GENERALIZED FACTOR ANALYSIS

PART I

Paul Horst

This Study Was Supported in Part by

Office of Naval Research Contract Nonr - 477(33)

and

Public Health Research Grant 2 R01 MH00743-14

Principal Investigator: Paul Horst

April 1969

University of Washington

Seattle, Washington

GENERALIZED FACTOR ANALYSIS

CHAPTER 1

THE DATA MATRIX

The problem of approximating a data matrix with one of lower rank is fundamental to all scientific investigation. This problem is embedded in most traditional data analysis techniques, such as multiple regression analysis, analysis of variance and covariance, configural analysis, pattern recognition, discriminant function analysis, factor analysis, etc. In all of these procedures we begin with an experimental data matrix. Transformations on the elements of the data matrix may then be carried out. A matrix which approximates in some sense the original or transformed data matrix is solved for. A residual matrix whose elements are the differences between the elements of the data matrix and those of the approximation matrix is calculated.

1.1 The Experimental Data

The experimental data matrix in its simplest form consists of rows which without loss of generality we may take to represent entities, observations, or cases, and columns which represent attributes, characteristics, or variables. These latter are also called variates. One may also have occasions and other categories, such as sets, instruments, conditions, and treatments, thus yielding multidimensional or multicategory data matrices. These extensions have been considered by Cattell (1957), Tucker (1963), and Horst (1965). In general it is possible, as shown by Horst (1965), to reduce multimode data matrices to two-mode data matrices in a number of ways. Tucker (1963) has presented the most sophisticated analytical procedures to date for analyzing multimode data matrices. In this report, however, we restrict ourselves to the two-mode data matrix, and for convenience we shall take rows as entities and columns as attributes, although this orientation is not necessary.

1.2 Transformation of the Data

A topic which has not been sufficiently considered in the past is that of mathematical transformations of experimental data before the more detailed analyses take place. The failure to recognize the importance of this topic has resulted in confusion between the disciplines of factor analysis and multidimensional scaling techniques. Much of the the work in multidimensional scaling can be regarded as special cases of factor analytic techniques. The generalized distance models in scaling theory reduce to the more conventional factor analysis models after appropriate transformations of the observed data have been made. It is not the purpose of this monograph to explore the general notion of transformations of the original data on the basis of theoretical formulations, or to relate the multidimensional scaling techniques to the more traditional factor analytic techniques. Ross and Cliff (1964) have suggested this relationship. However, they did not point out explicitly that their approach consists essentially of making a transformation of the original observations consistent with the distance hypothesis, and then treating the data by the more conventional factor analytic procedures. Coombs and Kao (1960) were among the first to suggest the relationship between the multidimensional scaling techniques and the conventional factor analytic techniques. It remained, however, for Ross and Cliff to indicate the explicit relationship between the two general approaches by showing that transformations of the original data consistent with the distance concept provide the basis for the more conventional factor analytic or lower rank data matrix approximation analyses.

In this section we shall consider four kinds of transformations. These are linear transformations, nonlinear transformations, single element transformations, and transformations involving combinations of variables.

In a linear transformation we have a scaling or multiplying constant and a location or additive constant. The variable to be transformed is expressed in the general form of

$$Y = A + BX$$

where the original variable is X, the transformed variable is Y, the location constant is A, and the scaling constant is B. We may have the special case where the additive constant A is zero and therefore the transformation consists simply of a change of scale. On the other hand, we may have the case where B is unity. In this case, we simply add a constant to the observed value. The transformation of raw data to deviation measures is a special case of a linear transformation where the additive constant is simply the negative of the mean of the variable, and the multiplying constant is unity. In standardized measures, the deviation measure is divided by the standard deviation of the sample so that the multiplying constant is the reciprocal of the standard deviation. Linear transformations of this sort are introduced early in introductory courses in statistics. However, the significance of transformation of this kind for factor analytic and data matrix approximation techniques are not so well recognized. It is one of the major objectives of this monograph to discuss in more detail the importance and implications of linear transformations of experimental data.

We have already considered the subject of multidimensional scaling and how these techniques involve the concept of data transformation. More specifically, the kinds of transformations involved in relating the multidimensional scaling techniques to the factor analytic techniques involve nonlinear transformations of the data. The types of transformations involved here are trigonometric. Nonlinear transformations may reflect the role of theory in data analysis. For example, it is in the distance theory of multidimensional scaling that the mathematical transformations of the data are suggested which convert distance models to factor

analytic models. Much of the mathematical models work in learning theory results in nonlinear transformations of data based on rational theories. It is quite probable that an explicit recognition of the role of nonlinear transformations of experimental data based on rational theories of learning could lead to a fruitful integration of mathematical models and factor analytic approaches in psychology. It is also probable that quantitative theory in other social science disciplines could lead to a better integration of methodologies, theories, and data analysis procedures.

In the preceding discussions of linear and nonlinear transformations, it was assumed that the transformations are on single variables. The same mathematical transformation applies to all elements of a single attribute vector. It is possible, however, to have transformations which involve several or more variables. An example of such a combination of variables is the image analysis model of Guttman (1953). An important case of combinations of variables consists of procedures where nonlinear combinations of variables are introduced. Perhaps by far the most common example of such nonlinear combinations is provided in the techniques of configural analysis or pattern recognition. These techniques involve multivariate polynomial transformations of the data in which new variables are generated that are products of subsets of the original data. We have discussed this approach elsewhere (Horst, 1968c). The generation of new variables that are product functions of the original variables may well contribute information not included in simple linear combinations of the data. Guttman (1955b) has recognized the importance of configural analysis. His concepts of the simplex, the radex, and the circumplex imply nonlinear combinations of the original variables.

Much remains, however, to be done to relate the configural analysis procedures to the more conventional data matrix approximation techniques. One of the unsolved problems in this approach is that of the disparate distribution phenomenon

which introduces artifactual dimensions into a data matrix. But the subject of nonlinear combinations of data, important as it is, will not be considered in detail in this monograph since it leads into problems which have not yet been adequately solved.

1.3 The Approximation Matrix

Assume now that we begin with either the original data matrix or a matrix in which the elements have been transformed, as indicated in the preceding discussion. We then wish to consider a matrix which approximates the original or transformed matrix but which in some sense is more simple than the original matrix. The subject of data matrix approximation has been extensively considered by many writers and has received detailed treatment by the author (Horst, 1963, 1965).

The approximation matrix is of lower rank than the data matrix or some transformation of it. It is the product of a factor score matrix by the transpose of a factor loading matrix. The number of columns in the factor score matrix is equal to the rank of the approximation matrix. This rank is the number of factors assumed or solved for. The factor score matrix is called basic because its rank is equal to its width or smaller dimension.

The factor loading matrix has as many columns as the number of factors and as many rows as the number of attributes in the data matrix. It is also basic so that its rank is equal to the number of factors or number of columns. Therefore both the factor loading matrix and the factor score matrix are basic matrices which cannot be expressed as the product of matrices whose common order is less than the number of factors or the rank of the approximation matrix. This implies, of course, that the number of factors is smaller than either the number of entities or the number of attributes, whichever is smaller. A more complete discussion of the factor loading matrix and the factor score matrix is provided elsewhere by the author (Horst, 1965).

We have defined the approximation matrix as the product of the factor score matrix postmultiplied by the transpose of the factor loading matrix. It can readily be shown (Horst, 1963, 1965) that a product of two matrices can be expressed as the product of an infinite number of different pairs of matrix factors. As a special case, we may consider the postmultiplication of the prefactor by any conformable square orthonormal matrix and the premultiplication of the postfactor by the transpose of this orthonormal matrix. The major product of these two matrices is the same as the major product of the original matrices since the product of the orthnormal matrix by its transpose is the identity matrix. It is also obvious that if the prefactor is postmultiplied by any nonvertical basic matrix, and the postfactor is premultiplied by the general inverse of this nonvertical matrix, then the major product of the two resulting matrices will be the same as for the original matrices. This nonuniqueness in the matrix factors of a product is considered in more detail in Chapter 9. That chapter develops a new model for a unique determination of the factor score and the factor loading matrices.

1.4 The Residual Matrix

The residual matrix is simply one whose elements are the difference between the corresponding elements of the data matrix and the approximation or product matrix. So far, we do not specify any constraints on the approximation matrix aside from those considered in the previous sections. Most factor analytic models, as well as the general multiple regression models, place certain constraints on the residual matrix as a basis for determining the factor loading and factor score matrices.

Most multivariate analysis procedures, including multiple regression, multiple discriminant function analysis, the multidimensional scaling techniques, and all of the varieties of factor analytic techniques, are concerned in some way with specifying properties of the residual matrix that are to be satisfied. We may

consider either the residual matrix itself or the covariance matrix which may be calculated from it. It is more convenient to begin with the residual covariance matrix than with the residual matrix itself. Two aspects of the residual covariance matrix may be considered in determining the factors in the product approximation matrix. The first of these concerns the elements of the covariance matrix to be included in any procedures of optimization. This matrix consists of the diagonal elements or residual variances and the offdiagonal elements or residual covariances. How we combine these will determine the solution for the factors in the approximation matrix. The second aspect of the residual covariance matrix concerns what particular function of the elements or combinations of elements is to be optimized by the solution for the factors of the approximation matrix. What combination of elements is included and what function of these elements is optimized is the subject of later chapters.

CHAPTER 2

ORIGIN TRANSFORMATION

In this monograph we shall restrict our discussion of the role of transformations of the elements in the data matrix to linear transformations involving only additive and scaling constants. Although in later chapters we shall restrict the transformations even further to those involving primarily scaling transformations, it is of interest to consider the subject of origin transformations or additive constants since these are also important for matrix approximation procedures. In data matrix transformation procedures, a major consideration is the determination of the transformation functions and parameters so as to optimize prespecified functions of the residual matrix. This monograph deals with determinations of scaling constants which with specified restrictions optimize prespecified functions of the residual matrix or its covariance matrix. Little has been done in the way of solving for origin or additive transformations that optimize such functions of the residual matrix. However, we have elsewhere considered (Horst, 1965) the effect of prespecified origin transformations on the basic structure of a matrix. In this chapter we shall review breifly the subject of origin transformations. We shall consider transformations by attributes, by entities, by entities and attributes, and then present briefly the current status of origin transformation techniques.

2.1 Origin Transformations by Attributes

By far the most common form of origin transformation is transformation by attributes. Here the constant, positive or negative, is added to each element of an attribute column. The constant may, and generally does, vary from one attribute to another.

The most common type of attribute transformation consists of subtracting the mean of a colum of attribute measures from each of the elements or measures. This, of course, results in the familiar deviation score matrix in which the sums of

column elements of the transformed matrix are all zero. The main reason for discussing this familiar origin transformation procedure is to emphasize that it is arbitrary and may not be appropriate for many kinds of analyses. There may well be better or more appropriate criteria for determining origin attribute transformations than the zero sum criterion. While the conventional multiple regression techniques give results invariant with respect to origin transformations, including the attribute centering transformations, such invariance does not hold in general for factor analytic techniques.

The attribute centering origin transformation is a special case of the more general attribute origin transformation. Another special case occurs when the additive constant is zero or when the raw data are not transformed by attribute origin. Tucker (1958) has considered cases where the raw measures may appropriately enter into factor analytic computations. However, the general case where the observed measures may be origin transformed by attributes has received little theoretical, empirical, or experimental consideration. If the raw measures may be regarded as in some sense absolute and the origins comparable from one attribute to another, then the zero origin transformation may be justified. But further rational or optimizing procedures are required for the general case of differential origin transformation for a set of attributes.

2.2 Origin Transformation by Entities

Just as origin transformations may be made by attributes, so also they may be made by entities, although this procedure is by no means as common as the attribute transformation. We can also have the two types of transformations by entities, namely, centering by entities and the more general origin transformation of which centering is a special case.

When the origin transformation is such as to center by entities, a constant is subtracted from attribute measures for each entity, such that the sum of the elements of each row is equal to zero. A special case of such centering occurs with ipsatized

variables, as in the case of forced choice personality instruments. It can be shown that such centering by entities also serves to center by attributes or columns. The subject of entity centering has not been considered extensively from a theoretical point of view and does not appear to have much justification, particularly since it can readily be shown that important information might be lost in such centering. For example, it is clear that if one has a data matrix of measures on a number of persons, when one centers by rows one obviously eliminates normative information from the data matrix. An extensive treatment of the subject of centering by rows has been given by Clemans (1966) in a discussion of normative and ipsative variables.

There may be more justification for a rational and more general transformation of origin by rows than for a mere centering transformation. Particularly in the case of ipsatively measured variables such as one finds in forced choice instruments, it may be desirable to change the origin by entities in order to satisfy optimizing functions in factor analytic or general matrix approximation techniques.

2.3 Transformation of Origin by Both Entities and Attributes

It is possible, and in some cases may be appropriate, to transform origins of a data matrix both by entities and by attributes. This can be done as a special case by a doubly centered, or right and left, centering operation. Here we may also have the general case, as in the centering or origin transformations by either attributes or entities.

In the doubly centered origin transformation, the elements in each row and in each column add up to zero in the transformed matrix. This procedure is followed when a two-way analysis of variance is applied to a matrix of observations and the effect of both row and column means is removed. Such an operation in the conventional two-way analysis of variance is not usually recognized explicitly as a doubly centering operation.

As in the case of the general origin transformation by either attributes or entities, we may transform the origin of both entities and attributes on the basis of any rationale which may be available.

2.4 Current Status of Origin Transformations

The basic structure of the data matrix or its covariance or correlation matrix is altered by origin transformations. While very little has been done in the way of developing general rationales for determining origin transformations, whether by entities or by attributes or both, considerable work has been done on the effect of any arbitrary origin transformation operations on the basic structure or latent roots and vectors of the covariance matrix. This work is presented in Chapter 13 of "Factor Analysis of Data Matrices" (Horst, 1965). It is shown that a root of a covariance matrix altered by an origin transformation must lie between adjacent roots of the original matrix. Procedures for solving for the latent roots and vectors of an origin-transformed matrix in terms of the original roots and vectors or basic orthonormals are presented in this reference, together with computational Fortran programs for effecting the transformations. These procedures indicate how one may pass from one origin transformation to another in terms of a solution of the roots of one as a function of the roots of the other. As one would guess, these are not closed solutions but require iterating computations. Usually, however, the solutions converge rapidly.

CHAPTER 3

Scale Transformation

In the previous chapter we have considered various methods by which one may transform a data matrix with reference to origin. In this chapter we consider the transformation of the matrix by a multiplying or scaling constant. It is possible, of course, to apply both origin and scaling constants but it is more convenient to consider the two separately. As we have seen in Chapter 2, the problems involved in transformation of origin have not been extensively considered in terms of matrix approximation, or in terms of optimal properties of the residual matrix. Only the effect of such transformations on the basic structure of the matrix has been considered in some detail (Horst 1965). The problems of scale transformation including rationales and procedures have been more extensively investigated, particularly in the area of factor analysis which of course is a special case of matrix approximation. We shall in this chapter consider briefly the scaling of attributes, the scaling of entities, and the scaling of both entities and attributes.

3.1 Scaling by Attributes

Here again, as in the case of transformation of origin, the scaling transformation has been much more extensively applied to attributes than to entities. The most obvious case of scaling by attributes is the transformation t standard measures, so that the standard deviations of all variables or attributes are unity. Such scaling is the most common among scaling procedures for factor analytic techniques. In scaling by attributes, we simply multiply the natural order of a data matrix on the right by a diagonal scaling matrix. In the case of the standardized data matrix, this scaling or diagonal matrix has the reciprocals of the standard deviations of the variables in the diagonal position.

One may also have other rationales for scaling a data matrix or making a scale transformation, but usually a decision must be made about scaling the attributes unless there is good evidence for assuming that all of the variables are measured in comparable units.

In general one would not consider scaling only one of the attributes in a data matrix, but this special case is of interest because it has interesting mathematical properties. One can readily relate the latent roots and vectors of a covariance matrix to another covariance matrix which has one of the elements rescaled. Formally, the mathematics is similar to the transformation of origin by attributes. However, mathematically it is just as simple to calculate new latent roots and vectors from the original ones when the origins for all of the variables are transformed as it is to calculate these when only a single variable is rescaled. To our knowledge, the mathematics substantiating this statement has not been previously presented in published works but it can readily be demonstrated.

Rationales for scaling all the variables in a data matrix could readily be found. A simple case is when all the variances are required to be equal or to be unity. However, the relationship between the latent roots and vectors of a covariance matrix and a generalized rescaling of the variables in the covariance matrix as functions of the new scaling parameters is extremely complicated and no simple relationships exist between the two. Even in the case of a rescaling of only two variables, the mathematics for expressing the relationships between the new and the old eigenvalues and eigenvectors is complicated. One can, of course, always determine the new ones empirically.

It is true that some types of multivariate analysis are independent of scale transformation by attributes. For example, in the case of multiple regression analysis, a simple relationship exists between scale transformations of the dependent and independent variables by attributes and the factor loading matrix. In this special case, the factor loading matrix can be shown to be (Horst, 1965) merely a supermatrix, the first matrix element of which is the identity matrix, and the second the matrix of regression coefficients. A rescaling of the submatrix of independent variables results simply in a reciprocal rescaling of the matrix of

regression coefficients. In the case of canonical correlation, it is also true that the solutions are independent of the scaling of the subsets of variables. Simple relationships exist between the scaling of the data submatrices by attributes and the scaling for the corresponding regression matrices.

3.2 Scaling by Entities

A procedure much less common than that of scaling by attributes is scaling by entities. Such data matrix transformations have rarely been used in practice and the conditions under which one is justified in using them do not appear to have been extensively considered. One justification for scale transformations of the data matrix by entities might be the assumption that some of the entities are more important than others in determining a solution for the approximation data matrix. Such assumptions of differential importance of the entities in determining a solution based on some prespecified criteria or rationale have not been generally utilized. In the theory of least squares, as applied to multiple regression analysis, some of the early theory utilizes the weighting of observations. If the loss function for matrix approximation has been adequately formulated in mathematical terms, then it should be possible to apply weighting functions to the entities to satisfy this loss function. Rationales of this type, however, must obviously place adequate restrictions on the entity scaling matrix. For example, the elements of the scaling matrix should probably all be positive and finite, and perhaps, some function of the weights should be a constant.

It is clear that in the multiple regression model, if all of the entity scaling weights were taken as zero except any subset equal in number to the number of independent variables, then the least squares loss function would be at its optimum or zero. This is equivalent to choosing a subset, in size equal to the number of independent variables, on the basis of which to determine the regression vector. Such a solution would of course always yield a regression vector which would

exactly reproduce the elements of the dependent variable in the sample. One might impose further restrictions on the scaling matrix such that the moments of the distributions of the estimated and actual dependent variables in the sample satisfy certain conditions. For example, one could specify that the weighting vector should be such as to yield a best approximation to a normal distribution for each of the independent variables and also for the estimates of these independent variables. To our knowledge, such rationales and mathematical formulations have not been experimented with.

Perhaps the most important distinction between entity and attribute scaling is evident in the multiple regression and canonical correlation approaches. We have seen that for these models the scaling of attributes is reflected in a simple manner in the multiple regression or weighting matrices. Obviously, this simple relationship cannot hold in the case of entity scaling since the data matrix and the regression matrix are not even conformable with respect to the entity order. It is possible that for some arbitrary scaling an interesting relationship might be found to relate the estimated dependent variables to those estimated without scaling as some simple function of the entity scaling matrix. However, these relationships may be of no more than academic interest.

What we have said about the effect of scaling on the basic structure of a matrix with reference to attributes applies also in the case of entities. The scaling or rescaling of a single entity results in a modification of all of the latent roots and vectors of the original data matrix. The relationship between the original eigenvectors and those resulting from the scaling of a single entity can be expressed in terms of upper and lower bounds. However, it is difficult to see of what practical importance such a single entity scaling would be. In general, one would not expect a practical problem to be concerned with the rescaling simply of a single entity selected arbitrarily, or even presumably on the basis of some rationale, from all of the entities in the sample.

For the scaling of several entities, the mathematics which indicate the relationship between the original and the final or rescaled eigenvalues is much more complicated than for a single entity. For more than one entity, therefore, it does not seem practical to consider the mathematical relationships between the matrices of the scaled and unscaled entities in terms of the eigenvalues and vectors of their covariance or correlation matrices.

As we shall see in Chapter 8, it is possible to set up scaling procedures so that the solution for the approximation matrix is independent of the original scaling of the data matrices. This is true for either entity or attribute scaling or both.

3.3 Scaling by Entities and Attributes

Just as we can have origin transformations by both entities and attributes for the data matrix, so also can we have scaling by both entities and attributes for any arbitrarily scaled or quantified matrix of observations. What we have said about the rationale for entity and attribute scaling applies equally well to any simultaneous scaling of both dimensions of the data matrix. Presumably any complete theory of scaling transformations should provide for both entity and attribute scaling. It should be possible to develop a rationale of scaling that takes into account both sides of the matrix. This would be an important contribution to the problem of metric in factor analysis specifically and in the analyses of data matrices in general. It is, however, beyond the scope of this report to consider in detail such dual scaling rationales.

CHAPTER 4

THE LOSS FUNCTION

Let us assume we have a data matrix which may have undergone some transformation, linear or nonlinear, by rows or columns or both, that we wish to approximate by the major product of two basic matrices with common order less than the smaller order of the data matrix. We indicate the deviated data matrix by Z, the factor score matrix by X, the factor loading matrix by A, and the residual matrix by e. We can then write

$$e = Z - XA' \tag{4.1}$$

The problem is to determine X and A so that some function of the elements of e will be optimized. Instead of considering the elements of e directly, we may consider the covariance matrix E given by

$$E = e'e \tag{4.2}$$

4.1 The Elements in the Loss Function

Without loss of generality we may assume the scaling of Z in Eq. 4.1 to have been such that we need not divide the right side of Eq. 4.2 by N, the number of entities. One of the simplest loss functions that has been commonly used, particularly in factor analytic work, utilizes only the diagonal elements of E in Eq. 4.2. Obviously, these elements are proportional to the variances of the residual column elements in e of Eq. 4.1. The function of these elements most commonly used in the loss function is simply their sum. This sum is simply the sum of squares of the residual elements in e. It can be shown that traditional multiple regression analysis with one or more independent variables is a special case of Eq. 4.1 in which the solution for A and X is constrained so that the elements vanish in the columns of e corresponding to the independent variables. Consequently, the corresponding diagonal elements of E in Eq. 4.2 are also zero. The X and A matrices are determined so that the sum of the diagonal elements in E is minimized. This formulation of the multiple regression model does not appear to have been generally obvious.

In the case of one type of factor analysis, which some call principal component analysis, only the diagonal elements in E are considered and A and X are determined so that the sum of these diagonal elements is minimized. Here again, we have simply the sum of squares of the elements in e. But in this case no constraints are put on any of the columns of e.

One may wish to utilize the offdiagonal elements of E or the covariances of e in some function in determining X and A so as to optimize that function. In this case we can write

$$\epsilon = E - D_E \tag{4.3}$$

where D_E is the diagonal of E, and hence the diagonal of ϵ is zero. We may, for example, wish to determine A and X so as to optimize some function of the elements of ϵ. In particular, we might wish to minimize the sums of squares of the elements of ϵ. This means that we wish to minimize only the sums of squares of residual covariances.

More generally, we may write

$$\epsilon = E - PD_E \tag{4.4}$$

where P may be some value between zero and one. It has been customary in selecting a loss function to take P as either zero or 1, but there appears to be no compelling reason to restrict it to these two values.

In the maximum likelihood method of Lawley (1940), the canonical method of Rao (1955), the minres method of Harmon (1967), and the alpha method of Kaiser and Caffrey (1965), P in Eq. 4.4 is taken as 1. As a matter of fact, many investigators insist that only covariance matrix factoring procedures using P = 1 may be called factor analysis. This convention has the sanction of usage but whether it is important or even justifiable is debatable. These investigators call factoring procedures which take P as zero "principal component analysis."

4.2 The Loss Function

It is clear from the previous section that the mathematical function of the residual matrix we wish to optimize depends on whether we consider the elements of the residual matrix itself or the covariance matrix derived from it. We have seen also that the sum of squares of the elements of the residual matrix is the sum of the diagonal elements of the residual covariance matrix.

It will doubtless be simpler and more useful to discuss the loss function in terms of ϵ, given by Eq. 4.4, whose offdiagonal elements are the covariances of the residual matrix and whose diagonal elements are proportional to but not greater than the residual variances. We therefore restrict our consideration of the elements entering into the loss function to the elements of the covariance matrix of the residual matrix e, where the diagonal variances have been reduced by the proportionality constant P.

In Chapter 8 we show that X in Eq. 4.1 can be determined so that

$$e'e = C - AA' \tag{4.5}$$

where

$$C = Z'Z \tag{4.6}$$

From Eqs. 4.2 through 4.5 we have

$$C - PD_E = AA' + \epsilon \tag{4.7}$$

Suppose now we write Eq. 4.7 in basic structure form as

$$C - PD_E = Q_m \delta_m Q_m' + Q_\beta \delta_\beta Q_\beta' - Q_\gamma \delta_\gamma Q_\gamma' \tag{4.8}$$

where

$$\min(\delta_m) > \max(\delta_\beta) \tag{4.9}$$

$$\min(\delta_m) > \max(\delta_\gamma) \tag{4.10}$$

We let

$$A = Q_m \delta_m^{\frac{1}{2}} \tag{4.11}$$

Therefore from Eqs. 4.7, 4.8, and 4.11

$$\epsilon = Q_\beta \delta_\beta Q_\beta' - Q_\gamma \delta_\gamma Q_\gamma' \tag{4.12}$$

If P in Eq. 4.4 is zero, then δ_γ in Eq. 4.12 will also be zero, as can be seen from the developments in Chapter 8. If P = 1, then according to the definition of ϵ the diagonal of ϵ will be zero. Therefore we can show from Eq. 4.12

$$\text{tr } \delta_\beta = \text{tr } \delta_\gamma \tag{4.13}$$

Perhaps the most obvious function of ϵ in Eq. 4.12 to minimize is

$$\phi = \text{tr } \epsilon^2 \tag{4.14}$$

From Eqs. 4.12 and 4.14 we can show that

$$\text{tr } \epsilon^2 = \text{tr } \delta_\beta^2 + \text{tr } \delta_\gamma^2 \tag{4.15}$$

However, the criterion of approximation should probably not be a function alone of the residual matrix but also of the total variance. Therefore we choose as a more rational criterion

$$\phi = 1 - \frac{\text{tr } (\delta_\beta^2 + \delta_\gamma^2)}{\text{tr } (\delta_m^2 + \delta_\beta^2 + \delta_\gamma^2)} \tag{4.16}$$

But from Eqs. 4.7, 4.8, and 4.16 we get

$$\phi = \frac{\text{tr } \delta_\beta^2}{\text{tr } (C - D_E)^2} \tag{4.17}$$

As a matter of fact, ϕ as given by Eq. 4.17 is the loss function we seek to maximize in Chapter 8. As is pointed out there, this function has the useful property that its maximum value is unity.

4.3 The Loss Function in the Maximum Likelihood Method of Factor Analysis

The type of loss function which is optimized in the maximum likelihood method of factor analysis is much more complicated than the function given in Eq. 4.17, although the procedure is believed by some to provide useful criteria for indicating

the number of factors to be solved for. We shall indicate how we may express the function of the ϵ matrix in Eq. 4.12 which is maximized. If we let δ_{β_i} be the i'th element of δ_β and δ_{γ_i} the i'th element of δ_γ, then the function maximized in maximum likelihood factor analysis can be shown to be

$$\phi = \pi (1 + \delta_{\beta_i}) \pi (1 - \delta_{\gamma_i}) \tag{4.18}$$

where the continued products include all the elements of δ_β and δ_γ. Because of the particular scaling which, as indicated in Chapter 5, is utilized by maximum likelihood analysis, the number 1 occurs within the inside parentheses of Eq. 4.18.

4.4 The Maximum Likelihood Equations and the Loss Function

We shall here consider a discussion by Jöreskog (1967) which appears to be of particular importance in considering possible loss functions. Using Jöreskog's notation we let

$$\Psi = D_S - D_{\Lambda\Lambda'} \tag{4.19}$$

$$(\Psi^{-\frac{1}{2}} S \Psi^{-\frac{1}{2}}) \Psi^{-\frac{1}{2}} \Lambda = \Psi^{-\frac{1}{2}} \Lambda (I + \Lambda' \Psi^{-1} \Lambda) \tag{4.20}$$

Equations 4.19 and 4.20 correspond to Jöreskog's (1967) equations 24 and 14 respectively. He states that the maximum likelihood estimates of Λ and Ψ are defined as the matrices satisfying Eqs. 4.19 and 4.20 or some equivalent ones. From Eq. 4.20 we get

$$(\Psi^{-\frac{1}{2}} S \Psi^{-\frac{1}{2}} - I) \Psi^{-\frac{1}{2}} \Lambda = \Psi^{-\frac{1}{2}} \Lambda (\Lambda' \Psi^{-1} \Lambda) \tag{4.21}$$

Premultiplying Eq. 4.21 by $\Lambda' \Psi^{-\frac{1}{2}}$ we get

$$\Lambda' \Psi^{-\frac{1}{2}} (\Psi^{-\frac{1}{2}} S \Psi^{-\frac{1}{2}} - I) \Psi^{-\frac{1}{2}} \Lambda = (\Lambda' \Psi^{-1} \Lambda)^2 \tag{4.22}$$

From Eqs. 4.21 and 4.22

$$(\Psi^{-\frac{1}{2}} S \Psi^{-\frac{1}{2}} - I) \Psi^{-\frac{1}{2}} \Lambda (\Lambda' \Psi^{-\frac{1}{2}} S \Psi^{-\frac{1}{2}} - I) \Psi^{-\frac{1}{2}} \Lambda)^{-\frac{1}{2}} = \Psi^{-\frac{1}{2}} \Lambda \tag{4.23}$$

Except for notation and a square orthonormal transformation h, this equation is identical to our equation 8.67 in Chapter 8.

But note that Eq. 4.20 can be obtained very simply without the use of the calculus. We write

$$S - \Lambda\Lambda' = E \tag{4.24}$$

We let

$$\Psi = \text{Diag}\,(E) \tag{4.25}$$

From Eq. 4.24 we get

$$\Psi^{-\frac{1}{2}}(S - \Lambda\Lambda')\Psi^{-\frac{1}{2}} = \Psi^{-\frac{1}{2}}E\Psi^{-\frac{1}{2}} \tag{4.26}$$

Let

$$\epsilon = \Psi^{-\frac{1}{2}}E\Psi^{-\frac{1}{2}} \tag{4.27}$$

From Eqs. 4.25 and 4.27

$$\text{Diag}\,(\epsilon) = I \tag{4.28}$$

From Eqs. 4.26 and 4.27

$$\Psi^{-\frac{1}{2}}(S - \Lambda\Lambda')\Psi^{-\frac{1}{2}} \cdot I = \epsilon - I \tag{4.29}$$

Let us require that Λ be determined so that

$$(\epsilon - I)\Psi^{-\frac{1}{2}}\Lambda = 0 \tag{4.30}$$

From Eqs. 4.29 and 4.30

$$(\Psi^{-\frac{1}{2}}(S - \Lambda\Lambda')\Psi^{-\frac{1}{2}} - I)\Psi^{-\frac{1}{2}}\Lambda = 0 \tag{4.31}$$

From Eq. 4.31

$$(\Psi^{-\frac{1}{2}}S\Psi^{-\frac{1}{2}})\ \Psi^{-\frac{1}{2}}\Lambda = \Psi^{-\frac{1}{2}}\Lambda\,(I + \Lambda'\Psi^{-1}\Lambda) \tag{4.32}$$

which is the same as Eq. 4.20. If it is true that the maximum likelihood estimates of Λ and Ψ are defined as the equations which satisfy Eqs. 4.19 and 4.20, or

Jöreskog's (1967) equations 14 and 24, then they can also be defined as the estimates which satisfy our Eq. 4.30.

Let us now write in basic structure form

$$(\Psi^{-\frac{1}{2}}S\Psi^{-\frac{1}{2}} - I) = Q_m \delta_m Q_m' + Q_\beta \delta_\beta Q_\beta' - Q_\gamma \delta_\gamma Q_\gamma' - q I_\sigma q' \tag{4.33}$$

where δ_m are the m largest roots of $\Psi^{-\frac{1}{2}}S\Psi^{-\frac{1}{2}} - I$, δ_β are the next β largest, and $-\delta_\gamma$ are the γ negative roots. If we let

$$\Psi^{-\frac{1}{2}}\Lambda = Q_m \delta_m^{\frac{1}{2}} \tag{4.34}$$

then

$$\text{tr } \Lambda'\Psi^{-\frac{1}{2}}(\Psi^{-\frac{1}{2}}S\Psi^{-\frac{1}{2}} - I)\Psi^{-\frac{1}{2}}\Lambda = \text{tr } \delta_m^2 \tag{4.35}$$

and

$$\text{tr } (\Psi^{-\frac{1}{2}}S\Psi^{-\frac{1}{2}} - I)^2 = \text{tr } (\delta_m^2 + \delta_\beta^2 + \delta_\gamma^2) \tag{4.36}$$

and the solution 4.23 minimizes

$$1 - \frac{\text{tr } (\epsilon - I)^2}{\text{tr } (\Psi^{-\frac{1}{2}}S\Psi^{-\frac{1}{2}} - I)^2} = \phi \tag{4.37}$$

This is equivalent to our equation 4.17 for the particular scaling of Eq. 4.19. If, as Jöreskog (1967) maintains, Eqs. 4.37 and 4.18 are not simultaneously minimized, then Eqs. 4.19 and 4.20 may be regarded only as necessary but not sufficient conditions to satisfy the maximum likelihood criterion. It could therefore not be said that all estimates of Λ and Ψ which satisfy Eqs. 4.19 and 4.20 are maximum likelihood estimates of these matrices even if the inequalities of Eqs. 4.9 and 4.10 are satisfied.

CHAPTER 5

SCALE FREE SCALING

5.1 Introduction

We have considered in Chapter 3 the case of scaling by entities or by attributes or both prior to matrix approximation procedures. We shall see in Chapter 8 that certain factor analysis procedures have an important invariance property with reference to the original scaling of the variables. These are called scale free methods. Actually the methods are not scale free because they involve or imply specific scaling procedures. But without loss of generality it is shown in Chapter 8 that for these methods we may begin with a data matrix of standardized scores or any other scaling. We have seen that the general matrix approximation equation is of the form

$$Z - XA' = e \tag{5.1}$$

where Z is the data matrix, X is the factor score matrix, A is the factor loading matrix, and e is the residual matrix. We have already considered the residual covariance matrix E which we may write

$$E = e'e \tag{5.2}$$

From Eqs. 5.1 and 5.2 we get

$$E = Z'Z - Z'XA' - AX'Z + AX'XA' \tag{5.3}$$

In Chapter 8 we show how X may be solved for so that for some determination of A we have

$$Z'XA' + AX'Z - AX'XA' = AA' \tag{5.4}$$

Therefore we may have from Eqs. 5.3 and 5.4

$$E = Z'Z - AA' \tag{5.5}$$

If we let

$$C = Z'Z \tag{5.6}$$

$$W = AA' \tag{5.7}$$

we have from Eqs. 5.5, 5.6, and 5.7

$$C - W - E = 0 \tag{5.8}$$

We may now designate the three terms in Eq. 5.8 as follows: We may call C the total covariance matrix, W the estimated covariance matrix, and E the residual covariance matrix. The data matrix X may be scaled in any way we please, including, of course, the original or arbitrary units of measurement yielded by the experimental procedures. For each of the covariance matrices in Eq. 5.8, we may consider the corresponding diagonal matrices D_C, D_W, and D_E, constructed from the diagonal elements of the covariance matrices. The general problem is to determine the A matrix so as to satisfy some constraint on some function of the elements of E. But since the scaling of the original variables has been arbitrary, we may insist that the determination of W be based on some rescaling of the variables. Any rescaling of the variables will of course affect the variances in the diagonals of the variance matrices in Eq. 5.8. Let us now consider a scaling matrix D and write from Eq. 5.8

$$D\,(C - W - E)\,D = 0 \tag{5.9}$$

Let

$$Y = DCD \tag{5.10}$$

$$\omega = DWD \tag{5.11}$$

$$\epsilon = DED \tag{5.12}$$

From Eqs. 5.9 through 5.12 we have

$$Y - \omega - \epsilon = 0 \tag{5.13}$$

Now for the diagonal matrices corresponding to the rescaled covariance matrices given by the terms of Eq. 5.13, we have D_Y, D_ω, and D_ϵ. These diagonal matrices consist of the rescaled variances of the total, the estimated, and the residual variances respectively. Let us consider now some interesting possibilities for the selection of the scaling matrix D.

5.2 Total Variance Scaling

We may determine D so that the variances of the total covariance matrix Y are all unity. If C has been calculated from the deviation data matrix X, its off diagonal elements are the covariances among the original variables and the diagonal elements are the variances. Therefore, if we wish to have

$$D_Y = I \tag{5.14}$$

it is obvious that we must have

$$D = D_C^{-\frac{1}{2}} \tag{5.15}$$

Therefore we have from Eqs. 5.10, 5.11, 5.12, and 5.15

$$Y = D_C^{-\frac{1}{2}} C D_C^{-\frac{1}{2}} \tag{5.16}$$

$$\omega = D_C^{-\frac{1}{2}} W D_C^{-\frac{1}{2}} \tag{5.17}$$

$$\epsilon = D_C^{-\frac{1}{2}} E D_C^{-\frac{1}{2}} \tag{5.18}$$

It is clear, therefore, that Y is simply the familiar matrix of correlation coefficients. This is of course the matrix from which traditionally most factor analyses have proceeded. It is the basis of most of the classical principal component analyses and more recently the minres analysis of Harmon (1967).

5.3 Estimated Variance Scaling

There is no compelling reason, however, for choosing the total variance scaling. We may wish to choose D so that the variances of the estimated covariance matrix W are all unity. This means that the off diagonal elements of the rescaled estimated covariance matrix ω are correlation coefficients. If we wish to have

$$D_\omega = I \tag{5.19}$$

we must have

$$D = D_W^{-\frac{1}{2}} \tag{5.20}$$

and we have from Eqs. 5.10, 5.11, 5.12, and 5.20

$$\gamma = D_W^{-\frac{1}{2}} C D_W^{-\frac{1}{2}} \tag{5.21}$$

$$\omega = D_W^{-\frac{1}{2}} W D_W^{-\frac{1}{2}} \tag{5.22}$$

$$\epsilon = D_W^{-\frac{1}{2}} E D_W^{-\frac{1}{2}} \tag{5.23}$$

If we substitute from Eq. 5.7 into Eq. 5.22 we have

$$\omega = D_W^{-\frac{1}{2}} A A' D_W^{-\frac{1}{2}} \tag{5.24}$$

We may let

$$\alpha = D_W^{-\frac{1}{2}} A \tag{5.25}$$

Now α is the factor loading matrix corresponding to the estimated variance scaling. It has the interesting property that the sum of squares of the factor loadings for each variable is unity. This scaling is used in the alpha factor analysis of Kaiser and Caffrey (1965) and in the communality scaling which we have discussed elsewhere (Horst, 1965).

5.4 Residual Variance Scaling

Instead of choosing D so that the total or the estimated variances are all unity, we may wish to choose it so that the residual variances are all unity. In this case, the rescaled residual covariance matrix would have correlations for offdiagonal elements. Here we select D so that

$$D_\epsilon = I \tag{5.26}$$

Therefore we must have

$$D = D_E^{-\frac{1}{2}} \tag{5.27}$$

and we have from Eqs. 5.10, 5.11, 5.12, and 5.27

$$\gamma = D_E^{-\frac{1}{2}} C D_E^{-\frac{1}{2}} \tag{5.28}$$

$$\omega = D_E^{-\frac{1}{2}} W D_E^{-\frac{1}{2}} \tag{5.29}$$

$$\epsilon = D_E^{-\frac{1}{2}} E D_E^{-\frac{1}{2}} \tag{5.30}$$

This scaling is used in the maximum likelihood factor analysis procedures of Lawley (1940) and the canonical factor analysis procedures of Rao (1955).

5.5 The Generalized Scale Free Method

We have seen that as special cases we may scale the data matrix so that the total, the estimated, or the residual covariance matrix is a correlation matrix. For each case we may begin with an arbitrary scaling. Therefore the three methods are called scale free. It is clear, however, from Eq. 5.8 that the total covariance matrix C is by definition the sum of the estimated and the residual covariance matrices. Therefore the total variance diagonal matrix is simply the sum of the estimated and the residual variance diagonal matrices. We may therefore consider a more general case of scaling in which the scaling matrix D is the reciprocal square root of a weighted sum of the estimated and the residual variance matrices. We may let

$$D = P_W D_W + P_E D_E)^{-\frac{1}{2}} \tag{5.31}$$

where P_W and P_E are weighting scalars. Suppose we let p be a value such that

$$0 \leq p \leq 1 \tag{5.32}$$

and

$$q = 1 - p \tag{5.33}$$

We now let

$$P_W = \frac{p}{1 - 2pq} \tag{5.34}$$

$$P_E = \frac{q}{1 - 2pq} \tag{5.35}$$

If we take p = .5 and substitute in Eqs. 5.34 and 5.35 respectively, we get

$$P_W = 1 \tag{5.36}$$

$$P_E = 1 \tag{5.37}$$

From Eq. 5.8 we have

$$D_C = D_W + D_E \tag{5.38}$$

From Eqs. 5.36, 5.37, and 5.38 in Eq. 5.31 we get

$$D = D_C^{-\frac{1}{2}} \tag{5.39}$$

which is the same as Eq. 5.15.

If we take p = 1, we get from Eqs. 5.31, 5.33, 5.34, and 5.35

$$D = D_W^{-\frac{1}{2}} \tag{5.40}$$

which is the same as Eq. 5.20.

If we take p = 0, we get from Eqs. 5.31, 5.33, 5.34, and 5.35

$$D = D_E^{-\frac{1}{2}} \tag{5.41}$$

which is the same as Eq. 5.27.

We see therefore that by taking the special cases for p = .5, 1, and 0 the scaling matrix given by Eq. 5.31 gives the scaling procedure utilized in various factor analytic rationales considered by previous investigators. However, we may let p take any value in the range indicated by Eq. 5.32 and the use of the scaling matrix D can still be regarded as a scale free procedure. This generalization of the scaling matrix will be developed more fully in Chapter 8..

CHAPTER 6

SIMPLE STRUCTURE

6.1 Criteria of Simple Structure

We have seen in Chapter 5 that we may write the matrix approximation equation in the form

$$Z - XA' - e = 0 \tag{6.1}$$

where Z is the data matrix, X is the factor score matrix, A is the factor loading matrix, and e is the residual matrix. We have also specified that X and A are basic and their common order is less than either order of Z. We have said that for any given A we shall define X so that the residual covariance matrix $e'e = E$ is given by

$$C - AA' = E \tag{6.2}$$

where

$$C = Z'Z \tag{6.3}$$

In Chapter 9 we show that the number of pairs of factors yielding the product XA' is infinite. We wish to put some restriction on A so that the solution becomes unique. We may consider some specified solution to X and A that optimizes a specified loss function, as discussed in Chapter 4. Further then, we may consider a square basic transformation X of A such that

$$B = Ah \tag{6.4}$$

We may now require that h be uniquely determined so that the elements of B will satisfy some predetermined set of criteria. This general problem was first considered by Thurstone (1947). He specified that the structure of the matrix B should be as simple as possible. He listed what he regarded as the criteria of simplicity. This concept he called simple structure. His criteria of simple structure were:

1. Each row of the factor matrix B should have at least one zero.
2. Each column of B should have at least m zero factor loadings, where m is the number of factors.

3. For every pair of columns of B there should be several tests whose entries vanish in one column but not the other.
4. For every pair of columns of B a large proportion of the tests should have zero entries in both columns.
5. For every pair of columns there should preferably be only a small number of tests with nonvanishing entries in both columns.

One of the difficulties with these criteria is that they are not stated in precise mathematical terms. Such statements would be necessary in order that mathematical functions could be optimized. However, it is possible to formulate mathematical functions of the elements of B such that, given A, the matrix h can be solved for which optimizes these functions. As indicated in Chapter 9, many attempts have been made to incorporate the consequences of at least some of Thurstone's criteria into mathematical functions which can be optimized by suitable determinations of the elements of h in Eq. 6.4. In addition to those that have been proposed by others, we present an analytical procedure in Chapter 9 that appears to have some advantages over others previously available. The chief advantage of the method is that it appears to work with a great variety of correlation matrices and factor analytic procedures.

6.2 Scaling of the Arbitrary Factor Matrix

Several problems arise in the transformation of an arbitrary factor matrix A to a simple structure matrix B, irrespective of what method of solution for A has been used and what rationale for determining the transformation matrix h is adopted. One of these concerns the scaling of the factor loading matrix A prior to transformation. Suppose we have a factor loading matrix A determined in some manner. Many methods are now available. A number of these we have considered in detail elsewhere (Horst, 1965). In Chapter 8 we discuss a general approach. Six specific cases of the general approach are identical or similar to methods that have

previously been proposed by others. In any case, with the great variety of methods available it is to be expected that the resulting A matrices could differ greatly for the same data. In particular, we may consider the diagonal matrix

$$D_{AA'} = \text{Diag}\,(AA') \tag{6.5}$$

Equation 6.5 may be regarded as the diagonal matrix of the estimated variances. In some contexts its elements are called the comunalities. Now these communalities may vary greatly not only from one method of analysis to another but also from one element to another for any given method of analysis. It is to be expected that in the sclution of B in Eq. 6.4 the variables witn the smallest communalities will have the least influence in the determination of h and hence the elements of B. The weight that a variable can have in the determination of h is then a function of its communality. It has been argued therefore that, for any function purporting to optimize simple structure criteria, the arbitrary matrix A should be rescaled by rows prior to the application of the analytical simple structure procedures. We may therefore write

$$\alpha = DA \tag{6.6}$$

and let

$$\beta = \alpha h \tag{6.7}$$

The simple structure criteria are now sought for β instead of B in Eq. 6.4. A reasonable rationale which has been rather generally adopted is that in simple structure solutions each variable should be given equal weight. This means that we should have

$$D_{\alpha\alpha'} = I \tag{6.8}$$

From Eqs. 6.6 and 6.8 therefore we have

$$\alpha = D_{AA'}^{-\frac{1}{2}} A \tag{6.9}$$

This simply means that the arbitrary factor matrix A is normalized by rows prior to transformation.

The usual procedure after the β matrix has been solved for is to descale the β matrix back to the B matrix by the equation

$$B = D_{AA'}^{2}\beta \tag{6.10}$$

However, one of the chief arguments in favor of the simple structure concept has been that not only does it provide a unique solution for the factor loading matrix but it also facilitates interpretation of the tests and the factors. This latter claim appears to have been well substantiated over the years, giving considerable justification for the taxonomic objectives of factor analysis. For purposes of interpretation it is still possible that the β matrix rather than the B matrix is generally more useful. However, complications arise when one attempts to use the β matrix in the solution of the factor score matrix. This topic is considered in Chapters 7 and 10.

It is of interest to note that the estimated variance scaling discussed in Chapter 5 can yield directly an A matrix whose rows are by definition normal vectors, i.e.

$$D_{AA'} = I \tag{6.11}$$

6.3 The Transformation Matrix

There has been considerable disagreement about constraints on the h matrix to be imposed in the simple structure solution. In general, most investigators agree that the matrix should be normal by columns so that

$$D_{h'h} = I \tag{6.12}$$

Some investigators require further that h be orthonormal so that

$$h'h = I \tag{6.13}$$

This issue has been considered at length by Harmon (1967) and Horst (1965). One advantage of the orthonormal constraint is that, for some types of solutions, uncorrelated or orthonormal factor scores will results. Another is that we have the equality

$$BB' = AA' \tag{6.14}$$

so that Eq. 6.2 can be written

$$C - BB' = E \tag{6.15}$$

Perhaps the chief advantage of relaxing the orthogonality constraint on h is that a more clearcut simple structure results in the B matrix and that the factors become more readily interpretable. This implies that the taxonomic objectives of factor analysis are more readily achieved by the oblique or nonorthogonal transformation than when orthogonality is imposed.

A disadvantage of the oblique procedures is that in general they have been much less satisfactory from a computational point of view. By far the best known and most used orthogonal procedure is the varimax method of Kaiser (1958) or variants of it (Horst, 1965).

Another disadvantage of the oblique procedures is that frequently it is difficult to keep one or more of the factors from collapsing into other factors. Nevertheless, it is probable that the constraint of orthonormality on the transformation matrix h is undesirably restrictive. Chapter 9 presents a method that does not impose this constraint and appears to work well with different types of data.

6.4 The Problem of Signs

One of the problems encountered in simple structure transformation procedures has to do with sign changes. Unfortunately, the importance of this problem has not been generally recognized. The sign problem has two distinct aspects. Suppose we have given a simple structure matrix B as in Eq. 6.4, obtained by one of the analytical methods available. Most of these methods optimize some function of the squares of the elements of B. Such a matrix of squared elements we may indicate by

$$b = B^{(2)} \tag{6.16}$$

where the superscript (2) means that each of the elements in B has been squared. Suppose now we let

$$\beta = i_L \, B \, i_R \tag{6.17}$$

where i_L and i_R are sign matrices. It is clear that whatever the i matrices we will have

$$b = \beta^{(2)} \tag{6.18}$$

$$\beta^{(2)} = B^{(2)} \tag{6.19}$$

Therefore for those methods of transformation which optimize a function of the elements of b, the corresponding matrix B may still require a pre- and post-multiplication by optimal sign matrices i_L and i_R respectively to give meaningful and interpretable simple structure factor loadings.

We indicated at the beginning of this chapter that the sign problem has two aspects. We may see now that one of these is the postmultiplier and one the pre-multiplier. Many investigators have found that in using an available transformation procedure some variables that should obviously have high positive loadings actually have high negative loadings and vice versa. It has been proposed that in such cases one merely reverses the sign of the loadings for all elements in the factor vector where the loadings of wrong signs occur. Certainly one may reverse signs of all elements in a given column of a matrix without affecting in any way the major product moment of the matrix. This would appear to be obvious but it is frequently overlooked. One is therefore at liberty to reverse signs by columns in either the arbitrary factor matrix A or the simple structure matrix B. But in many cases one cannot get rid of all high negative values in a column of B by reversing the signs of all elements in it, for the simple reason that the column may have both high positive and negative values in it.

This brings us then to a consideration of the left sign matrix i_L. Thurstone (1947) has emphasized that the simple structure concept and the positive manifold concept are independent. He defines the positive manifold simple structure factor matrix as one which has all positive elements or one in which the negative elements

are small in absolute magnitude. It is, however, possible in most cases to approximate the positive manifold for the simple structure factor matrix by appropriate pre- and postmultiplications by sign matrices. The interpretation of the left sign matrix may now be clarified. We have seen that the major product moment of the factor loading matrix is invariant with respect to postmultiplication of the factor loading matrix by a sign matrix. Suppose now in Eq. 6.1 we postmultiply by the sign matrix i_L. This gives

$$Zi_L - XA'i_L - ei_L = 0 \tag{6.20}$$

From Eqs. 6.2 and 6.3 we can also write

$$i_L Z'Zi_L - i_L AA'i_L = i_L Ei_L \tag{6.21}$$

Also from Eqs. 6.3 and 6.15 we may write

$$i_L Z'Zi_L - i_L BB'i_L = i_L Ei_L \tag{6.22}$$

Now it can be shown that any of the loss functions we have considered in Chapter 4 are invariant with respect to a pre- and postmultiplication of the residual variance matrix E by a sign matrix. We see further also from the first term on the left of Eqs. 6.20, 6.21, and 6.22 that premultiplication of either the A or B matrix by i_L implies postmultiplication of the data matrix by the same sign matrix. Suppose then we find an i_L and an i_R matrix in Eq. 6.17 which according to some acceptable criterion gives a best approximation to a positive manifold. We may then interpret i_L as a matrix that indicates by the position of its negative elements the columns in the data matrix Z whose elements should have their signs reversed. Such situations are encountered in the factor analysis of personality test items and other variables where the direction of the scale is not clear and has been arbitrarily specified by the scoring procedure.

There is, however, still some ambiguity in the determination of the i_L and i_R matrices. Assume that the i matrices have been determined to give a satisfactory

positive manifold for β in Eq. 6.17. We can write Eq. 6.17

$$\beta = (-i_L)\ B\ (-i_R) \tag{6.23}$$

The question then arises as to whether we should use the i matrices as given by some optimizing procedure or reverse the signs for both i matrices. The interpretation of the factors is usually based on an inspection of the variables having high loadings in them. It is immaterial whether for the right multiplier we use i_R or $-i_R$. For the left multiplier we must then decide whether interpretation will be simpler by reversing the scoring as indicated by i_L or $-i_L$. If one has some good a priori basis for deciding which is the "low" and which the "high" end of the scale for each variable and has provided scoring procedures accordingly, then presumably there should be very few negatives in i_L. In general, for lack of a better criterion, one would choose that i_L or $-i_L$ which has the fewest negatives in it and then choose the corresponding i_R or $-i_R$.

In any case, one may not willy-nilly change the signs of individual factor loadings to suit his fancy or preconceived notions. This procedure is not uncommon and is completely invalid. If the factor loading matrix is small and the simple structure clearcut, it is frequently possible to determine by inspection the optimal i matrices for approximating the positive manifold. However, for large numbers of variables and factors, inspectional procedures are impractical and objective mathematical and computational procedures are needed. Two of these we have given elsewhere (Horst 1965, 1968a), and the method of Chapter 9 attempts to take care of the sign problem.

CHAPTER 7

THE FACTOR SCORE MATRIX

7.1 The Role of Factor Scores

We have seen in previous chapters that the data matrix may be approximated by a lower rank matrix which is the major product of two basic matrices, one of which may be regarded as the factor score matrix and the other the factor loading matrix. We express this relationship by

$$Z - XA' - e = 0 \tag{7.1}$$

where as in previous chapters Z is the data matrix, X the factor score matrix, A the factor loading matrix, and e the residual matrix. The number of columns in X and A are presumed to be much less than in Z. Traditionally, there has been much greater interest in the determination of A or some transformation of it, B, as discussed in Chapters 6, 8, and 9, than in the matrix X. A study of the matrix B has been thought to yield interesting and useful information about the fundamental or "primary" variables of a scientific discipline. Equation 7.1 implies that the data matrix for a group of persons with respect to observed attributes can be approximated by appropriate linear combinations of a much smaller number of attributes. We have seen in Chapter 6 that the matrix A is usually transformed into a simple structure matrix B by some transformation matrix h so that

$$B = Ah \tag{7.2}$$

Now for Eq. 7.1 to hold identically when A is replaced by B, we first write

$$Z - Xh'^{-1}h'A' - e = 0 \tag{7.3}$$

If we let

$$Y = Xh'^{-1} \tag{7.4}$$

and use Eqs. 7.2 and 7.4 in Eq. 7.3, we get

$$Z - YB' - e = 0 \tag{7.5}$$

It is clear then that

$$YB' = XA' \tag{7.6}$$

A more general treatment is given in Chapter 9. But in any case, we may now regard Y in Eq. 7.5 as the simple structure factor score matrix. If h is a square orthonormal matrix as a special case, such as in the Kaiser (1958) varimax, then we have simply

$$Y = Xh \tag{7.7}$$

Although the major interest has traditionally been in the simple structure factor loading matrix B, in recent years much interest in the Y matrix of simple structure factor scores has also been growing. This is true not only in psychology where the factor techniques had their origin and greatest development but also in other scientific disciplines concerned about the basic or primary attributes of particular sets of entities under study, such as geographical units, educational institutions, members of governmental bodies, and so on. It seems reasonable that if one can discover or define adequately a relatively small number of primary variables of a discipline, then it could be useful to estimate the values of these variables from a much larger number of observed and arbitrarily defined variables. Such a procedure could yield a much more parsimonious characterization of the entities under study almost as completely as a much larger number of observed variables.

Furthermore, these primary variables could characterize the entities in terms that are objectively established by the techniques. This can make for a more objective, parsimonious, and unambiguous taxonomy as a basis for characterization and classification of entities or individuals within areas of human interest or activity.

But aside from the use of factor scores as a basis for parsimonious and unambiguous characterization of entities, these scores can also be utilized for increasing the accuracy of statistical prediction in a wide variety of situations and settings. The use of factor measures in prediction techniques has been considered by Horst (1941, 1965), Leiman (1951), and Burket (1964).

7.2 Estimation of Factor Scores

In Chapter 10 we shall consider in some detail the technical problems involved and procedures for estimating factor scores from the data matrix and the simple structure factor loading matrix. Rationales and computational procedures for calculating the simple structure factor loading matrix are presented in Chapters 8 and 9. Here we shall indicate some of the conditions that might be satisfied by the factor score matrix. Most of the work done in this area has been concerned with the X factor score matrix in Eq. 7.1 rather than the simple structure factor score matrix Y in Eq. 7.5. This work has been reviewed and amplified by Harris (1967) and by McDonald and Burr (1967). The treatment in both of these presentations has considered methods of approximating the X factor score matrix in Eq. 7.1 rather than the Y simple structure factor score matrix in Eq. 7.5. But it can readily be seen that if we have solved for the X factor score matrix and the simple structure factor loading matrix transformation h, we can solve for the simple structure factor score matrix by means of Eq. 7.4. We shall therefore consider the principles that appear relevant in determining the factor score matrix X. In the methods discussed by Harris and by McDonald and Burr, there is a confounding of estimation methods and scaling method. It is important that these be kept clearly separate. Harris lists five methods that have been proposed. Recalling that D_E is the diagonal matrix of residual variances, these methods are

$$X_1 = ZR^{-1}A \tag{7.8}$$

$$X_2 = ZD_E^{-1}A\,(A'D_E^{-1}A)^{-1} \tag{7.9}$$

$$X_3 = ZA(A'A)^{-1} \tag{7.10}$$

$$X_4 = ZA \tag{7.11}$$

$$X_5 = ZD_E^{-1}A(A'D_S^{-1}RD_E^{-1}A)^{-\frac{1}{2}} \tag{7.12}$$

Suppose now in Eq. 7.1 we assume that

$$Z'Z = R \tag{7.13}$$

where R is a correlation matrix. Let us then consider a postmultiplication of Eq. 7.1 by some scaling matrix D, thus:

$$(Z - XA' - e)\, D = 0 \tag{7.14}$$

From Eq. 7.14 we have

$$ZD - XA'D - eD = 0 \tag{7.15}$$

Let

$$ZD = U \tag{7.16}$$

$$DA = \alpha \tag{7.17}$$

$$eD = \epsilon \tag{7.18}$$

From Eqs. 7.16, 7.17, and 7.18 in Eq. 7.15 we have

$$U - X\alpha' - \epsilon = 0 \tag{7.19}$$

First we note that Eqs. 7.9 and 7.10 are not essentially different, for it can be readily shown that Eq. 7.10 minimizes tr $\epsilon'\epsilon$ for D = I and Eq. 7.9 minimizes this trace for $D = D_E$. This property of minimizing the sum of squares of residuals (or weighted residuals) has been regarded as a desirable property of the factor score matrix.

Next we note that Eq. 7.8 is independent of any scaling matrix D.

We let

$$X = U\,(U'U)^{-1}\alpha \tag{7.20}$$

From Eqs. 7.13, 7.16, and 7.17 in Eq. 7.20, we have

$$X = ZD(DRD)^{-1}DA \tag{7.21}$$

which becomes

$$X = ZR^{-1}A \tag{7.22}$$

and this is the same as Eq. 7.8. This property of independence of scale of a factor score matrix may also be regarded as desirable.

Harris (1967) regards Eq. 7.11 as a sort of "quick and dirty" method of estimating the factor score matrix. Be that as it may, this method may be generalized to a scaled data matrix by

$$X = ZDA \tag{7.23}$$

In particular, we may have $D = D_E^{-1}$ so that Eq. 7.23 becomes

$$X = ZD_E^{-1}A \tag{7.24}$$

The forms 7.11 and 7.23 appear to have little to recommend them. However, we shall see presently what happens when we consider another property of the factor score matrix which has been regarded as desirable. This is that X shall be orthonormal or

$$X'X = I \tag{7.25}$$

First, let us rewrite Eqs. 7.8, 7.10, and 7.11 in more general scaled form as the three equations

$$X_1 = ZR^{-1}A \tag{7.26}$$

$$X_2 = ZDA(A'DA)^{-1} \tag{7.27}$$

$$X_3 = ZDA \tag{7.28}$$

remembering that X_1 is independent of scale. Suppose now we try to find the best approximations to these three factor score matrices in the least square sense which satisfy Eq. 7.25. It is well known that these orthonormal approximation matrices are of the form

$$X = X(X'X)^{-\frac{1}{2}} \tag{7.29}$$

We may therefore write the three orthonormal approximations to Eqs. 7.26, 7.27, and 7.28 respectively as

$$X_4 = ZR^{-1}A(A'R^{-1}A)^{-\frac{1}{2}} \tag{7.30}$$

$$X_5 = ZDA(A'DA)^{-1}((A'DA)^{-1}A'DRDA(A'DA)^{-1})^{-\frac{1}{2}} \tag{7.31}$$

$$X_6 = ZDA(A'DRDA)^{-\frac{1}{2}} \tag{7.32}$$

It is interesting that Eq. 7.32 becomes identical to Eq. 7.12, which was given by Anderson and Rubin (1956), when D is taken as D_E^{-1}. Hence it appears that the "quick and dirty" method may be made sophisticated by means of residual variance scaling and least square orthonormalization.

The forms 7.30 and 7.31 have not to our knowledge been previously presented and these properties have not been investigated. It can, however, be shown that a square orthonormal matrix q exists such that

$$X_5 h = X_6 \tag{7.33}$$

and that h is given by

$$h = ((A'DA)^{-1}(A'DRDA)(A'DA)^{-1})^{\frac{1}{2}}ADA(A'DRDA)^{\frac{1}{2}} \tag{7.34}$$

It should now be clear that the estimates X_1 and X_4 of Eqs. 7.26 and 7.30 respectively are independent of scale, and the estimates of X_2, X_3, X_5, and X_6 of Eqs. 7.27, 7.28, 7.31, and 7.32 respectively depend on the scaling matrix D. In Chapter 5 we have considered the generalized scale free scaling matrix which is the basis of the scale free methods discussed in Chapter 8. The matrix D in these latter four estimates of X may be taken as the matrix of Eq. 5.31 of Chapter 5 where the parameter p takes any value between 0 and 1. In particular, we can have one of the three scalings

$$D = D_C^{-\frac{1}{2}}$$

$$D = D_W^{-\frac{1}{2}}$$

$$D = D_E^{-\frac{1}{2}}$$

which are discussed in Chapter 5.

7.3 Other Desirable Properties of the Factor Score Matrix

We have already suggested that one desirable property of the factor score matrix X is that it be scale free. Another property we have considered is that the sum of squares (or weighted sums of squares) of the residual matrix elements be minimized. This property implies that the residual matrix is orthogonal to the factor loading matrix, that is,

$$eDA = 0 \tag{7.35}$$

where D may be the identity.

However, this condition is not consistent with the variable loss function where P in Eq. 4.4 of Chapter 4 is other than zero. Therefore it is of questionable value except for this special case which yields the so-called principal components factor loading matrix. That the factor score matrix X should be orthonormal seems desirable but perhaps not at the cost of other properties.

Perhaps the most important property of the matrix X is that for a given A the covariance matrix of the residual matrix be given by

$$C - AA' - e'e = 0 \tag{7.36}$$

This is the solution proposed in Chapter 5, and the solution for the matrix that satisfies this condition is given in Chapter 8 and considered further in Chapter 10.

In addition to the condition implied in Eq. 7.36, it is also desirable that the factor score matrix be orthgonal to the residual matrix e. If this condition is satisfied, then we have from Eq. 7.1

$$X'(Z - XA') - X'e = 0 \tag{7.37}$$

If we have also that X is orthonormal as indicated in Eq. 7.25, then we have from Eq. 7.37

$$X'Z = A' \tag{7.38}$$

If Z is such that Eq. 7.13 holds, then the left side of Eq. 7.38 is a matrix of correlations of the factor scores with the test scores or observed variables and the factor loading matrix A can be interpreted directly as a matrix of these correlations. There has been much wringing of hands over the decades that factor scores cannot be calculated but only estimated. More recently, Guttman (1955b), Harris (1967), and others have recognized that the "true" factor scores cannot be uniquely calculated. Presumably "true" scores are those which satisfy Eqs. 7.37 and 7.38. It is surprising that the problem of uniqueness has been so frightening when so many have so courageously and ingeniously and profitably attacked the nonuniqueness problem for the factor loading matrix by the various simple structure transformation approaches. In Chapter 10, we suggest an approach to the uniqueness problem for "true scores."

A topic of considerable interest concerns covariance matrices involving the various proposed estimation methods. Those involving the X matrices in Eqs. 7.8 through 7.12 have been presented by Harris (1967) and by McDonald and Burr (1967). We shall not review them here. However, in Chapter 10 the covariance relationships involving the factor score matrices considered there will be presented.

We also leave to Chapter 10 a discussion of the covariance properties of the simple structure factor score matrices derived from the two types of factor score matrices derived in that chapter.

CHAPTER 8

GENERALIZED SCALING AND LOSS FUNCTION

8.1 The Residual Matrix

Suppose we let Z be an N x n basic vertical data matrix. We need make no assumptions about transformations applied to the raw data matrix which have yielded Z′ but it will be convenient to assume that transformations have been made such that

$$R = Z'Z \tag{8.1}$$

where R is the correlation matrix. We now consider an N x m basic matrix X where $m < n$, and an n x m basic matrix A. We indicate the vertical major product of these two matrices by U so that

$$U = XA' \tag{8.2}$$

Then U is of the same order as Z. Since A and X are both basic and their common order is m, U is of rank m and therefore nonbasic. Let us assume now that A and X are to be determined so that U is in some sense, to be subsequently specified, an approximation to Z. We then write the residual matrix e as

$$e = Z - U \tag{8.3}$$

We shall determine X and A so as to optimize some function of the elements of e in Eq. 8.3. More specifically, we shall begin by considering the covariance matrix E of e which is given by

$$E = e'e \tag{8.4}$$

From Eqs. 8.3 and 8.4

$$E = Z'Z - Z'U - U'Z + U'U \tag{8.5}$$

From Eq. 8.1, 8.2, and 8.5

$$E = R - Z'XA' - AX'Z + AX'XA' \tag{8.6}$$

Without at once specifying the solution for A, we shall require that the solution for X shall be some function of Z and A such that

$$AA' = Z'XA' + AX'Z - AX'XA' \tag{8.7}$$

We have then from Eqs. 8.6 and 8.7

$$E = R - AA' \tag{8.8}$$

8.2 The Factor Score Matrix

Next we shall consider the solution for X which satisfies Eq. 8.7.

We let

$$\sigma = A'R^{-1}A \tag{8.9}$$

We indicate the basic structure of σ by

$$Q_\sigma d_\sigma^2 Q_\sigma' = \sigma \tag{8.10}$$

and let

$$\Delta = (I - (I - d_\sigma^2)^{\frac{1}{2}}) d_\sigma^{-2} \tag{8.11}$$

Then the solution for X which satisfies Eq. 8.7 is

$$X = ZR^{-1}AQ_\sigma \Delta Q_\sigma' \tag{8.12}$$

To show that the solution 8.12 for X does satisfy Eq. 8.7 we have from Eqs. 8.2 and 8.12

$$Z'X = = AQ_\sigma \Delta Q_\sigma' \tag{8.13}$$

From Eqs. 8.2, 8.9, and 8.12 we have

$$X'X = Q_\sigma \Delta Q_\sigma' \sigma Q_\sigma \Delta Q_\sigma' \tag{8.14}$$

From Eqs. 8.10 and 8.14 we have

$$X'X = Q_\sigma d_\sigma^2 \Delta^2 Q_\sigma' \tag{8.15}$$

From Eq. 8.11

$$d_\sigma^2 \Delta^2 = (I - (I - d_\sigma^2)^{\frac{1}{2}})^2 d_\sigma^{-2} \tag{8.16}$$

But

$$(I - (I - d_\sigma^2)^{\frac{1}{2}})^2 = 2(I - (I - d_\sigma^2)^{\frac{1}{2}} - d_\sigma^2 \tag{8.17}$$

From Eqs. 8.11, 8.16, and 8.17 (8.18)

$$d_\sigma^2 \Delta^2 = (2\Delta - I) \tag{8.18}$$

From Eqs. 8.15 and 8.18

$$X'X = 2Q_\sigma \Delta Q_\sigma' - I \tag{8.19}$$

Substituting Eqs. 8.13 and 8.19 into Eq. 8.7 gives the identity. Hence the solution 8.12 for X satisfies Eq. 8.7 and therefore also Eq. 8.8.

We have now to show that the solution 8.11 for Δ is real and finite. To do this, we must show that the largest element of d_σ^2 is less than or equal to 1 and that the smallest element is greater than zero. From Eq. 8.8

$$R^{-1}ER^{-1} = R^{-1} - R^{-1}AA'R^{-1} \tag{8.20}$$

From Eq. 8.20

$$A'R^{-1}ER^{-1}A = A'R^{-1}A - A'R^{-1}AA'R^{-1}A \tag{8.21}$$

Let

$$F = eR^{-1}A \tag{8.22}$$

From Eqs. 8.3, 8.9, 8.21, and 8.22

$$F'F = \sigma - \sigma^2 \tag{8.23}$$

From Eqs. 8.10 and 8.23

$$(Q_\sigma' F') (FQ_\sigma) = d_\sigma^2 - d_\sigma^4 \tag{8.24}$$

The left side of Eq. 8.24 is Grammian since it is a product moment matrix and diagonal. Hence for all d_{σ_i} we must have

$$d_{\sigma_i}^2 (1 - d_{\sigma_i}^2) \geq 0 \tag{8.25}$$

therefore

$$1 \geq d_{\sigma_i}^2 \tag{8.26}$$

To show that all $d_{\sigma_i}^2$ are positive, we need only show that $A'R^{-1}A$ is basic. By definition, R^{-1} is basic. A general theorem for the rank of the product of

matrices states that the rank of a product of two matrices cannot be less than the sum of their ranks less their common order. If we let

$$\gamma = R^{-\frac{1}{2}}A \tag{8.27}$$

then the rank of γ must be equal to the rank of A which is basic. We have from Eq. 8.27

$$\gamma'\gamma = A'R^{-1}A \tag{8.28}$$

But the product moments of a matrix have the same rank as the matrix, hence the rank of Eq. 8.28 is its order and therefore for all $d_{\sigma_i}^{\ 2}$ we have

$$d_{\sigma_i}^{\ 2} > 0 \tag{8.29}$$

8.3 The Factor Loading Matrix

Let us now return to a solution for the matrix A. We seek a solution which will be scale free and which will have a variable loss function in the sense that it will allow for differential weighting of the variance and covariance elements in the covariance matrix E. We let

$$D_R = \text{Diag } (R)$$

$$D_A = \text{Diag } (AA')$$

$$D_E = \text{Diag } (E)$$

From Eq. 8.8 we write

$$E - P_W D_E = R - P_W D_E - AA' \tag{8.30}$$

where

$$0 \leq P_W \leq 1 \tag{8.31}$$

Let

$$0 \leq p \leq 1 \tag{8.32}$$

$$q = 1 - p \tag{8.33}$$

$$P_A = p/(1 - 2pq) \tag{8.34}$$

$$P_E = q/(1 - 2pq) \tag{8.35}$$

$$D^2 = (P_A D_A + P_E D_E)^{-1} \tag{8.36}$$

From Eq. 8.30 we may write

$$D (E - P_W D_E) D = D (R - P_W D_E - AA') D \tag{8.37}$$

We may now regard the matrix D in Eq. 8.36 as a generalized scaling matrix which may vary as the value p goes from 0 to 1. The parameters P_A and P_E in Eqs. 8.34 and 8.35 respectively have been constructed so that when p = 1, $D^2 = D_A^{-1}$; when p = 0, $D^2 = D_E^{-1}$; when p = .5, $D = D_R^{-1}$. We may refer to D_A^2 as a diagonal matrix of estimated variances, D_E^2 as a diagonal matrix of residual variances, and D_R^2 as a diagonal matrix of total variances. It is seen therefore that the inverse of the matrix D^2 in Eq. 8.36 is a linear combination of the estimated and the residual variances. The special case for p = 0 may be recognized as the scaling function adopted in what have come to be called maximum likelihood and canonical factor analysis. The special case of p = .5 is the scaling function adopted in what some refer to as principal component analysis, although this designation could apply equally well to other scalings. This case is also the scaling function adopted in what has been designated by Harmon (1967) as minres factor analysis. The special case of p = 1 is the scaling function used in Kaiser's (1965) alpha factor analyses.

Let us now consider the generalized loss matrix on the left of Eq. 8.30 or its generalized scaled form on the left of Eq. 8.37. When the parameter P_W in the loss matrix is unity, the loss matrix is the one used in what have somewhat arbitrarily come to be called factor analysis models. When P_W takes the value zero, the loss matrix is the one used in what has equally arbitrarily come to be called

the principal components model. It is seen then that the two special cases of the general loss function parameter P_W determine whether the analysis is called factor analysis or principal component analysis. And the three special cases of the generalized scaling parameter p determine what the corresponding factor analysis technique is called.

Suppose now we let

$$\epsilon = D\,(E - P_W\,D_E)\,D \tag{8.38}$$

We shall refer to ϵ as the generalized loss matrix since it involves both the scaling parameter p and the loss parameter P_W. We shall also let

$$S = R - P_W\,D_E \tag{8.39}$$

$$G = DSD \tag{8.40}$$

$$\alpha = DA \tag{8.41}$$

We may from Eq. 8.38 through Eq. 8.41 write

$$\epsilon = G - \alpha\alpha' \tag{8.42}$$

To solve for A we require that α be orthogonal to the generalized loss matrix ϵ, that is,

$$\epsilon\alpha = 0 \tag{8.43}$$

From Eqs. 8.42 and 8.43

$$G\alpha - \alpha\alpha'\alpha = 0 \tag{8.44}$$

From Eq. 8.44

$$(\alpha'G\alpha)^{\frac{1}{2}} = \alpha'\alpha \tag{8.45}$$

From Eq. 8.44 and 8.45

$$\alpha = G\alpha(\alpha'G\alpha)^{-\frac{1}{2}} \tag{8.46}$$

But from Eq. 8.42 we see that ϵ is independent of any square orthonormal transformation of α. We may therefore write Eq. 8.46 as

$$\alpha = G\alpha\,(\alpha' G\alpha)^{-\frac{1}{2}} h \tag{8.47}$$

where h is any conformable square orthonormal matrix. In particular, we let

$$tt' = (\alpha' G\alpha) \tag{8.48}$$

and choose h so that

$$t'^{-1} = (\alpha' G\alpha)^{-\frac{1}{2}} h \tag{8.49}$$

Therefore without loss of generality we may write

$$\alpha = G\alpha t'^{-1} \tag{8.50}$$

8.4 The Loss Function

We shall see presently that Eq. 8.50 provides the basis for an iterative procedure for solving for α and hence also A. First, however, let us examine in more detail the generalized loss matrix ϵ and the determination of A which will optimize some specified function of it. First we write the matrix G given by Eq. 8.40 in basic structure form as

$$G = Q_m \delta_m Q_m' + Q_\beta \delta_\beta Q_\beta' - Q_\gamma \delta_\gamma Q_\gamma' \tag{8.51}$$

where δ matrices are of order indicated by their subscripts. If P_W in Eq. 8.39 is zero, then obviously γ is zero and $m + \beta = n$, although this is not a necessary condition for γ to be zero. Suppose now we let

$$\alpha = Q_m \delta_m^{\frac{1}{2}} \tag{8.52}$$

From Eqs. 8.42 and 8.52

$$\epsilon = Q_\beta \delta_\beta Q_\beta' - Q_\gamma \delta_\gamma Q_\gamma' \tag{8.53}$$

Equations 8.51 and 8.53 are still perfectly general both with respect to the scaling parameter p and the loss parameter P_W. The loss function involving the loss matrix ϵ may be chosen in a number of ways. In the case of the scaling parameter $p = 0$ and the loss parameter $P_W = 1$, we have from Eqs. 8.39 and 8.40, and Eqs. 8.33, 8.35 and 8.36

$$G = D_E^{-\frac{1}{2}} R D_E^{-\frac{1}{2}} - 1 \tag{8.54}$$

Also by definition

$$D_\epsilon = 0 \tag{8.55}$$

From Eqs. 8.52 through 8.54

$$D_E^{-\frac{1}{2}} R D_E^{-\frac{1}{2}} - I - \alpha\alpha' = \epsilon \tag{8.56}$$

From Eq. 8.56

$$D_E^{-\frac{1}{2}} R D_E^{-\frac{1}{2}} - \alpha\alpha' = I + \epsilon \tag{8.57}$$

Now from Eqs. 8.53 and 8.57 we may write

$$Q_\beta \delta_\beta Q_\beta' - Q_\gamma \delta_\gamma Q_\gamma' = (Q_\beta, Q_\gamma, q_m) \begin{bmatrix} \delta_\beta & 0 & 0 \\ 0 & -\delta_\gamma & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} Q_\beta' \\ Q_\gamma' \\ q_m' \end{bmatrix} \tag{8.58}$$

where q is orthonormal and orthogonal to Q_β and Q_γ and $\beta + \gamma + m = n$.

From Eq. 8.58

$$I + \epsilon = (Q_\beta, Q_\gamma, q_m) \begin{bmatrix} I_\beta + \delta_\beta & 0 & 0 \\ 0 & I_\gamma - \delta_\gamma & 0 \\ 0 & 0 & I_m \end{bmatrix} \begin{bmatrix} Q_\beta' \\ Q_\gamma' \\ q_m' \end{bmatrix} \tag{8.59}$$

Now it can be proved that no element of δ_γ can be greater than 1 because R is Grammian and hence also $D_E^{-\frac{1}{2}} R D_E^{-\frac{1}{2}}$. From Eqs. 8.55 and 8.59

$$\text{tr}\,(I + \epsilon) = \text{tr}\,(I_\beta + \delta_\beta) + \text{tr}\,(I_\gamma - \delta_\gamma) + m \tag{8.60}$$

From Eq. 8.60

$$\text{tr}\,\delta_\beta = \text{tr}\,\delta_\gamma \tag{8.61}$$

But it is known (Bergmann, 1963) that the maximum likelihood solution for factor analysis maximizes the determinant of $I + \epsilon$. It is also known that the value of a determinant is equal to the product of its characteristic roots or the basic diagonal elements of the determinant of the matrix. Therefore the determinant of $I + \epsilon$ is given by

$$|I + \epsilon| = \prod_{i=1}^{\beta} (1 + \delta_{\beta_i}) \prod_{i=1}^{\gamma} (1 - \delta_{\gamma_i}) \tag{8.62}$$

With the constraints on δ_β and δ_γ, Eq. 8.62 evidently increases as their elements approach zero. In any case, Eq. 8.62 gives the loss function to be optimized in the case of so-called maximum likelihood factor analysis. As Jöreskog (1967) has pointed out, "The maximum likelihood estimates are obtained when the n - m smallest roots are as equal to one as possible in an approximate least square sense." This is tantamount to saying that the sum of the squares of the deviations of the roots of $I + \epsilon$ from unity shall be a minimum. But since the roots of $I + \epsilon$ are those of ϵ increased by one, Jöreskog's statement implies that tr $(\delta_\beta^2 + \delta_\gamma^2)$ shall be a minimum. But from Eq. 8.53

$$\text{tr } \epsilon^2 = \text{tr } \delta_\beta^2 + \text{tr } \delta_\gamma^2 \tag{8.63}$$

The foregoing discussion is based on the choice of the scaling and loss function parameters of $p = 0$ and $P_W = 1$ respectively, which are the ones adopted in the so-called maximum likelihood method of factor analysis. We may, however, regard Eq. 8.63 as a more generalized loss function appropriate for any and all values of the scaling and loss parameters p and P_W. However, it is important to note that tr ϵ^2 may be small in absolute value but could be large compared to tr G^2. Computationally, a better function to optimize is

$$\phi = 1 - \frac{\text{tr } \epsilon^2}{\text{tr } G^2} \tag{8.64}$$

But from Eqs. 8.51, 8.52, and 8.64

$$\phi = \frac{\text{tr } (\alpha\alpha')^2}{\text{tr } G^2} \tag{8.65}$$

From Eqs. 8.45 and 8.65

$$\phi = \frac{\text{tr } \alpha' G \alpha}{\text{tr } G^2} \tag{8.66}$$

We shall then take ϕ in Eq. 8.66 as the generalized loss function for any specified value of m, the rank of the approximation matrix. We therefore seek to determine α and hence A so that ϕ will be maximized. The maximum value ϕ can take is of course 1, in which case tr ϵ^2 vanishes.

8.4 The Computations for the Factor Loading Matrix

To solve for A, we return to Eq. 8.47. Substituting from Eqs. 8.39, 8.40, and 8.41 we have

$$DA = [D (R - P_W D_E) D] DA [A' D (D (R - P_W D_E) D) DA]^{-\frac{1}{2}} h \tag{8.67}$$

From Eq. 8.67

$$A = (R - P_W D_E) D^2 A [A' D^2 (R - P_W D_E) D^2 A]^{-\frac{1}{2}} h \tag{8.68}$$

Equation 8.68 suggests a convenient iterative set of algorithms for solving for A. We begin with some approximation to A, say ${}_0A$. We then calculate a first approximation ${}_1D_A$ to D_A by

$${}_1D_A = D_{0A_0A'} \tag{8.69}$$

Next we let

$${}_1D_E = D_R - {}_1D_A \tag{8.70}$$

For some prespecified values of p between 0 and 1, we calculate P_A and P_E from Eqs. 8.34 and 8.35. We then calculate

$${}_1D^2 = (P_A \, {}_1D_A + P_E \, {}_1D_E)^{-1} \tag{8.71}$$

We let

$${}_1U = {}_1D^2 \, {}_0A \tag{8.72}$$

$$ {}_1S = R - P_W\, {}_1D_E \tag{8.73} $$

where P_W is some prespecified value between 0 and 1. Then

$$ {}_1W = {}_1S\, {}_1U \tag{8.74} $$

We then calculate ${}_1U'\, {}_1W$ and set up the supermatrix $\begin{bmatrix} {}_1U'\, {}_1W \\ {}_1W \end{bmatrix}$.

A partial triangular factoring of this matrix gives

$$ \begin{bmatrix} {}_1t \\ {}_1A \end{bmatrix} {}_1t' = \begin{bmatrix} {}_1U'\, {}_1W \\ {}_1W \end{bmatrix} \tag{8.75} $$

We calculate the criterion

$$ \phi_0 = \frac{\mathrm{tr}\,({}_1U'\, {}_1W)}{\mathrm{tr}\,({}_1D\, {}_1S\, {}_1D^2)} \tag{8.76} $$

In general we have

$$ {}_kD_A = D_{k-1A_{k-1}A'} \tag{8.77} $$

$$ {}_kD_E = D_R - {}_kD_A \tag{8.78} $$

$$ {}_kD^2 = (\Gamma_A\, {}_kD_A + P_E\, {}_kD_E)^{-1} \tag{8.79} $$

$$ {}_kU = {}_kD^2\, {}_{k-1}A \tag{8.80} $$

$$ {}_kS = R - P_W\, {}_kD_E \tag{8.81} $$

$$ {}_kW = {}_kS\, {}_kU \tag{8.82} $$

$$\phi_k = \frac{\text{tr}\,({}_{k-1}U'\,{}_kW)}{\text{tr}\,({}_kD\,{}_kS\,{}_kD)^2} \tag{8.83}$$

$$\begin{bmatrix} {}_kt \\ \\ {}_kA \end{bmatrix} \; {}_kt' = \begin{bmatrix} {}_kU' \;\; {}_kW \\ \\ {}_kW \end{bmatrix} \tag{8.84}$$

We repeat Eqs. 8.77 through 8.84 until ϕ_k and ϕ_{k+1} are sufficiently close.

8.6 Alternative Computational Procedures

Obviously, many alternative solutions for A may be available for special values of the scaling and loss function parameters p and P_W respectively. For the case of p = 0 and P_W = 1, computational procedures have recently been presented by Jöreskog (1967) and also by Horst (1968b). Previously, other methods have been presented by Lawley (1940), Rao (1955), and Hemmerle (1965). There has been some debate as to the difference of the procedures among maximum likelihood, canonical correlation, and least square methods which utilize these scaling and loss function parameters but we shall not elaborate these issues.

For the case of p = .5 and P_W = 1, various computational procedures have been presented, among which is one by Comrey (1962) and more recently the minres method of Harmon (1967). For the case of p = 1. and P_W = 1, Kaiser and Caffrey (1965) in their alpha factor analysis have suggested a computational algorithm. For the most familiar case of all when p = .5 and P_W = 0, we have the principal axis or principal components method for which many computational methods too numerous to mention are available.

Browne (1967) has discussed several variations of the scaling and loss function parameters as well as variations of the loss function itself, and reports the development of computational algorithms and computer programs for these variations. However, to our knowledge, none of the previously available computing algorithms

or computer programs are readily adaptible for variable scaling and loss parameters.

The computing algorithms given in Eqs. 8.77 through 8.84 are obviously readily generalizable, as can be seen from Eq. 8.79 which involves the P_A and P_E functions of the scaling parameter p and Eq. 8.81 which involves the loss parameter P_W.

8.7 Special Problems

However, several important questions remain to be considered about the computational procedure. The first of these has to do with the loss function ϕ given by Eqs. 8.66 and 8.83. Obviously, both the numerator and denominator of this ratio are extremely complicated functions of A and it is probable that many stationary points or local optima may exist. Whether and under what conditions the solution indicated gives in the limit the absolute maximum is a most pertinent question. Certainly for the case of $p = .5$ and $P_W = 0$, the well known principal axis case, we have shown (Horst, 1965) that the solution converges to an absolute maximum. For the case of $p = 0$ and $P_W = 0$, Anderson and Rubin (1956) have shown that, unless constrained, the solution for A which maximizes ϕ is not unique.

Aside from the question of uniqueness of the solution or the attainment of the absolute maximum, we must also be concerned with the questions of whether the solution converges, how rapidly it converges, whether the residual variances given by D_E are positive, what will constitute a suitable first approximation for the A matrix, and the number of factors to be solved for. None of these questions has been completely adequately answered. However, for a number of different types of data that have been analyzed, the solution presented in this chapter appears to be reasonably satisfactory with respect to each of these questions.

As a first approximation to the number of factors, we have adopted the rule of Kaiser (1958) that the number of factors solved for shall be equal to the number of roots of the correlation matrix greater than unity. With some of the data which we have analyzed, this number appears to give one or several factors too few, while

with others it appears to give one or several too many. Therefore, it is probable that, lacking an adequate absolute criterion for the number of factors, the Kaiser rule may be taken as a first approximation. If some adjustment of the loss function is available that takes account of the number of factors m, one can then calculate these adjusted functions for each of some specified range of m which includes the Kaiser value. For example, one could calculate the function for the integers lying between $m_k - pm_k$ and $m_k + pm_k$ where $1 > p > 0$. Specifically, p might be .2 or .3. Jöreskog (1967) has suggested a method similar to this for the case of the scaling parameter $p = 0$ and the loss parameter $P_W = 1$. His loss function, however, is not identical with ours.

8.8 First Approximation to the Factor Loading Matrix

As a first approximation to the A matrix we could take the first m principal axis factors of the correlation matrix. This is the case of the scaling parameter $p = .5$ and the loss parameter $P_W = 0$. This procedure, however, has not yielded satisfactory results with some data. It can lead to a local maximum for the loss function rather than the absolute maximum. We have presented elsewhere (Horst, 1968b) a better first approximation.

We let

$$D_{R-1} = \text{Diag}\ (R^{-1}) \tag{8.85}$$

$${}_0D_E = D_{R-1}^{-1} \tag{8.86}$$

$${}_0D_A = I - {}_0D_E \tag{8.87}$$

$${}_0D^2 = (P_A\ {}_0D_A + P_E\ {}_0D_E)^{-1} \tag{8.88}$$

$$P_k = (P_W\ {}_0D^2\ {}_0D_E) \tag{8.89}$$

$${}_0G = {}_0D\ (R - P_W\ D_E)\ {}_0D \tag{8.90}$$

Let the basic structure of ${}_0G + P_k I$ be

$$q_m d_m q_m' + q_s d_s q_s' = {}_0G + P_k I \tag{8.91}$$

Then the initial approximation to A is

$${}_0A = {}_0D^{-1} q_m (d_m - P_k I)^{\frac{1}{2}} \tag{8.92}$$

Equations 8.69 through 8.84 indicate the successive approximations to A.

8.9 The Problem of Improper Solutions

The question of positive D_E values is important for the case $P_A = 0$. If it is not positive, then the scaling matrix D whose square is given by Eq. 8.36 may have imaginary or infinite elements. The conditions under which elements of D may become infinite or imaginary have not been adequately investigated. The methods of Jöreskog (1967) for the parameters $p = 0$, $P_W = 1$ prevent such cases, as does the minres method of Harmon for the parameters $p = .5$ and $P_W = 1$. In our own computing procedure, if any element of a D_E approximation is 1 or greater, the corresponding vector for that approximation of the A matrix is arbitrarily re-scaled to yield a D_E element less than 1 by some specified small number such as .0005. In the final approximation for A one can identify such variables by the fact that their D_E value is equal to this value. So far, no cases of real data have been encountered where any of the final D_E elements are at the constrained minimum with the exception of the solutions having the parameters $p = 0$, $P_W = 0$. For this case, one or more of the D_E values is always at the constrained minimum. This is to be expected as shown by the work of Rubin and Anderson (1956). An interesting and unanswered question is how for this case the variables reaching the minimum D_E values will vary according to the method of solution. Also of interest is how the D_E values of variables may approach the constrained minimum for $p = 0$ as P_W goes from 1 to 0.

No mathematical proof of the convergence of the loss function or the solution for A for the method here presented has been found. However, for all sets of data on which the method has been tried, satisfactory convergence does occur. It has been proved (Horst, 1965) that the method converges for the case of p = .5 and $P_W = 0$. This is of course the traditional principal axis solution for the correlation matrix with unity in the diagonals.

The rate of convergence for the sets of data subjected to the procedure varies and further evidence is given in Chapter 13. In general, the loss function at first rapidly approaches an asymptote and later the approach is much slower. For the case of p = 0 and $P_W = 1$, acceleration procedures have been introduced which greatly increase the rate of convergence (see Horst, 1968b).

8.9 Proof of Scale Free Property

We shall now prove that the generalized scaling and loss function procedure is independent of any scaling of the data matrix by attributes and hence also of any scaling of its covariance matrix. This proof supports the assertion that without loss of generality we can begin with the correlation matrix. To demonstrate this independence we let Δ be an arbitrary positive definite diagonal matrix. From Eq. 8.68 we can write

$$\Delta A = \Delta(R - P_W D_E)\, A(A^{-1}D^2A^{-1})\,\Delta A[A'\Delta(A^{-1}D^2A^{-1})\Delta(R - P_W D_E)\Delta(\Delta^{-1}D^2\Delta^{-1})\,\Delta A]^{-\frac{1}{2}}h \quad (8.93)$$

Let

$$C = \Delta R \Delta \quad (8.94)$$

$$a = \Delta A \quad (8.95)$$

$$F = C - aa' \quad (8.96)$$

$$d^2 = (P_A D_{aa'} + P_E D_F)^{-1} \quad (8.97)$$

From Eq. 8.95

$$D_{aa'} = A^2 D_A \quad (8.98)$$

From Eqs. 8.95 and 8.96

$$D_F = \Delta^2(D_R - D_A) \tag{8.99}$$

From Eq. 8.8

$$D_E = D_R - D_A \tag{8.100}$$

From Eq. 8.99 and 8.100

$$D_F = \Delta^2 D_E \tag{8.101}$$

From Eq. 8.97, 8.98, and 8.100

$$d^2 = (\Delta^2 (P_A D_A + P_E D_E))^{-1} \tag{8.102}$$

From Eqs. 8.36 and 8.102

$$\Delta^{-2} D^2 = d^2 \tag{8.103}$$

Substituting Eqs. 8.94, 8.95, 8.101, and 8.103 in Eq. 8.93

$$a = (C - P_W D_F)\, d^2 a\, (a' d^2 (C - P_W D_F)\, d^2 a)^{-\frac{1}{2}} h \tag{8.104}$$

But Eq. 8.104 is the same form as Eq. 8.68. Hence we may start with any covariance matrix C whose correlation matrix is R, and the solution of $\underline{a}$ satisfied by Eq. 8.104 will be related to the solution A obtained from the correlation matrix by the relation

$$A = \Delta^{-1} a \tag{8.105}$$

or by definition

$$A = D_C^{-\frac{1}{2}} a \tag{8.106}$$

where D_C is a diagonal matrix of variances of the arbitrarily scaled variables.

CHAPTER 9

THE SIMPLE STRUCTURE TRANSFORMATION

9.1 The Simple Structure Problem

We shall now return to Eq. 8.2:

$$U = XA' \tag{9.1}$$

The matrix U is the approximation to the data matrix Z and a generalized solution for it has been considered at length in the previous chapter. However, the solution is not unique as we can readily show. Suppose we let

$$B = Ah' \tag{9.2}$$

$$Y = X(h'h)^{-1}h' \tag{9.3}$$

where h is any nonhorizontal basic matrix. It can readily be shown from Eqs. 9.1, 9.2, and 9.3 that

$$U = YB' \tag{9.4}$$

The problem of finding an h matrix which yields a B matrix in Eq. 9.2 which in some sense optimizes certain prespecified criteria was first considered by Thurstone (1947) and called by him the problem of achieving simple structure. Traditionally, the matrix h has been taken as square so that the number of columns m in B is the same as in A. The criteria stated by Thurstone (1947), as given in Chapter 6, may be restated briefly:

1. There should be at least m elements in each column of B which in absolute value are very small or near zero.

2. There should be at least one very small or near-zero element in each row of B.

3. For every pair of columns there should be several or more rows in which both values are very small.

4. For every pair of columns there should be very few rows in which both values are large.

These criteria are not stated, of course, in analytical terms. Thurstone and many since then have attempted to formulate more objective analytical criteria whic

would tend to satisfy the descriptive criteria. Among the best known of these are the varimax criterion and procedures developed by Kaiser (1958).

Two general types of h matrices have been considered. One of these is the square orthonormal matrix used in the varimax procedures. The other type of methods utilizes a square basic transformation restricted only in that its columns are normalized. This type has been called an oblique transformation. For each type of transformation the h matrix should yield a B matrix such that some specified function of its elements will be optimized. The proponents of oblique transformations believe that these yield better simple structure than do orthonormal transformations. Many analytical methods for achieving simple structure B matrices have been presented and discussed by Horst (1965) and by Harmon (1967). In spite of the variety of methods now available, none of them has been consistently satisfactory for all types of data.

The generalized method of factor analysis which we have developed includes the special cases that we have already discussed. Some prefer one of these special cases and some another. It is probable that an adequate set of criteria for simple structure and methods for optimizing stable functions would provide a more objective and useful basis for evaluating the various special cases than are provided by the subjective rationalizations of the numerous investigators. We shall present a transformation rationale and procedures based on certain criteria which give some promise for achieving this objective. It also gives promise of yielding more satisfactory results for a wider variety of data than methods currently available.

9.2 The Rationale of the Criterion

We let A be an n x m arbitrary factor loading matrix. In particular, it may be a matrix solved for by the methods of Chapter 8. We let h be an m x m basic matrix and define the matrix B by

$$B = Ah \tag{9.5}$$

It will be convenient to regard h as normal by columns so that

$$D_h' h = I \tag{9.6}$$

We now define an exponent by

$$F = \frac{2W}{2W - 1} \tag{9.7}$$

where W is a positive integer. We note then that any number raised to the F power is a positive value and any number raised to the F + 1 power retains the original sign.

It will also be convenient to define a matrix γ such that γ_{ij} is +1 if B_{ij} is positive and -1 if B_{ij} is negative. We indicate the elemental product of two matrices by placing a dot between them, and the elemental power of a matrix by enclosing its exponent in parentheses. It is seen then that because of Eq. 9.7 the signs of the elements of $\gamma.B^{(F)}$ are the same as those of the corresponding elements of B.

Now instead of determining h so as to optimize some function of the elements of B, we shall consider a preliminary scaling of the columns of B by a diagonal matrix D and let

$$b = BD \tag{9.8}$$

We wish to determine D so that for each column of b the sum of the absolute values of the F + 1 powers of its elements is equal to the sum of the fourth powers of the elements. We let

$$d_G = D_b' b(3) \tag{9.9}$$

$$d_F = D_b' (\gamma.b^{(F)}) \tag{9.10}$$

We wish now to determine D in Eq. 9.8 so that

$$d_G = d_F \tag{9.11}$$

To determine D which satisfies Eq. 9.11 we let

$$D_G = D_{B'B(3)} \tag{9.12}$$

$$D_F = D_{B'(\gamma.B^{(F)})} \tag{9.13}$$

Considering the subscript on the right of Eq. 9.13, we note that although elemental multiplication as such is commutative, distributive, and associative, it does not have these properties in combination with standard matrix multiplication. In particular, the elemental products must be taken before the matrix products. It can now be proved that the D which satisfies Eq. 9.9 is given by

$$D = (D_G\ D_F^{-1})^{\frac{1}{F-3}} \tag{9.14}$$

To show this we have from Eq. 9.8 and from Eqs. 9.9 and 9.10 respectively

$$d_G = D_{DB'B(3)D^3} \tag{9.15}$$

$$d_F = D_{DB'(\gamma.B^{(F)})D^F} \tag{9.16}$$

From Eqs. 9.12 and 9.15 we have

$$d_G = D^4\ D_G \tag{9.17}$$

From Eqs. 9.13 and 9.16

$$d_F = D^{(F+1)}D_F \tag{9.18}$$

From Eqs. 9.11, 9.17, and 9.18

$$D^4 D_G = D^{(F+1)}D_F \tag{9.19}$$

From Eq. 9.19

$$D_G\ D_F^{-1} = D^{(F-3)} \tag{9.20}$$

From Eq. 9.20

$$D = (D_G \, D_F^{-1})^{\frac{1}{F-3}} \tag{9.21}$$

which is the same as Eq. 9.14.

We next define the two diagonal matrices

$$D_f = D_{b'b}(F) \tag{9.22}$$

$$D_g = D_{b'} D_{bb'} D_b \tag{9.23}$$

We note that D_f in Eq. 9.22 is the same as d_F in Eq. 9.10, except that the elemental factor γ has been omitted. We now let

$$\delta_f = d_F - D_f \tag{9.24}$$

$$\delta_g = D_g - d_G \tag{9.25}$$

Now the minimum that the set of values δ_f in Eq. 9.24 can take is given by $\delta_f = 0$. This occurs when all elements in γ are +1, in which case $d_F = D_f$. This is of course the case when all b_{ij} are non negative.

We have therefore

$$\delta_f \geq 0 \tag{9.26}$$

To determine lower bounds for the elements of δ_g in Eq. 9.25 we note first

$$D_{bb'} = \sum_{j=1}^{m} D_{b_{.j}}^2 \tag{9.27}$$

Therefore the k'th element of D_g in Eq. 9.23 can be written as

$$D_{g_k} = b_{.k}' D_{b_{.k}}^2 b_{.k} + b_{.k}' \left(\sum_{j=1}^{m} D_{b_{.j}}^2 - D_{b_{.k}}^2\right) b_{.k} \tag{9.28}$$

But the k'th element of d_G in Eq. 9.9 can be written

$$d_{G_k} = b_{.k}' D_{b_{.k}} b_{.k} \tag{9.29}$$

Then from Eqs. 9.25, 9.28, and 9.29, the k'th element of δ_g in Eq. 9.25 can be written

$$\delta_{g_k} = b_{.k}^{(2)} \left(\sum_{i=1}^{m} b_{.j}^{(2)} - b_{.k}^{(2)} \right) \tag{9.30}$$

From Eq. 9.30 it is clear that only when a $b_{.k}^{(2)}$ is orthogonal to the sum of the remaining $b_{.j}^{(2)}$ can the δ_{g_k} be zero. Otherwise it must be greater than zero. We have therefore that

$$\delta_g \geq 0 \tag{9.31}$$

We may now recognize that the nearer δ_f in Eq. 9.24 is to zero the closer the positive manifold criterion of Thurstone (1947) is satisfied. Also in the limiting case, when no two columns in B have nonvanishing elements in any row for either column, the δ_g will be zero.

9.3 Development of the Equations

We shall now make use of the two facts in the paragraph above in developing a criterion which will be optimized in our solution for h. We begin by writing from Eq. 9.24

$$D_f = d_F - \delta_f \tag{9.32}$$

From Eq. 9.11 and 9.25

$$D_g = d_F + \delta_g \tag{9.33}$$

We let

$$\Delta = D_f D_g^{-1} \tag{9.34}$$

$$\Psi = \text{tr } \Delta \tag{9.35}$$

From Eqs. 9.32 through 9.35

$$\Psi = \text{tr } \left((d_F - \delta_f) (d_F + \delta_g)^{-1} \right) \tag{9.36}$$

From Eq. 9.36 we see that Ψ increases as the elements of δ_f and δ_g decrease. As these approach zero, Ψ approaches m, the number of factors. Ψ is a function of h. If we differentiate Ψ with respect to h and equate the derivative to 0, we should

obtain an expression for h which gives an optimum solution for Ψ. We begin by taking the differential of Ψ. From Eqs. 9.34 and 9.35 we have

$$d\Psi = \mathrm{tr}\ (d\ (D_f)\ D_g^{-1} - d\ (D_g)\ D_f\ D_g^{-2}) \tag{9.37}$$

From Eqs. 9.34 and 9.37

$$d\Psi = \mathrm{tr}\ (d\ (D_f) - d\ (D_g)\ \Delta)\ D_g^{-1} \tag{9.38}$$

From Eq. 9.38

$$\frac{\partial \Psi}{\partial h'} = \left(\frac{\partial\ (D_f)}{\partial h'} - \frac{\partial\ (D_g)}{\partial h'}\ \Delta\right) D_g^{-1} \tag{9.39}$$

The differentiation of D_f and D_g with respect to h is extremely complicated. We shall not attempt this differentiation directly but shall proceed somewhat more simply. First we write from Eq. 9.22

$$D_f = D_{b'(b^{(F-1)}.b)} \tag{9.40}$$

From Eqs. 9.5, 9.8, and 9.40

$$D_f = D^2\ D_{h'A'(b^{(F-1)}.(Ah))} \tag{9.41}$$

From Eqs. 9.5, 9,8, and 9.23

$$D_g = D^2 D_{h'A'D_{bb}\cdot Ah} \tag{9.42}$$

Suppose we have some approximate solution for h satisfying Eq. 9.6 and we arrive at some fixed approximate solutions to D, $b^{(F-1)}$, and D_{bb}, by means of equations already presented. We substitute these fixed approximations in Eqs. 9.41 and 9.42. Then it can be shown that

$$\frac{\partial\ (D_f)}{\partial h} = 2D^2 A'(b^{(F-1)}.(Ah)) \tag{9.43}$$

$$\frac{\partial\ (D_g)}{\partial h} = 2D^2 A' D_{bb}\cdot Ah \tag{9.44}$$

We now let

$$\frac{\partial \Psi}{\partial h'} = 0 \tag{9.45}$$

From Eqs. 9.39, 9.43, 9.44, and 9.45

$$A'(b^{(F-1)} \cdot (Ah)) - A'D_{bb} \cdot Ah\Delta = 0 \tag{9.46}$$

9.4 The Computational Procedure

To set up an iterative set of algorithms to solve for h, we substitute for the unknown h in the first term of the left hand side of Eq. 9.46 the approximation to h by means of which we solved for the fixed matrix b. We have from Eqs. 9.5 and 9.8

$$Ah = bD^{-1} \tag{9.47}$$

From Eqs. 9.46 and 9.47

$$A'b^{(F)} - A'D_{bb} \cdot Ah\Delta D = 0 \tag{9.48}$$

We let

$$E = A'b^{(F)} \tag{9.49}$$

$$S = A'D_{bb} \cdot A \tag{9.50}$$

From Eqs. 9.48, 9.49, and 9.50

$$S^{-1}E = h\Delta D \tag{9.51}$$

Let

$$H = S^{-1}E \tag{9.52}$$

From Eqs. 9.51 and 9.52

$$H'H = \Delta Dh'hD\Delta \tag{9.53}$$

From Eqs. 9.6 and 9.53

$$\Delta = D_{H'H}^{\frac{1}{2}}D^{-1} \tag{9.54}$$

From Eqs. 9.6, 9.51, and 9.52

$$h = HD_{H'H}^{-\frac{1}{2}} \tag{9.55}$$

We are now ready to consider the iterative computational sequence for h and B. We begin with some approximation to h which satisfies the relation

$$D_{h'h} = I \tag{9.56}$$

Then we calculate

$$B = Ah \tag{9.57}$$

We let

$$F = \frac{2W}{2W - 1} \tag{9.58}$$

where W is a positive integer to be discussed later.

We calculate

$$D_G = D_{1'B}(4) \tag{9.59}$$

$$D_F = D_{1'|B|}(F+1) \tag{9.60}$$

where $|B|$ means the matrix of absolute values of the elements of B.

Next we calculate

$$D = (D_G\, D_F^{-1})^{\frac{1}{F-3}} \tag{9.61}$$

$$b = BD \tag{9.62}$$

$$D_1 = D_{b(2)_1} \tag{9.63}$$

$$S = A'D_1 A \tag{9.64}$$

$$E = A'b^{(F)} \tag{9.65}$$

$$H = S^{-1}E \tag{9.66}$$

$$D_2 = D_{1'H}(2) \tag{9.67}$$

$$h = HD_2^{-\frac{1}{2}} \tag{9.68}$$

$$\Delta = D_2^{\frac{1}{2}} D^{-1} \tag{9.69}$$

$$\Psi = (\mathrm{tr}\ \Delta)/m \tag{9.70}$$

$$B = Ah \tag{9.71}$$

For any given value of W in Eq. 9.58, the calculations 9.59 through 9.71 may be repeated until two successive values of Ψ in Eq. 9.70 are sufficiently close.

9.5 Special Problems

The rationale and procedures we have considered in this chapter make some assumptions about the solution for the A matrix. The research so far conducted with the method on experimental data has begun with A matrices calculated by the methods of Chapter 8. The computational algorithms calculate an α matrix which is actually a principal components or basic structure solution of a scaled correlation matrix with adjusted diagonal elements. In any case then, the α matrix is orthogonal. The A matrix, which is in effect a descaling of the α matrix, is not in general orthogonal. However, the solution for the A matrix is such that the first vector has all positive elements. Implicit in the transformation solution of this chapter is the assumption that the first principal axis of the α matrix has all positive elements. This amounts to the pre- and postmultiplication of a scaled symmetric matrix by a sign matrix such that its first basic orthonormal vector has all positive elements.

The A matrix may be operated upon directly or it may first be normalized by rows before the simple structure computations begin. The question of whether to normalize rows of the arbitrary factor matrix before applying simple structure procedures has arisen with other methods of transformation and was discussed in Chapter 6. Kaiser (1958) has recommended such a row scaling before the application of the varimax procedures, followed by a descaling of the simple structure matrix.

In the computer programs provided in this report, the option of either normalized or original scaling is available.

Beginning with an A matrix, either normally scaled by rows or not, we start with some approximation to the h matrix. The simplest procedure is to begin with h as the identity matrix. This in general is a very poor approximation if we have a principal axis or basic structure type solution for A. However, it is the one we use in the accompanying computer programs and it has appeared to give good results with data for which the simple structure factors have been rather well established.

We have attempted no proof that the method does converge. Intuitively it appears that it should. For data on which it has been tried, it appears to converge satisfactorily. Whether the convergence can be to a local maximum has not been proved and may well not be capable of proof. Again, however, the empirical results with data whose simple structure has been well established would indicate that the solutions are in general close to the absolute maximums for the Y values.

9.6 The Exponential Parameter

The determination of the integer W in the calculation of F in Eq. 9.58 leaves much to be desired from a theoretical point of view but empirically determined procedures appear reasonably adequate. The question may well be raised as to why F is not simply taken as 3, so that F + 1 would be 4, and thus bring the method into line with those of Kaiser (1958), Neuhaus and Wrigley (1954), Saunders (1953), and Carroll (1953) whose methods have emphasized 4th power terms. The answer is that variations of their methods, as well as the use of F = 3, have not given consistently good results for a wide variety of data types. Largely as a result of extensive empirical experimentation, we begin with W = 2 which gives F = 4/3. Iterations proceed with this value until the solution stabilizes. The integer is increased for subsequent solutions until the following condition obtains: One

or more columns of the stabilized B matrix has less than m negative values. Since the negative values are typically small, they are regarded as the near-zero values. When this condition is reached, the B value for the previous W value is taken as the final B matrix. The program always retains in storage this one previous B matrix. In some cases, even the B matrix for W = 2 does not have at least m negative values in each column. But typically, each column does have a number of small positive values so that even for W = 2 the number of negatives and near zeros in each column tends to exceed m. A limit is put on the value of W, such as 20, in case the criterion of negatives less than m is not reached sooner. Such cases are rare but one example is given by data set 10 in Chapter 12.

CHAPTER 10

SIMPLE STRUCTURE FACTOR SCORES

10.1 The Traditional Arbitrary Factor Score Matrices

In Chapter 7 we considered five of the methods that have been used for estimating the factor score matrix. We saw that only one of these was scale free and that two of them were identical except for a scaling matrix. Only one of the methods gave an orthonormal factor score matrix and this was shown to be the least square orthogonalization of a residual variance scaling for what Harris (1967) quite properly regards as a method that is "wrong most of the time." We showed that by generalizing the scaling of the variables and introducing orthgonalizations of the resulting estimates we have actually six methods. None of these, however, satisfies the desirable relationship that the residual covariance matrix is the difference between the original covariance matrix and the major product moment of the factor loading matrix. None of the methods presented in Chapter 7 yields matrices that are orthogonal to the residual data matrix.

10.2 The Exact Residual Covariance Solution

In Chapter 8 we presented a factor score matrix which does satisfy the condition that the residual covariance matrix be the difference between the total and the estimated covariance matrix, as discussed in Chapter 5. Since this matrix is the basis of the simple structure factor score matrix we shall develop later in this chapter, we shall consider it further at this time. Using a slightly different form than in Chapter 8 we let

$$\sigma = A'C^{-1}A \tag{10.1}$$

and indicate the basic structure of Eq. 10.1 as

$$Q_\sigma d_\sigma^2 Q_\sigma' = \sigma \tag{10.2}$$

We let

$$\Delta = (I - (I - d_\sigma^2)^{\frac{1}{2}})d_\sigma^{-2} \tag{10.3}$$

Then the factor score matrix is given by

$$X = ZC^{-1}AQ_\sigma \Delta Q_\sigma' \tag{10.4}$$

In the above, we use the covariance matrix C instead of the correlation matrix used in Chapter 8 to show that the estimate of X is scale free. We define Z so that

$$Z'Z = C \tag{10.5}$$

Suppose now we return to the fundamental matrix approximation equation

$$Z - XA' - e = 0 \tag{10.6}$$

Indicating the approximation matrix by U, we have

$$Z - U - e = 0 \tag{10.7}$$

From Eq. 10.7 we have

$$\begin{aligned} 0 = {} & Z'Z - Z'U - Z'e \\ & - U'Z + U'U - U'e \\ & - e'Z - e'U + e'e \end{aligned} \tag{10.8}$$

Now from Eqs. 10.1 through 10.4 it can be shown that the covariance matrix for X is

$$X'X = Q_\sigma(2\Delta - I)Q_\sigma' \tag{10.9}$$

The covariance matrices in Eq. 10.8 can readily be derived. The matrix C is of course by definition Z'Z. The others are:

$$Z'U = A Q_\sigma \Delta Q_\sigma' A' \tag{10.10}$$

$$Z'e = C - A Q_\sigma \Delta_\sigma \sigma_\sigma' A' \tag{10.11}$$

$$U'U = A Q_\sigma \Delta d_\sigma^2 Q_\sigma' A' \tag{10.12}$$

$$U'e = A Q_\sigma \Delta (I - d_\sigma^2) Q_\sigma' A' \tag{10.13}$$

$$e'e = C - AA' \tag{10.14}$$

It is obvious from Eq. 10.9 that X is not in general orthonormal but only when $\Delta = I$. But from Eq. 10.3, Δ cannot be the identity unless d_σ^2 is also the identity. From Eqs. 10.1 and 10.2 this can only be the case if A is some subset of the column

vectors in $Q\Delta$ where $Q\Delta^2Q'$ is the basic structure of C. In particular, the case of $P_W = 0$ in Chapter 8 gives one of these solutions, namely for the so-called "principal component" solution.

For the matrices of covariances of Z and e with X, we have

$$Z'X = A\,Q_\sigma\,\Delta\,Q_\sigma' \tag{10.15}$$

and

$$e'X = A\,Q_\sigma(I - \Delta)\,Q_\sigma \tag{10.16}$$

In Eq. 10.15 we see that $Z'X$ is equal to the factor loading matrix A only if $I = \Delta$, which would be the case if A were a "principal component" factor loading matrix. It is also clear from Eq. 10.16 that only if $I = \Delta$ is the factor score matrix orthogonal to the residual matrix. However, from Eq. 10.14 we see that the factor matrix given by Eq. 10.4 does give the total covariance matrix as the sum of the estimated and the residual covariance matrices, as discussed in Chapter 5.

10.3 The True Factor Score Matrix

We shall now define a true factor score matrix X as one that is orthonormal and is orthogonal to the residual data matrix. These conditions are:

$$X'X = I \tag{10.17}$$

$$X'e = 0 \tag{10.18}$$

From Eqs. 10.6, 10.17, and 10.18 we have also

$$Z'X = A' \tag{10.19}$$

and

$$C - AA' = e'e \tag{10.20}$$

Conditions 10.19 and 10.20 are those we have previously indicated as desirable and the latter we have seen is satisfied in the previous section.

Suppose now we let

$$V = ZR^{-1}A \tag{10.21}$$

We recognize the right side of Eq. 10.21 as the scale free estimate of X discussed in Chapter 7. However, it does not satisfy the conditions given in Eqs. 10.17 through 10.20. As a matter of fact, there is no right hand transformation of Z which in general does satisfy these conditions. Other investigators have pointed out that to find a matrix to satisfy these conditions we must go "into the people space as distinguished from the test space," and that can be done in a multiply infinite number of ways. Let us see what this somewhat mystic complaint means in terms of simple algebra.

Suppose we let

$$X = V - P\,(I - V'V)^{\frac{1}{2}} \tag{10.22}$$

where P is restricted by

$$P'P = I \tag{10.23}$$

and

$$P'Z = 0 \tag{10.24}$$

From Eqs. 10.22, 10.23, and 10.24, it can be shown that Eqs. 10.17 through 10.20 are satisfied. For Eq. 10.24 to be satisfied we must have (Horst, 1963)

$$N \geq (n + m) \tag{10.25}$$

where N is the number of entities, n the number of attributes, and m the number of factors. If the N is equal to the right of Eq. 10.25, then there are an infinite number of P matrices differing only by a square orthonormal transformation on the left which satisfies Eqs. 10.23 and 10.24. However, if N is greater than n + m, then the indeterminacy increases. In this case, an orthonormal matrix P of width N - n exists which satisfies Eq. 10.24, and any square orthonormal transformation on the right of any matrix subset of vectors from P of width m will satisfy Eq. 10.24. This is the indeterminacy problem which Guttman (1955b) first discussed and which has cast a pall over attempts to calculate factor score matrices. However, the situation doubtless does not call for so much pessimism. Since we have

such an overwhelming embarrassment of riches from which to choose P to satisfy Eqs. 10.23 and 10.24, why not consider some simple function of the elements of P and the time-honored though sometimes distrusted scale free estimate of X given by Eq. 10.21. For example, we may consider optimizing the function

$$\phi = \mathrm{tr}\,(P'V^{(k)}) \tag{10.26}$$

where the superscript in parentheses is a positive integer and means elemental exponentiation. We now set up the function

$$\Psi = \phi - P'Z\lambda - \tfrac{1}{2}P'P\gamma \tag{10.27}$$

where λ and γ are matrices of Lagrangian multipliers. Because of Eq. 10.23, it can be shown that

$$\gamma = \gamma' \tag{10.28}$$

Differentiating Eq. 10.27 symbolically with respect to P and equating to zero we have

$$\frac{\partial \Psi}{\partial B'} = V^{(k)} - Z\lambda - P\gamma = 0 \tag{10.29}$$

From Eqs. 10.5, 10.24, and 10.29, we have

$$Z'V^{(k)} - Z'Z\lambda = 0 \tag{10.30}$$

$$\lambda = R^{-1}Z'V^{(k)} \tag{10.31}$$

From Eqs. 10.29 and 10.31

$$(I - ZC^{-1}Z')V^{(k)}\gamma^{-1} - P = 0 \tag{10.32}$$

Let

$$W = (I - ZC^{-1}Z')V^{(k)} \tag{10.33}$$

It can be shown that the only γ which satisfies both Eq. 10.28 and 10.23 is

$$(W'W)^{\frac{1}{2}} = \gamma \tag{10.34}$$

From Eqs. 10.32, 10.33, and 10.34 we have

$$W(W'W)^{\frac{1}{2}} = P \tag{10.35}$$

From Eq. 10.35 we have

$$V^{(k)} \cdot W(W'W)^{-\frac{1}{2}} = V^{(k)}_P \tag{10.36}$$

But from Eq. 10.33

$$\cdot(\cdot)' W = W'W \tag{10.37}$$

From Eqs. 10.26, 10.36, and 10.37

$$\phi = \mathrm{tr}(W'W)^{\frac{1}{2}} \tag{10.38}$$

The question of appropriate rationales for the selection of the exponent k in Eq. 10.26 has not been investigated. As a matter of fact, more complicated functions of the V matrix than the elemental positive integral power functions might be investigated. In any case, it is probable that the function ϕ in Eq. 10.26 should be held to linear functions of the elements of P to avoid iterative type solutions. No attempts have been made to apply the proposed solution to experimental data.

10.4 The Simple Structure Factor Score Matrix

In Chapter 7 and in the previous sections of this chapter, we have considered mainly the factor score matrix corresponding to the factor loading matrix A which has not yet been transformed to a simple structure matrix. We have, however, in Chapters 7 and 9 indicated that if the factor loading matrix is transformed to a simple structure factor loading matrix B by a simple structure transformation matrix h, then the factor score matrix X must be transformed into the simple structure factor score matrix Y by the transformation h'^{-1}. Thus, if

$$B = Ah \tag{10.39}$$

$$Y = Xh'^{-1} \tag{10.40}$$

These relations we have seen enable us to write

$$Z - YB' - e = 0 \tag{10.41}$$

without altering the residual matrix e in Eq. 10.6. No matter how A has been determined, the covariance matrices involving the simple structure factor score matrix must be transformed accordingly. The covariance matrix of the simple structure factor score matrix is given by

$$Y'Y = h^{-1}X'Xh'^{-1} \tag{10.42}$$

In terms of the simple structure factor loading matrix, the residual covariance matrix must now be written

$$C - BSB' = e'e \tag{10.43}$$

where

$$S = (h'h)^{-1} \tag{10.44}$$

This can readily be verified by writing from Eq. 10.39

$$A = Bh^{-1} \tag{10.45}$$

Substituting Eq. 10.45 in Eq. 10.14

$$e'e = C - Bh^{-1}h'^{-1}B' \tag{10.46}$$

or

$$e'e = C - B(h'h)^{-1}B' \tag{10.47}$$

The matrix $h'h$ and the matrix S in Eq. 10.44 have been extensively discussed by Thurstone (1947), Thomson (1950), Harmon (1967), and others.

10.5 Computing the Simple Structure Factor Score Matrix

We shall assume that the factor loading matrix A has been computed by the methods of Chapter 8 and that a simple structure transformation matrix h has been computed by the methods of Chapter 9. Assuming that this matrix gives B as indicated in Eq. 10.39, we still have the problem of signs to consider, discussed in Chapter 6 Section 4. Suppose we have determined the right and left sign matrix multipliers i_R and i_L so that from Eq. 10.39 we get

$$\beta = i_L B i_R \tag{10.48}$$

From Eqs. 10.39 and 10.48

$$\beta = i_L A h i_R \tag{10.49}$$

Actually, in the methods described in Chapter 9 the solution is such that i_R is the identity but, as indicated in Chapter 6, this is not the case for some of the transformation procedures. However, the computer programs in Chapter 14 do solve for an i_L matrix during the computations for the matrix A by the methods of Chapter 8. Therefore it is necessary to incorporate this matrix in the calculation of the simple structure factor score matrix.

To date no computer programs for computing this simple structure factor score matrix have been written. However, the procedure can be outlined. We do not actually use the basic structure factor loading matrix by

$$a = i_L A \tag{10.50}$$

Presumably, the inverse of the correlation matrix R^{-1} is available since it has been calculated in Chapter 8 to get a first approximation to the residual variances. We next calculate

$$\sigma = a' R^{-1} a \tag{10.51}$$

The basic structure factors of σ, indicated by

$$Q_\sigma d_\sigma^2 Q_\sigma' = \sigma \tag{10.52}$$

and then computed.

Next we calculate the diagonal matrix Δ from the basic diagonal in Eq. 10.52 by

$$\Delta = (I - (I - d_\sigma^2)^{\frac{1}{2}}) d_\sigma^{-2} \tag{10.53}$$

Using Δ from Eq. 10.53 and the basic orthonormals of Eq. 10.52, we calculate

$$p = Q_\sigma \Delta Q_\sigma' \tag{10.54}$$

We now need the transpose of the inverse of the simple structure transformation matrix h. It could be calculated directly but usually its minor product moment is

desired to calculate the correlations or covariances among the "true" or ideal simple structure factor scores discussed in Section 10.3. This matrix of covariances is given by

$$S = (h'h)^{-1} \tag{10.55}$$

After the minor product moment of h and its inverse S are computed, we calculate

$$h'^{-1} = hS \tag{10.56}$$

Using Eqs. 10.54 and 10.56, we then calculate

$$G = \rho h'^{-1} \tag{10.57}$$

From Eqs. 10.50 and 10.57 we calculate

$$b = aG \tag{10.58}$$

Then we get

$$F = R^{-1}b \tag{10.59}$$

Since we have assumed throughout that the diagonals of $Z'Z$ are unity, it is usually desirable in actual practice to express F as

$$F = F(\sqrt{N}) \tag{10.60}$$

If the correlation matrix has been calculated from the raw score matrix, we may calculate X from the raw score matrix as follows:

Let

Z be the raw score matrix

M be the vector of means from Z

D_σ be the diagonal matrix of standard deviations

Calculate

$$f = D_\sigma^{-1}F \tag{10.61}$$

$$V' = M'f \tag{10.62}$$

Then the X matrix is given by

$$X = (Z, 1)\begin{bmatrix} f \\ -V' \end{bmatrix} \tag{10.63}$$

That Eqs. 10.61 and 10.63 do give the same results may readily be verified.

BIBLIOGRAPHY

Anderson, T. W., and Rubin, H. Statistical inference in factor analysis. In J. Neyman (Ed.), Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. V. Berkeley: University of California Press, 1956, 111-150.

Bargmann, R. Factor analysis program for 7090, preliminary version. Internal Document, 2-126, 1963. Yorktown Heights, N.Y.: IBM Center.

Bechtold, H. P. An empirical study of the factor analysis stability hypothesis. Psychometrika, 1961, 26, 405-432.

Browne, M. W. Fitting the factor analysis model. Research Bulletin, Jan. 1967. Princeton, N.J.: Educational Testing Service.

Burket, G. R. A study of reduced rank models for multiple prediction. Psychometric Monographs, No. 12, 1964.

Carroll, J. B. An analytical solution for approximating simple structure in factor analysis. Psychometrika, 1953, 18. 23-38.

Cattell, R. B. Personality and motivation: structure and measurement. New York: Harcourt, Brace, 1957.

Clemans, W. V. An analytical and empirical examination of some properties of ipsative measures. Psychometric Monographs, No. 14, 1966.

Comrey, A. L. The minimum residual method of factor analysis. Psychological Reports, 1962, 15-18.

Coombs, C. H., and Kao, R. C. On a connection between factor analysis and multidimensional unfolding. Pschometrika, 1960, 25, 219- 231.

Davis, F. B. Fundamental factors of comprehension in reading. Psychometrika, 1944, 9, 185-197.

Guttman, L. Image theory for the structure of quantitative variates. Psychometrika, 1953, 18, 277-296.

Guttman, L. A generalized simplex for factor analysis. Psychometrika, 1955(a), 20, 173-192.

Guttman, L. The determinacy of factor score matrices with implications for five other basic problems of common-factor theory. British Journal of Statistical Psychology, 1955(b), 8, 65-82.

Hammond, W. H. An application of Burt's multiple general factor analysis to the delineation of physical types. Man, 1942, 4-11.

Harmon, H. H. Modern factor analysis. Chicago: University of Chicago Press, 1967.

Harris, C. W. On factors and factor scores. Psychometrika, 1967, 32, 363-379.

Hemmerle, W. J. Obtaining maximum likelihood estimates of factor loadings and communalities using an easily implemented iterative procedure. Psychometrika, 1965, 30, 291-302.

Holzinger, K. J., and Harmon, H. H. Factor analysis. Chicago: University of Chicago Press, 1941.

Holzinger, K. J., and Swineford, F. A study in factor analysis: The stability of a bi-factor solution. Supplementary Educational Monographs, No. 48. Chicago: Dept. of Education, University of Chicago, 1939.

Horst, P. The prediction of personal adjustment. Social Science Research Bulletin 48, 1941.

Horst, P. Matrix algebra for social scientists. New York: Holt, Rinehart and Winston, 1963.

Horst, P. Factor analysis of data matrices. New York: Holt, Rinehart and Winston, 1965.

Horst, P. Personality: measurement of dimensions. San Francisco: Jossey-Bass, 1968(a).

Horst, P. Residual variance scaling and matrix approximation. Psychological Reports, 1968(b), 22, 415-430.

Horst, P. Configural analysis and pattern recognition. Journal of Clinical Psychology, 1968(c), 24, 383-405.

Jöreskog, K. G. Some contributions to maximum likelihood factor analysis. Psychometrika, 1967, 32, 443-482.

Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. Psychometrika, 1958, 23, 187-200.

Kaiser, H. F., and Caffrey, J. Alpha factor analysis. Psychometrika, 1965, 30, 1-14.

Lawley, D. N. The estimation of factor loadings by the method of maximum likelihood. Proceedings of the Royal Society of Edinburgh, Section A, 1940, 60, 64-82.

Leiman, J. M. The calculation of regression weights from common factor loadings. Unpublished doctoral dissertation, University of Washington, 1951.

Lord, F. M. A study of speed factors in tests and academic grades. Psychometrika, 1956, 21, 31-50.

McDonald, R. P. and Burr, E. J. A comparison of four methods of contructing factor scores. Psychometrika, 1967, 32, 381-401.

Mullen, F. Factors in the growth of girls seven to seventeen years of age. Unpublished doctoral dissertation, Dept. of Education, University of Chicago, 1939.

Neuhaus, J. O., and Wrigley, C. The quartimax method: an analytical approach to orthogonal simple structure. British Journal of Statistical Psychology, 1954, 7, 81-91.

Rao, C. R. Estimation and tests of significance in factor analysis. Psychometrika, 1955, 20, 93-111.

Ross, J., and Cliff, N. A generalization of the interpoint distance model. Psychometrika, 1964, 29, 167-176.

Saunders, D. R. An analytic method for rotation to orthogonal simple structure. Research Bulletin, RB 53-10, 1953. Princeton, N.J.: Educational Testing Service.

Thomson, G. H. The factorial analysis of human ability. Boston: Houghton Mifflin Co., 1950.

Thurstone, L. L. Factor analysis and body types. Psychometrika, 1946, 11, 15-21.

Thurstone, L. L. Multiple factor analysis. Chicago: University of Chicago Press, 1947.

Thurstone, L. L. and Thurstone, L. G. Factorial studies of intelligence. Psychometric Monographs No. 2, 1941.

Tucker, L. R. Determination of parameters of a functional relation by factor analysis. Psychometrika, 1958, 23, 19-23.

Tucker, L. R. Implications of factor analysis of three-way matrices for measurement of change. In Chester Harris, et al., Problems in measuring change, Madison, Wis.: University of Wisconsin Press, 1963, 122-137.

Security Classification

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
University of Washington, Seattle, Washington	None
	2b. GROUP: None

3. REPORT TITLE

Generalized Factor Analysis: Part I Rationale

4. DESCRIPTIVE NOTES (Type of report and, inclusive dates)

Does not apply

5. AUTHOR(S) (First name, middle initial, last name)

Paul Horst

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
April 1969	88	44

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
Office of Naval Research Contract Nonr-477(33)	None
b. PROJECT NO. None	
d. None	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) None

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
None	Office of Naval Research

13. ABSTRACT

This document is Part I of a two-part report. In this report the rationale for a generalized approach to scale free factor analysis is developed. Special cases of a scaling function parameter and a loss function parameter yield the canonical, minres, alpha and principal components models of factor analysis.

A new method of oblique transformation to simple structure factor loadings is developed. Rationales for the computation of factor scores are generalized and a new rationale is presented.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6811

Security Classification

4-31408

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Scale Free Factor Analysis						
Oblique Rotations						
Simple Structure						
Canonical Factor Analysis						
Minres Factor Analysis						
Alpha Factor Analysis						
Principal Axis Factor Analysis						
Factor Scores						