters were considered in deciding to pass the new Base Line 2 through point 7: first was the principle of a positive manifold; second was a matter of interpretation. With the decision to pass Base Line 2 through point 7, all points except point 2 were now positive with point 2 only trivially negative. This satisfies the positive manifold principle. Use of interpretation to guide factor transformations greatly weakens the strength of conclusions derived from a study. However, in the case of the example, reference to the attributes named in Table 10.5 indicates that factor 2 could be some sort of spatial factor with attributes 8 and 9. Faces and Mirror Reading, having small positive loadings. As noted above, this use of interpretations to guide factor transformations weakens conclusions which may be made from factor studies. Only conjectures as to the nature of factors appear to be justified. Follow up studies involving new measures based upon these tentative conclusions are imperative to strengthen interpretations. A program of studies should lead to confirmatory studies based on interpretations developed in the program.

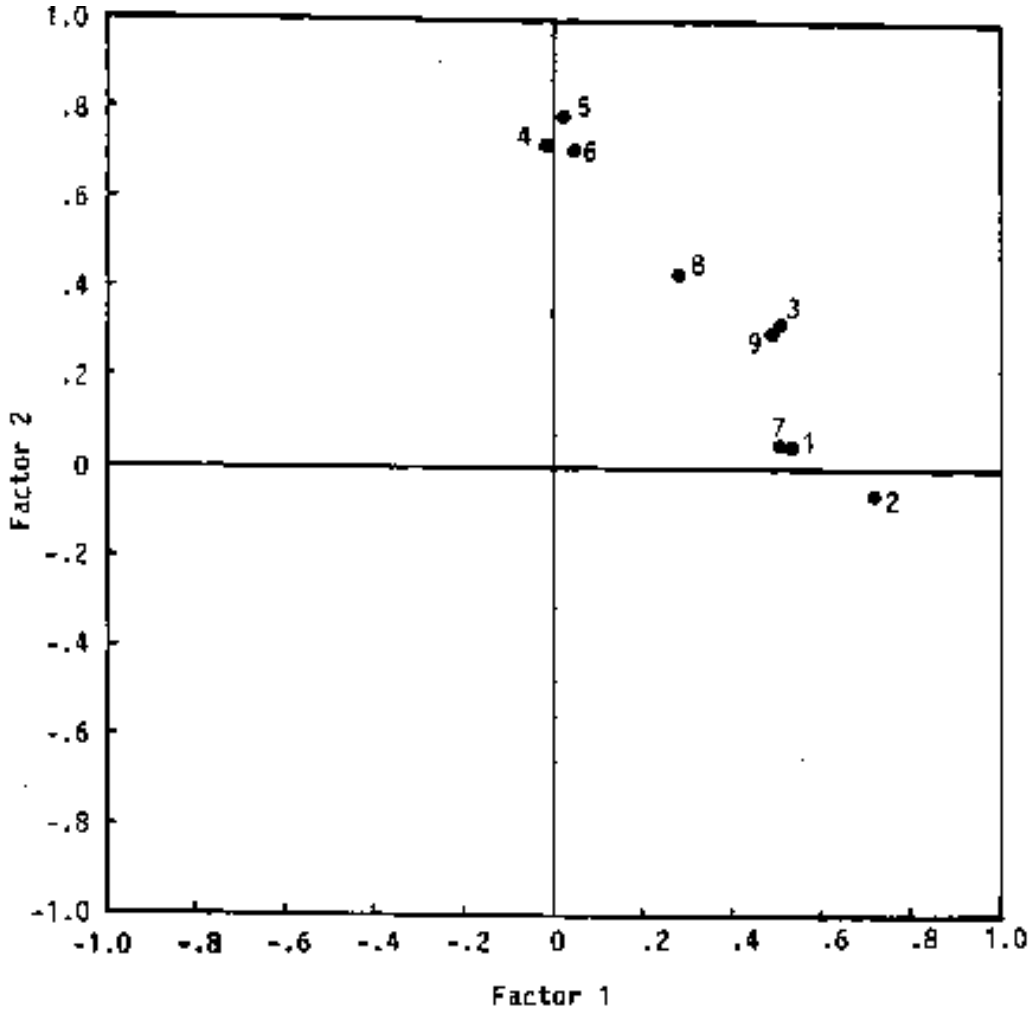

Figure 10.6a Nine Mental Tests: Rotation 1, Factor Plot for Factors 1 and 2.

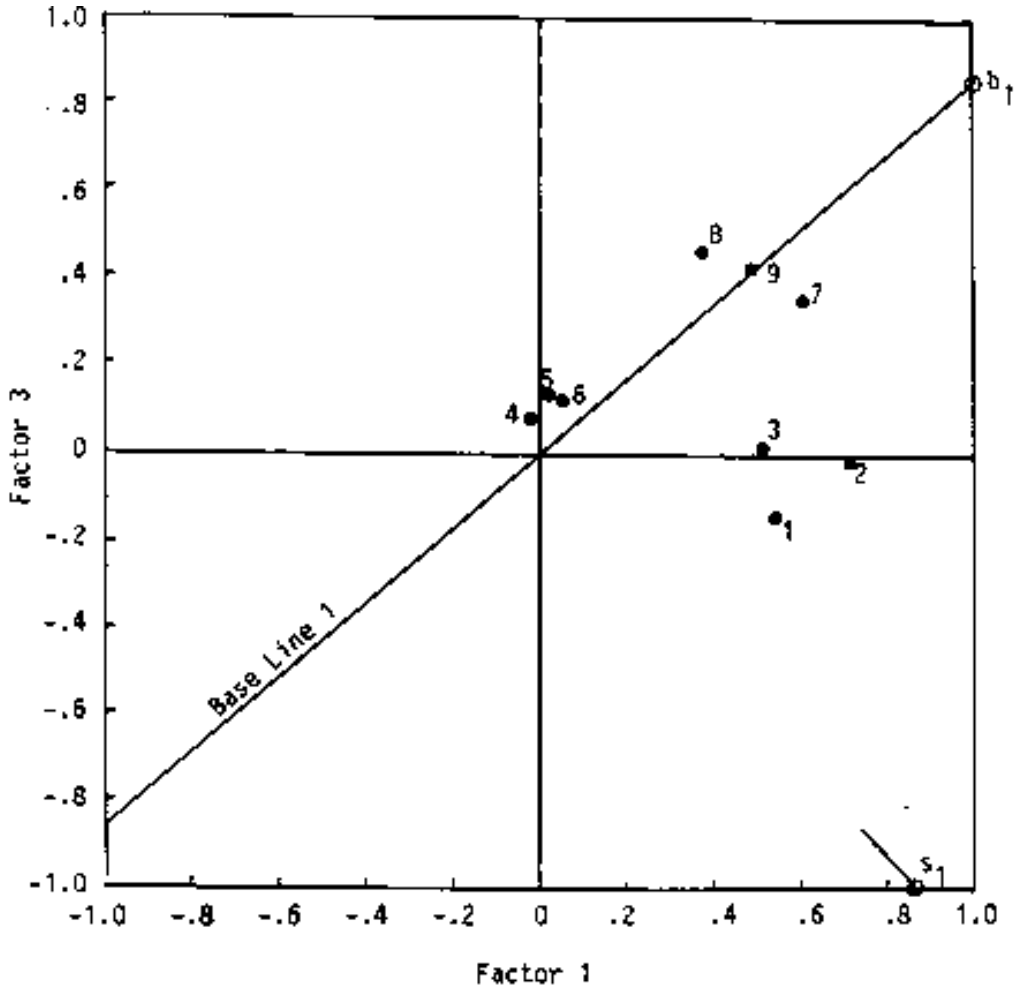

Figure 10.6b Nine Mental Tests: Rotation 1, Factor Plot for Factors 1 and 3.

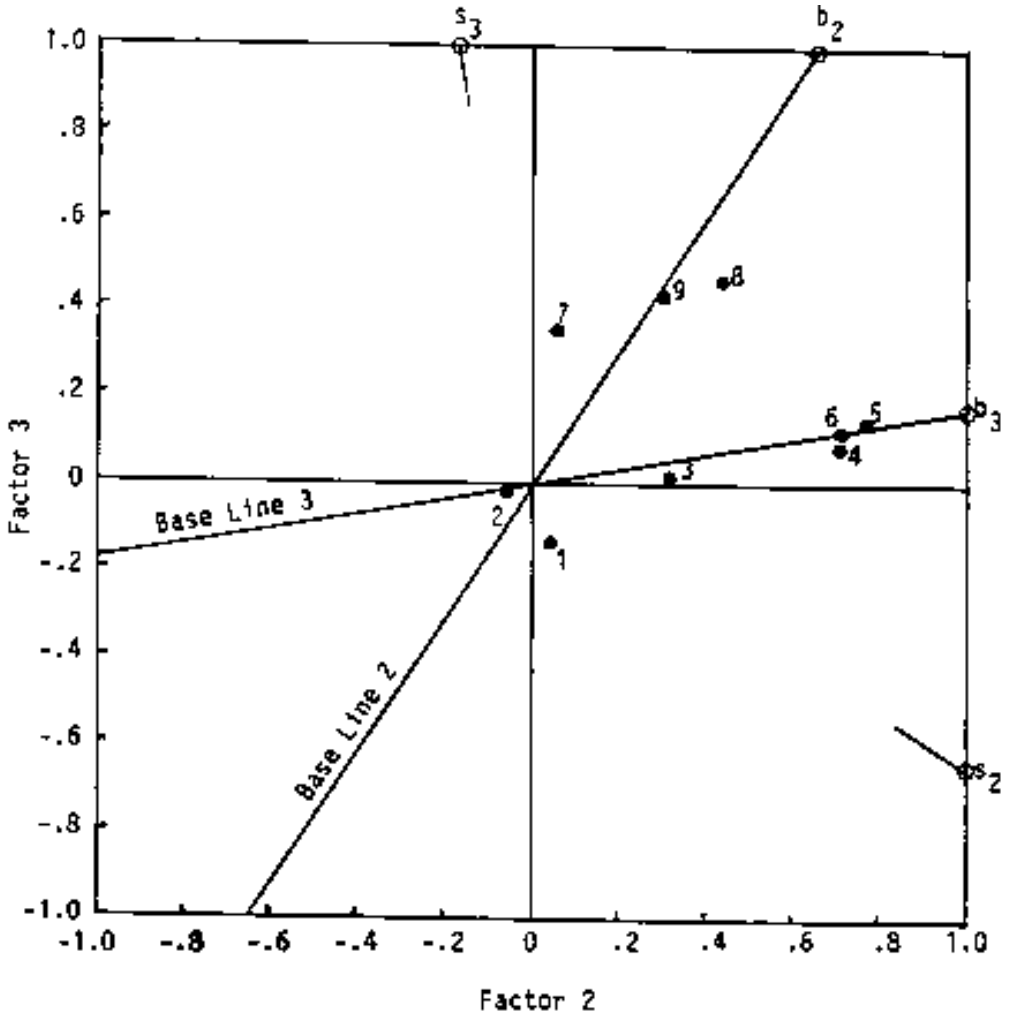

Figure 10.6c Nine Mental Tests: Rotation I, Factor Plot for Factors 2 and 3.

Table 10.7

Graphical Transformation 2
Nine Mental tests Example

Shift Matrix S_2

	1	2	3
1	.86	.00	-1.00
2	.00	1.00	-.66
3	.00	-.17	1.00

Raw Normals Matrix $\widetilde{F}_2$

	1	2	3	L^2	L
1	.232	.672	.952	1.4121	1.1883
2	.499	-.714	.629	1.1540	1.0743
3	.186	.121	-.952	.9564	.9779

Normal Matrix F_2

	1	2	3
1	.196	.565	.801
2	.464	-.665	.585
3	.190	.124	-.974

Cosines of Angles between Normals F_2F_2'

	1	2	3
1	1.00	.18	-.67
2	.18	1.00	-.56
3	-.67	-.56	1.00

Projections on Normals Matrix G_2

	1	2	3
1	.51	.12	-.15
2	.53	-.05	.00
3	.36	.29	-.05
4	-.08	.61	-.04
5	-.11	.64	.01
6	-.07	.58	.00
7	.15	-.17	.35
8	-.11	.11	.40
9	.01	.02	.38

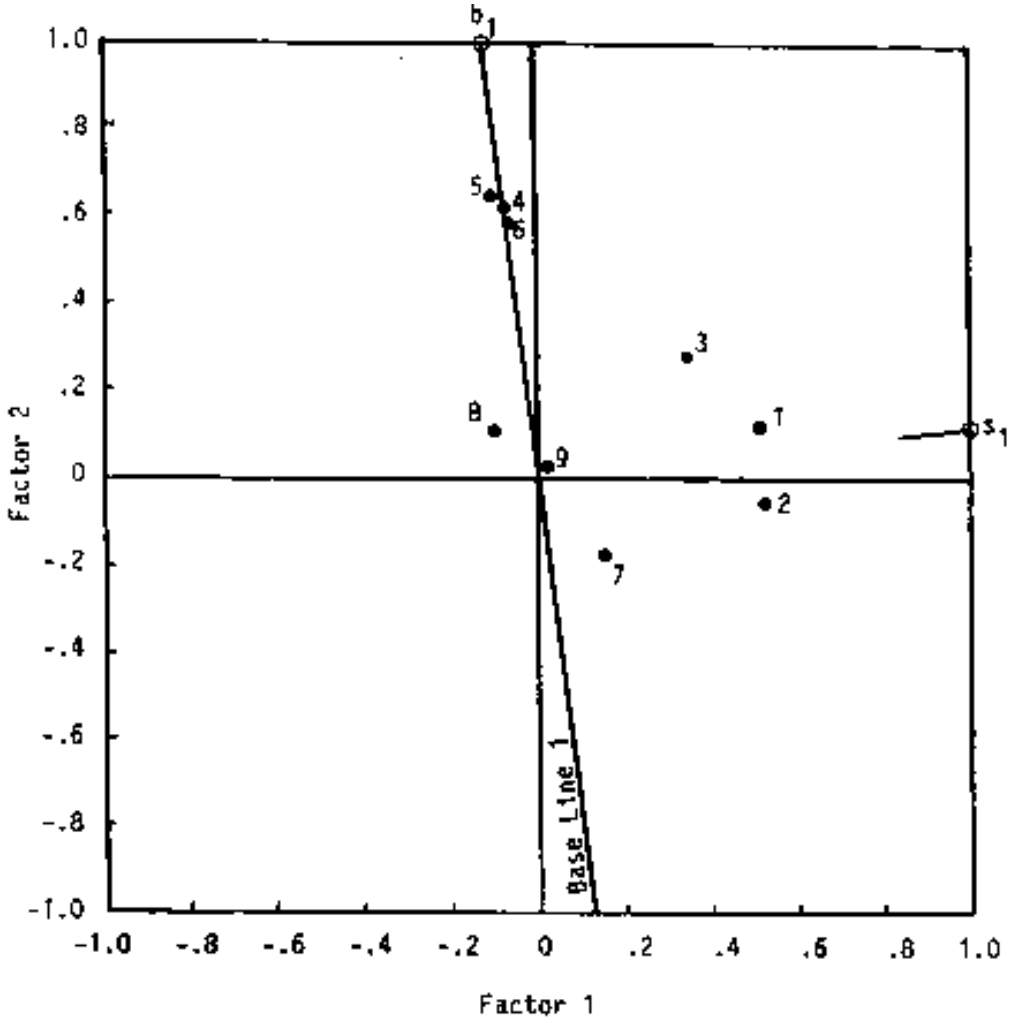

Figure 10.7a Nine Mental Tests: Rotation 2, Factor Plot for Factors 1 and 2.

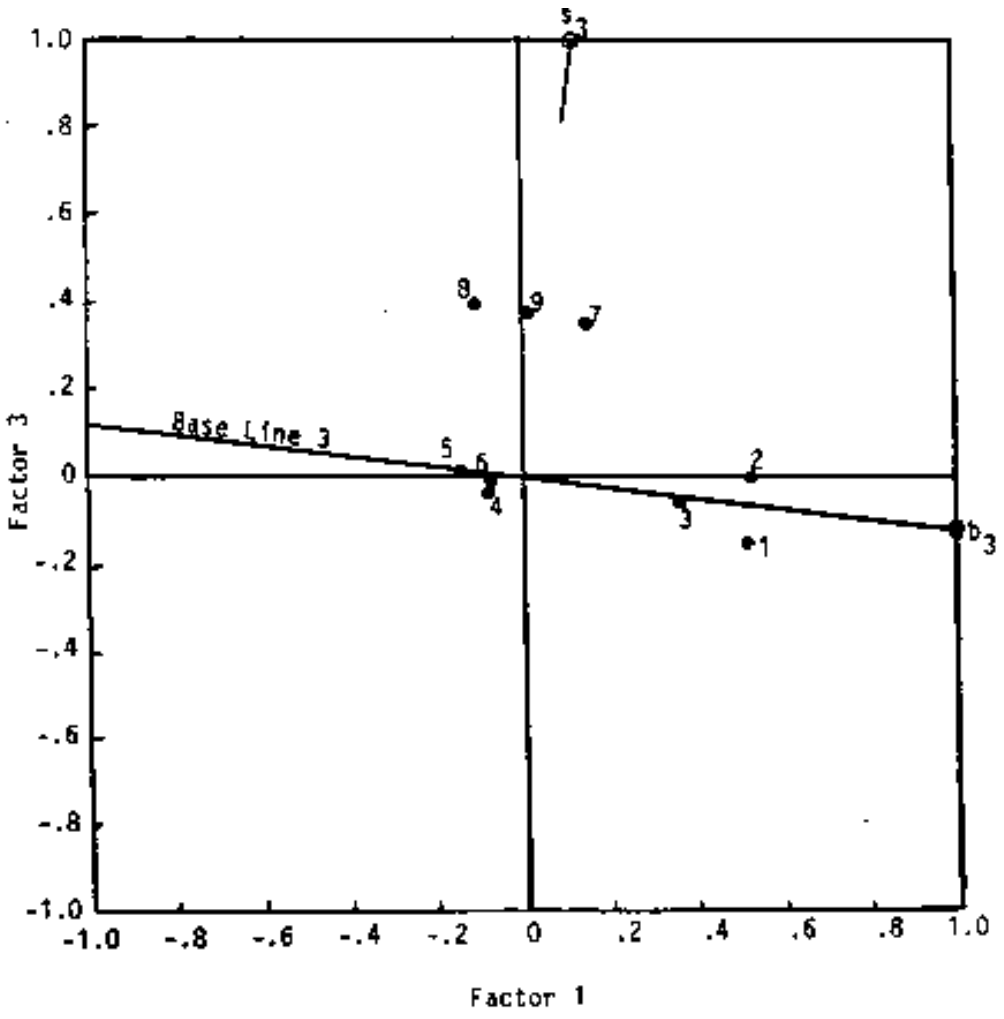

Figure 10.7b Nine Mental Tests: Rotation 2, Factor Plot for Factors I and 3.

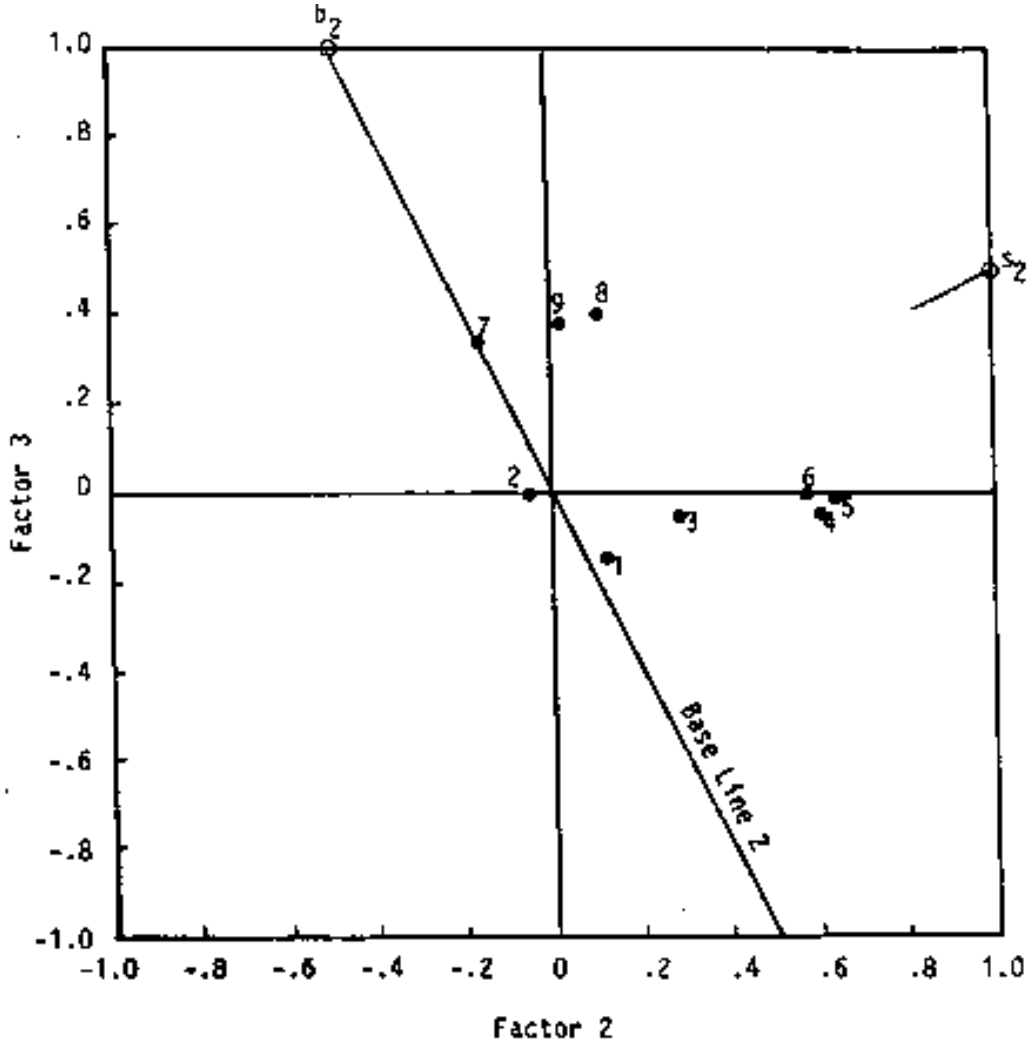

Figure 10.7c Nine Mental Tests: Rotation 2, Factor Plot for Factors 2 and 3.

Matrix $\boldsymbol{S_3}$ in Table 10.8 contains the coordinates of the shift points read from the factor plots for rotation 2 in Figures 10.7a, 10.7b, 10.7c. Raw normals matrix $\widetilde{\boldsymbol{F_3}}$ was computed by:

$$\widetilde{\boldsymbol{F_3}} = \boldsymbol{S_3 F_2}$$

and rows of $\widetilde{\boldsymbol{F_3}}$ were normalized to unit length vectors in normals matrix $\boldsymbol{F_3}$. Computations continued with matrix $\boldsymbol{F_3 F_3'}$ of cosines of angles between the normals and matrix $\boldsymbol{G_3}$ of projections on the normals.

$$\boldsymbol{G_3 = A F_3'} \ .$$

Factor plots for rotation 3 for the example are given in Figures 10.8a, 10.8b, 10.8c. These plots were inspected for further adjustments. Several quite small adjustments appeared but have not been made. Such small adjustments might be made until the analyst was completely satisfied. For the present purpose of an illustration these adjustments were not made. With this decision matrix $\boldsymbol{F_3}$ is considered to be the final normals matrix and matrix $\boldsymbol{G_3}$ is considered to be the final matrix of projections on the normals.

The general scheme of graphical transformations for rotation t is to inspect the factor plots from rotation (t $-$ 1) for adjustments to new base lines. The coordinates for these shift points are recorded as rows in matrix $\boldsymbol{S_t}$ and the new raw normals matrix $\widetilde{\boldsymbol{F_t}}$ is computed by;

$$\widetilde{\boldsymbol{F_t}} = \boldsymbol{S_t F_{(t-1)}} \ . \tag{10.14}$$

where $\boldsymbol{F_{(t-1)}}$ is the preceding rotation $\boldsymbol{F}$ matrix. Rows of $\widetilde{\boldsymbol{F_t}}$ are normalized to rows of the normals matrix $\boldsymbol{F_t}$. Projections on the new normals are in matrix $\boldsymbol{G_t}$.

$$\boldsymbol{G_t = A F_t'} \ . \tag{10.15}$$

When no more adjustments are to be made the last normals matrix becomes the final normals matrix.

Given a final normals matrix $\boldsymbol{F}$, the complete body of transformed matrices are to be computed. Table 10.9 gives the final matrices for the example. Matrix $\boldsymbol{F}$ was given from the final graphical transformation. The other matrices were computed by the following formulas.

$$\boldsymbol{D} = [Diag(\boldsymbol{F F'})^{-1}]^{-\frac{1}{2}} \tag{10.5}$$

$$\boldsymbol{T} = \boldsymbol{D}(\boldsymbol{F'})^{-1} \tag{10.3}$$

$$\boldsymbol{T} = \boldsymbol{D}(\boldsymbol{F F'})^{-1}\boldsymbol{F} \tag{10.16}$$

$$\boldsymbol{R_{bb}} = \boldsymbol{T T'} \tag{7.1}$$

Table 10.8

Graphical Transformation 3
Nine Mental tests Example

Shift Matrix S_3

	1	2	3
1	1.00	.12	.00
2	.00	1.00	.50
3	.12	.00	1.00

Projections on Normals Matrix G_3

	1	2	3
1	.51	.06	-.09
2	.51	-.06	.07
3	.39	.32	-.01
4	-.01	.71	-.06
5	-.03	.78	.00
6	.00	.70	-.01
7	.13	.01	.39
8	-.10	.38	.42
9	.01	.25	41

Raw Normals Matrix $\widetilde{F}_3$

	1	2	3	L^2	L
1	.251	.485	.872	1.0586	1.0289
2	.559	-.603	.098	.6858	.8281
3	.213	.192	-.878	.8528	.9235

Normal Matrix F_3

	1	2	3
1	.244	.472	.847
2	.675	-.728	.119
3	.231	.208	-.950

Cosines of Angles between Normals F_3F_3'

	1	2	3
1	1.00	-.08	-.65
2	-.08	1.00	-.11
3	-.65	-.11	1.00

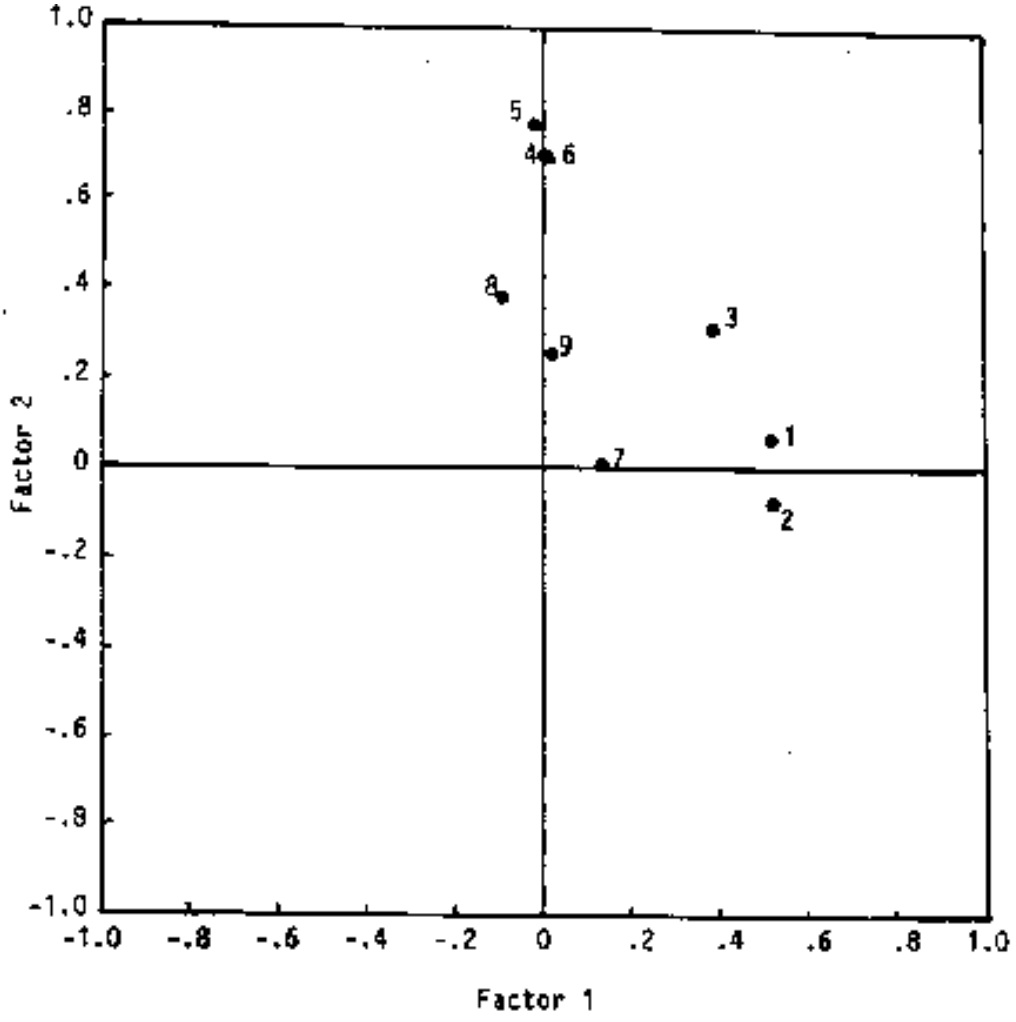

Figure 10.8a Nine Mental Tests: Rotation 3, Factor Plots for Factors 1 and 2.

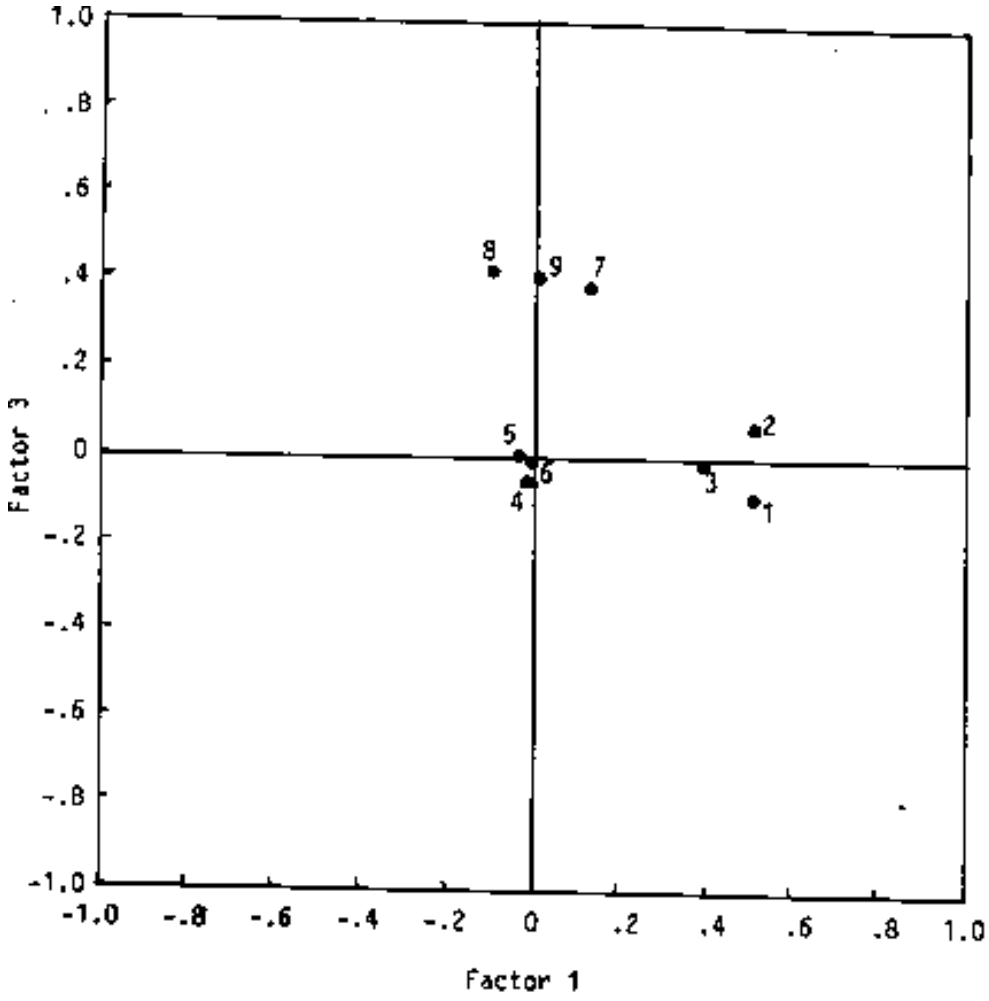

Figure 10.8b Nine Mental Tests: Rotation 3, Factor Plots for Factors 1 and 3.

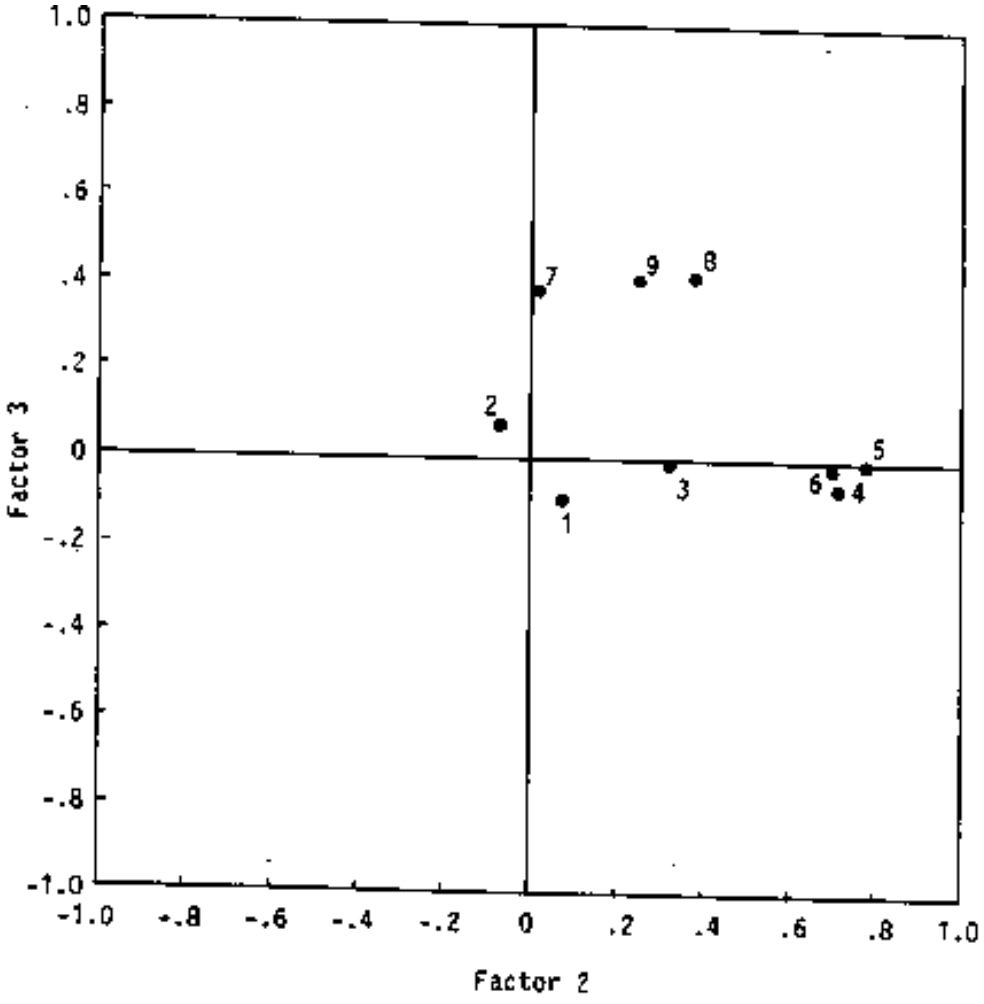

Figure 10.8c Nine Mental Tests: Rotation 3, Factor Plots for Factors 2 and 3.

Table 10.9

Graphical Transformation Matrices for Nine Mental Tests Example

Normals Matrix F

	1	2	3
1	.244	.472	.847
2	.675	-.728	.119
3	.231	.208	-.950

Traits Matrix T

	1	2	3
1	.671	.673	.310
2	.822	-.564	.077
3	.675	.545	-.498

Factor Correlation Matrix R_{bb}

	1	2	3
1	1.000	.196	.655
2	.196	1.000	.210
3	.665	.210	1.000

Cosines of Angles Between Normals and Trait Vectors

Matrix D

	1	2	3
1	.745	.000	.000
2	.000	.975	.000
3	.000	.000	.742

$$\boldsymbol{R_{bb} = D(FF')^{-1}D} \tag{10.6}$$

The final factor matrices for the example are given in Table 10.10. Formulas for these matrices are given below.

$$\boldsymbol{G = AF'} \tag{7.9}$$

$$\boldsymbol{B = AT^{-1}} \tag{7.4}$$

$$\boldsymbol{B = GD^{-1}} \tag{10.9}$$

$$\boldsymbol{Q = AT'} \tag{7.5}$$

$$\boldsymbol{Q = BR_{bb}} \tag{10.10}$$

Alternative formulas are given which facilitate the computations. Discussion of relations among these matrices occurs earlier in this chapter in the section for transformations for two factor studies.

10.3. Least Squares Hyperplane Fitting

While least squares hyperplane fitting involves a mathematical solution it is considered here as an adjunct to graphical transformations of factors. In order to avoid a number of small transformations of factors as well as to provide more precise definitions of the factor hyperplanes least squares solutions are obtained for projections on the normals for attribute points selected to be "in" the hyperplanes. Points "in" a hyperplane are those which have trivial projections on the normal to that hyperplane. A possible result from graphical transformations is to be able to specify which points are to be considered in each hyperplane. Then a final solution for each hyperplane can be obtained by least squares hyperplane fitting. Such a system was proposed by Tucker (1940). This method has wider applications which will not be considered here.

Before continuing with least squares hyperplane fitting a general weighted least squares proposition is required, Let $\boldsymbol{A}$, $n \times r$, be a given factor matrix on orthogonal axes. Let $\underline{\mathbf{f}}$ be a row vector with r entries which are to be determined. With $\underline{\mathbf{f}}$ as a normal vector the projections on this normal are in column vector $\mathbf{g}$ which has n entries. Vector $\mathbf{g}$ is defined by:

$$\boldsymbol{A\underline{f}' = g} \tag{10.17}$$

or, in expanded notation for entries g_i of $\mathbf{g}$:

Table 10.10

Graphically Transformed Factor Matrices for Nine Mental Tests Example

Matrix G of Projections on Normals

	1	2	3
1	.510	.055	-.094
2	.505	-.057	.069
3	.385	.318	-.006
4	-.009	.710	-.056
5	-.030	.781	-.002
6	.002	.704	-.006
7	.128	.008	.394
8	-.097	.378	.417
9	.008	.252	.407

Factor Weight Matrix B

	1	2	3
1	.685	.056	-.127
2	.678	-.058	.093
3	.518	.326	-.009
4	-.012	.729	-.076
5	-.041	.801	-.002
6	.002	.722	-.008
7	.172	.008	.530
8	-.130	.387	.561
9	.010	.259	.549

Matrix Q of Covariances of Modeled Attributes with Traits

	1	2	3
1	.611	.164	.340
2	.729	.094	.532
3	.576	.426	.404
4	.081	.710	.069
5	.115	.793	.139
6	.139	.721	.145
7	.526	.153	.646
8	.319	.479	.556
9	.426	.376	.610

$$g_i = \sum_{j=1}^{r} a_{ij} f_j \tag{10.18}$$

where a_{ij} is the ij entry in $\boldsymbol{A}$ and f_j is the j'th entry of $\underline{\mathbf{f}}$. Vector $\underline{\mathbf{f}}$ is restricted being a unit vector so that:

$$\sum_{j=1}^{r} f_j^2 = 1 \tag{10.19}$$

A non-negative weight, w_i, is given for each projection g_i and sum of weighted squares, designated θ, is given by:

$$\theta = \sum_{i=1}^{n} w_i g_i^2 \ . \tag{10.20}$$

This coefficient is to be minimized. However, the restriction of equation (10.19) is to be observed. To accomplish this restricted minimum, a revised criterion is written which involves an undetermined LaGrange multiplier, β:

$$\theta = \sum_{i=1}^{n} w_i g_i^2 - \beta \sum_{j=1}^{r} f_j^2 \tag{10.21}$$

To minimize θ the partial derivative is taken with respect to f_j and set equal to zero.

$$\frac{\partial \theta}{\partial f_j} = 2 \sum_{i=1}^{n} w_i g_i \frac{\partial g_i}{\partial f_j} - 2\beta f_j = 0 \ . \tag{10.22}$$

With substitution from equation (10.18) for g_i and the partial derivative of g_i with respect to f_j equation (10.22) becomes:

$$2 \sum_{j=1}^{n} w_i \sum_{k=1}^{r} a_{ik} f_k a_{ij} - 2\beta f_j = 0$$

or

$$\sum_{k=1}^{r} f_k \sum_{i=1}^{n} w_i a_{ij} a_{ik} = \beta f_j \ . \tag{10.23}$$

Define matrix $\boldsymbol{P}$ with entries p_{jk} by:

$$p_{jk} = \sum_{i=1}^{n} w_i a_{ij} a_{ik} \ . \tag{10.24}$$

Equation (10.23) can be written as

$$\sum_{k=1}^{r} \mathrm{f}_k p_{jk} = \beta \mathrm{f}_j$$

or, in matrix form:

$$\boldsymbol{P}\underline{\mathbf{f}}' = \beta\underline{\mathbf{f}}' \ . \tag{10.25}$$

Equation (10.25) may be written as:

$$(\boldsymbol{P} - \beta\boldsymbol{I})\underline{\mathbf{f}}' = \mathbf{0} \tag{10.26}$$

which is in the form of an eigen problem with eigenvalues β and corresponding eigenvectors $\underline{\mathbf{f}}'$. The remainder of the proposition relates β to coefficient θ .

Premultiplication of (10.24) by $\underline{\mathbf{f}}$ yields:

$$\underline{\mathbf{f}}\boldsymbol{P}\underline{\mathbf{f}}' = \beta\underline{\mathbf{f}}\underline{\mathbf{f}}'$$

which with the restriction of (10.19) yields:

$$\beta = \underline{\mathbf{f}}\boldsymbol{P}\underline{\mathbf{f}}' \tag{10.27}$$

From (10.20) and equation (10.18) coefficient θ may be written as:

$$\theta = \sum_{i=1}^{n} w_i \sum_{j=1}^{r} a_{ij}\mathrm{f}_j \sum_{k=1}^{r} a_{ik}\mathrm{f}_k = \sum_{j=1}^{r} \mathrm{f_j} \sum_{\mathrm{k=1}}^{r} \mathrm{f_k} \sum_{\mathrm{i=1}}^{\mathrm{n}} a_{\mathrm{ij}} w_i a_{ik}$$

which with equation (10.24) yields:

$$\theta = \sum_{j=1}^{r} \mathrm{f}_j \sum_{k=1}^{r} \mathrm{f}_k p_{jk}$$

or, in matrix form:

$$\theta = \underline{\mathbf{f}}\boldsymbol{P}\underline{\mathbf{f}}' \ .$$

Comparison with equation (10.27) yields:

$$\theta = \beta \ .$$

Consequently, minimum θ is obtained by using as f' the eigenvector corresponding to the least eigenvalue. Minimum θ is the value of the least eigenvalue.

The preceding development is stated in a more powerful form than will be used in least squares hyperplane fitting in that there is no restriction on the weights other than they be

nonnegative. In least squares hyperplane fitting the weights are restricted being either 0 or 1. In other contexts to be considered in the next chapter these weights will have other positive values.

Least squares hyperplane fitting is a single plane method in that it operates on each transformed factor individually. A separate solution is required for each of the factors in a study. The computations for a factor start from a selection vector of weights for the attributes these weights being restricted to values of 0 or 1 . A weight of 0 is assigned to each attribute judged to be "out of the hyperplane", that is, to have a significant projection on the derived normal. A weight of 1 is assigned to each attribute judged to be "in the hyperplane", that is, to have a trivial projection on the derived normal. Consider Table 10.11 which gives the results for the nine mental tests example. The first column of the selection vectors section gives the weights for the first factor. These selections are based on the final graphical transformation. Attributes 1, 2, 3, and 7 were judged to be "out" of the hyperplane so that they were given weights of 0. The remaining attributes were judged to be "in" the hyperplane and were given weights of 1. Note that these values are weights used in hyperplane fitting. An easy misconception is that this selection vector represents an anticipated pattern of loadings with l's being interpreted as high and 0's being interpreted as low. The reverse of this misconception is true. 0's represent attributes which are to have high valued projections and l's represent attributes which are to have low valued projections. 0 is for "out; 1 is for "in".

Column 1 of the remaining three sections of Table 10.11 give results for the first factor. There are three eigenvalues of the product matrix with a high first eigenvalue, a moderate second eigenvalue and a small third eigenvalue, this third eigenvalue being the sum of squared projections for the attributes "in" the plane. These results indicate a good fit of the plane for the selected attributes. Column 1 of the normals matrix contains the third eigenvector for the first factor product matrix. Projections on this normal are given in column 1 of the bottom matrix. Inspection of this first column of projections indicates a satisfactory solution. Attributes judged to be "in" the plane in the selection vector have projections in the interval bounded by $\pm.10$. Attributes judged to be "out" of the plane have projections greater than .10. Note, also, that all attributes "out" of the plane have positive projections. This satisfies the principle of a positive manifold.

Columns 2 and 3 of all sections of Table 10.11 give the results for factors 2 and 3 of the example. These results, also, appear to be satisfactory. These solutions provide a normals matrix $\boldsymbol{F}'$. Note that it is in transpose form. From this matrix the transformed factor solution can be computed as was done in the last part of the preceding section.

Two classes of problems may occur in least squares hyperplane fitting. First, not all attributes selected to be "in" a hyperplane end up with projections outside the limits for being in the hyperplane. This condition can be detected by the last eigenvalue not being small enough.

Table 10.11

Least Squares Hyperplane fitting
Nine Mental Tests Example

Selection Vectors
Entries used as Weights

	1	2	3
1	0	1	1
2	0	1	1
3	0	0	1
4	1	0	1
5	1	0	1
6	1	0	1
7	0	1	0
8	1	0	0
9	1	0	0

Eigenvalues of Product Matrix

	1	2	3
1	2.101	1.220	1.979
2	.446	.131	1.023
3	.006	.006	.015

Normals matrix F Transposed

	1	2	3
1	.295	.684	.255
2	.495	-.725	.187
3	.816	.084	-.949

Projections on Normals

	1	2	3
1	.531	.050	-.091
2	.537	-.057	.069
3	.415	.317	.005
4	.005	.710	-.034
5	-.011	.783	.023
6	.020	.706	.017
7	.169	.021	.397
8	-.057	.393	.431
9	.050	.267	.418

Also, the projections on the normal can be inspected for such points not within bounds. The second class of problems involves degenerate selection of attributes to be "in" the hyperplane. The selected points may be in a space of two or more dimensions. The hyperplane is not well defined by the selection of points. In this case two or more eigenvalues of the product matrix will be extremely small. Note that each eigenvalue is the sum of squared projections on the corresponding eigenvector. To illustrate this case a degenerate selection vector was written for the nine mental tests example. This selection vector is given in Table 10.12 along with associated computations. Note that the last two eigenvalues are very small which indicates that the selected attributes, 4, 5, and 6, form a cluster in the space of this example. All three eigenvectors and projections on these eigenvector are given. Note that the three selected attributes have projections in the range of $\pm.10$ on both of the last two eigenvectors. Figure 10.9 presents a plot of the projections on these last two eigenvectors. Points 4, 5, and 6, form a tight cluster at the origin. To eliminate such a degeneracy one or more points should be added to the selection of points to be "in" the hyperplane. In the example from Figure 10.9 a judgment might be made to add points 8 and 9. Such a move would result in the selection vector for factor 1 in Table 10.11. An alternative would be to add points 1, 2, and 3 which would result in the selection vector for factor 3 of the example as given in Table 10.11. Either of these selections would eliminate the degeneracy.

Selection vectors can be developed by other means than as a result from graphical transformations. In a tradition of confirmatory studies selection vectors may be hypothesized from previous studies and results studied toward confirmation of these hypotheses. Care must be taken in this class of use to guard against the problems described in the preceding paragraph. Another alternative is to develop selection vectors from other methods of factor transformation. One particular suggestion was by Kaiser and Cerny (1978) in an article entitled "Casey's method for fitting hyperplanes from an intermediate orthomax solution." In this method they started from an orthomax solution such as a VARIMAX transformation (to be described in the next chapter) and selected attributes to be "in" hyperplanes in terms of loadings being less in absolute values than computed critical values. They suggested a method for computing these critical values. Since this method has not stood up well in Monte Carlo studies, it will not be discussed further here.

Table 10.12

Degenerate Example of LSQHYP Fitting
Nine Mental tests Example

Selection Vectors

1	0
2	0
3	0
4	1
5	1
6	1
7	0
8	1
9	1

Eigenvalues of Product Matrix

1	1.657
2	.003
3	.000

Eigenvectors of Product Matrix

	1	2	3
1	.790	.561	.245
2	-.606	.659	.446
3	.089	-.501	.861

Projections on Eigenvectors

	1	2	3
1	.139	.333	.505
2	.058	.539	.494
3	.402	.353	.385
4	.714	-.045	.005
5	.793	.012	-.017
6	.720	.031	.013
7	.118	.638	.115
8	.456	.488	-.100
9	.347	.562	.001

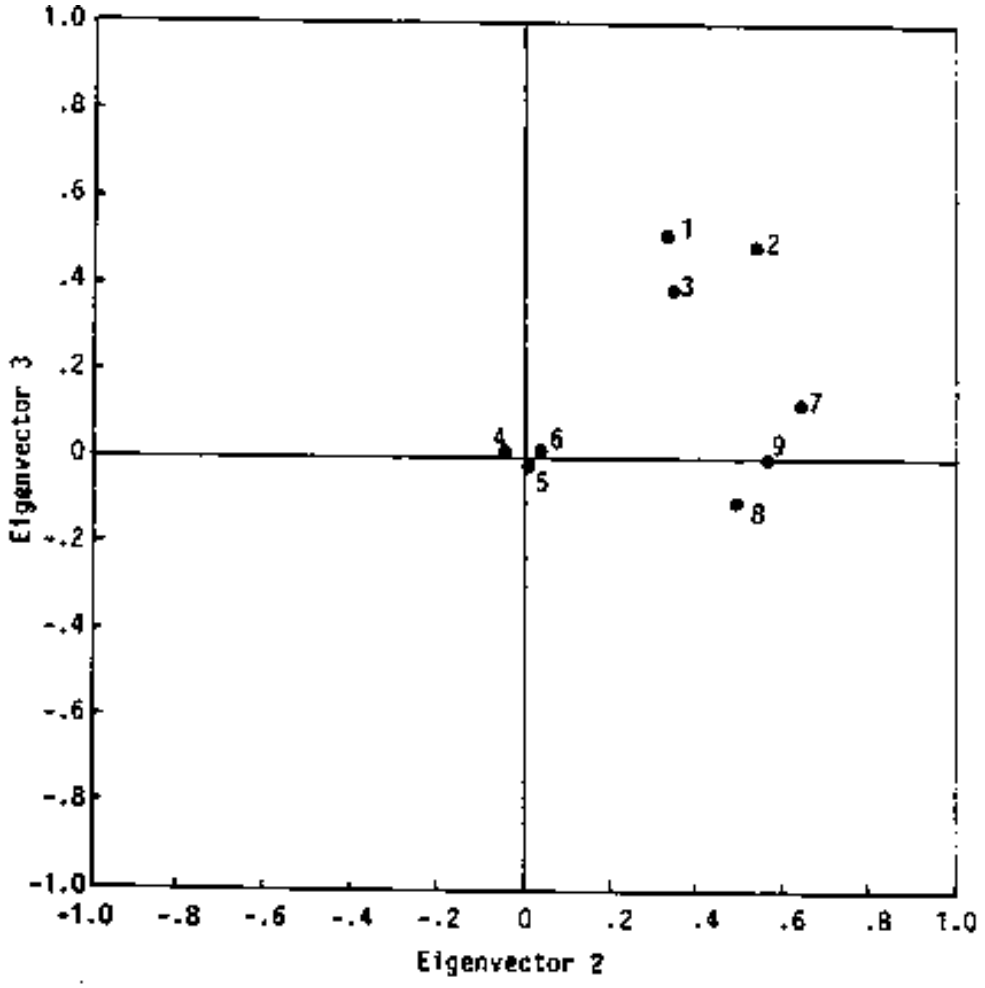

Figure 10.9 Plot of attribute point projections on eigenvectors 2 and 3 for degenerate LSQHYP fitting example, nine mental tests example.

Thurstone, L. L. Multiple Factor Analysis. Chicago: University of Chicago Press, 1947a.

Thurstone, L. L. & Thurstone, Thelma Gwinn. Factorial studies of intelligence. Psychometric Monograph No. 2. Chicago: University of Chicago Press, 1941.

Tucker, L. R. A rotational method based upon the mean principal axis of a sub-group of tests. Psychological Bulletin, 1940, 37, 578.

Kaiser, Henry F. & Cerny, Barbara A. Casey's method for fitting hyperplanes from an intermediate orthomax solution. Multivariate Behavioral Research, 1978, 13, 395-401.

CHAPTER 11
FACTOR TRANSFORMATIONS: ANALYTIC TRANSFORMATIONS

From
Exploratory Factor Analysis
Ledyard R Tucker
and
Robert C. MacCallum

CHAPTER 11

FACTOR TRANSFORMATIONS : ANALYTIC TRANSFORMATIONS

There has been a long felt need for automated factor transformation procedures both to eliminate the rather tedious graphical methods described in Chapter 10 and to provide objectivity in the results obtained. While the computations involved in the graphical transformation procedure can be programed readily for a computer and the factor graphs can be made mechanically, there remains the series of subjective judgments for each rotation. thus, the procedure had a stop and go characteristic which consumed considerable analysts time. Further, these judgments were sophisticated so that many individuals felt that they were not capable of making these judgments. A further and more serious complaint was the possible lack of objectivity in the results obtained. An important consideration is that the obtained results should not depend on the individual performaing the analysis. The same transformation for a particular factor matrix should be obtained by different analysts. That is: the transformation should not depend upon the analyst; it should be independent of the analyst so as to be repeatable at different laboratories. A variety of automated procedures will be discussed in this chapter.

Thurstone (1947) described several preliminary ideas for analytic approaches to simple structure. However, after some experimental trials these suggestions were not followed up. Several of the other developments refer to five criteria given by Thurstone in a section on "Uniqueness of simple structure in a given correlation matrix" (1947, pages 334-340). These criteria were to be applied after a transformation had been achieved to judge the uniqueness of the given solution. Further, attributes having complex factorial compositions were to be excluded from consideration. A preferable reference would be to Thurstone's suggsted equation for a simple structure (1947, pages 354-356). In Thurstone's notation, this equation is:

$$\phi = \sum_{j=1}^{n} \prod_{p=1}^{r} v_{jp}^2 \tag{11.1}$$

where the v's , for Thurstone, are the projections on the normals. However, in terms of subsequent developments, the v's could be interpreted either as projections on the normals (structure loadings) in our matrix G or as factor weights (pattern loadings) in our matrix B . In order for zero to be the least value considered, the coefficients v are squared. A product is obtained for each attribute of the squared coefficients of that attribute. Then, a sum is obtained of these products over all attributes. Thus, a minimum of the function is such as to maximize the number of near zero factor coefficients. This equation may be considered, with adaptation, as the basis for a number of proposed analytic procedures. Carroll (1953) proposed a criterion which

can be thouhgt of as an adaptation of equation (11.1). Instead of using the product of all coefficients for each attribute, Carroll used products of pairs of coefficients and obtained a sum of these products for each attribute. These sums of paired products were summed over all attributes. Let ϕ^* be Carroll's criterion, then:

$$\phi^* = \sum_{i=1}^{n} \sum_{j=1}^{r-1} \sum_{k=j+1}^{r} v_{ij}^2 v_{jk}^2 \quad . \tag{11.2}$$

Other interpreted properties of simple structure were used by some of the proposed techniques. The general concern was to maximize the zero or near zero factor coefficients in accorance with the concept of simple structure.

11.1 Orthogonal Transformations

Several transformation procedures restrict the transfomations to being orthogonal. For this class of transformations the matrix of normals is restricted to being orthonormal so that:

$$FF' = I \quad . \tag{11.3}$$

The matrix of projections on the normals is given by:

$$AF' = G \quad . \tag{7.9}$$

With the orthogonality restriction, the communality of each attribute i is given by:

$$h_i^2 = \sum_{j=1}^{r} a_{ij}^2 = \sum_{k=1}^{r} g_{ik}^2 \quad . \tag{11.4}$$

Carroll's criterion may be written in terms of the projections on the normals (structure loadings):

$$\phi^* = \sum_{i=1}^{n} \sum_{j=1}^{r-1} \sum_{k=j+1}^{r} g_{ij}^2 g_{ik}^2 \quad . \tag{11.5}$$

Several algebraic operations yield an interesting result which relates to other transformation criteria to be considered. The pairs of products of g's in equation (11.5) are as if the pairs below a diagonal of an r x r matrix. When all pairs of such a matrix, both sides of the diagonal as well as the diagonal, are considered equation (11.5) may be written as:

$$\phi^* = \frac{1}{2} \sum_{i=1}^{n} \left\{ \sum_{j=1}^{r} \sum_{k=1}^{r} g_{ij}^2 g_{jk}^2 - \sum_{j=1}^{r} g_{ij}^4 \right\} \tag{11.6}$$

with equation (11.4) this equation may be simplified to:

$$\phi^* = \frac{1}{2} \left\{ \sum_{i=1}^{n} (h_i^2)^2 - \sum_{i=1}^{n} \sum_{j=1}^{r} g_{ij}^4 \right\} \tag{11.7}$$

A constant C and function Q can be defined as below:

$$C = \sum_{i=1}^{n} (h_i^2)^2 \quad . \tag{11.8}$$

$$Q = \sum_{i=1}^{n} \sum_{j=1}^{r} g_{ij}^4 \quad . \tag{11.9}$$

Equation (11.7) becomes:

$$\phi^* = \tfrac{1}{2}C - \tfrac{1}{2}Q \quad . \tag{11.10}$$

Since the communalities are fixed by the initial factor matrix, C is a constant. Then, Carroll's criterion ϕ^* is a minimum when function Q is a maximum. This relation equates the solution for Carroll's criterion to the Quartimax criterion which is to be discussed next.

Shortly following Carroll's development, Saunders (1935) proposed a criterion for orthogonal transformations of factors based on the statistical concept of kurtosis of a distribution. First, Saunders constructed an extended factor matrix by appending vertically the given factor matrix with the negative of the given factor matrix. Thus, for any factor, original or transformed, the distribution of factor coefficients would be symmetric with mean zero. Each attribute would have two coefficients on any factor, a positive coefficient and a negative coefficient. Since such a distribution would be symmetric, the distribution would have zero skewness. Saunders observed for simple structure there would be a large number of attribute coefficients near zero with a smaller number of attribute coefficients trailing out to the tails of this distribution. Such a distribution should have a high value of kurtosis. Saunders proposed criterion S as:

$$S = \left\{ \sum_{i=1}^{n} \sum_{j=1}^{r} g_{ij}^4 \right\} \Big/ \left\{ \sum_{i=1}^{n} \sum_{j=1}^{r} g_{ij}^2 \right\}^2 \quad . \tag{11.11}$$

With the preceding equations Saunders' criterion becomes:

$$S = Q/C \tag{11.12}$$

which, since C is a constant, is maximized by maximizing function Q .

Neuhaus and Wrigley (1954) took a different statistical approach from that of Saunders. In order to eliminate the algebraic signs of the factor coefficients they considered the squared coefficients. They suggested that the transformation should result in a large number of small squared coefficients and a number of large squared coefficients. This result would occur with a large variance of the distribution of transformed, squared coefficients. Their function may be designated by the letter N and can be expressed as: .

$$N = \tfrac{1}{nr} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{r} (g_{ij}^2)^2 - \tfrac{1}{nr} \left[\sum_{i=1}^{n} \sum_{j=1}^{r} (g_{ij}^2) \right]^2 \right\} \tag{11.13}$$

With the preceding equations the Neuhaus and Wrigley function becomes:

$$N = \{(nr)Q - C\} \Big/ (nr)^2 \quad . \tag{11.14}$$

Since C is a constant, N is maximized by maximizing function Q .

Ferguson (1954) discussed factor transformations in terms of information theory and proposed that a function equivalent to Q would lead to a parsimonious result.

Since function Q involves the fourth power of the transfomed factor coefficients it became known as the Quartimax criterion. Computer programs were developed to maximize this function and experimental trials were run with a variety of original factor matrices. When the simple structure was strong and almost orthogonal the results were quite acceptable. However, when the configuration of attribute vectors was relatively complex and oblique the Quartimax solution was not acceptable. In thess types of cases the quartimax axes tended to be near the principal axes of the configuration. These conditions are illustrated by a small, artificial example. This example starts from a constructed factor weight matrix B given in Table 11.1 . There is a moderate simple structure with attributes 1 and 2 loading only on factor 1 while attributes 3 and 4 load only on factor 2 . Attributes 3 , 4 , and 5 are complex and span the space from factor 1 to factor 2 . Several examples are produced by choosing different correlations between the factors. This correlation is ϕ_{12} and is given values of .0 , .3 , .6 and .9 for the examples run. For the given matrix B and each of the chosen correlations between the factors, a transformation of the configuration was made to the principal axes of the configuration. This resulted in a matrix A for each of the examples. Table 11.2 presents the matrix A and the Quartimax transformed matrix for each of the chosen values of ϕ_{12} . For ϕ_{12} equal to .0 the Quartimax solution is quite near the input matrix B . When ϕ_{12} is .3 the Quartimax solution approaches the input matrix A . The trend increases markedly as the value of ϕ_{12} increases to .9 in which case the Quartimax matrix is almost identical with the input matrix A .

Following the experience with the Quartimax solution Kaiser (1958) proposed the Varimax orthogonal transformation. Kaiser pointed out that each of the Quartimax criteria worked to simplify the rows of the factor matrix. He proposed to concentrate on the columns of the transformed matrix with a criterion for each column j of the transformed matrix:

$$v_j = \left\{ n\sum_{i=1}^{n}(g_{ij}^2)^2 - \left[\sum_{i=1}^{n} g_{ij}^2\right]^2 \right\} \Big/ n^2 \qquad (11.15)$$

For a general function Kaiser summed the single column functions v_j over all columns of the transformed matrix to obtain function V .

$$V = \sum_{j=1}^{r}\left\{ n\sum_{i=1}^{n}(g_{ij}^2)^2 - \left[\sum_{i=1}^{n} g_{ij}^2\right]^2 \right\} \Big/ n^2$$

or

$$V = \left\{ n\sum_{j=1}^{r}\sum_{i=1}^{n} g_{ij}^4 - \sum_{j=1}^{r}\left[\sum_{i=1}^{n} g_{ij}^2\right]^2 \right\} \Big/ n^2 \quad . \qquad (11.16)$$

A further suggestion by Kaiser was that the rows of the input factor matrix should be normalized to unit length vectors before applying the Varimax solution. He noted that some transformed

Table 11.1

Input Matrix B for Artificial Example
for Orthogonal Transformations

	1	2
1	.7	.0
2	.5	.0
3	.0	.6
4	.0	.4
5	.5	.2
6	.4	.4
7	.3	.6

Table 11.2

Quartimax Transformation for Artificial Example

	$\phi_{12} = .0$		$\phi_{12} = .3$		$\phi_{12} = .6$		$\phi_{12} = .9$	
	Input Matrix A							
	1	2	1	2	1	2	1	2
1	.54	.45	.58	.39	.64	.30	.68	.15
2	.38	.32	.42	.28	.45	.21	.49	.11
3	.38	-.46	.47	-.38	.53	-.28	.58	-.14
4	.26	-.31	.31	-.25	.35	-.19	.39	-.09
5	.51	.17	.57	.15	.63	.12	.68	.06
6	.56	-.05	.64	-.03	.72	-.02	.78	-.01
7	.62	-.27	.72	-.21	.80	-.16	.88	-.08
	Quartimax Transformed Matrix							
	1	2	1	2	1	2	1	2
1	.70	.06	.41	.57	.63	.31	.68	.15
2	.50	.04	.29	.41	.45	.23	.49	.11
3	-.05	.60	.57	-.19	.54	-.27	.58	-.14
4	-.04	.40	.38	-.13	.36	-.18	.39	-.09
5	.48	.24	.48	.34	.63	.14	.68	.07
6	.36	.43	.61	.20	.72	.00	.78	.00
7	.25	.62	.75	.05	.81	-.13	.88	-.07

factors had few high loadings which resulted in reduced weighting for these factors. In order to counter this underweighting he suggested the normalization prodcedure in order to equalize the weights for the various transformed factors. To accomplish this step let H be a diagonal matrix containing the square roots of the communalities of the attributes. Let $\tilde{A}$ contain the normalized input factor matrix determined by:

$$\tilde{A} = H^{-1}A \quad . \tag{11.17}$$

Application of the Varimax procedure to $\tilde{A}$ results in a matrix $\tilde{G}$ for the normalized vectors. Return to the original lengthed vectors can be accomplished by:

$$G = H\tilde{G} \quad . \tag{11.18}$$

Transformation without the normalization procedure is termed the Raw Varimax solution. When the normalization procedure is used the solution is termed the Normal Varimax procedure.

Table 11.3 gives the results for the artificial example for the Raw Varimax and the Normal Varimax solutions. For the value of ϕ_{12} equal to .0 the Raw Varimax solution had a small rotation from the input B matrix while the Normal Varimax solution exactly reproduced the input B matrix. At ϕ_{12} equal to .3 the Raw varimax solution interchanged the two transformed factors. However, these factors are quite recognizable by low loadings for attributes 1 and 2 for factor 1 and for attributes 3 and 4 for factor 2. Note that the obliqueness has resulted in the input loadings of zero now being positive for the transformed factors. Common practice is to interpret small loadings as zero. For the Normal Varimax the loadings of these previously zero loadings are small positives. However, the transformed factors are quite identifiable to the input matrix B factors. By the time ϕ_{12} becomes .6 the previously small loadings have increased. For ϕ_{12} equals .9 the Raw Varimax solution has approached the principal factors of the input matrix A. The Normal Varimax solution balances the two input B factors with the previously zero loadings having become moderately positive. This balance of the factors for the Normal Varimax is shown by attribute 6, which started out in the input matrix B in Table 11.1 having equal loadings on the two factors. For the Normal Varimax solution for all values of ϕ_{12} this attribute 6 has equal loadings on the two transformed factors.

Figure 11.1 shows the locations of the transformed axes for $\phi_{12} = .6$ for all three orthogonal methods of tranformation discussed. The configuration of the seven attribute vector terminals are plotted. Note that the first Quartimax axis, Q_1, is almost identical with the first principal axis, A_1. This is the property of the Quartimax transformation discussed previously. The configuration of attribute points lies inside the two Normal Varimax axes. The Raw Varimax axes have drifted from the Normal Varimax axes toward the input principal axes. Axis R_1 has moved from axis N_1 toward input axis A_1. This figure illustrates some of the relative properties of the three methods of orthogonal transformations discussed in this section.

Table 11.3

Varimax Transformations for Artificial Example

	$\phi_{12} = .0$		$\phi_{12} = .3$		$\phi_{12} = .6$		$\phi_{12} = .9$	
Input Matrix A								
	1	2	1	2	1	2	1	2
1	.54	.45	.58	.39	.64	.30	.68	.15
2	.38	.32	.42	.28	.45	.21	.49	.11
3	.38	-.46	.47	-.38	.53	-.28	.58	-.14
4	.26	-.31	.31	-.25	.35	-.19	.39	-.09
5	.51	.17	.57	.15	.63	.12	.68	.06
6	.56	-.05	.64	-.03	.72	-.02	.78	-.01
7	.62	-.27	.72	-.21	.80	-.16	.88	-.08
Raw Varimax Transformed Matrix								
	1	2	1	2	1	2	1	2
1	.70	.06	.17	.68	.31	.63	.67	.20
2	.50	.04	.12	.49	.22	.45	.48	.14
3	-.05	.60	.60	.04	.59	.11	.59	-.10
4	-.04	.40	.40	.03	.39	.07	.39	-.07
5	.48	.24	.32	.50	.42	.48	.68	.11
6	.36	.43	.50	.41	.57	.43	.78	.04
7	.25	.62	.67	.33	.72	.38	.88	-.02
Normal Varimax Transformed Matrix								
	1	2	1	2	1	2	1	2
1	.70	.00	.69	.11	.66	.22	.59	.37
2	.50	.00	.49	.08	.47	.16	.42	.27
3	.00	.60	.09	.59	.19	.57	.32	.51
4	.00	.40	.06	.40	.13	.38	.21	.34
5	.50	.20	.52	.28	.54	.35	.53	.44
6	.40	.40	.46	.46	.51	.51	.55	.55
7	.30	.60	.39	.64	.48	.66	.57	.67

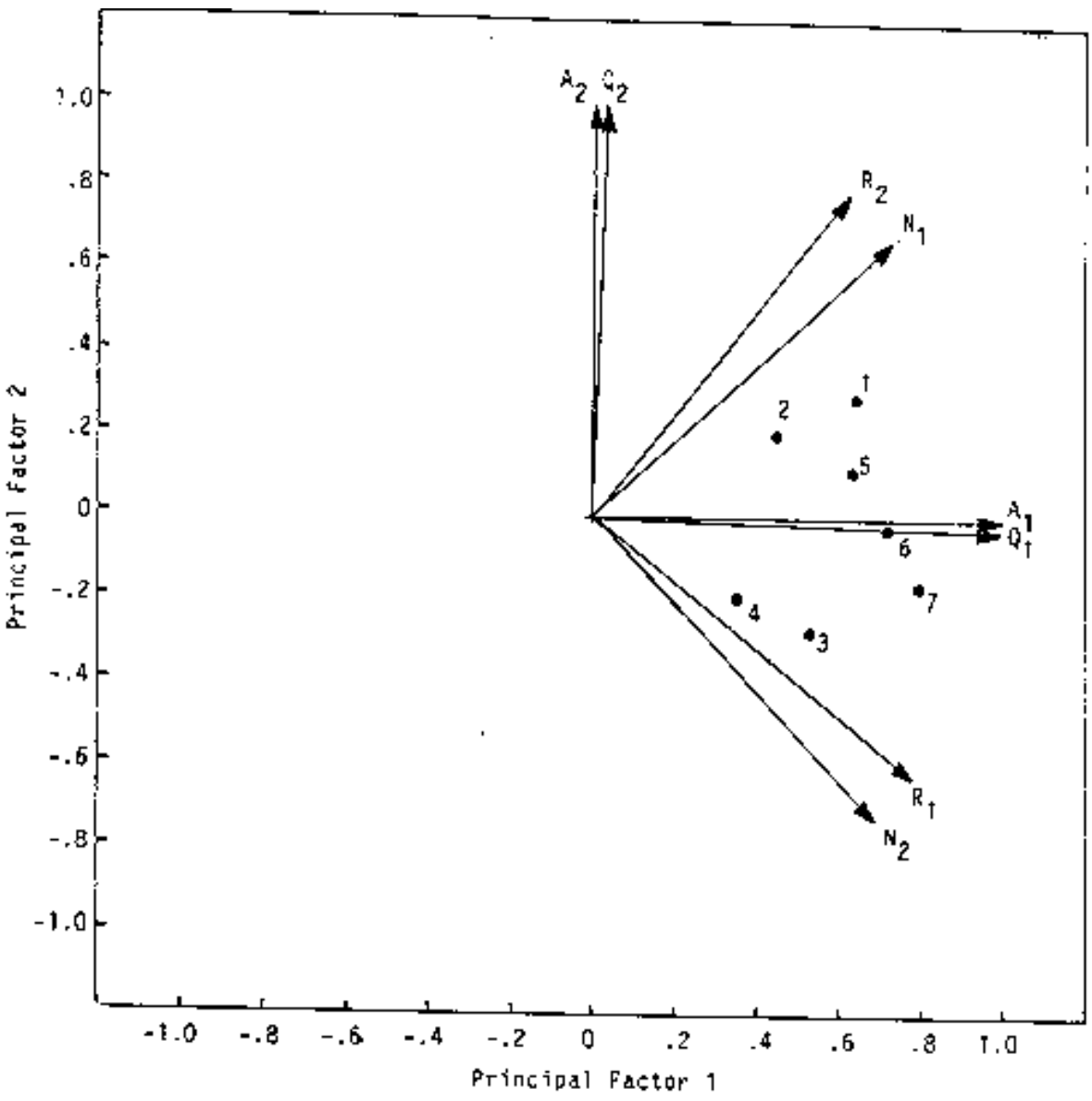

Figure 11.1 Factor Plot for Artificial Example, $\Phi_{12} = .6$, showing axes:

A_1 , A_2 for Input Matrix A;

N_1 , N_2 for Normal Varimax Transformation;

Q_1 , Q_2 for Quartimax Transformation;

R_1 , R_2 for Raw Varimax Transformation.

Results for transformation of the Nine Mental Tests Example by the analytic orthogonal transformation methods are given in Table 11.4. These results may be compared to the projections on the normals obtained by least squares hyperplane fitting in Table 10.11. Since the present results are restricted to orthogonal transformations they can not be as clean as the LSQHYP results that are not so restricted. However, the three transformed factors are quite recognizable. Even so, the interpretations can not be as clean and sharp as with the oblique transformation.

With the addition of a parameter γ the Quartimax and Varimax functions can be combined to a single function ξ which we may term Orthomax .

$$\xi = \left\{ n\sum_{j=1}^{r}\sum_{i=1}^{n} g_{ij}^4 - \gamma\sum_{j=1}^{r}\left[\sum_{i=1}^{n} g_{ij}^2\right]^2 \right\} \quad . \tag{11.19}$$

The denominator n^2 of the Varimax function has been dropped as unessential. for the Quartimax, γ equals 0 ; for the Varimax, γ equals 1 . There have been suggestions for other values of γ ; however, these have not lead to practical use. Computational procedures are based on function ξ .

An iterative type plan is followed in the computations which transform orthogonally each pair of factors in turn. These computations cycle through all pairs of factors solving for the minimum ξ for each pair. In each cycle of computations, a minimum is found for each pair of factors. These transformations are considered as conditional since, for any pair of factors, the transformations of other pairs will affect the solution for the given pair. Consequently, the cycles must be repeated until there are no significant transformations for all pairs. Consider the pair of factors k and m . The function may be written for this pair as:

$$\xi_{km} = n\left\{\sum_{i=1}^{n} g_{ik}^4 + \sum_{i=1}^{n} g_{im}^4\right\} - \gamma\left\{\left[\sum_{i=1}^{n} g_{ik}^2\right]^2 + \left[\sum_{i=1}^{n} g_{im}^2\right]^2\right\} \quad . \tag{11.20}$$

An orthogonal transformation of factors k and m is accomplished through an angle θ . The new factor coefficients are:

$$\tilde{g}_{ik} = g_{ik}\cos\theta + g_{im}\sin\theta \quad ; \tag{11.21a}$$

$$\tilde{g}_{im} = -g_{ik}\sin\theta + g_{im}\cos\theta. \tag{11.21b}$$

The transformed function is:

$$\tilde{\xi}_{km} = n\left\{\sum_{i=1}^{n} \tilde{g}_{ik}^4 + \sum_{i=1}^{n} \tilde{g}_{im}^4\right\} - \gamma\left\{\left[\sum_{i=1}^{n} \tilde{g}_{ik}^2\right]^2 + \left[\sum_{i=1}^{n} \tilde{g}_{im}^2\right]^2\right\} \quad . \tag{11.22}$$

After algebraic and trigonometric operations this function may be written as:

$$\tilde{\xi}_{km} = K + \tfrac{1}{4}C \bullet \cos 4\theta + \tfrac{1}{4}E \bullet \sin 4\theta \tag{11.23}$$

where K , C and E are temporary coefficients defined by:

Table 11.4

Orthogonal Analytic Transformations
Nine Mental Tests Example

Transformed Factor Matrices									
	Quartimax			Raw Varimax			Normal Varimax		
	1	2	3	1	2	3	1	2	3
1	.59	.09	-.15	.61	.08	.08	.61	.07	.12
2	.73	.00	.00	.69	-.03	.26	.67	-.05	.30
3	.55	.36	-.05	.53	.34	.19	.53	.32	.23
4	.03	.71	-.03	.04	.71	.06	.05	.71	.09
5	.06	.79	.02	.05	.78	.13	.06	.78	.16
6	.09	.72	.01	.08	.71	.13	.09	.70	.15
7	.56	.07	.34	.40	.01	.53	.37	-.01	.55
8	.33	.43	.40	.16	.37	.54	.14	.35	.56
9	.44	.31	.38	.28	.25	.54	.25	.23	.57

$$K = n\left\{ \tfrac{1}{2}\sum_{i=1}^{n}[g_{ik}^2 + g_{im}^2]^2 + \tfrac{1}{4}\sum_{i=1}^{n}[g_{ik}^2 - g_{im}^2]^2 + \sum_{i=1}^{n}[g_{ik}g_{im}]^2 \right\}$$
$$- \gamma\left\{ \tfrac{1}{2}\left[\sum_{i=1}^{n}(g_{ik}^2 + g_{im}^2)\right]^2 + \tfrac{1}{4}\left[\sum_{i=1}^{n}(g_{ik}^2 - g_{im}^2)\right]^2 + \left[\sum_{i=1}^{n}g_{ik}g_{im}\right]^2 \right\} ; \qquad (11.24)$$

$$C = n\left\{ \sum_{i=1}^{n}[g_{ik}^2 + g_{im}^2]^2 - 4\sum_{i=1}^{n}g_{ik}^2 g_{im}^2 \right\} - \gamma\left\{ \left[\sum_{i=1}^{n}(g_{ik}^2 - g_{im}^2)\right]^2 - 4\left[\sum_{i=1}^{n}g_{ik}g_{im}\right]^2 \right\}; \qquad (11.25)$$

$$E = 4\left\{ n\sum_{i=1}^{n}(g_{ik}^2 - g_{im}^2)(g_{ik}g_{im}) - \gamma\sum_{i=1}^{n}(g_{ik}^2 - g_{im}^2)\sum_{i=1}^{n}g_{ik}g_{im} \right\} \qquad (11.26)$$

The desired maximum of $\tilde{\xi}_{km}$ is obtained by setting the first derivative of $\tilde{\xi}_{km}$ equal to zero.

$$\frac{d\tilde{\xi}_{km}}{d\theta} = -C \bullet \sin 4\theta + E \bullet \cos 4\theta = 0 \quad . \qquad (11.27)$$

However, equation (11.27) only gives the condition for an optimum solution which may be either a minimum or a maximum. The second derivative must be investigated to determine the maximum solution.

$$\frac{d^2\tilde{\xi}_{km}}{d\theta^2} = -4C \bullet \cos 4\theta - 4E \bullet \sin 4\theta = -4[C \bullet \cos 4\theta + E \bullet \sin 4\theta] \quad .$$

For a maximum this second derivative must be negative which results in the following inequality.

$$[c \bullet \cos 4\theta + E \bullet \sin 4\theta] > 0 \quad . \qquad (11.28)$$

The desired solution must satisfy equations (11.27) and (11.28).

There are four conditions to be considered in terms of whether or not each of the coefficients C and E equals zero.

1) If $C = 0$ and $E = 0$,
   The angle of rotation is indeterminate and may be set at 0. Then:
   $\cos\theta = 1$; $\sin\theta = 0$.
2) If $C = 0$ and $E \neq 0$,
   from (11.27), $\cos 4\theta = 0$ so that $\sin 4\theta = \pm 1$; by (11.28), when $E > 0$, $\sin 4\theta = +1$ which results in $4\theta = \pi/2$ and $\theta = \pi/8$ so that:
   $\cos\theta = .923880$; $\sin\theta = .382683$.
   When $E < 0$, $\sin 4\theta = -1$ which results in $4\theta = -\pi/2$ and $\theta = -\pi/8$ so that:
   $\cos\theta = .923880$; $\sin\theta = -.382683$.
3) If $C \neq 0$ and $E = 0$,
   from (11.27), $\sin 4\theta = 0$ so that $\cos 4\theta = \pm 1$; by (11.28), when $c > 0$, $\cos 4\theta = +1$ which results in $4\theta = 0$ and $\theta = 0$ so that:
   $\cos\theta = 1$; $\sin\theta = 0$.

When $C < 0$, $\cos 4\theta = -1$ which results in $4\theta = \pi$ and $\theta = \pi/4$ so that:

$\cos\theta = .707107$; $\sin\theta = .707107$.

4) If $C \neq 0$ and $E \neq 0$,

algebraic operations on equation (11.27) yields:

$E \bullet \sin 4\theta = C \bullet \sin 4\theta$

so that:

$$\frac{E}{C} = \frac{\sin 4\theta}{\cos 4\theta} = \tan 4\theta \quad . \tag{11.30}$$

The inequality of equation (11.28) needs to be investigated in selection of a maximum solution. From (11.27)

$$\sin 4\theta = \frac{E}{C}\cos 4\theta$$

which may be substituted in (11.28) to obtain:

$$\frac{E^2}{C}\cos 4\theta + C \bullet \cos 4\theta > 0$$

or

$$\frac{(C^2 + E^2)}{C}\cos 4\theta > 0 \quad . \tag{11.31}$$

Since $(C^2 + E^2)$ must be positive, the $\cos 4\theta$ must have the same algebraic sign as coefficient C. This rule helps pick out the maximum solution. Let T be a temporary coefficient defined by the positive value of:

$$T = \sqrt{1 + \tan^2 4\theta} \quad .$$

$$\text{If } C > 0, \ \cos 4\theta = \frac{1}{T} \quad ; \tag{11.32a}$$

$$\text{if } C < 0, \ \cos 4\theta = \frac{-1}{T} \quad . \tag{11.32b}$$

The remainder of the solution follows. Note that positive square roots are appropriate in the following steps.

$$\sin 4\theta = \cos 4\theta \bullet \tan 4\theta \quad ; \tag{11.33}$$

$$\cos 2\theta = \sqrt{\frac{1 + \cos 4\theta}{2}} \quad ; \tag{11.34}$$

$$\sin 2\theta = \frac{\sin 4\theta}{2 \bullet \cos 2\theta} \quad ; \tag{11.35}$$

$$\cos\theta = \sqrt{\frac{1 + \cos 2\theta}{2}} \quad ; \tag{11.36}$$

$$\sin\theta = \frac{\sin 2\theta}{2 \bullet \cos\theta} \quad . \tag{11.37}$$

The desired transformation is given in equations (11.36) and (11.37). The foregoing steps are readily programmed into an efficient computer subroutine.

11.2 Oblique Algebraic Transformations, OBLIMIN

Oblique algebraic transformations started with Carroll's (1953) proposal which was expressed as equations (11.2) and (11.5) without the orthogonality restriction. Dur to the relation with the Quartimax procedure this unrestricted propossal has become termed the Quartimin

procedure. Kaiser in his unpublished Ph.D. dissertation (University of California at Berkeley, 1956) suggested an oblique counterpart to his Varimax criterion which is termed the Covarimin criterion. In experimental trials by Carroll and Kaiser the Quartimin procedure yielded transformed factors that were overly oblique for the trial factor matrices they concidered. In contrast, the Covarimin procedure yielded transformed factors that were not adequately oblique. A compromize suggested by Carroll (1957) was to take a transformation that was half way between the Quartimin and the Covarimin transformations. This lead to equation (11.38) with the use of parameter γ in a manner similar to equation (11.19).

$$\xi = \sum_{k=1}^{r-1} \sum_{m=k+1}^{r} \left\{ n \sum_{i=1}^{n} g_{ik}^2 g_{im}^2 - \gamma \left[\sum_{i=1}^{n} g_{ik}^2 \right] \left[\sum_{i=1}^{n} g_{im}^2 \right] \right\} \quad . \tag{11.38}$$

The three values of γ considered are given below along with interpretations.

Quartimin: $\gamma = 0$ said to be most oblique;

Biquartimin; $\gamma = .5$ said to be less oblique;

Covarimin: $\gamma = 1.0$ said to be least oblique.

Solutions for these three oblique criteria for the Nine Mental Tests example are given in Tables 11.5 and 11.6. The computations follow an iterative scheme to be discussed later may be started from any selected initial transformation. In Table 11.5 an intial transformation was performed with a Normal Varimax solution. Note for $\gamma = 0.0$ the cosines of the angles of normals 1 and 2 with normal 3 are very negative; that is: very oblique. This obliqueness has reduced for $\gamma = 0.5$. However, for $\gamma = 1.0$ the cosine of the angless among the normals have turned positive so that the trend is not from most oblique to least oblique but from most negative cosines of angles among the normals to positive cosines among the angles. Table 11.6 illustrates a common problem experienced early with a variety of studies. Note that the cosine of the angle between transformed factor 1 and 2 has become 1.00 in the iteration. These two tranformed factors have become identical. This is a most unsatisfactory result. This problem occurs also for $\gamma = 1.0$. Another property of the solution is illustrated with these two tables. The obtained results depend upon the starting position. As a consequence of the problems illustrated here, this approach to oblique algebraic transformations is not considered for practical use.

An elemental step in the solution to minimize ξ is to transform normal k (row k of F) in the direction of normal m (row m of F) so that ξ is reduced as much as possible. A cycle of computations will involve taking each normal in turn as k and to transform it in the direction of each other normal in turn as normal m . These cycles are continued until there are no significant transformations for any normal. The oblimin criterion can be written for normal k as;

Table 11.5

Structure OBLIMIN Transformation
Nine Mental Tests Example
Start from Normal Varimax Solution

$\gamma = 0.0$ Quartimin				$\gamma = 0.5$ Biquartimin				$\gamma = 1.0$ Covarimin			
Projections on Normals											
	1	2	3		1	2	3		1	2	3
1	.46	.07	-.12	1	.60	.04	-.05	1	.61	.09	.28
2	.42	-.10	.03	2	.67	-.09	.12	2	.72	-.01	.46
3	.34	.24	-.03	3	.50	.30	.05	3	.55	.35	.38
4	.02	.59	-.04	4	-.01	.70	-.02	4	.04	.71	.12
5	-.01	.62	.02	5	-.01	.77	.05	5	.07	.79	.20
6	.02	.56	.01	6	.03	.69	.04	6	.09	.71	.20
7	.04	-.22	.37	7	.36	-.05	.44	7	.48	.05	.63
8	-.16	.07	.41	8	.11	.32	.46	8	.25	.41	.59
9	-.07	-.03	.39	9	.23	.20	.45	9	.36	.29	.62
Cosines of Angles Between Normals											
	1	2	3		1	2	3		1	2	3
1	1.00	.36	-.79	1	1.00	-.11	-.25	1	1.00	.01	.47
2	.36	1.00	-.59	2	-.11	1.00	-.16	2	.01	1.00	.13
3	-.79	-.59	1.00	3	-.25	-.16	1.00	3	.47	.13	1.00
8 Cycles $\xi = .24025$				14 Cycles $\xi = -1.10194$				14 Cycles $\xi = -4.15831$			

Table 11.6

Structure OBLIMIN Transformation
Nine Mental Tests Example
Start from Principal Fators

$\gamma = 0.0$ Quartimin				$\gamma = 0.5$ Biquartimin				$\gamma = 1.0$ Covarimin			
Projections on Normals											
	1	2	3		1	2	3		1	2	3
1	.33	.32	-.06	1	.34	.34	.05	1	.61	.09	.28
2	.23	.22	.14	2	.24	.25	.27	2	.72	-.01	.46
3	.21	.21	-.08	3	.23	.23	-.02	3	.55	.35	.38
4	.03	.04	-.31	4	.05	.04	-.39	4	.04	.71	.12
5	-.01	.00	-.29	5	.01	.00	-.37	5	.07	.79	.20
6	.01	.01	-.26	6	.03	.02	-.33	6	.09	.71	.20
7	-.15	-.16	.39	7	-.14	-.14	.43	7	.48	.05	.63
8	-.29	-.29	.25	8	-.27	-.27	.21	8	.25	.41	.59
9	-.23	-.23	.30	9	-.21	-.21	.29	9	.36	.29	.62
Cosines of Angles Between Normals											
	1	2	3		1	2	3		1	2	3
1	1.00	1.00	-.78	1	1.00	1.00	-.62	1	1.00	.01	.47
2	1.00	1.00	-.80	2	1.00	1.00	-.60	2	.01	1.00	.13
3	-.78	-.80	1.00	3	-.62	-.60	1.00	3	.47	.13	1.00
50 Cycles* $\xi = .50114$				50 Cycles* $\xi = .16493$				18 Cycles $\xi = -4.15831$			

* Convergence was not achieved in 50 cycles.

$$\xi_k = \sum_{\substack{j=1 \\ j\neq k}}^{r} \left\{ n\sum_{i=1}^{n} g_{ik}^2 g_{ij}^2 - \gamma \left[\sum_{i=1}^{n} g_{ik}^2\right] \left[\sum_{i=1}^{n} g_{ij}^2\right] \right\} \quad . \tag{11.39}$$

Let q_{ik} be a coefficient for each attribute i in terms of normal k :

$$q_{ik} = \sum_{\substack{j=1 \\ j\neq k}}^{r} g_{jk}^2 \quad . \tag{11.40}$$

Then equation (11.39) may be written as:

$$\xi_k = n\sum_{i=1}^{n} [g_{i\ ik}^2 q_{ik}] - \gamma \left\{\sum_{i=1}^{n} g_{ik}^2\right\} \left\{\sum_{i=1}^{n} q_{ik}\right\} \quad . \tag{11.41}$$

Let $\underline{f}_k$ be row k of matrix F (the row for normal k) and $\underline{f}_m$ be row m of matrix F (the row for normal m). As in graphical transformations, an unnormalized, transformed normal k may be noted as $\underline{f}_k^*$ which is obtained by:

$$\underline{f}_k^* = \underline{f}_k + z_{km}\underline{f}_m \tag{11.42}$$

where z_{km} is the raw shift coefficient. Normalization of $\underline{f}_k^*$ to the new, normalized normal, $\underline{\tilde{f}}_k$ yields:

$$\underline{\tilde{f}}_k = \{\underline{f}_k + z_{km}\underline{f}_m\} \Big/ \sqrt{1 + 2z_{km}p_{km} + z_{km}^2} \tag{11.43}$$

where z_{km} is the cosine of the angle between normals k and m . The projections on the tranformed normal are:

$$\tilde{g}_{ik} = \{g_{ik} + z_{km}g_{im}\} \Big/ \sqrt{1 + 2z_{km}p_{km} + z_{km}^2} \quad . \tag{11.44}$$

Coefficient z_{km} is the km cell of matrix P defined by:

$$P = FF' \quad . \tag{11.45}$$

The transformed criterion is $\tilde{\xi}_{km}$ which is obtained by substitution of $\tilde{g}_{ik}$ for the g_{ik} in equation (11.41):

$$\tilde{\xi}_{km} = n\sum_{i=1}^{n} [\tilde{g}_{ik}^2 q_{ik}] - \gamma \left\{\sum_{i=1}^{n} \tilde{g}_{ik}^2\right\} \left\{\sum_{i=1}^{n} q_{ik}\right\} \quad . \tag{11.46}$$

To obtain a minimum of $\tilde{\xi}_{km}$ the derivative with respect to z_{km} is set equal to zero; this yields:

$$\frac{d\tilde{\xi}_{km}}{dz_{km}} = \frac{a + bz_{km} + cz_{km}^2}{[1 + 2p_{km}z_{km} + z_{km}^2]^2} = 0 \tag{11.47}$$

where:

$$a = n\sum_{i=1}^{n} [g_{im}g_{ik} - g_{ik}^2 p_{km}]q_{ij} - \gamma \left\{\sum_{i=1}^{n} [g_{im}g_{ik} - g_{ik}^2 p_{km}]\right\} \left\{\sum_{i=1}^{n} q_{ik}\right\} \quad ; \tag{11.48}$$

$$b = n\sum_{i=1}^{n} [g_{im}^2 - g_{ik}^2]q_{ik} - \gamma \left\{\sum_{i=1}^{n} [g_{im}^2 - g_{ik}^2]\right\} \left\{\sum_{i=1}^{n} q_{ik}\right\} \quad ; \tag{11.49}$$

$$c = n\sum_{i=1}^{n} [g_{im}^2 p_{km} - g_{ik}g_{im}]q_{ik} - \gamma \left\{\sum_{i=1}^{n} [g_{im}^2 p_{km} - g_{ik}g_{im}]\right\} \left\{\sum_{i=1}^{n} q_{ik}\right\} \quad . \tag{11.50}$$

Since the denominator of (11.47) is necessarily positive (being a square) and approaches infinity only when z_{km} approaches infinity (which would make (11.47) approach zero), attention is given to the numerator equalling zero for the derivative to equal zero. Let y designate the numerator:

$$y = a + bz_{km} + cz_{km}^2 \quad . \tag{11.51}$$

If $c = 0$,

$$z_{km} = -a/b \quad . \tag{11.52}$$

If $c \neq 0$, the roots of equation (11.51 are given by the quadratic equation:

$$z_{km} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2c} \quad . \tag{11.53}$$

There is a problem as to which sign to use. Investigation of the nature of the function in equation (11.46) indicates that the positive sign is always correct so that:

$$z_{km} = \frac{-b + \sqrt{b^2 - 4ac}}{2c} \quad . \tag{11.54}$$

The preceding material provides a basis for a computer program to obtain a solution for the Oblimin function. However, the obtained result may depend on the starting position as illustrated in Tables 11.5 and 11.6. There is no guarantee that there is a single minimum.

11.3 Oblique Transformations, DIRECT OBLIMIN

The DIRECT OBLIMIN transformations were initiated by Jennrich and Sampson (1966) in their article on rotation to simple structure in which they adapted Carrol's criterion to involving the squared factor weights (pattern loadings) in our matrix B. Let the Jennrich and Sampson criterion be designated J and be:

$$J = \sum_{j=1}^{r-1} \sum_{k=j+1}^{r} \sum_{i=1}^{n} b_{ij}^2 b_{ik}^2 \quad . \tag{11.55}$$

This criterion is to be minimized. As described previously, the Oblimin procedures yielded unsatisfactory results with some data matrices in that the transformation yielded singular normals matrices and factor matrices. A dimension had been lost. Jennrich and Sampson overcame this problem in that the b's approach infinity any time the factor correlation matrix approaches singularity. Jennrich and Sampson preferred to apply a transfomration, T , directly to the factor scores which results in applying the inverse of this transformation to the factor weight matrix; thus:

$$B = AT^{-1} \quad . \tag{7.4}$$

Other relevant equations from Chapter 7 folow.

$$R_{bb} = TT' \quad . \tag{7.1}$$

$$D = \{\text{Diag } R_{bb}^{-1}\}^{-1/2} \quad . \tag{7.8}$$

Then:

$$B = GD^{-1} \quad . \tag{11.56}$$

When the factor correlation matrix becomes singular due to highly oblique factors, the diagonal entries in the inverse become very large as do the diagonal entries in D^{-1}. Then, as indicated previously, the entries in the factor weight matrix B also become very large. This works against the minimization of the direct oblimin function

The Direct Oblimin function has been expanded to involve a parameter γ similar to the expansion of the Quartimin criterion to the Oblimin criterion given in equation (11.38). The expanded criterion may be designated ψ and is given below:

$$\psi = \sum_{j=1}^{r-1} \sum_{k=j+1}^{r} \left\{ n \sum_{i=1}^{n} b_{ij}^2 b_{ik}^2 - \gamma \left[\sum_{i=1}^{n} b_{ij}^2 \right] \left[\sum_{i=1}^{n} b_{ik}^2 \right] \right\} \quad . \tag{11.57}$$

Table 11.7 gives results of the Direct Oblimin solution for the Nine Mental Tests Example. Values of γ range from -1.0 to 0.5. In contrast to the Oblimin situation in which positive values of γ gave the better results, preferable results are obtained for the Direct Oblimin criterion with negative values of γ. In fact, the computations broke down with a value of γ of $+1.0$. With γ equal to 0.5, there were very few near zero factor loadings which was accompanied by high factor correlations. Jennrich (1979), in a study on the admissible values of γ in direct oblimin rotation, indicated that there might be trouble in trying to use a positive value of γ. A minimum value of ψ might not exist. He recommended that positive values of γ be avoided. The simple loadings of Jennrich and Sampson have γ equal to 0. Note that there are only small changes for this example in the near zero loadings as γ progresses from 0 to -1.0. However, the factor correlations decrease in this range. A phenomena not shown in this example is that when very negative values of γ are used the transformed factors approach the principal factors. A recomendation is that several small negative values of γ be tried and a selection made for the best results.

Comments on analysis procedures follow. In equation (11.57) ψ is the sum, as it were, of entries on one side of a square matirx. A useful alternative equation takes one half of the sum of all cells of such a matrix less the diagonal cells. Thus, an alternative equation for ψ is:

$$\psi = \tfrac{1}{2} \left\{ n \left[\sum_{i=1}^{n} \left\{ \sum_{k=1}^{r} b_{ik}^2 \right\}^2 - \sum_{i=1}^{n} \sum_{k=1}^{r} b_{ik}^4 \right] - \gamma \left[\left\{ \sum_{i=1}^{n} \sum_{k=1}^{r} b_{ik}^2 \right\}^2 - \sum_{k=1}^{r} \left\{ \sum_{i=1}^{n} b_{ik}^2 \right\}^2 \right] \right\} \tag{11.58}$$

An elemental step in solution to minimize ψ is to transform trait vestor k (row k of T) in the direction of trait vector m (row m of T) so as to reduce ψ as much as possible. There will be a number of cycles with each trait vector in turn being k which is paired with each other vector m in turn. The series of cycles is continued until only trivial transformations are performed

Table 11.7

Direct OBLIMIN Transformations
Nine Mental Tests Example

γ = -1.0			γ = -.5			γ = 0.0			γ = 0.5		
Factor Weight Matrices B											
1	2	3		1	2	3	1	2	3			
52	.06	-.02	1	.63	.06	-.03	1	.65	.06	-.08	1	1.08	-.08	-.50
55	-.10	.18	2	.65	-.10	.17	2	.65	-.11	.15	2	1.02	-.44	-.04
50	.30	.09	3	.50	.30	.08	3	.51	.29	.05	3	.73	.24	-.28
01	.72	-.01	4	.01	.72	-.01	4	.02	.72	-.03	4	-.15	1.05	-.33
01	.77	.06	5	-.01	.77	.06	5	-.01	.77	.05	5	-.24	1.09	-.20
03	.70	.06	6	.03	.70	.05	6	.03	.70	.04	6	-.15	.97	-.20
25	-.10	.53	7	.24	-.12	.54	7	.19	-.17	.57	7	.08	-.56	.98
01	.27	.55	8	-.02	.25	.56	8	-.08	.20	.61	8	-.48	.04	1.04
1	.14	.55	9	.10	.13	.56	9	.05	.08	.60	9	-.23	-.18	1.00
Factor Correlation Matrices R_{bb}														
1	2	3		1	2	3	1	2	3		1	2	3	
00	.11	.45	1	1.00	.12	.48	1	1.00	.13	.59	1	1.00	.73	.91
1	1.00	.29	2	.12	1.00	.32	2	.13	1.00	.40	2	.73	1.00	.84
45	.29	1.00	3	.48	.32	1.00	3	.59	.40	1.00	3	.91	.84	1.00
ψ = 5.667			ψ = 3.278			ψ = 0.800			ψ = -4.534					

within a cycle.

Let $\underline{t}_k$ and $\underline{t}_m$ be trait vectors k and m. Let $\underline{\tilde{t}}_k$ be the new trait vector k which is defined by:

$$\underline{\tilde{t}}_k = s_{kk}\underline{t}_k + s_{km}\underline{t}_m \tag{11.59}$$

where s_{kk} and s_{km} are the transformation coefficients. A restriction that $\underline{\tilde{t}}_k$ be a unit vector leads to the following restriction on the trasformation coefficients.

$$s_{kk}^2 + 2s_{kk}s_{km}\phi_{km} + s_{km}^2 = 1 \tag{11.60}$$

where ϕ_{km} is the correlation between factors k and m, that is, the km cell of matrix R_{bb} given in equation (7.1). Let $\tilde{T}$ be the new traits matrix given by:

$$\tilde{T} = ST \tag{11.61}$$

where the transformation matrix S has s_{kk} in the kk cell and s_{km} in the km cell with all other off-diagonal cells being zero and all other diagonal cells being unity. Let $\tilde{B}$ be the new matrix of factor weights which may be obtained by an equation similar to equation (7.4).

$$\tilde{B} = A\tilde{T}^{-1} \quad . \tag{11.62}$$

Algebraic operations yield:

$$\tilde{B} = BS^{-1} \quad . \tag{11.63}$$

Let $\underline{t}^k$ and $\underline{t}^m$ be the k'th and m'th columns of T^{-1}. Similarly, let $\underline{\tilde{t}}^k$ and $\underline{\tilde{t}}^m$ be the k'th and m'th columns of $\tilde{T}^{-1}$. Then equation (11.62) yields:

$$\underline{\tilde{t}}^k = \frac{1}{s_{kk}}\underline{t}^k \quad ; \tag{11.64}$$

$$\underline{\tilde{t}}^m = \frac{-s_{km}}{s_{kk}}\underline{t}^k + \underline{t}^m \quad . \tag{11.65}$$

Let $\underline{b}_k$ and $\underline{b}_m$ be the k'th and m'th columns of matrix B. Similarly, let $\underline{\tilde{b}}_k$ and $\underline{\tilde{b}}_m$ be the k'th and m'th columns of matrix $\tilde{B}$. Then, from equation (11.63);

$$\underline{\tilde{b}}_k = \frac{1}{s_{kk}}\underline{b}_k \quad ; \tag{11.66}$$

$$\underline{\tilde{b}}_m = \frac{-s_{km}}{s_{kk}}\underline{b}_k + \underline{b}_m \quad . \tag{11.67}$$

For $j \neq k, m$:

$$\underline{\tilde{b}}_j = \underline{b}_j \quad . \tag{11.68}$$

Let:

$$z_{km} = \frac{s_{km}}{s_{kk}} \quad . \tag{11.69}$$

From equations (11.60) and (11.69):

$$\frac{1}{s_{kk}^2} = 1 + 2z_{km}\phi_{km} + z_{km}^2 \quad . \tag{11.70}$$

From the preceding equations:

$$\tilde{b}_{ik}^2 = b_{ik}^2\{1 + 2z_{km}\phi_{km} + z_{km}^2\} \quad ; \tag{11.71}$$

$$\tilde{b}_{im}^2 = z_{km}^2 b_{ik}^2 - 2z_{km}b_{ik}b_{im} + b_{im}^2 \quad . \tag{11.72}$$

In terms of the transformed criterion $\tilde{\psi}_{km}$ can be written as:

$$\widetilde{\psi}_{km} = \tfrac{1}{2}\{nu_n - \gamma u_\gamma\} \tag{11.73}$$

where:

$$u_n = \sum_{i=1}^{n}\left[\widetilde{b}_{ik}^2 + \widetilde{b}_{im}^2 + \sum_{\substack{j=1\\ j\neq k,m}}^{r} b_{ij}^2\right]^2 - \sum_{i=1}^{n}\left[\widetilde{b}_{ik}^4 + \widetilde{b}_{im}^4 + \sum_{\substack{j=1\\ j\neq k,m}}^{r} b^4 ij\right] \quad ; \tag{11.74.n}$$

$$u_\gamma = \left\{\sum_{i=1}^{n}\left[\widetilde{b}_{ik}^2 + \widetilde{b}_{im}^2 + \sum_{\substack{j=1\\ j\neq k,m}}^{r} b_{ij}^2\right]\right\}^2 - \left\{\left[\sum_{i=1}^{n}\widetilde{b}_{ik}^2\right]^2 + \left[\sum_{i=1}^{n}\widetilde{b}_{im}^2\right]^2 + \sum_{\substack{j=1\\ j\neq k,m}}^{r}\left[\sum_{i=1}^{n} b_{ij}^2\right]^2\right\} \quad . \tag{11.74.\gamma}$$

With substitution from equations (11.71) and (11.72) into equation (11.73) the criterion $\widetilde{\psi}_{km}$ after algebraic operations may be written as:

$$\widetilde{\psi}_{km} = c_0 + c_1 z_{km} + c_2 z_{km}^2 + c_3 z_{km}^3 + c_4 z_{km}^4 \quad . \tag{11.75}$$

Coefficient c_0 equals the untransformed criterion obtained from equations (11.73), (11.74.n) (11.74.ψ) using the untransformed b's instead of the transformed $\widetilde{b}$'s. Coefficients c_1 through c_4 are obtained by the following equations. For notational conveience in the present context let q_i be defined by:

$$q_i = \sum_{\substack{j=1\\ j\neq k,m}}^{r} b_{ij}^2 \quad , \tag{11.76}$$

For p = 1, 2, 3, and 4, let:

$$c_p = \tfrac{1}{2}\{nc_{pn} + \gamma c_{p\gamma}\} \quad . \tag{11.77}$$

$$c_{1n} = 4\phi_{km}\left\{\sum_{i=1}^{n} b_{ik}^2 b_{im}^2 + \sum_{i=1}^{n} b_{ik}^2 q_i\right\} - 4\left\{\sum_{i=1}^{n} b_{ik}^3 b_{im} + \sum_{i=1}^{n} b_{ik} b_{im} q_i\right\} \quad . \tag{11.78.1n}$$

$$c_{1\gamma} = 4\phi_{km}\left\{\left[\sum_{i=1}^{n} b_{ik}^2\right]\left[\sum_{i=1}^{n} b_{im}^2\right] + \left[\sum_{i=1}^{n} b_{ik}^2\right]\left[\sum_{i=1}^{n} q_i\right]\right\} - 4\left\{\left[\sum_{i=1}^{n} b_{ik}^2\right]\left[\sum_{i=1}^{n} b_{ik} b_{im}\right] + \left[\sum_{i=1}^{n} b_{ik} b_{im}\right]\left[\sum_{i=1}^{n} q_i\right]\right\} \quad . \tag{11.78.1\gamma}$$

$$c_{2n} = 2\left\{\sum_{i=1}^{n} b_{ik}^4 + \sum_{i=1}^{n} b_{ik}^2 b_{im}^2 + 2\sum_{i=1}^{n} b_{ik}^2 q_i - 4\phi_{km}\sum_{i=1}^{n} b_{ik}^3 b_{im}\right\} \quad ; \tag{11.78.2n}$$

$$c_{2\gamma} = 2\left\{\left[\sum_{i=1}^{n} b_{ik}^2\right]^2 + \left[\sum_{i=1}^{n} b_{ik}^2\right]\left[\sum_{i=1}^{n} b_{im}^2\right] + 2\left[\sum_{i=1}^{n} b_{ik}^2\right]\left[\sum_{i=1}^{n} q_i\right]\right. \tag{11.78.2\gamma}$$

$$c_{3n} = 4\left\{\phi_{km}\sum_{i=1}^{n} b_{ik}^4 - \sum_{i=1}^{n} b_{ik}^3 b_{im}\right\} \quad ; \tag{11.78.3n}$$

$$c_{3\gamma} = 4\left\{\phi_{km}\left[\sum_{i=1}^{n} b_{ik}^2\right]^2 - \left[\sum_{i=1}^{n} b_{ik}^2\right]\left[\sum_{i=1}^{n} b_{ik} b_{im}\right]\right\} \tag{11.78.3\gamma}$$

$$c_{4n} = 2\sum_{i=1}^{n} b_{ik}^4 \quad ; \tag{11.78.4n}$$

$$c_{4\gamma} = 2\left[\sum_{i=1}^{n} b_{ik}^2\right]^2 \quad . \qquad (11.78.4\gamma)$$

The minimum of $\widetilde{\psi}_{km}$ may be obtained by equating the dreivative os $\widetilde{\psi}_{km}$ with respect to z_{km} equal to zero and finding the roots of the resulting equation.

$$\widetilde{\psi}_{km} = c_1 + \tfrac{1}{2}c_2 z_{km} + \tfrac{1}{3}c_3 z_{km}^2 + \tfrac{1}{4}c_4 z_{km}^3 = 0 \quad . \qquad (11.79)$$

A first special case to be considered involves a negative value for c_4 which, from inspection of equations (11.77), (11.78.4n) and (11.78.4γ), will occur only for positive values of γ. As previously noted, undesirable transformations may occur with positive values of γ. Therefore, a conclusion should be reached to exclude the solution whenever c_4 is negative. Otherwise, there are three zero roots to equation (11.79) whith the middle root being for a maximum $\widetilde{\psi}_{km}$ and the least root and greatest root being for minimum values of $\widetilde{\psi}_{km}$. The value of $\widetilde{\psi}_{km}$ is to be determined for these two roots and the one yielding the lesser value of $\widetilde{\psi}_{km}$ is to be chosen as the desired solution .

11.4 Orthoblique Transformations

Harris and Kaiser (1964) made a most interesting and stimulating contribution to factor transformations in their article on Oblique Factor Analysis Solutions by Orhtogonal Transformations. In this contribution they presented a general theory with two special cases that might be used with given original factor matrices. They showed that all transformations, orthogonal or oblique, could be generated with orthogonal transformations. In describing this contribution, as much as reasonable of their notation will be used.

Some preliminary matters are to be considered first. Let A_0 be any original factor matrix obtained by any of the factor extraction methods. Harris and Kaiser start from a factor matrix with the principal factors identification. It is not necessary for A_0 to be a principal factors matrix. These preliminary matters concern a transformation from any A_0 to a matrix A with the principal factors identification. Two procedures are to be considered: first the procedure followed by Harris and Kaiser and second an equivalent procedure which facilitates the application of any transformation applied to A to be applicable to the original matrix A_0. Harris and Kaiser used matrix R^* defined by:

$$A^* = A_0 A_0' \quad . \qquad (11.80)$$

The principal factors solution is obtained for R^* as in Chapter 8 by obtaining an eigen solution which is symbolized here as, folowing the Harris and Kaiser notation:

$$R^* = QM^2Q' \quad . \qquad (11.81)$$

Q is an n x r matrix of eigenvectors and, thus is orthonormal by columns and M^2 is a diagonal matrix of eigenvalues. Note the use of the square for the eigenvalues; this is to facilitate following equations. Then, the principal factors matrix A is given by:

$$A = QM \quad . \tag{11.82}$$

An alternative procedure yields a transformation of A_0 to A. This is desirable to be able to apply all transformations on A to be applicable to A_0. An eigen solution is obtained for the product matrix $A_0'A_0$:

$$A_0'A_0 = V_0\Lambda_0V_0' \quad . \tag{11.83}$$

Then:

$$A = A_0V_0 \quad ; \tag{11.84}$$

$$M^2 = \Lambda_0 \quad . \tag{11.85}$$

It can be shown that:

$$AA' = R^* \quad ;$$

$$A'A = \Lambda_0 = M^2 \quad . \tag{11.86}$$

Then:

$$Q = AM^{-1} \quad . \tag{11.87}$$

A following relation is that:

$$Q'Q = I \quad .$$

These relations are necessary for the results of this procedure to yield identical results to the first procedure.

Let $\tilde{T}$ be any r x r, nonsingular transformation matrix in the set of all possible such matrices. Note that the rows of $\tilde{T}$ are not normalized. Let, in a more general sense than equation (7.4):

$$A\tilde{T}^{-1} = \tilde{B} \quad . \tag{11.88}$$

Note that:

$$A_0V_0\tilde{T}^{-1} = \tilde{B} \quad . \tag{11.89}$$

Let:

$$V_0\tilde{T}^{-1} = \tilde{T}_0^{-1} \quad ;$$

$$\tilde{T}_0 = \tilde{T}V_0' \quad . \tag{1190}$$

Then:

$$A_0\tilde{T}_0^{-1} = \tilde{B} \quad . \tag{11.91}$$

Since V_0 is square, nonsingular, $\tilde{T}_0$, as given in equation (11.90), is an element in the set of all possible r x r, nonsingular matrices. Any generated $\tilde{T}$ yields a $\tilde{T}_0$ which may be applied to A_0 to yield $\tilde{B}$, as per equation (11.91).

Let $\tilde{\Phi}$ be a generalized factor covariance matrix defined, in parrallel whith equation (7.1), by:

$$\tilde{T}\tilde{T}' = \tilde{\Phi} \quad . \tag{11.92}$$

The rows of $\tilde{T}$ are normalized to T by:

$$T = D_1^{-1}\tilde{T} \tag{11.93}$$

where D_1 is defined by:

$$D_1^2 = \text{Diag}(\tilde{\Phi}) \quad . \tag{11.94}$$

In the present context, the factor correlation matrix is symbolized by Φ and is given by equation (7.1) as:

$$TT' = \Phi \quad . \tag{11.95}$$

From (11.92) and (11.93):

$$\tilde{\Phi} = D_1 \Phi D_1 \quad . \tag{11.96}$$

The factor weight matrix for normalized factors is given in equation (7.4):

$$AT^{-1} = B \quad . \tag{7.4}$$

With equations (11.88) and (11.93), equation (7.4) yields:

$$\tilde{B} = BD_1^{-1} \quad . \tag{11.97}$$

The preceding discussion has focused on the trnaformational relations involving an original factor matrix A_0 and a principal factors matrix A as well as generalized transformations in $\tilde{T}$ with row wise normalized transformations in T. Next to be considered are the orthoblique transformations suggested by Harris and Kaiser. They presented the esscence of their suggextion in equation (2) (1964 page 350). We find that this understanding this development is enhanced by considering an Eckart-Young decomposition (Eckart and Young 1936; Johnson 1963) of matrix $\tilde{T}$:

$$\tilde{T} = T_1' D_2^{-2} T_2' \tag{11.98}$$

where T_1 and T_2 are r x r orthonormal matrices; D_2 is an r x r, positive, nonsingular diagonal matrix. In the following developments it is important to remember with T_1 and T_2 being square orthogonal matrices that:

$$T_1 T_1' = T_1' T_1 = I \quad ; \tag{11.99.1}$$

$$T_2 T_2' = T_2' T_2 = I \quad . \tag{11.99.2}$$

A reversal of directions is considered to develop $\tilde{T}$ from matrices T_1, T_2 and D_2. To develop all possible $\tilde{T}$, orthonormal matrices T_1 and T_2 range independently over all elemenst in the set of all possible r x r orthonormal matrices. Also, diagonal matrix D_2 ranges over all possible postive, nonsingular, r x r diagonal matrices. For special cases, one or two of these matrices may be set at prescribed values. Matrix $\tilde{\Phi}$ is obtained from equations (11.92) and (11.98):

$$\tilde{\Phi} = T_1' D_2^{-1} T_2' T_2 D_2^{-1} T_1 = T_1' D_2^{-1} T_1 \quad . \tag{11.100}$$

Matrices D_1, T and Φ may be obtained from equations (11.94), (11.93) and (11.96). The foregoing material provides the essence of the Harris-Kaiser orthoblique transformations in that all transformations, orthogonal or oblique, can be generated from orthogonal transformations T_1 and T_2 along with diagonal matrix D_2. Special cases are to be considered next.

A special category of transformations is that of orthogonal transformations. To produce orthogonal transformations matrix D_2 is set to an identity matrix. Equation (11.100) reduces to:

$$\tilde{\Phi} = T_1' I T_1 = I \quad . \tag{11.101}$$

Then:

$$\tilde{T} = T_1' T_2' \quad . \tag{11.102}$$

Since the product of orthonormal matrices is also an orthonormal matrix, $\tilde{T}$ is an orthonormal matrix. Since $\tilde{\Phi}$ is an identity matrix, D_1 is an identity matrix so that T equals $\tilde{T}$ and Φ equals an identity matrix. Thus, the obtained matrix T may be used in equation (7.4) to yield orthogonally transformations to factor weight matrix B.

Harris and Kaiser suggested that a special class of solutions might yield interesting practical procedures. In this class the number of parameters in equation (11.98) is reduced by defining:

$$T_2 = I \quad ; \tag{11.103}$$

$$D_2^{-1} = M^p \quad . \tag{11.104}$$

Then, from equation (11.98):

$$\tilde{T} = T_1' M^p \tag{11.105}$$

and from equation (11.100):

$$\tilde{\Phi} = T_1' M^{2p} T_1 \quad . \tag{11.106}$$

Also, from equations (11.88) and (11.105):

$$\tilde{B} = A M^{-p} T_1 = Q M M^{-p} T_1 \quad . \tag{11.107}$$

With these definitions, harris and Kaiser considered several cases generated from definitions of p.

A first case involves $p = 0$. With the convention that $M^0 = I$, $\tilde{T} = T_1'$ and $\tilde{\Phi} = I$. Thus, this case reduces to the orthogonal transformations discussed in a preceding paragraph.

A second case involves $p = 1/2$. From equations (11.105), (11.106) and (11.107):

$$\tilde{T} = T_1' M^{1/2} \quad ; \tag{11.108}$$

$$\tilde{\Phi} = T_1' M T_1 \quad ; \tag{11.109}$$

$$\tilde{B} = A M^{-1/2} T_1 \quad . \tag{11.110}$$

For this case, the product matrix $B'B$ is of interest. From equation (11.110)

$$\widetilde{B}'\widetilde{B} = T_1'M^{-1/2}A'AM^{-1/2}T_1$$

which becomes with equation 11.86)

$$\widetilde{B}'\widetilde{B} = T_1'MT_1 \quad . \tag{11.111}$$

Note that $\widetilde{B}'\widetilde{B}$ equals $\widetilde{\phi}$. Using equations (11.96) and (11.97), equation (11.111) yields:

$$B'B = D_1^2\Phi D_1^2 \quad . \tag{11.112}$$

This case can be termed the $\underline{\Phi \text{ proportional to } B'B}$ case. (Note: we use B where Harris and Kaiser used A.) Examples of transformation using this case are described in a subsequent paragraph.

A third case involves $p = 1$. From equations (11.105), (11.106) and (11.107):

$$\widetilde{T} = T_1'M \quad ; \tag{11.113}$$

$$\widetilde{\Phi} = T_1'M^2T_1 \quad ; \tag{11.114}$$

$$\widetilde{B} = AM^{-1}T_1 \quad . \tag{11.115}$$

As for the preceding case, the product matrix $\widetilde{B}'\widetilde{B}$ is of interest. From equation (11.115):

$$\widetilde{B}'\widetilde{B} = T_1'M^{-1}A'AM^{-1}T_1$$

which yields with equations (11.86) and ((11.99.1):

$$\widetilde{B}'\widetilde{B} = I \tag{11.116}$$

so that:

$$B'B = D_1^2 \tag{11.117}$$

which is a diagonal matrix. Note, from equation (11.82):

$$\widetilde{B} = QT_1 \quad . \tag{11.118}$$

so that the orthogonal tranxformation T_1 is applied directly to matrix Q of eigenvectors. Thus, for all possible orthogonal transformations, $B'B$ will be a diagonal matrix. As discussed in a subsequent paragraph, this case is termed the $\underline{\text{Independent Clusters}}$ case.

Examples of the Orthoblique transformations are given in Tables 11.8 and 11.9. Table 11.8 gives results for the artificial example used to illustrate the Orthomax tranformations with the parameter matrix B given in Table 11.1. The illustrations in Table 11.8 use the parameter $\phi_{12} = .6$. The Quartimax transformation was chosen to yield matrix T_1 for each of the cases of Orthoblique transformations . The input parameter matrix B is repeated from Table 11.1 at the left of Table 11.8 for comparison purposes. With $p = 0$ the Orthogonal transformation is identical to that given in Table 11.2 for $\phi_{12} = .6$. The parameter $\phi_{12} = .6$ was chosen to provide an example of the change in results using the Φ Proportional to $B'B$ case for which $p = 1/2$. The results obtained are moderately near to the input parameters. There has been a distinct improvement over the Orthogonal transformation results. For $p = 1$, the independent clusters case, the results have deteriorated. This is not surprising since the input configuration was not a

Table 11.8

Orthoblique Transformations for Artificial Example, $\phi_{12} = .6$
Using Quartimax Transformations

	Input Parameters		p = 0 Orthogonal		p = 1/2 Φ Proportional To B'B		p = 1 Independent Clusters	
				Factor Weight Matrices B				
	1	2	1	2	1	2	1	2
1	.70	.00	.63	.31	.61	.15	.77	-.09
2	.50	.00	.45	.23	.44	.11	.55	-.06
3	.00	.60	.54	-.27	-.04	.62	-.20	.75
4	.00	.40	.36	-.18	-.03	.41	-.14	.50
5	.50	.20	.63	.14	.42	.32	.48	.19
6	.40	.40	.72	.00	.32	.50	.30	.45
7	.30	.60	.81	-.13	.22	.69	.13	.71
				Matrices B'B				
	1	2	1	2	1	2	1	2
1	1.24	.44	2.56	.07	1.41	.55	1.57	.00
2	.44	1.08	.07	.29	.55	.90	.00	1.28
				Matrices $R_{bb} = \Phi$				
	1	2	1	2	1	2	1	2
1	1.00	.60	1.00	.00	1.00	.49	1.00	.80
2	.60	1.00	.00	1.00	.49	1.00	.80	1.00

Table 11.9

Orthoblique Transformations for Nine Mental Tests Example
Using Normal Varimax Transformation

p = 0.0 Orthogonal			p = 1/2 Φ Proportional To B'B			p =1.0 Independent Clusters		

Factor Weight Matrices B

	1	2	3	1	2	3	1	2	3
1	.61	.07	.12	.61	.05	.00	.70	.01	-.13
2	.67	-.05	.30	.65	-.10	.19	.68	-.18	.12
3	.53	.32	.23	.50	.29	.12	.54	.25	.02
4	.05	.71	.09	.01	.71	.02	.02	.73	-.05
5	.06	.78	.16	.00	.77	.09	-.01	.78	.04
6	.09	.70	.15	.04	.69	.09	.04	.70	.03
7	.37	-.01	.55	.27	-.10	.53	.15	-.22	.61
8	.14	.35	.56	.01	.27	.55	-.14	.18	.66
9	.25	.23	.57	.13	.14	.55	.00	.04	.64

Matrices B'B

	1	2	3	1	2	3	1	2	3
1	1.32	.41	.84	1.13	.13	.41	1.28	.00	.00
2	.41	1.88	.68	.13	1.76	.33	.00	1.80	.00
3	.84	.68	1.15	.41	.33	.95	.00	.00	1.26

Matrices $R_{bb} = \Phi$

	1	2	3	1	2	3	1	2	3
1	1.00	.00	.00	1.00	.09	.39	1.00	.24	.69
2	.00	1.00	.00	.09	1.00	.25	.24	1.00	.46
3	.00	.00	1.00	.39	.25	1.00	.69	.46	1.00

independent clusters situation. Note, matrix B′B is a diagonal matrix for $p = 1$. This agrees with equation 11.117. In proposing the Φ Proportional to B′B case, $p = 1/2$, Harris and Kaiser (1964) noted that this was approximately true for a large number of examples of real data. This case works moderately well in a variety of studies.

An extreme case of simple structure has a cluster of attributes for each transformed factor with no intermediate attributes Thus, since each attribute has a loading on one and only one transformed factor, the product of paired loadings for each pair of transformed factors will eaual zero so that the off-diagonal entries in B′B will be zero so that B′B will be a diagonal matrix. This corresponds to the result in equation (11.117) so that the third special case applies. The transformation problem is to develop transformation matrix T_1. Any of the Orthomax procedures described earlier appear to be usable. A Quartimax transformation was used in Table 11.8.

Table 11.9 presents results for the Nine Mental Tests Example. A Normal Varimax transformation was used to generate matrices T_1. Results may be compaired with the Matrix B of Table 10.10. While the transformed factors are quite recognizable for the Φ Proportional to B′B case, the definitions of the hyperplanes may not be as clean for near zero loadings as was true for the graphical transformations. There are several undesirable negative loadings in the B matrix for the Independent Clusters solution, a result that should not be surprising since the Nine Mental Tests Example did not possess an independent clusters configuration.

11.5 PROMAX Transformations

An intuitively appealing suggestion was made by Hendrickson and White (1964) as a procedure to transform an orthogonal tranformation to an oblique solution. They named this procedure as the PROMAX transformation. Starting from an orthogonal transformation to loadings x_{ij}, n oblique solution was desired which would reduce smaller loadings toward zero while increasing the absolute value of the higher loadings. They proposed to generate target loadings, y_{ij}, dependent on the x_{ij} which increased the relative ratio of the absolute values of the high loadings to the low loadings. Once having such a target, a least squares fitting procedure would be applied to obtain the oblique transformation. To accomplish this, Hendrickson and White suggested the following power type function. Let:

$$\delta_{ij} = +1 \quad \text{if } x_{ij} \geq 0 \quad ; \tag{11.119a}$$

$$\delta_{ij} = -1 \quad \text{if } x_{ij} < 0 \quad . \tag{11.119b}$$

then Hendrickson and white suggest that:

$$y_{ij} = \delta_{ij} |x_{ij}^{p}| \tag{11.120}$$

where p is a power parameter to be greater than unity. This function raises the x_{ij} to the power p and assigns the algebraic sign of x_{ij} to the y_{ij}. There remains a question as to the value of the power p. A least squares type fit of the oblique loadings to the y_{ij} is the next step. Let $\widetilde{F}'$ be a non-normalized transformation to matrix $\widetilde{G}$:

$$A\widetilde{F}' = \widetilde{G} \quad . \tag{11.121}$$

This transformation is taken as non-normalized to allow the derived factor loadings to have a proportional relation to the best fit to the target matrix Y with elements y_{ij}. Consider:

$$E = Y - \widetilde{G} = Y - A\widetilde{F}' \quad . \tag{11.122}$$

By standard least squares theory:

$$\widetilde{F}' = (A'A)^{-1}A'Y \quad . \tag{11.123}$$

Matrix $\widetilde{F}'$ is normalized by columns to matrix F' of normals to the hyperplanes To accomplish this:

$$D^{-2} = \text{diag}(\widetilde{F}\widetilde{F}') \tag{11.124}$$

$$F' = \widetilde{F}'D \quad . \tag{11.125}$$

Hendrickson and White provided three examples of the use of Promax; Harman 24 Tests Example (1960), Cattell and Dickman Ball Problem (1962) and Thurstone Box Problem (1947). In the first two of these examples Hendrickson and White found that a power coefficient, p, equal to 4 worked best. For the Thurstone Box Problem their finding was that p equal to 2 worked best. Use of Promax for the Nine Mental Tests Example is given in Table 11.10. The Least Squares Hyperplane Fitting, LSQHYP, from Chapter 10, are given at the left for comparison. Results for two values of p are given: $p = 2$ and $p = 4$. For factors 1 and 3 the results for $p = 4$ appear to be superior, being quite close to the LSQHYP results. For factor 2, the results for $p = 2$ appear to be better, the quite negative loadings for attributes 2 and 7 on factor 2 for $p = 4$ are undesirable. Maybe, an analyst should try both $p = 2$ and $p = 4$ and select the preferable results. It appears that this selection might be made factor by factor. General experience has indicated that Promax yields good to excellent results.

11.6 Weighted Least Squares Hyperplane Fitting

The general form for least squares hyperplane fitting is similar to Thurstone's (1947) suggested equation for simple structure given in equation (11.1). The equation for this form is:

$$\xi = \sum_{i=1}^{n}\sum_{j=1}^{r} v_{ij}^2 w_{ij} \quad . \tag{11.126}$$

where the v_{ij}'s may be either the projections on normals (structure loadings) in our matrix G or the factor weights (pattern loadings) in our matrix B. The new feature consists of the weights w_{ij} which may be determined by various judgments and/or techniques. Given the weights the

Table 11.10

Promax Transformations for Nine Mental Tests Example
Using Normal Varimax to Develop Target

	LSQHYP Transformation			Promax Transformation p = 2			Promax Transformation p = 4		
	Normal Matrices F'								
	1	2	3	1	2	3	1	2	3
1	.30	.68	.26	.35	.55	.36	.27	.48	.28
2	.50	-.73	.19	.56	-.77	.26	.51	-.79	.21
3	.82	.08	-.95	.75	.32	-.90	.82	.39	-.94
	Projections on Normals, Matrices G								
	1	2	3	1	2	3	1	2	3
1	.53	.05	-.09	.56	.04	-.01	.53	.03	-.07
2	.54	-.06	.07	.59	-.11	.16	.53	-.14	.09
3	.42	.32	.01	.45	.27	.09	.40	.24	.03
4	.01	.71	-.03	.00	.68	-.01	-.01	.66	-.03
5	-.01	.78	.02	-.01	.73	.06	-.03	.70	.03
6	.02	.71	.02	.02	.66	.05	.00	.63	.02
7	.17	.02	.40	.24	-.11	.46	.16	-.17	.42
8	-.06	.39	.43	-.01	.25	.48	-.07	.18	.44
9	.05	.27	.42	.11	.12	.48	.04	.06	.43

analytic problem is to establish the factor transformation to minimize ξ. The analytic procedure for the projections on the normals was described in Chapter 10, section 10.3. A special case for the weights was discussed there for which the weights were restricted to being 0 or 1. With this restriction a procedure LSQHYP was discussed. A suggestion was made that these weights could be defined by the analyst as a result of judgments from graphical rotatations. Other determinations of weights for the projections on normals case will be described subsequently. Development of an anlytic solution for the case of factor weights is discussed next.

For the case of the factor weights, b_{ij}, equation (11.126) is written as:

$$\xi = \sum_{i=1}^{n} \sum_{j=1}^{r} b_{ij}^2 w_{ij} \quad . \tag{11.127}$$

There appears to be no simple, general solution for the transformation matrix T to minimize function ξ. Instead, a series of transformations of each factor with respect to each other factor is employed for each of a series of cycles. This is similar to the solution for the Direct Oblimin criterion. Consider factor k which, successively, takes on the value of each factor. For each value of k, each of the other factors is taken as factor m and a series of solutions is computed for transforming k in the direction of each m. For each pair of factors, k and m, the criterion may be written as:

$$\tilde{\xi}_{km} = \sum_{i=1}^{n} \tilde{b}_{ik}^2 w_{ik} + \sum_{i=1}^{n} \tilde{b}_{im}^2 w_{im} + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq k,m}}^{r} b_{ij}^2 w_{ij} \quad . \tag{11.128}$$

The transformation is defined in equation (11.59) as was done for the Direct Oblimin solution. The restriction on coefficients s_{kk} and s_{km} is given in equation (11.60). Parameter z_{km} is defined in terms of s_{kk} and s_{km} in equation (11.69). The transformed $\tilde{b}_{ik}^2$ and $\tilde{b}_{im}^2$ are given in equations (11.71) and (11.72). Substitution fromequations (11.71) and (11.72) into equation (11.128) yields:

$$\tilde{\xi}_{km} = c_0 + c_1 z_{km} + c_2 z_{km}^2 \tag{11.129}$$

where:

$$c_0 = \left\{ \sum_{i=1}^{n} b_{ik}^2 w_{ik} + \sum_{i=1}^{n} b_{im}^2 w_{im} + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq k,m}}^{r} b_{ij}^2 w_{ij} \right\} \quad ; \tag{11.130.0}$$

$$c_1 = 2 \left\{ \sum_{i=1}^{n} b_{ik}^2 w_{ik} \phi_{km} - \sum_{i=1}^{n} b_{ik} b_{im} w_{im} \right\} \quad ; \tag{11.130.1}$$

$$c_2 = \left\{ \sum_{i=1}^{n} b_{ik}^2 w_{ik} + \sum_{i=1}^{n} b_{ik}^2 w_{im} \right\} \quad . \tag{11.130.2}$$

With the weights w_{ik} and w_{im} being limited to being non-negative, coefficient c_2 must be

non-negative. The possibility that c_2 equalling zero is so unlikely that this possibility may be ignored. Since equation (11.129) is a quadratic the solution for an optimum is:

$$z_{km} = -c_1/2c_2 \quad . \tag{11.131}$$

With c_2 taken as positive, equation (11,131) yields the desired minimum value of $\tilde{\xi}_{km}$. Then, from equation (11.70):

$$s_{kk} = 1 \Big/ \sqrt{1 + 2z_{km}\phi_{km} + z_{km}^2} \quad . \tag{11.132}$$

From equation (11.69):

$$s_{km} = s_{kk} z_{km} \quad . \tag{11.133}$$

Rvised values of t_{hk} , h = 1,r, can be obtained from equation (1159) as:

$$\tilde{t}_{hk} = s_{kk} t_{hk} + s_{km} t_{hm} \quad . \tag{11.133}$$

Revised values of ϕ_{hk} are obtained from:

$$\tilde{\phi}_{hk} = \tilde{\phi}_{kh} = s_{kk}\phi_{hk} + s_{km}\phi_{hm} \quad \text{for } h \neq k \tag{11.134}$$

$$\tilde{\phi}_{kk} = 1 \quad . \tag{11.135}$$

Revised values of b_{ik} and b_{im} may be obtained from eauations (11.66) and (11.67). A computer program can start from initial matrices T , Φ , and B ; cycle through the factors as k and, within each k , cycle through all factors except k as m . Such a program would continue until none of the transformations within a cycle are greater than some minimum value. A possible good starting transformation can be obtained from the LSQHYP solution described in Chapter 10. There remains the problem of determining the weights, w_{ij} to be used.

Examples of least squares hyperplane fitting are given in Table 11.11 with the weights being 0 or 1 as selected from graphical rotations as in the LSQHYP procedure discussed in Chapter 10, section 10.3. The least weighted least squares solution for the projections on the normals is given on the left of Table 11.11 with the weights for the projections being given as superscripts. These results are the same as given in Table 10.11. The resulting factor weights are given at the lower right of Table 11.11 for comparison with results for the weighted least squares solutionfor the factor weights which are given at the lower right. Again, the weights used are given as superscripts to the factor weights. These weights are identical to those used on the projections on the normals. In this example there are only small changes in the factor weights from the weighted least squares solution on the projections to the solution on the factor weights. In general, there may be only slight differences between these two solutions so that the weighted least squares solution on the projections on the normals might be recommended when the weights are given before the solution is accomplished.

Weighted least squares hyperplane fitting provides a mechanism to automate in a simulated fashion some of the detailed judgments made by analysts in graphical rotations of

Table 11.11
Weighted Least Squares Hyperplane Fitting
Nine Mental Tests Example

	Least Squares Projections			Least Squares Factor Weights		
	Projections on Normals (Structure Loadings)					
	1	2	3	1	2	3
1	.53[0]	.05[1]	-.09[1]	.52	.05	-.08
2	.54[0]	-.06[1]	.07[1]	.54	-.06	.08
3	.42[0]	.32[0]	.01[1]	.41	.32	.01
4	.01[1]	.71[0]	-.03[1]	.00	.71	-.04
5	-.01[1]	.78[0]	.02[1]	-.01	.78	.02
6	.02[1]	.71[0]	.02[1]	.02	.71	.02
7	.17[0]	.02[1]	.40[0]	.18	.03	.41
8	-.06[1]	.39[0]	.43[0]	-.04	.40	.44
9	.05[1]	.27[0]	.42[0]	.06	.27	.42
	Factor Weights (Pattern Loadings)					
	1	2	3	1	2	3
1	.68	.05	-.12	.66[0]	.05[1]	-.10[1]
2	.68	-.06	.09	.68[0]	-.06[1]	.10[1]
3	.53	.32	.01	.52[0]	.32[0]	.02[1]
4	.01	.72	-.04	.00[1]	.72[0]	-.04[1]
5	-.01	.79	.03	-.02[1]	.79[0]	.03[1]
6	.03	.71	.02	.02[0]	.72[0]	.02[1]
7	.22	.02	.50	.23[0]	.03[1]	.49[0]
8	-.07	.40	.55	-.05[1]	.40[0]	.53[0]
9	.06	.27	.53	.08[1]	.28[0]	.52[0]

[0] Weights = 0.
[1] Weights = 1.

factors. An experienced analyst develops judgments as to the likelihood that an attribute vector will be in a particular hyperplane. Such a judgment may be termed the analyst's personal probability Tucker and Finkbeiner (1981) proposed a simulated representation of such judgments based on the loadings on a factor at some point in the graphical rotation procedure. They termed this an artificial personal probability and suggested a convenient equation:

$$APP_{ij} = 1 \Big/ \{1 + a|v_{ij}|^{c}\} \qquad (11.136)$$

where APP_{ij} is the value of the artificial personal probability for attribute i on factor j and v_{ij} is the loading of attribute i on factor j. Parameters of this function are a and c to be set by the analyst from properties of the data being analysed. Values of a and c are suggested here for the analysis of correlation matrices. Figure 11.2 illustrates the form of this function in terms of the factor weights b_{ij} as the v_{ij} for what Tucker and Finkbeiner termed the two sided condition, this being the most general condition when meaningful factor loadings may be either positive or negative. There is a short distance about a factor loading of zero when the value of APP is approximately unity, this representing a judgment that attributes having such small loadings would be in the hyperplane. On either side of the central interval the APP function slopes off until becoming approximately zero. This function as illustrated is different from a step function that symbolize judgments that attributes are either in a hyperplane or out of a hyperplane. Tucker and Finkbeiner suggested that there should be a transition between in a hyperplane and being out of a hyperplane.

Tucker and Finkbeiner considered the case of what has been called the positive manifold for which meaningful factor loadings are judged to be zero or positive. Such a case arrises for ability measures where factors should not have negative effects on the performance of any attribute. They termed this the one sided case with the form of the APP function illustrated in Figure 11.3. In order to force loadings away from being negative the valuse of APP are set to unity. A judgment is required from the analyst before using the transformation procedure as to whether a one sided or a two sided function is to be used. This judgment should depend on the nature of the data being analysed. One point is that all measures should be made in a positive direction for which a higher value is better. Attributes, such as reaction time, for which low values are better should be reversed in direction by changing the signs of the loadings for such an attribute in the original factor matrix. On the positive side, the APP function is the same for the one sided case as for the two sided case.

Values of the parameters a and c of the APP function of equation 11.136 may be related to the value of v for $APP = .5$ and $APP = .1$, these values being selected for analyses involving correlation matrises. Let $v_{.5}$ be the positive value of v for $APP = .5$ and $v_{.1}$ be the

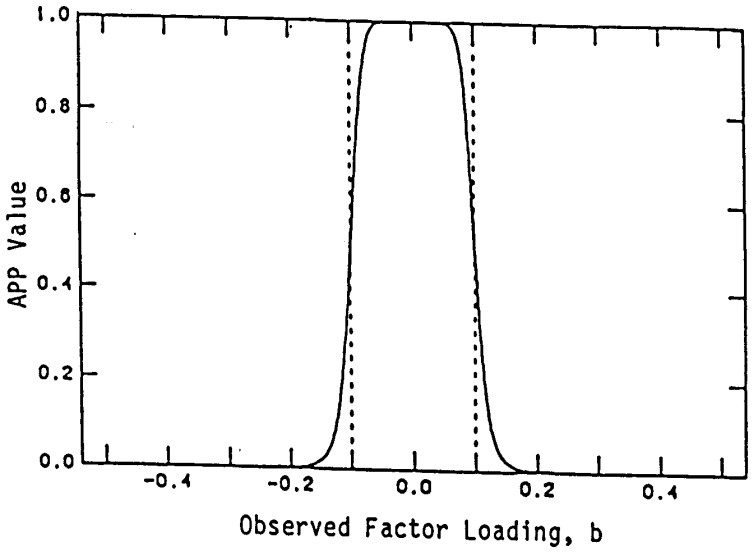

Figure 11.2. Illustration of two sided APP function:

$a = 10^9$; $c = 9$.

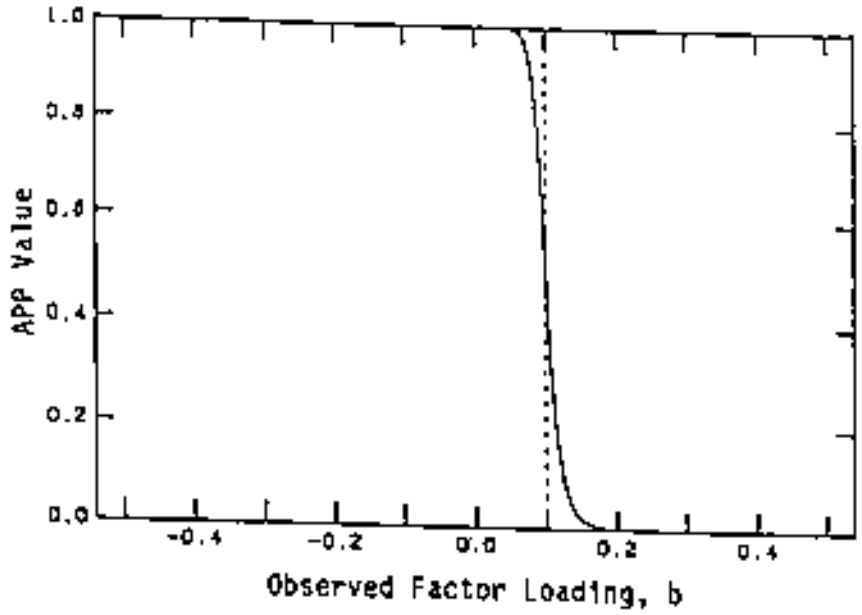

Figure 11.3. Illustration of one sided APP function: $a = 10^9$; $c = 9$; if $b < 0$, APP = I .

positive value of v for APP = .1 . The ratio of $b_{.1}$ to b_5 is related to the slope of the APP function. At a ratio of unity the function would be a step function; as the ratio increases, the slope of the function flattens out. The following equations give the relations of the parameters to these values of v :

$$c = \ln(9) \Big/ [\ln(v_{.1}) - \ln(v_{.5}] \quad ; \tag{11.137}$$

$$a = (v_{.5})^{-c} \quad . \tag{11.138}$$

Tucker and Finkbeiner found with many experimental trials the a value of $c = 9$ gave best results. The value of $v_{.5}$ can be selected by the analyst and the value of c computed by equation 11.138. A computational method suggested by Tucker and Finkbeiner will be discussed subsequently.

A first attempt to use APP in an iterative fashion involved the projections on the normals. In this procedure the v_{ij} were defined as the projections on the normals, g_{ij} . A starting transformation was necessary which might be obtained by a VARIMAX transformation. Each of the factors was transformed independently with trial weights being the APP values computed from the trial projections on the normals. This was a successive trial type procedure with new weights being computed as the APP values after each solution for new normals and projections on these normals by the least squares hyperplane fitting procedure described previously. While there was no explicit criterion to be minimized or maximized, each normal appeared to come to a stable position after a few trials. With some bodies of data the results obtained appeared excellent; there was, however, a problem with some other bodies of data. This procedure suffered the same problem as other single plane methods: two or more factors arrived at the same stable position even though they had different starting positions. Consequently, this procedure was discarded.

Tucker and Finkbeiner reported excellent results with the APP applied to the factor weights. They termed this procedure DAPPFR for direct artificial probability factor rotation. Results for the nine mental tests example are given in Table 11.12. Since these attributes are ability tests, a prior judgment is indicated that a one sided solution would be appropriate. This solution is shown on the left of Table 11.12 and, indeed, these results are very close to the LSQHYP solution given in Table 11.11. For a contrast, a two sided solution was run with results given at the right of Table 11.12. While factors 1 and 3 results are close to the LSQHYP results, the results for factor 2 are quite unsatisfactory. The negative loading for attribute 7 indicates that these results should be rejected for ability measures. Tucker and Finkbeiner (personal correspondence) reported that the results from extensive simulation and Monte Carlo studies that DAPPFR gave, on the average, the best of any automated transformation procedure tried including Direct Oblimin. Second place was PROMAX with a power constant of 2. Such

Table 11.12
Dappfr Transformations Factor Weights
Nine Mental Tests Example

	One Sided Solution			Two Sided Solution		
	1	2	3	1	2	3
1	.66	.05	-.10	.65	.24	-.13
2	.67	-.05	.10	.68	.01	.11
3	.52	.33	.01	.52	.42	.00
4	.00	.72	-.04	.00	.74	-.05
5	-.02	.80	.03	-.02	.78	.03
6	.02	.71	.02	.02	.70	.02
7	.24	.02	.49	.29	-.25	.59
8	-.06	.40	.54	.00	.05	.64
9	.08	.28	.52	.14	-.04	.62

judgments are possible with simulation and Monte Carlo studies since a criterion is available from the generation of the data. Trials with real data can only lead to a judgment that the obtained solutions are liked or are not liked. But, there is no concrete criterion for a judgment of superiority. The DAPPFR procedure appears to work very well both for real data and simulated data.

Computational procedures start from some initial location of the trait vectors in matrix T and factor weights in matrix B. Then, a series of cycles of weighted least squares hyperplanes fitting are performed with the APP values used as weights. These APP values are computed by equation 11.136 using the factor weights b_{ij} as the v_{ij}. Initial values of APP are computed from the initial matrix B. These weights are revised after each cycle. While there is no explicit algebraic function to be optimized the procedure does arrive at a stable position, usually in 20 cycles or fewer. Tucker and Finkbeiner commented that a value of $b_{.5}$ equal to .15 gave good results when data being analysed were correlation matrices and attributes have high communalities. They suggested an alternative computational procedure dependent on the mean absolute value of the entreis in the B matrix. Let $\bar{b}$ be the mean absolute value of the entries in the trial matrix B. For the first cycle use:

$$a = (\bar{b})^{-9} \quad . \tag{11.139}$$

For all subsequent cycles use:

$$a = \frac{r}{6}(\bar{b})^{-9} \tag{11.140}$$

where r is the number of factors. A good starting transformation involves VARIMAX followed by PROMAX with a power constant of 2.

CHAPTER 12
FACTOR OPERATIONS FOR MULTIPLE DATA

From
Exploratory Factor Analysis
Ledyard R Tucker
and
Robert C. MacCallum

CHAPTER 12
FACTOR OPERATIONS FOR MULTIPLE DATA

A number of situations exist which involve operations with two or more bodies of data. This is in contrast to the preceding considerations of operations for a single body of data. The sections in this chapter consider such situations as when after a factor analysis has been performed for one battery of attributes data is available for additional attributes such as course grades following the analysis of a battery of pre-course aptitudes data. A second example involves transformations of an original factor matrix to hypothetical loadings. Also to be considered are analyses and tranformations for covariance matrices and factor matrices from several samples.

Scales of measurement of attribute measures and of factors are important mattesr when dealing with data from several samples. The common factor model in the sample was discussed in Chapter 4. Equation (4.19) indicates that population factor weights are multiplied by sample factor standard deviations to obtain the sample factor weights. These factor weights are for analyses of covariance matrices. When analyses are made of sample correlation matrices the factor weights for each attribute are divided by the sample standard deviation for that attribute as indicated in equation (4.34). Thus, there are column wisea and402 row wise constants of proportionality for the factor loadings when analyses are made of correlation matrices. These scaling constants must be considered in the analyses involving multiple samples.

Analyses in several populations present extended problems as compared with analyses in samples from a single population. This topic is relevant to several of the sections in this chapter and especially to the section on factor analyes in several populations. A most important point is that the attribute measures should be expressed in a common scale across the populations. For a small example see Table 12.1 and Table 12.2 which involves measures on five attributes for a sample of first-year medical students and a sample of forth-year medical students. In that differential selection and different experiences have occured for the students in these two samples they may be considered to have been drawn from two populations. For the first three tests the fourth-year students scored lower and had markedly higher standard deviations than did the first-year students. There is a need to pool the data for these two samples to obtain some type of standardization similar to converting each covariance matrix to a correlation matrix. One possibility is to obtain the unweighted mean covariance matrix to obtain pooled variancse from which pooled standard deviations may be obtained. Some individuals would like to weight each covariance matrix by the sample size in obtaining a mean. This would be reasonable if the two samples are considered to be drawn from the same population. However, when there are two

Table 12.1

Multiple Populations Sample Means and Standard Deviations
Medical Students Data*

		First-Year Students (N = 141)		Fourth-Year Students (N = 75)	
	Tests	Mean	SD	Mean	SD
1	Logical Diagrams	10.8	2.68	10.4	2.95
2	Locations	9.9	2.74	8.7	3.43
3	Resourceful Arithmetic	21.5	3.02	19.3	4.30
4	Vocabulary	23.0	5.70	24.5	5.80
5	Word Meaning	16.0	4.79	16.3	4.61

*Data on cognitive process tests in study by Frederiksen, et al, (1980) on "Development of methods for selection and evaluation in undergraduate medical education."

Table 12.2
Multiple Populations Sample covariance matrices
Medical Students Data

Raw Covariance Matrices

First –Year Students, N = 141

	1	2	3	4	5
1	7.2	2.5	2.4	1.3	.8
2	2.5	7.5	1.9	-2.0	-1.3
3	2.4	1.9	9.1	-.5	.3
4	1.3	-2.0	-.5	32.5	20.4
5	.8	-1.3	.3	20.4	22.0

Fourth –Year Students, N = 141

	1	2	3	4	5
1	8.7	3.2	3.8	5.4	3.9
2	3.2	11.8	6.1	3.7	3.0
3	3.8	6.1	18.5	5.3	4.2
4	5.4	3.7	5.3	33.6	20.8
5	3.9	3.0	4.2	20.8	21.2

Mean Covariance Matrices

	1	2	3	4	5
1	8.0	2.9	3.1	3.4	2.4
2	2.9	9.6	4.0	.8	.8
3	3.1	4.0	13.8	2.4	2.2
4	3.4	.8	2.4	33.0	20.6
5	2.4	.8	2.2	20.6	21.6

Scaled Covariance Matrices

First –Year Students, N = 141

	1	2	3	4	5
1	.91	.29	.22	.08	.06
2	.29	.78	.16	-.11	-.09
3	.22	.16	.66	-.02	.02
4	.08	-.11	-.02	.98	.76
5	.06	-.09	.02	.76	1.02

Fourth –Year Students, N = 141

	1	2	3	4	5
1	1.09	.37	.36	.33	.30
2	.37	1.22	.53	.21	.21
3	.36	.53	1.34	.25	.24
4	.33	.21	.25	1.02	.78
5	.30	.21	.24	.78	.98

Mean Covariance Matrices

	1	2	3	4	5
1	1.00	.33	.29	.21	.18
2	.33	1.00	.34	.05	.06
3	.29	.34	1.00	.11	.13
4	.21	.05	.11	1.00	.77
5	.18	06	.13	.77	1.00

different populations there is a question as to what covariance matrix is being estimated. There would be a desirability not to over weight a larger sample so as to overshadow relations appropriate for the population from which the smaller sample was drawn. The reciprocals of the pooled standard deviations are used as row and column multipliers of the sample covariance matrices to obtain the scaled covariance matrics given in Table 12.2 . Note that the diagonals of the mean scaled covariance matrices are unity.

12.1 Factor Extension Techniques

A number of factor analytic studies involve a major battery of attributes followed by some supplementary attributes to which the analysis of the major battery is to be extended. For an example consider a battery of aptitude measures followed by measures of some criteria. There is a desire to extend the analysis of the aptitude battery to the criteria measures so as to find the dependence of the criteria on the aptitude factors. Data for a small example are based on a study conducted by French, et al, (1952) on entrance tests and course grades at the United States Coast Guard Accademy. The correlations given in Table 12.3 are based on notes by Tucker from his participation in this study. There had been extensive selection of the cadets in the class based on entrance tests; however, the aptitude tests used in the illustation were not used in these selections. Never the less, the statistics for the tests in the example may have been influenced by incidental selection, a point to consider during inspection of results of the example. Following the five aptitude measures in the example are course grades in four first term courses. The purpose in our example is to relate the course grades to the factors from analysis of the aptitude measures. A second type of example for factor extension involves a battery of aptitude measures followed by some personality measures. In this type of example the purpose is to investigate the dependence of the personality measures on the aptitude factors. There are a number of other types of examples. The main purpose of factor extension is to relate supplementary measures to factors determined from a major battery of measures.

Chapter 7 provides the basic framework for factor fitting as given in equation (7.16). For the present purposes this equation may be written as:

$$\Delta = (C - U^2) - AA' \qquad (12.1)$$

where Δ is a matrix of residuals. With the two batteries matrix C is partitioned into the four sections illustrated in Table 12.3. Matrix Δ is similarly partitioned. Furthter, there are sections of the factor matrix A : A_1 and A_2 for the two batteries. These partitionings result in the following equations:

$$\Delta_{11} = (C_{11} - U_1^2) - A_1A_2' \quad ; \qquad (12.2.11)$$

$$\Delta_{12} = C_{12} - A_1A_2' \quad ; \qquad (12.2.12)$$

Table 12.3
Illustration of Factor Extension Procedure
Correlation Matrix*

Section C_{11}

	1	2	3	4	5
1	1.00	.59	.31	.22	.14
2	.59	1.00	.16	.10	.03
3	.31	.16	1.00	.47	.39
4	.22	.10	.47	1.00	.47
5	.14	.03	.39	.47	1.00

Section C_{12}

	6	7	8	9
1	.25	.22	.09	.09
2	.38	.33	.11	.10
3	-.05	.06	.32	.35
4	-.10	.06	.31	.34
5	-.05	.02	.35	.41

Section C_{21}

	1	2	3	4	5
6	.25	.38	-.05	-.10	-.05
7	.22	.33	.06	.06	.02
8	.09	.11	.32	.31	.35
9	.09	.10	.35	.34	.41

Section C_{22}

	6	7	8	9
6	1.00	.56	.31	.24
7	.56	1.00	.42	.36
8	.31	.42	1.00	.76
9	.24	.36	.76	1.00

* These correlations are based on the study by French, et al, (1952) on entrance tests and course grades at he United States Coast Guard Academy. The attributes are:

1. American Council on Education Test, Linguistics Section.
2. College Entrance Examination Board, SAT – Verbal.
3. College Entrance Examination Board, Secondary School Physics.
4. American Council on Education Test, Quantitative Section.
5. College Entrance Examination Board, SAT – Mathematics.
6. First Term Grade in History and Literature.
7. First Term Grade in English Composition.
8. First Term Grade in Physics.
9. First Term Grade in Analytic Geometry and Calculus.

$$\Delta_{21} = C_{21} - A_2 A_1' = \Delta_{12}' \qquad (12.2.21)$$

$$\Delta_{22} = (C_{22} - U_2^2) - A_2 A_2' \quad .; \qquad (12.2.22)$$

Diagonal matrices U_1^2 and U_2^2 are the uniqueness for batteries 1 and 2. Note that the off-diagonal sections do not involve the uniqueness in this formal model; however, some cases with real data, this model may have to be adjusted to include the uniqueness of some attributes in batteries 1 and 2. These cases will be discussed subsequaently.

Factor extension involves a least squares solution for the entries in matrix Δ_{21} (equivalently in matrix Δ_{12}) which results, according general least squares solutions, in:

$$A_2 = C_{21} A_1 (A_1' A_1)^{-1} \quad . \qquad (12.3)$$

When this formula is applied to a body of data the results should be inspected for possible violations of the assumptions implied by equations (12.2). Interpretations of these results may be seriously effected by these violations of assumptions. Possible such problematic results will be discussed in terms of the example using the Coast Guard Accademy data.

As previously indicated, the correlation matrix for the Coast Guard Accademy data is given in Table 12.3. This matrix is sectioned into matrices C_{11} , C_{21} , C_{12} , C_{22} . A principal factors solution of C_{11} using the Jöreskog and Goldberger (1972) initial uniqueness (see Chapter 9, equation 9.97) is given as matrix A_1 in Table 12.4 . Extension matrix A_2 was obtained by equation (12.3). The residual matrix with sections given in equation (12.2.11 through (12.2.22) is given in Table 12.5. These matrices are to be inspected for violations of the assumptions of the factor extension model. Consider section Δ_{21} , these residuals appear quite small indicating a good application of the extension model. However, the column of negative residuals for attribute 1 with grades 6 - 9 along with the column of positive residuals for attribute 2 with theses grades and the column of positive residuals for attribute 5 indicates some small general effect which may have arisen from the selective process for this class. Such an effect might be noted in the interpretation of the results. An effect which might have been anticipated did not occur. This effect was a possible high residual between the two measures of physics performance: attribute 3, Secondary School Physics, and attribute 8, Forst Term Grade in Physics. If this residual had been high, the interpretation would be that there was an overlaping specific between the two attributes. Section Δ_{22} has high residuals among the grades attributes which indicates a further factor among these attributes which does not overlap with the aptitude battery. This appears to be the 'grades factor' found in the analysis by French, et al, (1952). A factor extracted from section Δ_{22} would be appended to matrix A_2 with zero loadings for matrix A_1 .

Results of a factor transformation of matrix A_1 by the DAPPFR method are given in

Table 12.4
Illustration of Factor Extension Procedure
Original Factor Matrices

Factor Matrix A_1*

	1	2
1	.61	.45
2	.47	.56
3	.63	-.21
4	.61	-.34
5	.52	-.39

Factor Matrix A_2

	1	2
6	.13	.47
7	.23	.30
8	.42	-.25
9	.46	-.30

*Principal factors of C_{11} using the Joreskog and Goldberger (1972) initial uniqueness (see Chapter 9, equation 9.97).

Table 12.5
Illustration of Factor Extension Procedure
Residual Matrix

Section Δ_{11}

	1	2	3	4	5
1	.42	.05	.02	.00	.00
2	.05	.47	-.02	.00	.00
3	.02	-.02	.56	.01	-.02
4	.00	.00	.01	.51	.02
5	.00	.00	-.02	.02	.58

Section Δ_{12}

	6	7	8	9
1	-.04	-.05	-.05	-.05
2	.06	.06	.05	.06
3	-.03	-.02	.01	.00
4	-.02	.02	-.03	-.04
5	.07	.02	.04	.06

Section Δ_{21}

	1	2	3	4	5
6	-.04	.06	-.03	-.02	.07
7	-.05	.06	-.02	.02	.02
8	-.05	.05	.01	-.03	.04
9	-.05	.06	.00	-.04	.06

Section Δ_{22}

	6	7	8	9
6	.76	.39	.37	.32
7	.39	.86	.40	.35
8	.37	.40	.77	.50
9	.32	.35	.50	.70

factor extracted from section Δ_{22} would be appended to matrix A_2 with zero loadings for matrix A_1.

Results of a factor transformation of matrix A_1 by the DAPPFR method are given in Table 12.6. The structure transformation results are given at the left with the pattern transformations results being given at the right. This transformation is carried on to matrix A_2. Factor 1 appears to be a linguistics factor while factor 2 appears to be a quantitative factor. These two factors carry over nicely from the aptitude battery to the grades battery. There is one surprize: the large negative loading for attribute 5, First Term Grades in History and Literature. The interpretation of this negative loading would be a challenge to the analyst. It will be left as an observation in the present context.

A possible extension of equation (12.3) would results in matrix G_2 directly from matrix C_{21} using matrices G_1 and product matrix FF'. This will not be pursued here; the simpler solution appears to apply the extension procedure to obtain matrix A_2 followed by the factor transformation used in the example. A similar possible extension of equation (12.3) would results in matrix B_2 directly from matrix C_{21} using matrices B_1 and product matrix TT'. Again, this extension will not be pursued here.

12.2 Factor Transformations to an Hypothesized Matrix of Loadings: Procrustes Transformations

The class of situations considered in this section concerns the transformation of an original factor matrix A to approximate hypothetical loadings defined by the analyst. These hypothetical loadings are in a matrix H. The analyst may specify the hypothetical loadings from previous experience such as the analysis of the battery of attributes in a previous study. Alternatively, the analyst may define the hypothetical loadings in terms of some theory. In a sense these operations verge on confirmatory analysis in which the accuracy of the hypothetical loadings is checked. A variety of techniques will be considered in this section. A major division between techniques concerns whether or not the transformed factors are restricted to being uncorrelated. A first subsection will be concerned with oblique factor transformations and a second subsection will consider orthogonal factor transformations.

Following is an outline of general notoation to be employed and conditions to be considered.

A is original, uncorrelated factor matrix with: n attributes ; r factors.

H is matrix of hypothetical factor loadings. There are to be n rows in H ; however, in cases (to be noted) there may be fewer than r columns of H. Let $\underline{h}_m$ be the m'th column of H.

Table 12.6
Illustration of Factor Extension Procedure
Factor Transformation*

	Structure Transformation Normals, F'			Pattern Transformation Traits, T'	
	1	2		1	2
1	.47	.69	1	.72	.88
2	.88	-.72	2	.69	-.47

Transformed Factor Loadings

	Projections on Normals G_1			Factor Weights B_1	
	1	2		1	2
1	.69	.09	1	.73	.10
2	.71	-.08	2	.75	-.09
3	.12	.59	3	.12	.62
4	-.01	.67	4	-.01	.70
5	-.09	.64	5	-.10	.67

	G_2			B_2	
	1	2		1	2
6	.48	-.25	6	.50	-.27
7	.37	-.06	7	.39	-.06
8	-.02	.47	8	-.02	.49
9	-.05	.53	9	-.05	.56

Product Matrices

	FF'			$\Phi = TT'$	
	1	2		1	2
1	1.00	-.31	1	1.00	.31
2	-.31	1.00	2	.31	1.00

* DAPPFR transformation of A_1.

Some or all entries in each $\underline{h}_m$ may be specified by the analyst. Let I_m be the set of attributes, i, for which the entries in $\underline{h}_m$ are specified. A sum over attributes in I_m is symbolized by: $\sum_{i \in I_m}$. A matrix $\tilde{A}_m$ is obtained by deleting from A rows for which the entries in $\underline{h}_m$ are not specified. Let $\tilde{\underline{h}}_m$ be the column of $\underline{h}_m$ with the unspecified entries deleted.

W is a matrix of weights to be defined to minimize a given least squares criterion. There is to be a column, $\underline{w}_m$, for each column of $\underline{h}_m$.

V is a matrix of derived factor loadings with columns $\underline{v}_m$.

$$AW = V \quad ; \tag{12.4M}$$

$$A\underline{w}_m = \underline{v}_m \quad . \tag{12.4V}$$

Entries in V are to be matched with entries in H in a least squares sense according to some chosen criterion.

12.2.1 Oblique Factor Transformations

A first issue for oblique factor transformations concerns the type of factor loadings. There are three cases. First is the case where there is no given restriction on the weights. This is the raw weights case. Second is the case where each weight vector is restricted to unit length so as to be the normal to an hyperplane with the factor loadings being projections on the normal (structure loadings). This is the normals case. Third is the case where each column of V is to be a column of T^{-1} so that the factor loadings are factor weights in matrix B (pattern loadings). This is the trait vectors case.

For the first two cases there may be fewer than r columns in H so that this matrix is incomplete. Any remaining columns of W and V are to be defined separately so as to provide a complete transformation. For the third case of factor weights (pattern loadings) matrix H must be complete in the sense of having r columns.

A second issue is whether the entries in H are to be taken in a fixed sense or are to be taken as known only within an undefined constant of proportionality for each column of H. This idea of undefined proportionality is supported by two matters discussed below.

(1) To permit freedom to the analyst to utilize hypothesized loadings defined within a constant of proportionality and not require the analyst to define these loadings at an exact level.

(2) To take into account the effects of sampling described in Chapter 4 whereby sample factor weights are proportional by columns to population weights.

An important point related to sampling sffects discussed in Chapter 4 is that in several samples the measures for each attribute should be measured on a common scale across the samples. This may be accomplished, instead of using a correlation matrix for each sample, by standardizing the covariance matrix for each sample using a combined standard deviation.

An undefined constant of proportionality for each factor is symbolized as c_m . For each of the weights vector cases there are two hypothetical loadings cases: the fixed hypothetical loadings case and the proportional hypothetical loadings case.

A 3 x 2 type design of matching criteria are considered in this note: three types of loadings by two conditions whether or not proportionality of columns of H are considered. In addition, a canonical type criterion is considered.

Operations for the oblique factor transformations will be illustrated using the Medical Students Data described earlier. The scaled covariance matrices given in Table 12.2 are used to meet the joint standardization requirement described earlier in this section. Table 12.7 gives the principal factor matrices for the two samples. Hypothetical loadings are derived from a transformation of the First Year Students Data; see Table 12.8. While the DAPPFR transformation was not restricted to being orthogonal, this transformation turned out to being so close to orthogonal that it will be considered here as an orthogonal transformation with the projections on the normals (structure loadings) being taken as the hypothetical loadings for all cases to be considered for oblique factor transformations. The factor transformations to these hypothetical loadings will be applied to the Fourth Year Students principal factors given in Table 12.7.

12.2.1.1 Raw Weights, Fixed Hypothetical Loadings

This appears to be the first, workable solution to the transformation to hypotheized loadings. Mosier (1959) proposed this as an approximate solution to his development for Normals, Fixed Hypothesized Loadungs which is discussed in section 12.2.1.3. Hurley and Cattell (1962) used this procedure in their deveopment: "The Procrustes program: producing direct rotation to test a hypothesized factor structure." They adopted the term "Procrustes Transformation" following a Greek mythology concerning a highwayman who made all his victims fit his bed. This term has become adopted to apply to all the various procedures to obtain a transformation to hypothesized factor loadings.

A separate solution is obtained for each existant column of H. The criterion for this case is:

$$\xi_{fm} = \sum_{i \in I_m} (v_{im} - h_{im})^2 \quad . \tag{12.5}$$

which may be written with equations (12.4):

Table 12.7

Principal Factors of Scaled Covariance Matrices*
Medical Students Data

	First Year Students			Fourth Year Students	
	Factor Loadings			Factor Loadings	
	1	2		1	2
1	.74	.569	1	.535	.212
2	-1.27	.485	2	.518	.465
3	-.006	.393	3	.552	.449
4	.844	.004	4	.770	-.397
5	.851	.022	5	.749	-.395

*Principal factors of scaled covariance matrices using Joreskog and Goldberger (1972) initial uniqueness.

Table 12.8
Factor Transformation* of First year medical Students Data

Structure Loadings Solution

Normal Vectors

	1	2
1	-.001	1.000
2	1.000	.029

Projections on Normals (Structure Loadings)

	1	2
1	.568	.090
2	.486	-.113
3	.393	.005
4	-.005	.844
5	.013	.852

Cosines of Angles Between Normals

	1	2
1	1.000	.018
2	.018	1.000

Pattern Loadings Solution

Trait Vectors

	1	2
1	-.029	1.000
2	1.000	.011

Factor Weights (Pattern Loadings)

	1	2
1	.568	.091
2	.486	-.113
3	.393	.005
4	-.005	.844
5	.013	.852

Factor Intercorrelations

	1	2
1	1.000	-.018
2	-.018	1.000

*Dapper transformation(see Chapter 11).

$$\xi_{fm} = \sum_{i \in I_m} \left(\sum_{j=1}^{r} a_{ij} w_{jm} - h_{im} \right)^2 \quad . \tag{12.6}$$

This is an ordinary least squares problem for which the standard solution for the weights vector is:

$$\underline{w}_m = \left(\tilde{A}'_m \tilde{A}_m \right)^{-1} \left(\tilde{A}'_m \underline{\tilde{h}}_m \right) \quad . \tag{12.7}$$

Note that there is an implied restriction that $\left(\tilde{A}'_m \tilde{A}_m \right)$ not be singular so as to possess an inverse. Also, a degenerate solution is obtained when all defined hypothetical loadings for one column of H are zero. In this case the weights vector $\underline{w}_m$ will contain zeros.

This solution is frequently used with an arbitrary additional step of scaling the derived vector of transformation weights, $\underline{w}_m$, either to unit length so as to be a normal to an hyperplane or to being a column of matrix T^{-1}. This rescaling of the vector $\underline{w}_m$ violates the least squares solution of criterion ξ_{fm}. However, this combination of operations may be justified with the canonical type criterion to be described subsequently.

Table 12.9 presents result for the Fourth Year Medical Students Data with the hypothetical loadings being given at the left. The raw weights solution is given in the middle of Table 12.9 with these weights being obtained by equation 12.7. The squared lengths of the raw weights are given in the diagonal of the product matrix. Note that the squared length of the first weights vector is markedly less than unity while the squared length of the second weights vector is larger than unity. The normalized weights solution on the right is obtained by scaling the weights vectors to unit length as shown in the diagonal of the product matrix. The sums of squared differences are given in the bottom line of the table. Note that the normalization procedure has increased these sums of squared differences. An important point is that the solution is no longer orthogonal as shown in the off-diagonal entry in the product matrix. There has been a change in the obliqueness of the structure from the first year students to the fourth year students. This effect will be noted for the orthogonal transformations to follow.

As discussed previously, this criterion is not scale free. Consider that the hypothesized weights were doubled. The raw transformation weights would be doubled as would be the loadings on these raw weights. However, the normalization procedure would compensate for these changes and yield the same solution as before the doubling. While the normalization process is an arbitrary addition to the solution, it does provide a scale free aspect to the revised solution.

12.2.1.2 Raw Weights, Proportional Hypothetical Loadings.

Table 12.9
Solution For: Raw Weights, Fixed Hypothetical Loadings
Fourth Year Medical Students Data

Hypothetical Loadings

Transformatin Results

Raw Weights

Transformation Weights Matrices

	1	2
1	.386	.636
2	.669	-.906

Normalized Weights

Transformation Weights Matrices

	1	2
1	.500	.575
2	.866	-.818

Given Loadings

	1	2
1	.568	.090
2	.486	-.113
3	.393	.005
4	-.005	.844
5	.013	.852

Transformed Loadings

	1	2
1	.349	.148
2	.511	-.092
3	.514	-.055
4	.032	.850
5	.025	.834

Transformed Loadings

	1	2
1	.451	.134
2	.662	-.083
3	.665	-.050
4	.041	.767
5	.033	.754

Product Matrices Of
Transformation Weights Matrices

	1	2
1	.597	-.360
2	-.360	1.225

	1	2
1	1.000	-.421
2	-.421	1.000

Sums of Squared Differences

1	2
.065	.008

1	2
.121	.021

A separate solution could be contemplated for each existant column of H. The criterion for this case is:

$$\xi_{pm} = \sum_{i \in I_m} (v_{im} - h_{im} c_m)^2 \tag{12.8}$$

where c_m is an unkown constant of proportionality to be determined in the solution. The minimum solution is a degenerate case with vector $\underline{w}_m = 0$ and $c_m = 0$. This is an unusable criterion.

12.2.1.3 Normals, Fixed Hypothetical Loadings.

This is the case considered by Mosier (1959) in his article "Determining a simple structure when loadfings for certain tests are known". A separate solution is obtained for each existant column of H. The criterion for this situation is the same as for Raw Weights, Fixed Hypothetical Loaings with a constraint that the weight vector be of unit length. Similar to equation (12.6):

$$\zeta_{fm} = \sum_{i \in I_m} \left(\sum_{j=1}^{r} a_{ij} w_{jm} - h_{im} \right)^2 \tag{12.9}$$

with the constraint that:

$$\sum_{j=1}^{r} w_{jm}^2 = 1 \quad . \tag{12.10}$$

A criterion which combines the functions in equations (12.9) and (12.10) utilizes a LaGrange multiplier, β_m, is:

$$\overset{\#}{\zeta}_{fm} = \sum_{i \in I_m} \left(\sum_{j=1}^{r} a_{ij} w_{jm} - h_{im} \right)^2 - \beta_m \left(\sum_{j=1}^{r} w_{jm}^2 - 1 \right) \quad . \tag{12.11}$$

This equation may be expressed in matrix form as:

$$\begin{aligned} \overset{\#}{\zeta}_{fm} &= \underline{w}_m' \tilde{A}_m' \tilde{A}_m \underline{w}_m - 2\underline{w}_m \tilde{A}_m' \tilde{h}_m + \tilde{h}_m' \tilde{h}_m - \underline{w}_m' \beta_m \underline{w}_m + \beta_m \quad , \\ &= \underline{w}_m' \left(\tilde{A}_m' \tilde{A}_m - \beta_m I \right) \underline{w}_m - 2\underline{w}_m' \tilde{A}_m' \tilde{h}_m + \tilde{h}_m' \tilde{h}_m + \beta_m \quad . \end{aligned} \tag{12.12}$$

To minimize function $\overset{\#}{\zeta}_{fm}$, the partial derivatives are found with respect to the elements, w_{jm}, of the weights vector and are set equal to zero.

$$\frac{\partial \overset{\#}{\zeta}_{fm}}{\partial w_{jm}} = 2 \sum_{i \in I_m} \left(\sum_{k=1}^{r} a_{ik} w_{km} - h_{im} \right) a_{ij} - 2\beta_m w_{jm} = 0 \quad . \tag{12.13}$$

This equation may be written in matrix form for all entries in $\underline{w}_m$ as:

$$2\tilde{A}_m' \tilde{A}_m \underline{w}_m - 2\tilde{A}_m' \tilde{h}_m - 2\beta_m \underline{w}_m = 0 \quad . \tag{12.14}$$

With algebraic operations this equation may be written as:

$$\left(\tilde{A}_m' \tilde{A}_m - \beta_m I \right) \underline{w}_m = \overset{\#'}{A}_m \tilde{h}_m \quad . \tag{12.15}$$

Conveient definitions follow. Let:

$$X_m = \tilde{A}'_m \tilde{A}_m \quad ; \tag{12.16}$$

$$\underline{y}_m = \tilde{A}'_m \underline{\tilde{h}}_m \quad . \tag{12.17}$$

Then:

$$(X_m - \beta_m I)\, \underline{w}_m = \underline{y}_m \quad . \tag{12.18}$$

The solution for the restricted criterion $\overset{\#}{\zeta}_{fm}$ is:

$$\underline{w}_m = (X_m - \beta_m I)^{-1} \underline{y}_m \quad . \tag{12.19}$$

The Lgrange multiplier, β_m, is to be determined to statisfy the restriction in equation (12.10).

In the solution for β_m, a continuous function q_m is defined by:

$$q_m = \underline{w}'_m \underline{w}_m$$

which with equation (12.19) becomes:

$$q_m = \underline{y}'_m (X_m - \beta_m I)^{-2} \underline{y}_m \quad . \tag{12.20}$$

Thus, q_m is a function of β_m. Solution of this function for the values of β_m for which q_m equals unity as per equation (12.10) is facilitated by a transformation defined from the eigensolution for matrix X_m given in the following equation:

$$X_m = V_m \Lambda_m V'_m \tag{12.21}$$

where Λ_m is a diagonal matrix of eigenvalues and V_m is a square, orthonormal matrix of eigenvectors. Note that:

$$(X_m - \beta_m I) = V_m (\Lambda_m - \beta_m I) V'_m \quad ; \tag{12.22}$$

$$(X_m - \beta_m I)^{-2} = V_m (\Lambda_m - \beta_m I)^{-2} V'_m \quad . \tag{12.23}$$

Equation (12.20) may be written as:

$$q_m = \underline{y}'_m V_m (\Lambda_m - \beta_m I)^{-2} V'_m \underline{y}_m \quad . \tag{12.24}$$

A notational simplification involves defining $\underline{z}_m$ by:

$$\underline{z}_m = V'_m \underline{y}_m \quad . \tag{12.25}$$

Then:

$$q_m = \underline{z}'_m (\Lambda_m - \beta_m I)^{-2} \underline{z}_m \quad . \tag{12.26}$$

With $(\Lambda_m - \beta_m I)$ being a diagonal matrix, equation (12.26) may be converted to sumational notation:

$$q_m = \sum_{j=1}^{r} z_{jm}^2 \Big/ (\lambda_{jm} - \beta_m)^2 \quad . \tag{12.27}$$

Note that q_m is the sum of r terms each of which must be non-negative due to the squares of the numerator and the denominator. Thus, q_m must be non-negative. Further inspection of this equation indicates that q_m will approach zero when β_m approaches either $+\infty$ or $-\infty$. Another point is that term j in the sum will approach $+\infty$ when β_m approaches λ_{jm} for $j = 1$

to r . At these points, q_m approaches $+\infty$. Further properties of this function is provided by the derivative of q_m with respect to β_m :

$$\frac{d\, q_m}{d\, \beta_m} = 2\sum_{j=1}^{r} z_{jm}^2 \Big/ (\lambda_{jm} - \beta_m)^3 \quad . \tag{12.28}$$

When β_m is less than λ_{rm} (λ_{rm} being the least eigenvalue of X_m) this derivative is positive so that the function in the range of β_m from $-\infty$ to λ_{rm} starts equal to zero and increases to ∞ as β_m approaches λ_{rm} . Consequently, there is a solution for q_m equal to unity in this range. There is another solution in the range of β_m from λ_{1m} to ∞ . Also, there are possible solutions in the spaces of β_m between consecutive eigenvalues. Study of these possible solutions leads to rejecting them for the desired minimization of the criterion $\overset{\#}{\zeta}_{fm}$. The desired solution is for β_m in the range from $-\infty$ to λ_{rm} . An iterative solution in this range is described in the following paragraph.

For the iterative procedure, equation (12.27) is written as:

$$q_m = z_{rm}^2 \Big/ (\lambda_{rm} - \beta_m)^2 + \sum_{j=1}^{r-1} z_{jm}^2 \Big/ (\lambda_{jm} - \beta_m)^2 \tag{12.29}$$

in which the final term of the sum is pulled out as a separate term. This equation is somewhat similar to an hyperbolic function which may be written as:

$$y = \frac{a}{(b-x)^2} + c \tag{12.30}$$

in which q_m is replaced by y , z_{rm}^2 is replaced by a, λ_{rm} is replaced by b , and β_m is replaced by x . The final term of equation (12.29) is considered, approximately, to be a constant c . The range of x is considered to be from $-\infty$ to λ_{rm} . Note that as $x \to -\infty$, $y \to c$ so that there is a left hand asymptote of c . The derivative of y with respect to x is:

$$\frac{dy}{dx} = y' = \frac{2a}{(b-x)^3} \quad . \tag{12.31}$$

The plan of the iterative procedure is to start from some trial value of x , termed x_t , then to define the parameters for the function in equation (12.30) for this value of x_t . Given the function of equation (12.30), a solution for x_1 is made for $y_1 = 1$. The obtained x_1 becomes the next trial value of x_t . Due to the similarity of equation (12.30) to the function for q_m in equation (12.29) this procedure should converge to the desired value of β_m . Parameter b is set equal to λ_{rm} . Trial values of parameters a and c are designated by a_t and c_t which are developed from solution of equations (12.30) and (12.31) for trial values y_t and y_t' derived from a trial value of $\beta_m = x_t$. y_t is the value of q_m obtained from equation (12.27) with the trial value of β_m and y_t' , the trial value of the derivative in equation (12.28) for the trial value of β_m . Then:

$$a_t = \tfrac{1}{2}\, y_t' (b - x_t)^3 \quad ; \tag{12.32}$$

$$c_t = y_t - \frac{a_t}{(b - x_t)^2} \quad . \tag{12.33}$$

The solution for x_1 when $y_1 = 1$ involves a solution from equation (12.30):

$$x_1 = b - \sqrt{\frac{a_t}{(1 - c_t)}} \quad . \tag{12.34}$$

In the range of $\beta_m = x_t$ from $-\infty$ to λ_{rm} the derivative in equation (12.28) must be positive and the term $(b-x_t)^3$ of equation (12.32) must be positive so that a_t given in equation (12.32) must be positive. It is imperative that $(1-c_t)$ in equation (12.34) be positive. A computer protection when $(1-c_t)$ is negative is to set it equal to a small positive value. A further condition in the program is to make sure that the obtained value of y_1 is nearer to 1 than the input trial value, y_t. When this is not so, a smaller change in x_t should be tried.

Table 12.10 presents the illustration of this transformation for the Fourth Year Medical Students Data. These results are similar to those presented in Table 12.9 for the Raw Weights, Fixed Hypothetical Loadings case with the sums of squared differences between those for the raw weights and the normalized weights given in Table 12.9. The squared lengths of the weight vectors given in the product matrix are unity corresponding to the constraint given in equation (12.10).

Mosier (1959) suggested an approximate solution obtained by setting β_m equal to zero and then normalizing the resulting weight vectors. This is identical to the procedure given earlier in section 12.2.1.1 for the Raw Weights, Fixed Hypothetical Loadings case.

For the special case when all defined hypothetical loadings for a factor equal zero the solution degenerates to the least square hyperplane fitting described in Chapter 10. In this case vector $\underline{\tilde{h}}_m$ equals zero as does vector $\underline{y}_m$ as defined in equation (12.17). Then equation (12.18) becomes the eigen problem of equation (10.26) with β_m being the least eigenvalue of X_m and $\underline{w}_m$ being the corresponding eigenvector.

12.2.1.4 Normals, Proportional Hypothetical Loadings.

This is a companion case to the preceding case with, now, the hypothetical loadings being subject to a possible constant of proportionality for each transformed factor. A separate solution is obtained for each existant column of H. The major criterion for this case is:

$$\zeta_{pm} = \sum_{i \in I_m} \left(\sum_{j=1}^{r} a_{ij} w_{jm} - h_{im} c_m \right)^2 \tag{12.35}$$

where c_m is a constant of proportionality. The constraint that vector $\underline{w}_m$ is of unit length is given in equation (12.10). Note the similarity of this criterion with the criterion for the preceding case in equations (12.9) and (12.10). The added feature is the constant of proportionality for the hypothetical loadings. A criterion which combines the functions in equation (12.35) and (12.10) utilizes a LaGrange multiplier, β_m, is:

Table 12.10
Solution For: Normals, Fixed hypothetical Loadings*
Fourth Year Medical Students Data

Hypothetical Loadings

Given Loadings

	1	2
1	.568	.090
2	.486	-.113
3	.393	.005
4	-.005	.844
5	.013	.852

Transformatin Results

Transformation Weights
(Normal Vectors)

	1	2
1	.429	.604
2	.903	-.797

Projections on Normals
(Structure Loadings)

	1	2
1	.421	.154
2	.642	-.057
3	.643	-.024
4	-.028	.782
5	-.035	.767

Cosines of Angles
Between Normals

	1	2
1	1.000	-.460
2	-.460	1.000

Sums of Squared Differences

1	2
.111	.019

*Mosier's (1939) exact solution.

$$\#\zeta_{pm} = \sum_{i\in I_m}\left(\sum_{j=1}^{r} a_{ij}w_{jm} - h_{im}c_m\right)^2 - \beta_m\left(\sum_{j=1}^{r} w_{jm}^2 - 1\right) \quad ; \tag{12.36}$$

$$= \sum_{i\in I_m}\left(\sum_{j=1}^{r} a_{ij}w_{jm}\right)^2 - 2\sum_{i\in I_m}\left(\sum_{j=1}^{r} a_{ij}w_{jm}\right) h_{im}c_m + \sum_{i\in I_m} h_{im}^2 c_m^2 - \beta_m\left(\sum_{j=1}^{r} w_{jm}^2 - 1\right) \quad . \tag{12.37}$$

This equation may be expressed in matrix form as:

$$\overset{\#}{\zeta}_{pm} = \underline{w}_m'\tilde{A}_m'\tilde{A}_m\underline{w}_m - 2c_m\underline{\tilde{h}}_m'\tilde{A}_m\underline{w}_m + c_m^2\underline{\tilde{h}}_m'\underline{\tilde{h}}_m - \beta_m(\underline{w}_m'\underline{w}_m - 1) \tag{12.38}$$

Minimization of $\overset{\#}{\zeta}_{pm}$ is accomplished using the derivative with respect to c_m. This derivative is set equal to zero.

$$\frac{d\overset{\#}{\zeta}_{pm}}{dc_m} = -2\sum_{i\in I_m}\left(\sum_{i=1}^{r} a_{ij}w_{jm}\right) h_{im} + 2c_m\sum_{i\in I_m} h_{im}^2 = 0 \quad . \tag{12.39}$$

Then:

$$c_m = \frac{1}{\sum_{i\in I_m} h_{im}^2}\sum_{i\in I_m}\left(\sum_{j=1}^{r} a_{ij}w_{jm}\right) h_{im} \quad . \tag{12.40}$$

In matrix form:

$$c_m = \frac{1}{\underline{\tilde{h}}_m'\underline{\tilde{h}}_m}\underline{\tilde{h}}_m'\tilde{A}_m\underline{w}_m \quad . \tag{12.41}$$

With the nature of equations (12.40) and (12.41) a restriction on the use of this criterion is that the sum of squares of the hypothesized loadings for each factor must be greater than zero. The next step is to obtain the partial derivative of $\overset{\#}{\zeta}_{pm}$ with respect to element w_{jm} of the weight vector. This partial derivative is set equal to zero.

$$\frac{\partial\tilde{\zeta}_{pm}}{\partial w_{jm}} = 2\sum_{i\in I_m}\left(\sum_{k=1}^{r} a_{ik}w_{km}\right) a_{ij} - 2\sum_{i\in I_m} a_{ij}h_{im}c_m - 2\beta_m w_{jm} = 0 \quad . \tag{12.42}$$

Let $\underline{\tilde{a}}_{jm}$ be column vector of the j'th column of $\tilde{A}_m$. Then equation (12.42) becomes, in matrix form:

$$\underline{\tilde{a}}_{jm}'\tilde{A}_m\underline{w}_m - c_m\underline{\tilde{a}}_{jm}'\underline{\tilde{h}}_m - \beta_m w_{jm} = 0 \quad . \tag{12.43}$$

For all values of j :

$$\tilde{A}_m'\tilde{A}_m\underline{w}_m - c_m\tilde{A}_m'\underline{\tilde{h}}_m - \beta_m\underline{w}_m = 0 \quad . \tag{12.44}$$

Substitution from equation (12.41) into equation (12.44) yields:

$$\tilde{A}_m'\tilde{A}_m\underline{w}_m - \frac{1}{\underline{\tilde{h}}_m'\underline{\tilde{h}}_m}\tilde{A}_m'\underline{\tilde{h}}_m\underline{\tilde{h}}_m'\tilde{A}_m\underline{w}_m - \beta_m\underline{w}_m = 0 \quad . \tag{12.45}$$

Or:

$$\left(\tilde{A}_m'\tilde{A}_m - \frac{1}{\underline{\tilde{h}}_m'\underline{\tilde{h}}_m}\tilde{A}_m'\underline{\tilde{h}}_m\underline{\tilde{h}}_m'\tilde{A}_m - \beta_m I\right)\underline{w}_m = 0 \quad . \tag{12.46}$$

A useful definition is:

$$X_m = \tilde{A}_m' \tilde{A}_m - \frac{1}{\tilde{h}_m' \tilde{h}_m} \tilde{A}_m' \tilde{h}_m \tilde{h}_m' A_m \quad . \tag{12.47}$$

Equation (12.46) becomes:

$$(X_m - \beta_m I)\, \underline{w}_m = 0 \quad . \tag{12.48}$$

This equation is in the form of an eigen problem with β_m being an eigenvalue of matrix X_m and $\underline{w}_m$ being the corresponding eigenvector. To satisfy the constraint of equation (12.10) $\underline{w}_m$ is to be scaled to unit length. There is, now, the problem of selecting the desired eigenvalue to minimize $\overset{\#}{\zeta}_{pm}$. Substitute the value of c_m from equation (12.41) into equation (12.38) to obtain:

$$\begin{aligned} \overset{\#}{\zeta}_{pm} &= \underline{w}_m' \tilde{A}_m' \tilde{A}_m \underline{w}_m - 2 \frac{1}{\tilde{h}_m' \tilde{h}_m} \underline{w}_m' \tilde{A}_m' \tilde{h}_m \tilde{h}_m' \tilde{A}_m \underline{w}_m \\ &\quad + \frac{1}{\left(\tilde{h}_m' \tilde{h}_m\right)^2} \underline{w}_m' \tilde{A}_m' \tilde{h}_m' \tilde{h}_m \tilde{A}_m \underline{w}_m \tilde{h}_m' \tilde{h}_m + \beta_m (\underline{w}_m' \underline{w}_m - 1) \\ &= \underline{w}_m' \left(\tilde{A}_m' \tilde{A}_m - \frac{1}{\tilde{h}_m' \tilde{h}_m} \tilde{A}_m' \tilde{h}_m \tilde{h}_m' A_m \right) \underline{w}_m + \beta_m (\underline{w}_m' \underline{w}_m - 1) \quad . \end{aligned} \tag{12.49}$$

With the definition of X_m in equation (12.47) and with $\underline{w}_m$ being a unit length vector and the relation between ζ_{pm} and $\overset{\#}{\zeta}_{pm}$ from equations (12.35) and (12.36) equation (12.49) yields:

$$\zeta_{pm} = \overset{\#}{\zeta}_{pm} = \underline{w}_m' X_m \underline{w}_m \quad . \tag{12.50}$$

From the properties of the eigen solution:

$$\zeta_{pm} = \beta_m \quad . \tag{12.51}$$

Then, a minimum value of ζ_{pm} is obtained by using the least eigenvalue of X_m and the corresponding unit length eigenvector as the desired weight vector $\underline{w}_m$. This least eigenvalue is the value of the criterion. Note that the eigenvector may be reversed in sign and still be of unit length. A sugestion is to choose the direction of the eigenvector so that the sum of the transformed loadings will be positive.

Results for the solution for this criterion are given in Table 12.11. Adjustment of the hypothetical loadings is accomplished by the constants of proportionality given at the top of the middle section of the table. Note that the first factor has a constant greater than unity so that the hypothetical loadings are increased while the constant for the second factor is less than unity so that the hypothetical loadings are decreased in absolute value. The transformation results are very similar to results obtained by preceding criteria. As should be expected, the inclusion of the constants of proportionality has reduced the sums of squared differences from those in Table 12.10 for Normals, Fixed Hypothetical Loadings.

12.2.1.5 Trait Vectors, Fixed Hypothetical Loadings.

Table 12.11
Solution For: Noramls, Proportional Hypothetical Loadings
Fourth Year Medical Students Data

Hypothetical Loadings

Transformation Results

Raw Weights

Normalized Weights

Transformation Weights Matrices

Constants of Proportionality

	1	2
	1.155	.898

Transformation Weights (Normal Vectors)

	1	2
1	.469	.573
2	.883	-.820

Given Loadings

	1	2
1	.568	.090
2	.486	-.113
3	.393	.005
4	-.005	.844
5	.013	.852

Adjusted Loadings

	1	2
1	.656	.081
2	.561	-.101
3	.454	.004
4	-.006	.758
5	.015	.765

Projections on Normals (Structure Loadings)

	1	2
1	.438	.133
2	.654	-.084
3	.655	-.052
4	.010	.767
5	.002	.753

Cosines of Angles Between Normals

	1	2
1	1.000	-.455
2	-.455	1.000

Sums of Squared Differences

1	2
.097	.006

With this function there is a switch from minimizing differences between projections on normals (structure loadings) and the hypothetical loadings to differences between factor weights (pattern loadings) and the hypothetical loadings. For the present case the major criterion is:

$$\Psi_f = \sum_{m=1}^{r} \left\{ \sum_{i \in I_m} (b_{im} - h_{im})^2 \right\} \tag{12.52}$$

where:

$$B = AT^{-1} \quad ; \tag{7.4}$$

$$R_{bb} = TT' = \Phi \tag{12.53}$$

Rows of matrix T contain the trait vectors as described in Chapter 7 with the trnasfomred factor weights (pattern loadings) being in matrix B. All columns of the hypothetical loadings in matrix H must be defined. However, not all entries in each column of H have to be specified. Matrix R_{bb} of correlations among the transformed factors is defined in equation (7.1) and is renamed here for conveience as Φ. Since this is a correlation matrix the diagonal entries must be unity so that:

$$\text{Diag}(\Phi) = I \quad . \tag{12.54}$$

Criterion Ψ_f is to be minimized under the constraint given in equation (12.54).

Gruvaeus (1970) described a procedure to accomplish the minimization using a series of Fletcher & Powell (1963) minimizing transformations. Browne (1972) described an alternative method to obtain the minimum solution using a series of elemental transformations. In each of these elemental transformations a single trait is considered as a pivot trait while another trait is considered as a spoke trait. The pivot trait is transformed in the direction of the spoke trait. A subcycle consists of transformations of the pivot trait in the direction of each of the other traits as the spoke trait. A cycle of elemental transformations consists of considering each of the traits in turn as the pivot trait. These elemental transformations are continued until every elemental transformation in a cycle is less than some prescribed minimum transformation. Following is a description of a single elemental transformation.

Let the pivot trait be deignated j and the spoke trait be designated k. Then the criterion Ψ_f can be written as:

$$\Psi_f = \sum_{i \in I_j} (b_{ij} - h_{ij})^2 + \sum_{i \in I_k} (b_{ik} - h_{ik})^2 + \sum_{\substack{m=1 \\ m \neq j,k}}^{r} \left\{ \sum_{i \in I_m} (b_{im} - h_{im})^2 \right\} \tag{12.55}$$

Let $\underline{t}_j$ and $\underline{t}_k$ be trait vectors j and k which are rows j and k of the trait matrix T. Also, let $\underline{\tilde{t}}_j$ be the transformed trait vector j. The transformation is given below

$$s_{jj}\underline{t}_j + s_{jk}\underline{t}_k = \underline{\tilde{t}}_j \tag{12.56}$$

where s_{jj} and s_{jk} are the transformation coefficients. This transformation is similarr to the transformation in the direct oblimin procedure described in Chapter 11, see equation (11.59). The restriction that $\tilde{\underline{t}}_j$ be a unit vector leads to the following restriction on the transformation coefficients.

$$s_{jj}^2 + 2s_{jj}s_{jk}\phi_{jk} + s_{jk}^2 = 1 \qquad (12.57)$$

where ϕ_{jk} is the correlation between traits j and k in matrix Φ. These coefficients can be recorded in row j and columns j and k of a 'shift' matrix S which has unities in all other diagonal cells and zeros in all other off-diagonal cells. This transformation is illustrated in expanded form for a 4 factor case with $j = 1$ and $k = 2$.

$$\begin{bmatrix} s_{11} & s_{12} & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix} \begin{bmatrix} \underline{t}_1 \\ \underline{t}_2 \\ \underline{t}_3 \\ \underline{t}_4 \end{bmatrix} = \begin{bmatrix} \tilde{\underline{t}}_1 \\ \underline{t}_2 \\ \underline{t}_3 \\ \underline{t}_4 \end{bmatrix}$$

In matrix equation form:

$$ST = \tilde{T} \quad . \qquad (12.58)$$

The transformed factor weights matrix, $\tilde{B}$, is given by:

$$\tilde{B} = A\tilde{T}^{-1}$$

which with equations (7.4) and (12.58) becomes

$$\tilde{B} = AT^{-1}S^{-1} = B\,S^{-1} \quad . \qquad (12.59)$$

Matrices S and S^{-1} are illustrated in expanded form for the 4 factor situation used previously.

$$\begin{bmatrix} s_{11} & s_{12} & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{s_{11}} & \frac{-s_{12}}{s_{11}} & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}$$

Columns of matrices B and $\tilde{B}$ are treated as column vectors such as $\underline{b}_j$ and $\tilde{\underline{b}}_j$. Then, equation (12.59) may be illustrated as follows.

$$[\underline{b}_1 \quad \underline{b}_2 \quad \underline{b}_3 \quad \underline{b}_4] \begin{bmatrix} \frac{1}{s_{11}} & \frac{-s_{12}}{s_{11}} & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix} = [\tilde{\underline{b}}_1 \quad \tilde{\underline{b}}_2 \quad \underline{b}_3 \quad \underline{b}_4]$$

Note in the illustration that factor weights are transformed for both factor 1 and 2. In the general case:

$$\underline{b}_j \frac{1}{s_{jj}} = \tilde{\underline{b}}_j \quad ; \qquad (12.60)$$

$$\underline{b}_j \frac{-s_{jk}}{s_{jj}} + \underline{b}_k = \tilde{\underline{b}}_k \quad . \qquad (12.61)$$

The restriction that the transformed trait vector $\tilde{\underline{t}}_j$ be of unit length resulted in the restriction of equation (12.57). A transformation of variables is used to implement this restriction. Define:

$$z_{jk} = \frac{s_{jk}}{s_{jj}} \quad . \qquad (12.62)$$

Solutions for s_{jj} and s_{jk} follow using equation (12.57).

$$\frac{1}{s_{jj}^2} = 1 + 2\phi_{jk}z_{jk} + z_{jk}^2 \quad ; \tag{12.63}$$

$$\frac{1}{s_{jj}} = \sqrt{1 + 2\phi_{jk}z_{jk} + z_{jk}^2} \quad . \tag{12.64}$$

In order to maintain the orientation of trait j the positive square root is used.

$$s_{jk} = s_{jj}z_{jk} \quad . \tag{12.65}$$

Elements of the transformed factor weights follow.

$$\tilde{b}_{ij} = b_{ij}\sqrt{1 + 2\phi_{jk}z_{jk} + z_{jk}^2} \quad . \tag{12.66}$$

$$\tilde{b}_{ik} = b_{ik} - b_{ij}z_{jk} \quad . \tag{12.67}$$

The function to be minimized for this transformation is writen from equation (12.55). (Since the last term is not affected by the transformation of trait j in terms of trait k this term is dropped.)

$$\tilde{\psi}_{fjk} = \sum_{i\in I_j}\left(\tilde{b}_{ij} - h_{ij}\right)^2 + \sum_{i\in I_k}\left(\tilde{b}_{ik} - h_{ik}\right)^2 \quad . \tag{12.68}$$

Expansion of the squared terms with substitution from equations (12.66) and (12.67) with algebraic operations leads to:

$$\tilde{\psi}_{fjk} = a_2 z_{jk}^2 + a_1 z_{jk} + a_* \Big/ s_{jj} + a_0 \tag{12.69}$$

where:

$$a_2 = \sum_{i\in I_j} b_{ij}^2 + \sum_{i\in I_k} b_{ij}^2 \quad ; \tag{12.70.2}$$

$$a_1 = 2\left(\sum_{i\in I_j} b_{ij}^2\phi_{jk} - \sum_{i\in I_k} b_{ij}b_{ik} + \sum_{i\in I_k} b_{ij}h_{ik}\right) \quad ; \tag{12.70.1}$$

$$a_* = -2\sum_{i\in I_j} b_{ij}h_{ij} \quad ; \tag{12.70.*}$$

$$a_0 = \sum_{i\in I_j} b_{ij}^2 + \sum_{i\in I_k} b_{ik}^2 - 2\sum_{i\in I_k} b_{ik}h_{ik} + \sum_{i\in I_j} h_{ij}^2 + \sum_{i\in I_k} h_{ik}^2 \quad . \tag{12.70.0}$$

Since coefficient a_2 involves only sums of squares of trial factor weights it, almost always, will be positive. Only in an extremely rare case will a_2 be zero and never negative. Consequently, a_2 will be taken as positive. Coefficient a_1 may be either positive or negative. Coefficient a_* involves the sum of products between the trial factor weights and hypothetical loadings. The sum of squared differences between the trial factor weights and the hypothetical loadings to be small, the sum of products between the trial factor weights and the hypothetical loadings should be positive. If this is not the case, the trial trait vector j should be reversed in sign. This reversal in sign will result in changes in signs of the entries in row j and column j of matrix Φ. With the changes in both the row and the column, the diagonalt entry ϕ_{jj} will remain a

positive unity. Also, the entries in column j of T^{-1} will change. Consequently, coefficient a_* will be considered as being negative.

An important special case occurs when a_* equals zero, then equation (12.69) reduces to a parabola with an optimum at:

$$z_{jk} = -a_1 \Big/ 2a_2 \quad \text{when } a_* = 0 \quad . \tag{12.71}$$

This case is most likely to occur when all specified hypothetical loadings are zero in which case the present procedure becomes the complement of the least square hyperplane fitting procedure discussed in Chapter 10. The least squares is applied to the factor weights rather than to the projections on the normals. This is a transformation discussed by Lawley and Maxwell (1964) in their article on "Factor Transformation Methods." This solution may be termed "Least Squares Zero Pattern Fitting" (LSQZPF).

Solution of equation (12.69) for z_{jk} to yield a minimum $\widetilde{\psi}_{fjk}$ is moderately complex. Except for the case when a_* equals zero there may be two minima so that there is a problem of selecting the most desirable minimum. Browne (1972) suggests using Bailey's (see McCalla, 1967, p. 90) modification of the Newton-Raphson iterative procedure. For this method the first three derivatives of $\widetilde{\psi}_{fjk}$ with respect to z_{fjk} are needed. Note: s_{jj} is determined by equation (12.64) using the positive square root. Let:

$$\psi' = \frac{d\widetilde{\psi}_{fjk}}{dz_{jk}} = 2a_2 z_{jk} + a_1 + a_*(\phi_{jk} + z_{jk})\, s_{jj} \quad ; \tag{12.72}$$

$$\psi'' = \frac{d\psi'}{dz_{jk}} = 2a_2 + a_*\left(1 - \phi_{jk}^2\right) s_{jj}^3 \quad ; \tag{12.73}$$

$$\psi''' = \frac{d\psi''}{dz_{jk}} = -3a_*\left(1 - \phi_{jk}^2\right)(\phi_{jk} + z_{jk})\, s_{jj}^5 \quad . \tag{12.74}$$

Consider trial t given z_{jkt} . The value of s_{jj} is determined by equation (12.64). Derivatives ψ'_t, ψ''_t and ψ'''_t are determined from equations (12.72), (12.73) and (12.74) using the value of z_{jkt} . The resulting trial z is $z_{jk(t+1)}$ biven by:

$$z_{jk(t+1)} = z_{jkt} - \frac{\psi'_t}{\left\{\psi''_t - \psi'_t\psi'''_t \Big/ 2\psi''_t\right\}} \quad . \tag{12.75}$$

An inspection of the equations yields profitable relations. From equation (12.64). the coefficient s_{jj} approaches 0 as z_{jk} approaches either $-\infty$ or $+\infty$ so that at these extremes the criterion $\widetilde{\psi}_{|fjk}$ approaches zero as seen form equation (12.69). However, as indicated in equation (12.68), this criterion must be positive between these two extremes. For a minimum, the second derivative, ψ'' must be positive. A minimum value of this derivative is found by setting the third derivative, ψ''' , equal to zero which occurs when $z_{jk} = -\phi_{jk}$. Thus, the second derivative is positive at all times when the value of ψ'' is positive at this minimum. Then there is only one minimum for the function $\widetilde{\psi}_{fjk}$. While there may be no worry that there is a maximum of $\widetilde{\psi}_{fjk}$

between two minima, the program should check the algebraic sign of ψ'' at each trial and if it is negative to take a gradient type step toward a minimum.

Solution for a minimum value of criterion Ψ_f starts from some desirable first approximation and makes improvements for each pair of traits as pivot trait and spoke trait until no improvement occurs. A good first approximation may be to obtain trial normals by the simple procedure for raw weight vectors with fixed hypothetical loadings. In case all specified hypothetical loadings for a factor are zero, the normal can be computed by the LSQHYP procedure described in Chapter 10. This trial normals matrix may be converted to a trials trait matrix by the standard formula:

$$T = D\,(F')^{-1} \tag{10.3}$$

where:

$$D = \left[\text{Diag}(FF')^{-1}\right]^{-1/2} \quad . \tag{10.5}$$

Table 12.12 presents an illustrationof the results for the Trait Vectors, Fixed Hypothetical Loadings transformation for the Fourth Year Medical Students Data. Hypothetical loadings given on the left are from the DAPPFR solution for the First Year Medical Students Data given in Table 12.8. While the correlation between factors 1 and 2 was almost zero for the First Year Medical Students, this correlation given in Table 12.12 is .360 for the Fourth Year Medical Students. Thus, while the factors were almost orthogonal for the First Year Medical Students, the factors are fairly correlated for the Fourth Year Medical Students. This is a result to be considered for the orthogonal transformation procedures.

12.2.1.6 Trait Vectors, Proportional Hypothetical Loadings.

This function is very similar to the previous function with the only change being that the hypothetical loadings are multiplied by a constant of proportionality for each factor.

$$\Psi_p = \sum_{m=1}^{r} \left\{ \sum_{i \in I_m} (b_{im} - c_m h_{im})^2 \right\} \tag{12.76}$$

where c_m is a constant of proportionality for each factor applied to the hypothetical loadings for that factor. The discussion following equation (12.52) applies to the present function with the b_{im} being entries in the matrix B and there being a restriction on the trait vectors given in equation (12.54). Again, each trait vector in turn will be considered a s a pivot trait. For each pivot trait, each of the other trait vectors will be considered in turn as a spoke trait. For each pair of a pivot trait and a spoke trait a transformation will be accomplished to a minimum of function Ψ_p. Computations will cycle through the combinations of pivot trait and spioke trait until only

Table 12.12
Solution For: Trait Vectors, Fixed Hypothetical Loadings
Fourth Year Medical Students Data

Hypothetical Loadings

Given Loadings

	1	2
1	.568	.090
2	.486	-.113
3	.393	.005
4	-.005	.844
5	.013	.852

Transformation Results

Trait Vectors

	1	2
1	.469	.573
2	.883	-.820

Factor Loadings (Pattern Loadings)

	1	2
1	.475	.196
2	.704	-.023
3	.707	.013
4	.024	.857
5	.015	.841

Factor Intercorrelations

	1	2
1	1.000	.360
2	.360	1.000

Sums of Squared Differences

1	2
.155	.020

Criterion ψ_f = .175

minimal changes occur. Convergence is guaranteed since the function is positive, or at most zero, in nature.

Let the pivot trait be designated j and the spoke trait be designated k. Then the criterion Ψ_p can be written as:

$$\Psi_p = \sum_{i\in I_j}(b_{ij} - c_j h_{ij})^2 + \sum_{i\in I_k}(b_{ik} - c_k h_{ik})^2 + \sum_{\substack{m=1 \\ m\neq j,k}}^{r} \left\{ \sum_{i\in I_m}(b_{im} - c_m h_{im})^2 \right\} \tag{12.77}$$

The transformation of trait vector $\underline{t}_j$ with respect to trait vector $\underline{t}_k$ to the transformed trait vector $\underline{\tilde{t}}_j$ involving transformation coefficients s_{jj} and s_{jk} is given in equation (12.56). The constraint on these coefficients is such that the transformed trait vector $\underline{\tilde{t}}_j$ is of unit length is given in equation (12.57). Transformation of the factor weights $\underline{b}_j$ and $\underline{b}_k$ to transformed factor weights $\underline{\tilde{b}}_j$ and $\underline{\tilde{b}}_k$ is given in equations (12.60) and (12.61). A derived coefficient z_{jk} is defined in equation (12.62) with the relations of s_{jj} and s_{jk} to z_{jk} being given in equations (12.63), (12.64), and (12.65). The transformed factor weights are given in terms of z_{jk} are given in equations (12.66) and (12.67). As in the preceding solution, the function to be minimized for the present transformation is written from equation (12.77). (Since the last term of equation (12.77) is not affected by the transformation of trait j in terms of trait k, this term is dropped.)

$$\tilde{\psi}_{pjk} = \sum_{i\in I_j}\left(\tilde{b}_{ij} - c_j h_{ij}\right)^2 + \sum_{i\in I_k}\left(\tilde{b}_{ik} - c_k k_{ik}\right)^2 \quad . \tag{12.78}$$

Substitutions from equations (12.66) and (12.67) with algebraic operations leads to:

$$\begin{aligned}\tilde{\psi}_{pjk} = {} & \sum_{i\in I_j} b_{ij}^2 + 2\phi_{jk} z_{jk} \sum_{i\in I_j} b_{ij}^2 + z_{jk}^2 \sum_{i\in I_j} b_{ij}^2 - 2\, c_j \sqrt{1 + 2\phi_{jk} z_{jk} + z_{jk}^2} \sum_{i\in I_j} b_{ij} h_{ij} + c_j^2 \sum_{i\in I_j} h_{ij}^2 \\ & + \sum_{i\in I_k} b_{ik}^2 - 2 z_{jk} \sum_{i\in I_k} b_{ij} b_{ik} + z_{jk}^2 \sum_{i\in I_k} b_{ij}^2 - 2 c_k \sum_{i\in I_k} b_{ik} h_{ik} + 2 c_k z_{jk} \sum_{i\in I_k} b_{ij} h_{ik} + c_k^2 \sum_{i\in I_k} h_{ik}^2 \quad .\end{aligned} \tag{12.79}$$

A solution for the constants of proportionality in terms of z_{jk} is obtained by setting the derivatives of $\tilde{\psi}_{pjk}$ with respect to these constants of proportionality equal to zero.

$$\frac{d\tilde{\psi}_{pjk}}{dc_j} = -2\sqrt{1 + 2\phi_{jk} z_{jk} + z_{jk}^2} \sum_{i\in I_j} b_{ij} h_{ij} + 2 c_j \sum_{i\in I_j} h_{ij}^2 = 0 \quad . \tag{12.80}$$

$$\frac{d\tilde{\psi}_{pjk}}{dc_k} = 2\left(z_{jk} \sum_{i\in I_k} b_{ij} h_{ik} - \sum_{i\in I_k} b_{ik} h_{ik} \right) + 2 c_k \sum_{i\in I_k} h_{ik}^2 = 0 \quad . \tag{12.81}$$

These equations lead to:

$$c_j = \sqrt{1 + 2\phi_{jk} z_{jk} + z_{jk}^2} \sum_{i\in I_j} b_{ij} h_{ij} \Bigg/ \sum_{i\in I_j} h_{ij}^2 \quad ; \tag{12.82}$$

$$c_k = \left(\sum_{i\in I_k} b_{ik} h_{ik} - z_{jk} \sum_{i\in I_k} b_{ij} h_{ik} \right) \Bigg/ \sum_{i\in I_k} h_{ik}^2 \quad . \tag{12.83}$$

Substitution of these results into equation (12.79) with algebraic operations leads to;

$$\widetilde{\psi}_{pjk} = a_2 z_{jk}^2 + a_1 z_{jk} + a_0 \tag{12.84}$$

where:

$$a_2 = \sum_{i\in I_j} b_{ij}^2 + \sum_{i\in I_k} b_{ij}^2 - \frac{\left(\sum_{i\in I_j} b_{ij}h_{ij}\right)^2}{\sum_{i\in I_j} h_{ij}^2} - \frac{\left(\sum_{i\in I_k} b_{ij}h_{ik}\right)^2}{\sum_{i\in I_k} h_{ik}^2} \quad ; \tag{12.85.2}$$

$$a_1 = 2\left\{\phi_{jk}\sum_{i\in I_j} b_{ij}^2 - \sum_{i\in I_k} b_{ij}b_{ik} - \phi_{jk}\frac{\left(\sum_{i\in I_j} b_{ij}h_{ij}\right)^2}{\sum_{i\in I_j} h_{ij}^2} + \frac{\left(\sum_{i\in I_k} b_{ik}h_{ik}\right)\left(\sum_{i\in I_k} b_{ij}h_{ik}\right)}{\sum_{i\in I_k} h_{ik}^2}\right\}; \tag{12.85.1}$$

$$a_0 = \sum_{i\in I_j} b_{ij}^2 + \sum_{i\in I_k} b_{ik}^2 - \frac{\left(\sum_{i\in I_j} b_{ij}h_{ij}\right)^2}{\sum_{i\in I_j} h_{ij}^2} - \frac{\left(\sum_{i\in I_k} b_{ik}h_{ik}\right)^2}{\sum_{i\in I_k} h_{ik}^2} \quad . \tag{12.85.0}$$

With equation (12.84) being that of a parabola the minimum solution for $\widetilde{\psi}_{pjk}$ occurs at:

$$z_{jk} = -a_1 \Big/ 2a_2 \quad . \tag{12.86}$$

Note that the sum of squared hypothetical loadings for each factor must be greater than zero; that is:

$$\sum_{i\in I_m} > 0 \quad \text{for } m = 1, r\,. \tag{12.87}$$

An illustration of the results for the trait vectors, proportional hypothetical loadings transformation is given in Table 12.13. The given hypothetical loadings are the same as used in Table 12.12 from transformation for First Year Medical Students in Table 12.8.. The constants of proportionality obtained in the solution are given at the top of the center section of the Table 12.13 with the adjusted hypothetical loadings being given in the middle of this center section. Note that the hypothetical loadings for the first factor are increased by the constant of proportionality of 1.278 while the hypothetical loadings for the second factor remain practically unchanged. The transformation results are quite similar to those in Table 12.12 for Trait Vectors, Fixed Hypothetical Loadings with the two factors having a correlation of .395. Use of the constants of proportionality has reduced materially the sums of squared differences from Table 12.12 to Table 12.13. Use of the constants of proportionality appears to have a very desirable effect.

12.2.1.7 Canonical Factor Matching.

Table 12.13
Solution For: Trait Vector, Proportional Hypothetical Loadings
Fourth Year Medical Students Data

Hypothetical Loadings

Given Loadings

	1	2
1	.568	.090
2	.486	-.113
3	.393	.005
4	-.005	.844
5	.013	.852

Transformation Results

Constants of Proportionality

1	2
1.278	.992

Adjusted Loadings

	1	2
1	.726	.089
2	.621	-.112
3	.502	.005
4	-.006	.838
5	.017	.846

Trait Vectors

	1	2
1	.799	.867
2	.601	-.498

Factor Weights (Pattern Loadings)

	1	2
1	.490	.165
2	.719	-.066
3	.723	-.029
4	.043	.849
5	.033	.833

Factor Intercorrelations

	1	2
1	1.000	.394
2	.394	1.000

Sums of Squared Differences

1	2
.117	.009

Criterion $\psi_p = .126$

The canonical factor matching function is a normed least squares function with no restraints on the transformation weights nor the constant of proportionality applied to the hypothetical loadings. All six of the preceding matching functions can be classified as raw least squares functions with various constraints on the transformation weights and constants of proportionality. The canonical function is scale free so that the results may be scaled to satisfy various demands. The transformation weights may be scaled to unit vectors so as to be normals to hyperplanes or theses weights may be scaled to be columns of matrix T^{-1} to yield columns of factor weight matrix B. Equation (12.88) gives the canonical factor matching function.

$$\theta_m = \frac{\sum_{i \in I_m} \left(\sum_{j=1}^{r} a_{ij} w_{jm} - h_{im} c_m \right)^2}{\sum_{i \in I_m} \left(\sum_{j=1}^{r} a_{ij} w_{jm} + h_{im} c_m \right)^2} \tag{12.88}$$

The numerator is the same as the raw weights, proportional hypothetical loadings criterion which was an unacceptable criterion; however, the norming by the denominator yields a very acceptable criterion with desirable results. This function follows the development by Tucker (1951) in his report on the synthesis of factor analytic studies in which he used this function in a more complex form as an index of congruence. As will be seen subsequently, this index of congruence is closely related to the coefficient of congruence which Tucker introduced in his report. A solution is made for each factor individually so that an overall coefficient could be considered by summing over the individual values of θ_m when there is a column of hypothetical loadings for every transformed factor.

The matter of scale freeness is considered first using a scaling constant s_m which is any undefined, nonzero constant. Let:

$$\overset{\#}{w}_{jn} = w_{jm} s_m \quad ; \tag{12.89}$$

$$\overset{\#}{c}_m = c_m s_m \quad . \tag{12.90}$$

Also:

$$\overset{\#}{\theta}_m = \frac{\sum_{i \in I_m} \left(\sum_{j=1}^{r} a_{ij} \overset{\#}{w}_{jm} - h_{im} \overset{\#}{c}_m \right)^2}{\sum_{i \in I_m} \left(\sum_{j=1}^{r} a_{ij} \overset{\#}{w}_{jm} + h_{im} \overset{\#}{c}_m \right)^2} \quad . \tag{12.91}$$

With substitution from equations (12.89) and (12.90) a result is obtained that:

$$\overset{\#}{\theta}_m = \theta_m \quad . \tag{12.92}$$

Thus, the function is independent of the scaling defined in equation (12.89) and (12.90).

With the scale free property of the canonical function, this function is not completely identified. A convenient possibility is to define the original constant c_m equal to unity so that the function becomes:

$$\theta_m = \frac{\sum_{i\in I_m}\left(\sum_{j=1}^{r} a_{ij} w_{jm} - h_{im}\right)^2}{\sum_{i\in I_m}\left(\sum_{j=1}^{r} a_{ij} w_{jm} + h_{im}\right)^2} \tag{12.93}$$

The numerator now is the same as the raw weights, fixed hypothetical case. To obtain a minimum solution the partial derivatives of θ_m with respest to the w_{jm}'s are set equal to zero.

$$\frac{\partial \theta_m}{\partial w_{jm}} = \frac{\left\{\left[\sum_{i\in I_m}\left(\sum_{k=1}^{r} a_{ik} w_{km} + h_{im}\right)^2\right]\left[2\sum_{i\in I_m}\left(\sum_{k=1}^{r} a_{ik} w_{km} - h_{im}\right) a_{ij}\right] - \left[\sum_{i\in I_m}\left(\sum_{k=1}^{r} a_{ik} w_{km} - h_{im}\right)^2\right]\left[2\sum_{i\in I_m}\left(\sum_{k=1}^{r} a_{ik} w_{km} + h_{im}\right) a_{ij}\right]\right\}}{\left[\sum_{i\in I_m}\left(\sum_{k=1}^{r} a_{ik} w_{km} + h_{im}\right)^2\right]^2} \tag{12.94}$$

Equating these derivatives to zero accompanied with algebraic operations yields:

$$\left\{\begin{array}{cc} \left[\sum_{i\in I_m}\left(\sum_{k=1}^{r} a_{ik} w_{km}\right)^2 + \sum_{i\in I_m} h_{im}^2\right] & \left[\sum_{i\in I_m} a_{ij} h_{im}\right] \\ -2\left[\sum_{i\in I_m} h_{im}\sum_{k=1}^{r} a_{ik} w_{km}\right] & \left[\sum_{i\in I_m} a_{ij}\sum_{k=1}^{r} a_{ik} w_{km}\right] \end{array}\right\} = 0 \quad . \tag{12.95}$$

Then:

$$\sum_{i\in I_m} a_{ij}\sum_{k=1}^{r} a_{ik} w_{km} = \frac{\left[\sum_{i\in I_m}\left(\sum_{k=1}^{r} a_{ik} w_{km}\right)^2 + \sum_{i\in I_m} h_{im}^2\right]}{2\left[\sum_{i\in I_m} h_{im}\sum_{k=1}^{r} a_{ik} w_{km}\right]} \sum_{i\in I_m} a_{ij} h_{im} \quad . \tag{12.96}$$

Since the fraction on the left of equation (12.96) is a ratio of scalars, this fraction can be replaced by a coefficient defined in terms of:

$$g_m = \frac{2\left[\sum_{i\in I_m} h_{im}\sum_{k=1}^{r} a_{ik} w_{km}\right]}{\left[\sum_{i\in I_m}\left(\sum_{k=1}^{r} a_{ik} w_{km}\right)^2 + \sum_{i\in I_m} h_{im}^2\right]} \quad . \tag{12.97}$$

Then:

$$\sum_{i\in I_m} a_{ij}\sum_{k=1}^{r} a_{ik} w_{km} = \frac{1}{g_m}\sum_{i\in I_m} a_{ij} h_{im} \tag{12.98}$$

which may be written in matrix form as:

$$\tilde{A}_m' \tilde{A}_m \underline{w}_m = \frac{1}{g_m}\tilde{A}_m' \underline{\tilde{h}}_m \quad . \tag{12.99}$$

The solution for $\underline{w}_m$ is given by:

$$\underline{w}_m = \frac{1}{g_m}\left(\tilde{A}_m' \tilde{A}_m\right)^{-1}\left(\tilde{A}_m' \underline{\tilde{h}}_m\right) \quad . \tag{12.100}$$

Note that these transformation weights have the raw weights for the raw weights fixed hypothetical loadings case divided by the coefficient g_m. When the weights are scaled such as to unit length this matter of proportionality is of no consequence. Thus, the present approach justifies the scaling process which was considered to be an arbitrary addition for the raw weights,

fixed hypothetical loadings case. The common practice has been justified. Further, since the factor weights in matrix B are proportional by columns to the projections on the normals, the scaling of the weight vectors, $\underline{w}_m$, to columns in matrix T^{-1} is proper.

There are several very interesting relations to be discussed. Following are several useful relations involving conversion to matrix notation and substitution for the value of $\underline{w}_m$ from equation (12.101).

$$\sum_{i\in I_m} h_{im} \sum_{k=1}^{r} a_{ik} w_{km} = \underline{\tilde{h}}_m' \tilde{A}_m \underline{w}_m = \frac{1}{g_m} \left(\underline{\tilde{h}}_m' \tilde{A}_m\right)\left(\tilde{A}_m' \tilde{A}_m\right)^{-1}\left(\tilde{A}_m' \underline{\tilde{h}}_m\right) \quad . \tag{12.101}$$

$$\sum_{i\in I_m} \left(\sum_{k=1}^{r} a_{ik} w_{km}\right)^2 = \sum_{i\in I_m} \left(\sum_{k=1}^{r} w_{km}\right)\left(\sum_{l=1}^{r} a_{il} w_{lm}\right) = \underline{w}_m' \tilde{A}_m' \tilde{A}_m \underline{w}_m \quad ;$$

$$= \frac{1}{g_m^2} \left(\underline{\tilde{h}}_m' \tilde{A}_m\right)\left(\tilde{A}_m' \tilde{A}_m\right)^{-1}\left(\tilde{A}_m' \underline{\tilde{h}}_m\right) \quad . \tag{12.102}$$

$$\sum_{i\in I_m} h_{im} \sum_{k=1}^{r} a_{ik} w_{im} = g_m \sum_{i\in I_m} \left(\sum_{k=1}^{r} a_{ik} w_{im}\right)^2 \quad . \tag{12.103}$$

An important relation is derived by substituting from equation (12.103) into equation (12.97).

$$g_m = \frac{2g_m \left[\sum_{i\in I_m} \left(\sum_{k=1}^{r} a_{ik} w_{km}\right)^2\right]}{\left[\sum_{i\in I_m} \left(\sum_{k=1}^{r} a_{ik} w_{km}\right)^2 + \sum_{i\in I_m} h_{im}^2\right]} \quad .$$

$$\sum_{i\in I_m} \left(\sum_{k=1}^{r} a_{ik} w_{km}\right)^2 + \sum_{i\in I_m} h_{im}^2 = 2\sum_{i\in I_m} \left(\sum_{k=1}^{r} a_{ik} w_{km}\right)^2 \quad .$$

$$\sum_{i\in I_m} h_{im}^2 = \sum_{i\in I_m} \left(\sum_{k=1}^{r} a_{ik} w_{km}\right)^2 \quad . \tag{12.104}$$

This relation states that the sum of squared hypothetical loadings equals the sum of squared raw loadings. This relation is scale free when the scaling relations in equations (12.89) and (12.90) are considered. Remeber that c_m was set equal to zero. With the application of the scaling operations:

$$\sum_{i\in I_m} \left(\overset{\#}{c}_m h_{im}\right)^2 = \sum_{i\in I_m} \left(\sum_{k=1}^{r} a_{ik} \overset{\#}{w}_{km}\right)^2 \quad . \tag{12.105}$$

This equation verifies the scale free nature of the solution. With the result in equation (12.104) the coefficient g_m as given in equation (12.97) may be written as:

$$g_m = \frac{\sum_{i\in I_m} h_{im} \sum_{k=1}^{r} a_{ik} w_{km}}{\sum_{i\in I_m} \left(\sum_{k=1}^{r} a_{ik} w_{km}\right)^2} = \frac{\sum_{i\in I_m} h_{im} \sum_{k=1}^{r} a_{ik} w_{km}}{\sum_{i\in I_m} h_{im}^2} \quad . \tag{12.106}$$

An important alternative is:

$$g_m = \frac{\sum_{i\in I_m} h_{im} \sum_{k=1}^{r} a_{ik} w_{km}}{\sqrt{\sum_{i\in I_m} \left(\sum_{k=1}^{r} a_{ik} w_{km}\right)^2}\sqrt{\sum_{i\in I_m} h_{im}^2}} \quad . \tag{12.107}$$

With equation (12.4.V):

$$g_m = \frac{\sum_{i \in I_m} h_{im} v_{im}}{\sqrt{\sum_{i \in I_m} v_{im}^2} \sqrt{\sum_{i \in I_m} h_{im}^2}} \quad . \tag{12.108}$$

This is the coefficient of congruence, defined by Tucker (1951), between the transformed factor loadings and the hypothetical loadings. Burt (1948) termed this coefficient as the "unadjusted correlation" between two sets of factor coefficients.

There is an important relation between the index of congruence, θ_m, and the coefficient of congruence, g_m. Expanding equation (12.93):

$$\theta_m = \frac{\sum_{i \in I_m} \left(\sum_{j=1}^{r} a_{ij} w_{jm} \right)^2 - 2 \sum_{i \in I_m} h_{im} \sum_{j=1}^{r} a_{ij} w_{jm} + \sum_{i \in I_m} h_{im}^2}{\sum_{i \in I_m} \left(\sum_{j=1}^{r} a_{ij} w_{jm} \right)^2 + 2 \sum_{i \in I_m} h_{im} \sum_{j=1}^{r} a_{ij} w_{jm} + \sum_{i \in I_m} h_{im}^2} \quad ;$$

which with equation (12.106) and algebraic operations becomes:

$$\theta_m = \frac{1 - g_m}{1 + g_m} \quad . \tag{12.109}$$

Thus, θ_m and g_m are obversely related in that as g_m increases θ_m decreases so that a maximum value of g_m corresponds to a minimum value of θ_m and vis-a-vers. Korth and Tucker (1976) presented solution for "Procrustes Matching by Congruence Coefficients' in which the coefficient of congrunce was maximized. This is a completely identical solution to the present one.

This approach is termed "Canonical factor matching" due to the considerable analogy to regular canonical analysis in which two batteries of measured attributes are related by weighted composites of the two batteries. The analogy for the present factor matching procedure is to the special case when the second battery is constituted by a single measured attribute. In regular canonical analysis a sample of entities are measured on the attributes in the two batteries. For factor matching the entities are the attributes in the battery being analysed while the measured attributes are the original factors. A major difference is that in regular canonical analysis the measures are adjusted for the mean measures over the sample of entities while in factor matching there is no adjustment for mean values. This shows up in the coefficient of congruence. In the index of congruence given in equation (12.88) the weights for the first battery, the original factor matrix, are the transformation weights while the weight for the second battery, the hypothetical loadings, is the constant of proportionality. For the special case where there is only one measured attribute in the second battery the canonical weights for the first battery are proportional to the multiple regression weights. In the present case the transformation weights are proportional to the least squares solution weights, as given for the Raw weights, fixed hypothetical loadings case.

An aid in the interpretation of the strength of matching is given by a transformation of the coefficient of congruence to an F statistic for random hypothetical loadings. This development involves a series of transformations. Subscripts and special designations are droppeed for the sake of simplicity. Matrix A is taken to involve only rows for which hypothetical loadings are specified. Also, vector h includes only specified entries. From equations (12.102) and (12.104);

$$g^2 = \frac{(h'A)(A'A)^{-1}(A'h)}{h'h} \quad . \tag{12.110}$$

A singular value representation of matrix A follows:

$$A = [P_1 \quad P_2] \begin{bmatrix} \Delta_1 \\ 0 \end{bmatrix} Q' = P\Delta Q' = P_1\Delta_1 Q' \tag{12.111}$$

Note that P is an n x n orthonormal matrix with sections P_1, n x r, and P_2, n x (n−r), with Δ_1 being an r x r diagonal matrix. Matrix Q is an r x r orthonormal matrix. Usual operations have:

$$A'A = Q\Delta_1^2 Q' \tag{12.112}$$

which is eigen solutione for matrix A′A. Matrix P_1 is obtained by:

$$AQ\Delta_1^{-1} = P_1 \tag{12.113}$$

with matrix P_2 being set such that:

$$P_1'P_2 = 0 \quad ; \tag{12.114.1}$$

$$P_2'P_2 = I \quad . \tag{12.114.2}$$

With the definition of P_1 in equation (12.113) it can be shown that:

$$P_1'P_1 = I \tag{12.115}$$

so that matrix P is orthonormal as indicated following equation (12.111). With the eigen solution of equation (12.112):

$$(A'A)^{-1} = Q\Delta_1^{-2}Q'$$

so that, with algebraic operations:

$$A(A'A)^{-1}A' = P_1P_1' \quad . \tag{12.116}$$

The hypothetical loadings vector may be transformed to a vector $\underline{x}$ by:

$$\underline{x} = P'\underline{h}$$

or:

$$\begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \end{bmatrix} = \begin{bmatrix} P_1' \\ P_2' \end{bmatrix} \begin{bmatrix} \underline{h} \end{bmatrix} \quad . \tag{12.117}$$

With this definition and equation (12.116):

$$\underline{h}'A(A'A)^{-1}A'\underline{h} = \underline{x}_1'\underline{x}_1 \tag{12.118}$$

and

$$\underline{h}'\underline{h} = \underline{x}'\underline{x} = \underline{x}_1'\underline{x}_1 + \underline{x}_2'\underline{x}_2 \quad . \tag{12.119}$$

For the present purposes, vector $\underline{h}$ may be considered to be a random vector from a multidimensional normal density function with mean vector equal to 0 and covariance matrix equal to an identity matrix. With this definition, vector $\underline{h}$ is a vector in a random direction in the n dimensional space. The transformation of equation (12.117) expresses this vector in terms of another set of axes. Use of the multidimensional density function the individual entries in vector $\underline{h}$ are independently distributed random normal deviates. After the transformation of equation (12.117) the same is true of the entries in vector $\underline{x}$ so that sums of squares of these entries are distributed as chi squares with degrees of freedom equal to the number of entries entered into the sums. Let:

$$S_1 = \underline{x}_1'\underline{x} = \sum_{i=1}^{r} x_i^2 \qquad (12.120.1)$$

$$S_2 = \underline{x}_2'\underline{x}_2 = \sum_{i=r+1}^{n} x_i^2 \qquad (12.120.2)$$

From equation (12.110) and equations (12.118) throuhg (12.120.2):

$$g^2 = \frac{S_1}{S_1 + S_2} \quad . \qquad (12.121)$$

Algebraic operations yield:

$$\frac{S_1}{S_2} = \frac{g^2}{1 - g^2} \qquad (12.122)$$

As seen in equations (12.120.1) and (12.120.2) S_1 and S_2 are sums of independent normal deviates and are, thus, chi squares with r and (n−r) degrees of freedom. From this an F statistic may be defined as:

$$F_{r,n-r} = \frac{(n-r)}{r} \frac{g^2}{1 - g^2} \quad . \qquad (12.123)$$

This implies a distribution generated by random vectors $\underline{x}$ and may be used to compare observed coefficients of congruence with such random coefficients. Note that the n is the number of specified hypothetical loadings for a transformed factor.

The preceding F statistic depends on the assumption of the $\underline{h}$ vector being drawn from a multidimensional normal density function. This assumption may not be appropriate and some other multidimensional density function could be used. However, the F statistic may not be accurate, even as an approximation. An alternative for some defined statistical function is the use of Monte Carlo type procedures. One possibility is to run a number of replications using randomly reordered observed $\underline{h}$ vectors. A distribution of coefficients of congruence would be obtained for comparison with the observed coefficient.

Results for an illustrative solution for the canonical transformation system are given in Tables 12.14 and 12.15. In Table 12.14 the hypothetical loadings are the same as those used previously. On the rightof Table 12.14 are the raw transformation results. Of interest are the indices of congruence given toward the bottom of Table 12.14. These appear to be quite small

Table 12.14
Canonical Factor Matching: Raw Results
Fourth Year Medical Students Data

Hypothetical Loadings

Normal Vectors

Given Loadings

	1	2
1	.568	.090
2	.486	-.113
3	.393	.005
4	-.005	.844
5	.013	.852

Transformation Results

Trait Vectors

	1	2
1	.405	.638
2	.702	-.908

Raw Loadings

	1	2
1	.366	.149
2	.536	-.092
3	.539	-.056
4	.034	.852
5	.026	.837

Coefficients Obtained

	1	2
Index of Congruence	.024	.001
Coefficient of Congruence	.953	.997
$F_{2,3}$	14.8	249
p	.027	.0005

Table 12.15
Canonical Factor Matching: Scaled Results
Fourth Year Medical Students Data

Hypothetical Loadings

Normal Vectors

	1	2
1	.500	.575
2	.866	-.818

Transformation Results

Trait Vectors

	1	2
1	.818	.866
2	.575	-.500

Projections on Normals
(Structure Loadings)

	1	2
1	.451	.134
2	.662	-.083
3	.665	-.050
4	.041	.767
5	.033	.754

Factor Loadings
(Pattern Loadings)

	1	2
1	.497	.148
2	.730	-.091
3	.733	-.055
4	.046	.846
5	.036	.831

Cosines of Angles
Between Normls

	1	2
1	1.000	-.421
2	-.421	1.000

Factor
Intercorrelations

	1	2
1	1.000	.421
2	.421	1.000

which could indicate a good fit of the obtained raw loadings to the hypothetical loadings. Also of interest are the coefficients of congruence which look quite high. However, these coefficients might be compared with the results from the F distribution of equation (12.123). The values of F along with the corresponding p's (proportions above the value of F in the F distribution) are given at the bottom of the table. The F for factor 1 is significant at the 5% level while the F for the second factor is significant well beyond .1% level. The match for factor 1 is not very strong while the match for factor 2 is extremely strong. An interesting alternative is to solve equation (12.123) for the value of the coefficient of congruence for a given level of significance. The value of $F_{2,3}$ for the 1% level of confidence is 30.82 which converts to a g of .977 . This appears to be a quite high value and indicates that the factor matching is very biased to high values. Randon results appear much too good. There should be considerable care taken in interpreting the outcome of factor matching operations. One might conjecture that this observation might apply to all of the methods of factor transformations to hypothetical loadings.

As indicated in preceding discussions the canonical transformation results may be scaled so that the raw weights become normals to hyperplanes or become columns in matrix T^{-1}. The loadings become either projections on the normals (structure loadings) or factor weights (pattern loadings). These scalings involve only constants of proportionality for the various columns of loadings and do not change the coefficients of congruence. The canonical transformation procedure is scale free. Table 12.15 presents the results for the Fourth Year Medical Students data. These are legitimate results due to the scale free property of the canonical transformation approach. Note thatthe structure solution given on the left is identical to the normalized results for the Raw Weights, Fixed Hypothetical Loadings solution given in Table 12.9. Thus, this approach makes legitimate the frequently used Raw Weights, Fixed Hypothetical Loadings procedure.

12.2.2 Orthogonal Factor Transformations

A number of analysts prefer orthogonal transformations to uncorrelated factors due to the mathematical simplicity of the resulting system. Further some analysts argue for statistically independent factors. With uncorrelated factors there is no distinction between normals and trait vectors since these are identical. Also, the structure loadings and pattern loadings are identical. In this case, the matrix Φ is an identity matrix so that:

$$\Phi = TT' = I \quad . \tag{12.124}$$

Then:

$$T' = T^{-1} \tag{12.125}$$

and

$$B = AT^{-1} = AT' \quad . \tag{12.126}$$

In summation form:

$$\phi_{km} = \sum_{j=1}^{r} t_{kj} t_{mj} \tag{12.127}$$

and

$$b_{im} = \sum_{j=1}^{r} a_{ij} t_{mj} \quad . \tag{12.128}$$

Two cases are to be considered: first when the hypothetical loadings are considered to be fixed as given by the analyst and second when each column of hypothetical loadings may be multiplied by a constant of proportionality.

12.2.2.1 Orthogonal Trait Vectors, Fixed Hypothetical Loadings

These transformations started with Green's (1952) development of orthogonal approximations of oblique structures in factor analysis which was followed by Cliff (1966) and Schönemann (1966) who applied Green's results to the approximation of given hypothetical loadings. These procedures applied to fully specified hypothetical loadings matrices. Browne (1972) developed a procedure for orthogonal rotation to a partially specified target (hypothetical loadings matrix). Schönemann's development has several interesting features and will be discussed first to be followed by the discussion of Browne's development.

For the case for a fully specified hypothetical loadings matrix, H, nxr, has all entries in each column specified; thus the, summations over entities in column j may be expressed as a simple sum. The criterion to be minimized is closely related to the criterion for trait vectors, fixed hypothetical loadings and is designated $\tilde{\Psi}_f$ which is a revision of the criterion in equation (12.52).

$$\tilde{\Psi}_f = \sum_{m=1}^{r} \left\{ \sum_{i=1}^{n} (b_{im} - h_{im})^2) \right\} \quad . \tag{12.129}$$

This minimization is to be constrained by the condition of equation (12.124) which is implemented by defining a function $\tilde{\Psi}_\lambda$:

$$\tilde{\Psi}_\lambda = \sum_{k=1}^{r} \sum_{m=1}^{r} \lambda_{km} (\phi_{km} - \delta_{km}) \tag{12.130}$$

where the λ_{km}'s are Lagrange multipliers and the δ_{km}'s are Kronecker deltas with:

$$\delta_{kk} = 1 \; ; \; \delta_{km} = 0 \quad \text{for } k \neq m \quad . \tag{12.131}$$

The combined criterion is:

$$\tilde{\Psi}_t = \tilde{\Psi}_f + \tilde{\Psi}_\lambda \quad . \tag{12.132}$$

To obtain a minimum of $\tilde{\Psi}_t$ the partial derivative of $\tilde{\Psi}_t$ with respect to element t_{gj} of matrix T is set equal to zero.

$$\frac{\partial \tilde{\Psi}_t}{\partial t_{gj}} = 2\sum_{i=1}^{n}(b_{ig} - h_{ig})a_{ij} + \sum_{m=1}^{r}\lambda_{gm}(t_{mj}) + \sum_{k=1}^{r}\lambda_{kg}(t_{kj}) = 0 \quad . \tag{12.133}$$

With substitution for b_{jg} from equation (12.128), this equation may be written in matrix form for all entries in T as:

$$TA'A - H'A + \tfrac{1}{2}(\Lambda + \Lambda')T = 0 \quad . \tag{12.134}$$

Postmultiply by T', note from equation (12.125 that $TT' = I$, and define:

$$P = A'A = P' \quad ; \tag{12.135}$$

$$S = A'H \quad ; \tag{12.136}$$

$$Q = \tfrac{1}{2}(\Lambda + \Lambda') = Q' \quad . \tag{12.137}$$

Then, equation (12.134) may be written as:

$$TPT' + Q = S'T' = TS \quad . \tag{12.138}$$

The last part of this equation is written since both P and Q are symetric. The solution is facilitated using an Eckart and Young decomposition of matrix S (see Johnsom(1963) for the form used here)

$$S = WDV' \tag{12.139}$$

where W and V are r x r orthonormal marices and D is a diagonal matrix. This decomposition is also termed a singular value solution. Then:

$$TS = TWDV'; \tag{12.140}$$

$$S'T' = VDW'T'$$

With equation (12.138):

$$TWDV' = VDW'T'$$

and

$$WDV'T = T'VDW' \quad .$$

An equivalence is obtained by setting:

$$T'V = W$$

which results in:

$$T' = WV' \quad . \tag{12.141}$$

An interesting comparison is between equations (12.139) and (12.141). Diagonal matrix D is eliminated from equation (12.139) and matrix S becomes matrix T'.

To complete the solution for T an eigen solution is obtained for the product matrix $S'S$ which yields:

$$S'S = VD^2V' \quad . \tag{12.142}$$

With equation (12.139):

$$W = SVD^{-1} \quad . \tag{12.143}$$

A similar solution may be obtained using the product matrix $S\,S'$. There remains a problem as to the algebraic sign to be used in going from D^2 to D and D^{-1}. From equation (12.129):

$$\widetilde{\Psi}_f = \sum_{m=1}^{r} \left\{ \sum_{i=1}^{n} b_{im}^2 - 2 \sum_{i=1}^{n} b_{im} h_{im} + \sum_{i=1}^{n} h_{im}^2 \right\} \tag{12.144}$$

In matrix form:

$$\widetilde{\Psi}_f = tr(B'B) - 2tr(B'H) + tr(H'H) \quad . \tag{12.145}$$

With the orthogonal transformation:

$$tr(B'B) = tr(A'A) \quad . \tag{12.146}$$

With equation (12.126), (12.136), (12.140), and (12.141):

$$tr(B'H) = tr(TA'H) = tr(TS) = tr(TWDV') = tr(VW'WDV')$$

from which:

$$tr(B'H) = tr(D) \quad . \tag{12.147}$$

Then, with equations (12.146) and (12.147), the value of the criterion from equation (12.145) becomes:

$$\widetilde{\Psi}_f = tr(A'A) + tr(H'H) - 2tr(D) \quad . \tag{12.148}$$

With $tr(A'A)$ being the sum of squares of the entries in matrix A and $tr(H'H)$ being the sum of squares in matrix H, both of which are given and positive, criterion $\widetilde{\Psi}_f$ is made as small as possible by making $tr(D)$ as large as possible. This is accomplished by choosing the positive square roots of D^2 from equation (12.142).

While the preceding development was for a completely specified hypothetical loadings matrix, Brown (1972) considered the problem of orthogonal rotation to a partially specified target in which not every element in each column of hypothetical loadings needed to be specified. As in the section for Trait Vectors, Fixed Hypothetical Loadings, the set of specified elements in column j of hypothetical loadings will be designated by I_j. The overall criterion considered here is the same as equation (12.52) with the constraint that the trait vectors be orthogonal.

$$\Psi_f = \sum_{m=1}^{r} \left\{ \sum_{i \in I_m} (b_{im} - h_{im})^2 \right\} \quad . \tag{12.52}$$

Instead of a general solution, Browne presented a procedure involving a series of two dimensional rotations for pairs of factors. There was a series of cycles with each cycle involving rotation for every pair of factors. However, before sarting these rotations, Browne presented a system for ordering the factors in matrix A and of reflection of these factors to a most advantageous situation. In the present discussion, the matter of interchanging factors in each pair and the reflection of these factors will be considered before a rotation for the factors in the pair

instead of the general reordering of the initial factors. The following material will concentrate on the treatment for a given pair of factors, j and k.

The treatment of factor pair j and k starts from trial loadings b_{ij} and b_{ik}, progresses through tentative loadings $\overset{\#}{b}_{ij}$ and $\overset{\#}{b}_{ik}$ to transformed loadings $\tilde{b}_{ij}$ and $\tilde{b}_{ik}$. The first step from trial loadings to tentative loadings involves possible interchange of the two factors and reflection of the factors. The second step from tentative loadings to transformed loadings invloves the orthogonal rotation of the tentative loadings. For the procedure for interchange and reflection, four coefficients delta are defined:

$$\delta_j = \pm 1 \; ; \quad \delta_k = \pm 1 \; ; \quad \delta_j^* = \pm 1 \quad \delta_k^* = \pm 1 \quad .$$

Two conditions are considered: first that the factors will not be interchanged and second that the factors will be interchanged. Whichever of these conditions apply will be determined later. Following are definitions for each of these conditions.

Without interchange:

$$\overset{\#}{b}_{ij} = \delta_j b_{ij} \; ; \text{ and } \overset{\#}{b}_{ik} = \delta_k b_{ik} \quad . \tag{12.149}$$

With interchange:

$$\overset{\#}{b}_{ij} = \delta_j^* b_{ik} \; ; \text{ and } \overset{\#}{b}_{ik} = \delta_k^* b_{ij} \quad . \tag{12.150}$$

Following is the criterion for the tentative loadings for factor pair j and k:

$$\overset{\#}{\Psi}_{fjk} = \sum_{i \in I_j} (\overset{\#}{b}_{ij} - h_{ij})^2 + \sum_{i \in I_k} (\overset{\#}{b}_{ik} - h_{ik})^2 \quad . \tag{12.151}$$

The criterion $\overset{\#}{\Psi}_{fjk}$ is written without reflection and with reflection as follows. Note that the squared deltas equal +1.

Without reflection:

$$\begin{aligned} \overset{\#}{\Psi}_{fjk} &= \sum_{i \in I_j} (\delta_j b_{ij} - h_{ij})^2 + \sum_{i \in I_k} (\delta_k b_{ik} - h_{ik})^2 \\ &= \sum_{i \in I_j} b_{ij}^2 - 2\delta_j \sum_{i \in I_j} b_{ij} h_{ij} + \sum_{i \in I_j} h_{ij}^2 + \sum_{i \in I_k} b_{ik}^2 - 2\delta_k \sum_{i \in I_k} b_{ik} h_{ik} + \sum_{i \in I_k} h_{ik}^2 \quad . \end{aligned} \tag{12.152}$$

With reflection:

$$\begin{aligned} \overset{\#}{\Psi}_{fjk}^* &= \sum_{i \in I_j} (\delta_j^* b_{ik} - h_{ij})^2 + \sum_{i \in I_k} (\delta_k^* b_{ij} - h_{ik})^2 \\ &= \sum_{i \in I_j} b_{ik}^2 - 2\delta_j^* \sum_{i \in I_j} b_{ik} h_{ij} + \sum_{i \in I_j} h_{ij}^2 + \sum_{i \in I_k} b_{ij}^2 - 2\delta_k^* \sum_{i \in I_k} b_{ij} h_{ik} + \sum_{i \in I_k} h_{ik}^2 \quad . \end{aligned} \tag{12.153}$$

In order to make criteria $\overset{\#}{\Psi}_{fjk}$ and $\overset{\#}{\Psi}_{fjk}^*$ as small as possible the algebraic signs of the deltas are selected as follow:

$$\text{sign}(\delta_j) = \text{sign}\left(\sum_{i \in I_j} b_{ij} h_{ij} \right) \; ; \quad \text{sign}(\delta_k) = \text{sign}\left(\sum_{i \in I_k} b_{ik} h_{ik} \right) \; ; \tag{12.154}$$

$$\text{sign}(\delta_j^*) = \text{sign}\left(\sum_{i\in I_j} b_{ik}h_{ij}\right) \quad ; \quad \text{sign}(\delta_k^*) = \text{sign}\left(\sum_{i\in I_k} b_{ij}h_{ik}\right) \quad . \tag{12.155}$$

Whether to interchange or not to interchange depends on selecting the lesser of criteria $\overset{\#}{\Psi}_{fjk}$ and $\overset{\#}{\Psi}^*_{fjk}$:

$$\text{if } \overset{\#}{\Psi}_{fjk} \leq \overset{\#}{\Psi}^*_{fjk}\text{, do not interchange and use equation (12.149) for the } \overset{\#}{b}\text{'s ;} \tag{12.156}$$

$$\text{if } \overset{\#}{\Psi}_{fjk} > \overset{\#}{\Psi}^*_{fjk}\text{, interchange the factors and use equation (12.150) for the } \overset{\#}{b}\text{'s .} \tag{12.157}$$

The transformation from the tentative loadings to the transformaed loadings is performed by a two dimensional orthogonal rotation illustrated below.

$$\begin{bmatrix} \overset{\#}{b}_{1j} & \overset{\#}{b}_{1k} \\ \overset{\#}{b}_{2j} & \overset{\#}{b}_{2k} \\ \vdots & \vdots \\ \overset{\#}{b}_{nj} & \overset{\#}{b}_{nk} \end{bmatrix} \begin{bmatrix} c & -s \\ s & c \end{bmatrix} = \begin{bmatrix} \tilde{b}_{1j} & \tilde{b}_{1k} \\ \tilde{b}_{2j} & \tilde{b}_{2k} \\ \vdots & \vdots \\ \tilde{b}_{nj} & \tilde{b}_{nk} \end{bmatrix}$$

In this transformation c is the cosine of the angle of rotation and s is the sine of the angle of rotation. In equation form:

$$\tilde{b}_{ij} = c\overset{\#}{b}_{ij} + s\overset{\#}{b}_{ik} \quad ; \tag{12.158}$$

$$\tilde{b}_{ik} = -s\overset{\#}{b}_{ij} + c\overset{\#}{b}_{ik} \quad . \tag{12.159}$$

Coefficient c may be restricted to a positive value with the relation between c and s being:

$$c^2 + s^2 = 1 \quad . \tag{12.160}$$

Coefficient c may be considered to be dependent on s, using the positive square root:

$$c = \sqrt{1 - s^2} \quad . \tag{12.161}$$

The criterion for this transformation is:

$$\tilde{\Psi}_{fjk} = \sum_{i\in I_j}(\tilde{b}_{ij} - h_{ij})^2 + \sum_{i\in I_k}(\tilde{b}_{ik} - h_{ik})^2 \tag{12.162}$$

which may be written, with substitution from the preceding equations, as:

$$\tilde{\Psi}_{fjk} = a_0 + a_1 s^2 + a_2 cs + a_3 c + a_4 s \tag{12.163}$$

where:

$$a_0 = \left\{\sum_{i\in I_j}\overset{\#2}{b}_{ij} + \sum_{i\in I_j} h_{ij}^2 + \sum_{i\in I_k}\overset{\#2}{b}_{ik} + \sum_{i\in I_k} h_{ik}^2\right\} \quad ; \tag{12.164}$$

$$a_1 = \left\{\sum_{i\in I_j}\overset{\#2}{b}_{ik} + \sum_{i\in I_k}\overset{\#2}{b}_{ij} - \sum_{i\in I_j}\overset{\#2}{b}_{ij} - \sum_{i\in I_k}\overset{\#2}{b}_{ik}\right\} \quad ; \tag{12.165}$$

$$a_2 = 2\left\{\sum_{i\in I_j}\overset{\#}{b}_{ij}\overset{\#}{b}_{ik} - \sum_{i\in I_k}\overset{\#}{b}_{ij}\overset{\#}{b}_{ik}\right\} \quad ; \tag{12.166}$$

$$a_3 = -2 \left\{ \sum_{i \in I_j}^{\#} b_{ij} h_{ij} + \sum_{i \in I_k}^{\#} b_{ik} h_{ik} \right\} \quad ; \tag{12.167}$$

$$a_4 = 2 \left\{ \sum_{i \in I_k}^{\#} b_{ij} h_{ik} - \sum_{i \in I_j}^{\#} b_{ik} h_{ij} \right\} \quad . \tag{12.168}$$

Browne (1972) recommends an iterative solution using Bailey's modification of the Newton-Raphson method (see: McCalla, 1967, p. 90). With s being considered as the independent variable, this method involves the first three derivatives of $\tilde{\Psi}_{fjk}$ with respect to s. Note for these derivatives that:

$$\frac{dc}{ds} = \frac{-s}{\sqrt{1 - s^2}} = \frac{-s}{c} \quad . \tag{12.169}$$

Then:

$$\Psi' = \frac{d\tilde{\Psi}_{fjk}}{ds} = 2a_1 s - a_2 \frac{(1 - 2c^2)}{c} - a_3 \frac{s}{c} + a_4 \tag{12.170}$$

$$\Psi'' = \frac{d\Psi'}{ds} = 2a_1 - a_2 s \frac{(1 + 2c^2)}{c^3} - \frac{a_3}{c^3} \quad . \tag{12.171}$$

$$\Psi''' = \frac{d\Psi''}{ds} = -3 \frac{(a_2 + a_3 s)}{c^5} \quad . \tag{12.172}$$

Given trial t of s, s_t, the value of s for trial $t+1$ is given by:

$$s_{t+1} = s_t - \frac{\Psi'}{\left\{ \Psi'' - \Psi' \Psi''' \Big/ 2\Psi'' \right\}} \quad . \tag{12.173}$$

The trials are continued until there is a minimal change from s_t to s_{t+1}.

As indicated earlier, the computations involves a series of cycles with each cycle involving a transformation for each pair of factors. Factor j ranges from 1 to $r-1$ with k ranging for each j from $j+1$ to r. The computations are stopped when there are no changes for all pairs of factors in a cycle.

There are two special cases of interest. The first case is when all specified hypothetical loadings are zero, a case closely related to hyperplane fitting. In this case coefficients a_3 and a_4 equal zero. Note that this eliminates the last two terms from equation (12.162) for the criterion and from the first derivative in equation (12.170). Note, also, the changes in the equations for the second and third derivatives in equations (12.171) and 912.172). From the interchange and reflection portion of transformation for a pair of factors, the reflections of factors are indeterminant since all terms $(\sum bh)$ equal zero. A further feature after the interchange: $a_1 \geq 0$.

There are two special situations as follow.

When $a_2 = 0$, coefficients s and c may be set as:

$s = 0$ and $c = 1$.

When $a_1 = 0$ and $|a_2| > 0$ (assume, as will be developed, that $|c| > 0$) from the first derivative, equation (12.170):

$1 - 2c^2 = 0$

so that:

$c = \sqrt{1/2}$

$s = \pm\sqrt{1/2}$

with the sign of s set so that the term $a_2 s\frac{(1+2c^2)}{c^3}$ in the second derivative is negative so as to make the contribution of this term positive to the second derivative; that is:

$\text{sign}(s) = \text{sign}(-a_2)$.

The general solution is considered using two conditions to control the sizes of numbers in the computations.

When $|a_1| \geq |a_2| > 0$, define, in consideration of setting the first derivative, equation (12.170), equal to zero:

$$x = -\frac{|a_2|}{|a_1|} = \frac{2\delta^* cs}{(1-2c^2)} \qquad (12.174)$$

with $\delta^* = \pm 1$, the sign being selected such that the term $a_2 s\frac{(1+2c^2)}{c^3}$ in the second derivative is negative so as to make a positive contribution to the second derivative. Then:

$1 + x^2 = 1\Big/(1-2c^2)^2$

or, using the positive square root and $\delta = \pm 1$:

$(1-2c^2) = \delta\Big/\sqrt{1+x^2}$.

Then:

$c^2 = \frac{1}{2}\left(1 + \delta\Big/\sqrt{1+x^2}\right)$.

For $c^2 \geq \frac{1}{2}$, $\delta = +1$ so that, using a positive square root:

$c = \sqrt{\frac{1}{2}\left(1 + 1\Big/\sqrt{1+x^2}\right)}$.

From equation (12.174):

$s = \delta^* x\Big/2c\sqrt{1+x^2}$

with the algebraic sign of δ^* being:

$\text{sign}(\delta^*) = \text{sign}(a_2)$.

When $|a_2| > |a_1| > 0$, define:

$$y = -\frac{|a_1|}{|a_2|} = \frac{\delta^*(1-2c^2)}{2cs} \qquad (12.175)$$

with $\delta^* \pm 1$, the sign being selected such that the term $a_2 s\frac{(1+2c^2)}{c^3}$ in the second derivative is negative so as to make the contribution positive to the second derivative. Then:

$1 + y^2 = 1\Big/4c^2s^2$

and

$\frac{y^2}{1+y^2} = (1-2c^2)^2$.

Using the positive square root and a coefficient $\delta = \pm 1$:

$$1 - 2c^2 = \delta\sqrt{\frac{y^2}{1+y^2}} \qquad (12.176)$$

so that:

$$c^2 = \frac{1}{2}\left[1 - \delta\sqrt{\frac{y^2}{1+y^2}}\right] \quad .$$

In order for $c^2 \geq \frac{1}{2}$, coefficient δ is set to equal -1. The, using the positive square root:

$$c = \sqrt{\frac{1}{2}\left[1 + \sqrt{\frac{y^2}{1+y^2}}\right]}$$

From equations (12.175) and (12.176):

$$s = \delta^* \Big/ 2c\sqrt{1+y^2} \quad .$$

As indicated above, the algebraic sign of δ^* is to be set as follows:

$$\text{sign}(\delta^*) = \text{sign}(-a_2) \quad .$$

The second special case occurs when all the specified hypothetical loadings for factors j and k are for the same entities; that is when $I_j = I_k$. The most collmon situation is that the matrix of hypothetical loadings is completely specified. In this case coefficients a_1 and a_2 equal zero. then, from equations (12.163), (12.170), (12.171):

$$\tilde{\Psi} = a_0 + a_3 c + a_4 s \quad ;$$

$$\Psi' = -a_3 \frac{s}{c} + a_4 \quad ;$$

$$\Psi'' = -\frac{a_3}{c^3} \quad .$$

After setting the first derivative equal to zero algebraic operations yield:

$$\frac{s}{c} = \frac{a_4}{a_3}$$

$$c^2 = \frac{a_3^2}{a_3^2 + a_4^2} \quad .$$

Then, using the positive square root and $\delta = \pm 1$:

$$c = \frac{\delta\, a_3}{\sqrt{a_3^2 + a_4^2}} \quad ;$$

$$s = \frac{\delta\, a_4}{\sqrt{a_3^2 + a_4^2}} \quad .$$

With the interchange and reflections a_3 is negative, see equation (12.167), so that for the second derivative to be positive, as is necessary for a minimum solution:

$$\delta = -1 \quad .$$

All of the preceding solutions yield a matrix B but not the transformation matrix T. In order to obtain the matrix T, a solution from equation (12.126) yields:

$$T' = (A'A)^{-1}(A'B) \quad .$$

Table 12.16 gives the results for orthogonal trait vectors, fixed hypothetical loadings for the fourth year medical students data. As in previous tables, the matrix of hypothetical loadings is given on the left with the transformation results being given on the right. These results are to

Table 12.16
Solution For: Orthogonal Trait Vectors, Fixed Hypothetical Loadings
Fourth Year Medical Students Data

Hypothetical Loadings

Given Loadings

	1	2
1	.568	.090
2	.486	-.113
3	.393	.005
4	-.005	.844
5	.013	.852

Transformatin Results

Trait Vectors

	1	2
1	.635	.772
2	.772	-.635

Factor Loadings

	1	2
1	.504	.278
2	.688	.105
3	.697	.141
4	.183	.847
5	.171	.829

Sums of Squared Differences

1	2
.198	.102

Criterion ψ_f = .300

be compared with those given in Table 12.12 when the trait vectors were permitted to be oblique. Most notable for the factor loadings is that the small loadings, now, are measurably positive rather than being near zero, some of the previous loadings being negative. This is the result of the shift in the transformed factors from having a correlation of .360 to the restricted correlation of zero. Further, the sums of squared differences have been increased with the criterion Ψ increasing from .175 for the oblique case to .300 for the orthogonal case. A question to be considered is whether the loss of goodness of fit is worth the reduction in number of parameters due to the restriction on the correlation between the factors.

12.2.2.2 Orthogonal Trait Vectors, Proportional Hypothetical Loadings

The option of using proportional hypothetical loadings is available when the analyst consideres that the hypothetical loadings are given only within constants of proportionality, one such constant for each factor. However, the trait vectors are to be orthogonal. A restriction is that at least one of the specified hypothetical loadings for each factor must be greater than zero in absolute value. The criterion for this case is the same as for the case permitting oblique factors:

$$\widetilde{\Psi}_p = \sum_{m=1}^{r} \left\{ \sum_{i \in I_m} (b_{im} - c_m h_{im})^2 \right\} \quad . \tag{12.76}$$

where c_m is a constant of proportionality for each factor applied to the hypothetical loadings for that factor. However, now, this criterion is to be minimized under the constraint of equation (12.124) that the trait vectors be orthogonal. This constrain is implemented by the definition of a function $\widetilde{\Psi}_\lambda$ defined in equation (12.130)

$$\widetilde{\Psi}_\lambda = \sum_{k=1}^{r} \sum_{m=1}^{r} \lambda_{km}(\phi_{km} - \delta_{km}) \tag{12.130}$$

where the λ_{km}'s are Lagrange multipliers and the δ_{km}'s are Kronecker deltas defined in equations (12.131):

$$\delta_{kk} = 1 \; ; \quad \delta_{km} = 0 \quad \text{for } k \neq m \quad . \tag{12.131}$$

The combined criterion is:

$$\widetilde{\Psi}_t = \widetilde{\Psi}_p + \widetilde{\Psi}_\lambda \quad . \tag{12.132}$$

Solutions for the c_m's are considered first for given trail trait vectors and loadings in matrix B. The partial derivative of $\widetilde{\Psi}_t$ with respect to the c_m is set equal to zero.

$$\frac{\partial \widetilde{\Psi}_t}{\partial c_m} = 2 \sum_{i \in I_m} (b_{im} - c_m h_{im})\, h_{im} = 0 \quad . \tag{12.177}$$

This yields:

$$c_m = \left(\sum_{i \in I_m} b_{im} h_{im} \right) \Big/ \left(\sum_{i \in I_m} h_{im}^2 \right) \quad . \tag{12.178}$$

Adjusted hypothetical loadings are obtained by:

$$\ddot{h}_{im} = c_m h_{im} \quad . \tag{12.179}$$

An alternating type of solution is utilized starting from initial trial coefficients c_m, possibly using all c_m equal to unity. Adjusted hypothetical loadings are computed by equation (12.179) and initial trial loadings in matrix B are determined using the adjusted hypothetical loadings as if they were fixed and the solution for orthogonal trait vectors, fixed hypothetical loadings in the preceding section. Having the trial loadings in matrix B, new trial coefficients c_m are computed by equation (12.179). New adjusted hypothetical loadings are computed by equation (12.179) and new trial loadings in matrix B are computed using the adjusted hypothetical loadings as if they are fixed and the solution for orthogonal trait vectors, fixed hypothetical loadings. This alternation between trial c_m's and trial matrices B is continued until there is a minimal change in the constants of proportionality, c_m, from trial to trial.

Table 12.17 presents the results for this criterion for the fourth year mdeical students. As before, the hypothetical loadings are given onthe left with the given hypothetical loadings at the far left. In the second section for the hypothetical loadings are the obtained constants of proportionality and the adjusted hypothetical loadings. The transformation results are given at the right with the trait vectors, factor loadings, and sums of squared differences between the factor loadings and the adjusted hypothetical loadings. A comparison of these results given in Table 12.13 for the case when the trait vectors were permitted to be oblique indicates that the constants of proportionality in these two tables are very similar. The first factor hypothetical loadings are adjusted by a constant more than 1.25 while the constants of proportionality for the second factor are slightly lexx than unity. Again, as for the preceding section in Table 12.16, the factor loadings for small given hypothetical loadings are markedly positive. The sums of squared differences for the orthogonal trait vectors, proportional hypothetical loadings are less than for the case in Table 12.16 for fixed hypothetical loadings. However, these sums of squared differences are greater than for the case in Table 12.13 when the trait vectors were permitted to be oblique. Again, there is a question for the fourth year medical students data whether or not the restriction for orthogonal trait vectors is more desirable than the case when the trait vectors were permitted to be oblique.

12.3 Factor Analysis of Several Batteries in One Sample

Two major purposes are served by factor analyses of several batteries in one sample. The firstn purpose is to confirm generalization of factors from one battery to a second battery. Tucker (1958) introduced inter-battery factor analysis for the purpose of generalizing factors across

Table 12.17

Solution For: Orthogonal Trait Vector, Proportional Hypothetical Loadings
Fourth Year Medical Students Data

Hypothetical Loadings

Transformation Results

Constants of Proportionality

1	2
1.267	.977

Trait Vectors

	1	2
1	.659	.752
2	.752	-.659

Given Loadings

	1	2
1	.568	.090
2	.486	-.113
3	.393	.005
4	-.005	.844
5	.013	.852

Adjusted Loadings

	1	2
1	.720	.088
2	.615	-.110
3	.498	.005
4	-.006	.825
5	.016	.833

Factor Loadings

	1	2
1	.512	.263
2	.691	.083
3	.701	.119
4	.209	.841
5	.196	.824

Sums of Squared Differences

1	2
.169	.081

Criterion $\psi_p = .250$

selections of attributes measures in contrast to generalization of results for a given battery over samplinsg of individuals. The second purpose is to determine factors which are common to several batteries separate from factors which appear in single batterise. For an example consider one battery of course grades in secondary school, a second battery of entrance examinations to college, and a third battery of college grades. Here the questions could involve what factors are common to pairs of these batteries and are common through all three of the batteries. For an example, there may be factors common to secondary school grades and college grades which are not involved in the entrance examinations.

The model for inter-battery factor analysis is to be considered first followed by discussion of the generalized model for three or more batteries. Equation (12.180) gives the model in the population for two batteries where Σ is the covariance matrix for all attributes. There are n_1 attributes in battery 1 and n_2 attributes in battery 2. This matrix is partitioned into sections for the intercovariances for each battery and covariance matrices between the two batteries.

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} \mathbb{A}_1 & \mathbb{S}_1 & 0 \\ \mathbb{A}_2 & 0 & \mathbb{S}_2 \end{bmatrix} \begin{bmatrix} \mathbb{A}_1' & \mathbb{A}_2' \\ \mathbb{S}_1' & 0 \\ 0 & \mathbb{S}_2' \end{bmatrix} \tag{12.180}$$

Matrices $\mathbb{A}_1$ and $\mathbb{A}_2$ are the common factor matrices for factors common to the two batteries while matrices $\mathbb{S}_1$ and $\mathbb{S}_2$ contain the factors specific to each of the two matrices. Note that the attribute specific factors are included in matrices $\mathbb{S}_1$ and $\mathbb{S}_2$. There are r inter-battery common factors. In a sample, the model for covariance matrix C is:

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} A_1 & S_1 & 0 \\ A_2 & 0 & S_2 \end{bmatrix} \begin{bmatrix} A_1' & A_2' \\ S_1' & 0 \\ o & S_2' \end{bmatrix} + \begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{bmatrix} \Delta_{12} \tag{12.181}$$

where matrices A_1 and A_2 contain the inter-battery common factors in the sample while matrics S_1 and S_2 contain the battery specific factors in the sample. Matrix Δ contains the discrepancies of fit. An interesting feature of this model is that matrices C_{12} and C_{21} do not involve the battery specific factors.

$$C_{12} = A_1 A_2' + \Delta_{12} = C_{21}' \quad . \tag{12.182}$$

Tucker (1958) proposed a least squares solution for the inter-battery factors using an Eckart and Young (1936) decomposition (This is the same as a singular value solution.). Equation (12.183) symbolizes this solution.

$$C_{12} = V_1 D V_2' = \begin{bmatrix} V_{1r} & V_{1\delta} & V_{10} \end{bmatrix} \begin{bmatrix} D_r & 0 & 0 \\ 0 & D_\delta & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} V_{2r}' \\ V_{2\delta}' \\ V_{20}' \end{bmatrix} \tag{12.183}$$

where V_1 and V_2 are orthonormal matrices, V_1 being n_1 x n_1 and V_2 being n_2 x n_2. Matrix D has diagonal sections D_r and D_δ. There are r columns in matrix sections V_{1r} and V_{2r} with r being the number inter-battery factors retained in the analysis. D_r has r entries.

Let there be R non-zero entries in the diagonal of D, then the number of columns in $V_{1\delta}$, $V_{2\delta}$ are $R - r$ and the number of entries in D_δ is $R - r$. The value chosen for r is discussed below. The direction of the column vectors in V_1 and V_2 are to be chosen such that the diagonal entries in D_r and D_δ are positive. Initial matrices A_1 and A_2 are given by:

$$A_1 = V_{1r}D_r^{1/2} \quad ; \tag{12.184.1}$$

$$A_2 = V_{2r}D_r^{1/2} \quad . \tag{12.184.1}$$

The diecrepancy matrix Δ_{12} is given by:

$$\Delta_{12} = V_{1\delta}D_\delta V'_{2\delta} \quad . \tag{12.185}$$

The sum of squares of entries in Δ_{12} is given by:

$$SSQ(\Delta_{12}) = \sum_{j=r+1}^{R} d_j^2 \quad . \tag{12.186}$$

An alternative is:

$$SSQ(\Delta_{12}) = SSQ(C_{12}) - \sum_{j=1}^{r} d_j^2 \quad . \tag{12.187}$$

The series of entries d_j^2 , j = 1,R , may be inspected to determine the number of inter-battery factors, r, such that the last dimension accepted has the last d_j^2 of acceptable size.

The initial matrices A_1 and A_2 determined in equations (12.184.1) and (12.184.2) are subject to a general transformation. Let T , r x r , be a general, nonsingular transformation matrix. Then transformed inter-battery matrices, $\tilde{A}_1$ and $\tilde{A}_2$, are obtained by:

$$\tilde{A}_1 = A_1T \quad ; \tag{12.188.1}$$

$$\tilde{A}_2 = A_2(T')^{-1} \quad . \tag{12.188.2}$$

Then, $\tilde{A}_1$ and $\tilde{A}_2$ may be substituted in equation (12.182) for A_1 and A_2 . Note that this transformation is more geneneral than those usually encountered in factor analysis, there is no restriction on the lengths of the column vectors of T . This presents a very difficult problem in the treatment of inter-battery factor analysis.

Browne (1979) described a maximum likelyhood solution in inter-battery factor analysis. He noted, also, the general transformation described in the preceding paragraph.

The inter-battery model may be extended to three or more batteries (the number of batteries being designated by p) as indicated in the following equation for the population.

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1p} \\ \Sigma_{12} & \Sigma_{22} & \cdots & \Sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{p1} & \Sigma_{p2} & \vdots & \Sigma_{pp} \end{bmatrix} = \begin{bmatrix} \mathbb{A}_1 & \mathbb{S}_1 & 0 & \cdots & 0 \\ \mathbb{A}_2 & 0 & \mathbb{S}_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbb{A}_p & 0 & 0 & \cdots & \mathbb{S}_p \end{bmatrix} \begin{bmatrix} \mathbb{A}'_1 & \mathbb{A}'_2 & \cdots & \mathbb{A}'_p \\ \mathbb{S}'_1 & 0 & \cdots & 0 \\ 0 & \mathbb{S}'_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbb{S}'_p \end{bmatrix} \tag{12.189}.$$