

CHAPTER 8
FACTOR EXTRACTION BY MATRIX FACTORING TECHNIQUES

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

©1997

CHAPTER 8
FACTOR EXTRACTION BY MATRIX FACTORING TECHNIQUES

In this chapter we delve into a number of intricacies of factor extraction by matrix factoring techniques. These methods are the older ones followed in many early studies when there were considerable computation limitations. These methods were developed prior to the advent of large scale computers. However, some of the techniques were not utilized due to the computing requirements but were considered as highly desirable. With modern computers these methods are quite feasible. For a general framework in considering the factor extraction techniques we refer to Guttman's (1944) general theory and methods for matrix factoring which provides a theoretic foundation for the matrix factoring techniques which had been in use or had been being considered for some time. The methods to be considered in this chapter maximize some function of the obtained factor loadings. Consideration of residuals is only tangential. Two areas of problems are closely related to the factor extraction procedures, these are: the problem of "guessed communalities" and the problem of number of factors to be extracted. After presentation of the general theory of matrix factoring, a section will discuss the subject of "guessed communalities". Determination of the number of factors appears to be closely related to the method of factor extraction and will be discussed with each such technique.

Before discussion of details of matrix factoring techniques there are several preliminary matters to be considered. These techniques are not scale free when applied to covariance matrices in general in that results vary with scaling of the attributes. The property of a factor extraction technique being scale free may be explained in reference to the following equations. Let C be an original covariance matrix to which a scaling diagonal matrix D is applied, D being finite, positive, non-singular. Let \tilde{C} be the rescaled covariance matrix.

$$\tilde{C} = DCD$$

A selected factor extraction technique is applied to \tilde{C} to yield a common factor matrix \tilde{A} which is scaled back to A_c for the original covariance matrix by:

$$A_c = D^{-1}\tilde{A}$$

For the factor extraction technique to be scale free, matrix A_c must be invariant with use of different scaling matrices D . As is well known, a correlation matrix is independent of the scaling of a covariance matrix.

$$R = S^{-1}CS^{-1} = \tilde{S}^{-1}\tilde{C}\tilde{S}^{-1}$$

where

$$S^2 = \text{Diag}(C)$$

$$\tilde{S}^2 = \text{Diag}(\tilde{C})$$

Note that:

$$\tilde{S} = DS$$

A common factor matrix A obtained from R may be scaled back to \tilde{A} and A_c by:

$$\tilde{A} = \tilde{S} A$$

$$A_c = SA$$

In order to avoid scaling problems we follow the tradition of applying these techniques to correlation matrices. In a sense, this usage does make these techniques scale free when the obtained factor matrix may be scaled back to apply to the original attribute scales.

For notational convenience, the subscripts of the observed correlation matrix R_{yy} are dropped so that the observed correlation matrix is indicated by R . Also, matrix Δ_R will be indicated by Δ . Equation (7.17) becomes:

$$R = AA' + U^2 + \Delta \quad (8.1)$$

This is the basic equation considered in this chapter. An alternative equation is obtained by defining matrix R_h with adjusted diagonal entries with "guessed communalities" by:

$$R_h = R - U^2 \quad (8.2)$$

From equation (8.1):

$$R_h = AA' + \Delta \quad (8.3)$$

Matrix Δ contains residual correlations. Several letters used in transformations of factors will be used in the present context on a temporary basis to designate other matrices in this chapter.

8.1. General Theory of Matrix Factoring

Guttman (1944) developed a general theory of matrix factoring with which he described several existing methods of factor extraction. This theory applies, strictly in the present context, to Gramian matrices. These matrices need not be of full rank; however, for the present purposes they must not have imaginary dimensions (the least eigenvalue must be non-negative). A correlation matrix satisfies these conditions since it is the product of a score matrix times its transpose, an original definition of a Gramian matrix. Usually, however, matrix R_h with adjusted

diagonal entries containing "guessed communalities" is not Gramian. Nevertheless, the theory of matrix factoring is applied to matrix R_h . In practice, with a few exceptions, this use appears to work satisfactorily.

The general procedure starts from a Gramian matrix G_1 which is $n \times n$ and of rank r , greater than 0 and equal to or less than n . A matrix W , $n \times m$ with m greater than 0 and equal to or less than n . W is to contain real numbers and satisfy a restriction stated later. Matrix Q is defined by:

$$Q = G_1 W \quad (8.4)$$

Note that Q is $n \times m$. Matrix P , $m \times m$, is defined by:

$$P = W' G_1 W = W' Q \quad (8.5)$$

Matrix P is Gramian and must be of full rank; this is the restriction on matrix W . A square decomposition matrix F is determined such that:

$$F F' = P \quad (8.6)$$

Any of several techniques may be used to determine F . A section of a factor matrix, $n \times m$, on orthogonal axes is defined by:

$$A_1 = Q (F^{-1})' \quad (8.7)$$

and a residual matrix G_2 is defined by:

$$G_2 = G_1 - A_1 A_1' \quad (8.8)$$

There is a dual problem of proof. First, that the rank of G_2 is $(r - m)$. Second, that A_1 is a section with m columns of a complete, orthogonal factor matrix of G_1 .

The required proofs are expedited by considering a complete decomposition of G_1 to a matrix B , $n \times r$, such that:

$$G_1 = B B' \quad (8.9)$$

B is a factor matrix on orthogonal axes and can be obtained by any of a number of procedures such as the procedure developed by Commandant A. L. Cholesky of the French Navy around 1915 and described by Dwyer (1944) as the square root method. There are many other possible procedures to obtain this decomposition. With equation (8.9), matrices Q and P become:

$$Q = B B' W \quad (8.10)$$

$$P = W' B B' W \quad (8.11)$$

Then, matrix A_1 becomes:.

$$A_1 = BB'W(F^{-1})' \quad (8.12)$$

Define a matrix T_1 , $r \times m$:

$$T_1 = B'W(F^{-1})' \quad (8.13)$$

Then, matrix A_1 becomes:

$$A_1 = BT_1 \quad (8.14)$$

Matrix T_1 is column wise orthonormal as shown by the following:

$$\begin{aligned} T_1 T_1' &= F^{-1} W' B B' (F^{-1})' \\ &= F^{-1} P (F^{-1})' \\ &= F^{-1} F F' (F^{-1})' \\ &= I \end{aligned} \quad (8.15)$$

Matrix T , $r \times r$ orthonormal, is completed by adjoining section T_2 , $r \times (r - m)$ to T_1 .

$$T = [T_1, T_2] \quad (8.16)$$

T_2 is column wise orthonormal and orthogonal by columns to T_1 . Matrix B is rotated orthogonally by T to yield matrix A with sections A_1 and A_2 .

$$BT = B[T_1, T_2] = [BT_1, BT_2] = [A_1, A_2] = A \quad (8.17)$$

Since T is an orthonormal rotation and from equation (8.9):

$$G_1 = AA' = A_1 A_1' + A_2 A_2' \quad (8.18)$$

Then with equation (8.8):

$$G_2 = A_2 A_2' \quad (8.19)$$

Matrix G_2 is of rank $(r - m)$. The derivation of A_1 has removed this section from the complete matrix A . This completes the needed proof.

A point of interest is that G_2 is orthogonal to W as shown as follows.

$$G_2 W = (G_1 - A_1 A_1') W = G_1 W - A_1 A_1' W \quad (8.20)$$

Note, from equation (8.7), that:

$$A_1 A_1' = Q(F^{-1})' F^{-1} Q' = Q P^{-1} Q' \quad (8.21)$$

With the results of equation (8.21), equation (8.20) becomes:

$$\begin{aligned}
 G_2W &= G_1W - QP^{-1}Q'W \\
 &= Q - QP^{-1}P \\
 &= 0
 \end{aligned}
 \tag{8.22}$$

A second point of interest is the relation between matrix W and the obtained factor weight matrix A_1 . From equations (8.5), (8.6), and (8.7)

$$\begin{aligned}
 W'A_1 &= W'Q(F^{-1})' \\
 &= P(F^{-1})' \\
 &= FF'(F^{-1})' \\
 &= F
 \end{aligned}
 \tag{8.23}$$

This result is important in the factor extraction methods.

Equations (8.4) through (8.8) are the bases of major steps in the factor extraction techniques to be considered in this chapter. These methods differ in the determination of matrix W . Each of these techniques starts with a correlation matrix R_{h1} , having adjusted diagonal entries, determines a weight matrix W , computes a section of a factor matrix A_1 , then computes a residual matrix R_{h2} by equation (8.8). This process is repeated with the residual matrix to obtain the next section of the extracted factor matrix. This process is repeated with the succession of residual matrices until the full extracted factor matrix is obtained. As indicated earlier, the number of factors to be extracted is a decision frequently made from information obtained during the factor extraction process. The basis of this decision depends on the factor extraction technique employed.

8.2. The "Guessed Communalities" Problem

A preliminary operation in matrix factoring is to establish entries in the diagonals of the correlation matrix. The theory of common factor analysis presented in the preceding chapters establishes a basis for this operation. Note that the principal components procedure ignores the issue of common vs. unique variance and leaves unities in the diagonal of the correlation matrix. An early operational procedure using communality type values was the "Highest R" technique developed by Thurstone (1947) and was used in many of his studies as well as by many other individuals following Thurstone's lead. However, early in applications of digital computers there was a proposal to use principal components as an easy approximation to factor analysis. Kaiser (1960) described such a technique followed by VARIMAX transformation of the weight matrix. This procedure, which became known as the "Little Jiffy" after a suggestion by Chester Harris, is retained as an alternative in a number of computer packages. We can not recommend this

procedure and are concerned that many unwary individuals have been misled by the ease of operations. Serious problems exist.

A very simple example is presented in Table 8.1 with the constructed correlation matrix being given at the upper left. This correlation matrix was computed from a single common factor with uniqueness so that the theoretic communalities were known and have been inserted into the matrix on the upper right. There are neither sampling nor lack of fit discrepancies. A principal components analysis is given on the left and a principal factors analysis is given on the right. The principal factors procedure will be considered later in detail. For the principal components analysis an eigen solution (see Appendix A on matrix algebra for a discussion of the eigen problem) was obtained of the correlation matrix having unities in the diagonal. The series of eigenvalues are given on the left along with the component weights for one dimension. These weights are the entries in the first eigenvector times the square root of the first eigenvalue. The matrix of residuals after removing this first component is given at the bottom left. The principal factors analysis followed the same procedure as the principal components analysis but is applied to the correlation matrix having communalities in the diagonal. For the principal factors, both the weights for one factor and the uniqueness are given. These are the values used in constructing the correlation matrix.

There are a number of points to note. First, the eigenvalues for the principal components after the first value do not vanish as do the corresponding eigenvalues for the principal factors analysis. If one followed the procedure for the principal components of retaining only dimensions for which eigenvalues were greater than unity, only the first principal component would be used, this being the number of dimensions used in this example. For the principal factors analysis, only one factor existed since there was only one eigenvalue that did not vanish. A more important comparison concerns the obtained weights. All of the principal component weights are greater than the corresponding principal factor weights. This is especially true for the low to medium sized weights. Use of the principal components procedure exaggerates the values of the obtained weights. A further comparison is provided by the residual matrices. For the principal components analysis on the left, the diagonal values might be taken to be unique variances (a common interpretation). However, note the negative residuals. The component weights have removed too much from the correlations leaving a peculiar residual structure. For the principal factors analysis, all of the residual entries vanish in this example. The principal factors analysis yields a proper representation of the correlation matrix. A conclusion from this example is that leaving unities in the diagonals of a correlation matrix leads to a questionable representation of the structure of the correlation matrix.

Tables 8.2 and 8.3 provide a comparison when lack of fit is not included and is included in a constructed correlation matrix. The correlation matrix in Table 8.2 was computed from the

Table 8.1

Comparison of factor Extraction from a Correlation Matrix with Unities in the Diagonal
versus Communalities in the Diagonal

Correlation Matrices

Unities in Diagonal (RW1)				
	1	2	3	4
1	<u>1.00</u>	.35	.21	.07
2	.35	<u>1.00</u>	.15	.05
3	.21	.15	<u>1.00</u>	.03
4	.07	.05	.03	<u>1.00</u>

Communalities in Diagonal (RWH)				
	1	2	3	4
1	<u>.49</u>	.35	.21	.07
2	.35	<u>.25</u>	.15	.05
3	.21	.15	<u>.09</u>	.03
4	.07	.05	.03	<u>.01</u>

Eigenvalues of (RW1)	
1	1.50
2	.99
3	.87
4	.64

Eigenvalues of (RWH)	
1	.84
2	.00
3	.00
4	.00

Principal Components	
Weights	
1	.78
2	.73
3	.56
4	.22

Principal Factors		
	Weights	Uniqueness
1	.70	.51
2	.50	.75
3	.30	.91
4	.10	.99

Residual Matrix from (RW1)				
	1	2	3	4
1	<u>.40</u>	-.22	-.23	-.10
2	-.22	<u>.46</u>	-.26	-.11
3	-.23	-.26	<u>.69</u>	-.09
4	-.10	-.11	-.09	<u>.95</u>

Residual Matrix from (RWH)				
	1	2	3	4
1	<u>.00</u>	.00	.00	.00
2	.00	<u>.00</u>	.00	.00
3	.00	.00	<u>.00</u>	.00
4	.00	.00	.00	<u>.00</u>

Table 8.2
 Illustration of “Communality” Type Values
 Simulated Correlation matrix without Discrepancies of Fit

	Major Domain Matrix		Variance Components		
	1	2	Major	Unique	Minor
1	.8	.0	.64	.36	.00
2	.6	.0	.36	.64	.00
3	.5	.1	.26	.74	.00
4	.2	.3	.13	.87	.00
5	.0	.2	.04	.96	.00
6	.0	.6	.36	.64	.00

Simulated Population Correlation Matrix

	1	2	3	4	5	6
1	1.00					
2	.48	1.00				
3	.40	.30	1.00			
4	.16	.12	.13	1.00		
5	.00	.00	.02	.06	1.00	
6	.00	.00	.06	.18	.12	1.00

“Communality” Type Values

Method	Attribute					
	1	2	3	4	5	6
Major Domain Variance	.64	.36	.26	.13	.04	.36
Highest R	.48	.48	.40	.18	.12	.18
Squared Multiple Correlation	.31	.25	.18	.07	.02	.05
Iterated Centroid Factors	.64	.36	.26	.13	.04	.36
Iterated Principal Factors	.64	.36	.26	.13	.04	.36
Alpha Factor Analysis	.64	.36	.26	.13	.04	.36
Unrestricted Maximum Likelihood	.64	.36	.26	.13	.04	.36

Table 8.3
 Illustration of “Communality” Type Values
 simulated Correlation matrix without Discrepancies of Fit

	Major Domain Matrix		Variance Components		
	1	2	Major	Unique	Minor
1	.8	.0	.64	.36	.10
2	.6	.0	.36	.54	.10
3	.5	.1	.26	.64	.10
4	.2	.3	.13	.77	.10
5	.0	.2	.04	.86	.10
6	.0	.6	.36	.54	.10

Simulated Population Correlation Matrix

	1	2	3	4	5	6
1	1.00					
2	.49	1.00				
3	.37	.28	1.00			
4	.17	.16	.15	1.00		
5	-.04	-.02	.03	.02	1.00	
6	-.03	.02	.07	.21	.10	1.00

“Communality” Type Values

Method	Attribute					
	1	2	3	4	5	6
Major Domain Variance	.64	.36	.26	.13	.04	.36
Highest R	.49	.49	.37	.21	.10	.21
Squared Multiple Correlation	.31	.25	.16	.09	.01	.06
Iterated Centroid Factors	.67	.36	.23	.12	.01	.73
Iterated Principal Factors	.65	.37	.23	.15	.02	.50
Alpha Factor Analysis	.69	.37	.23	.11	.02	.63
Unrestricted Maximum Likelihood	.65	.37	.23	.16	.02	.45

major domain matrix and the listed uniqueness. This is a perfect population correlation matrix. In contrast, the technique described in Chapter 3, section 3.9, was used to add lack of fit discrepancies to the generated correlation matrix given in Table 8.3. The bottom section of each of these tables presents communality type values determined by a number of techniques. Discussion will compare results by these techniques between the cases when there are no discrepancies of fit and when there are discrepancies of fit.

Consider the case in Table 8.2 when no discrepancies of fit were included. The first row in the bottom section contains the major domain variances which may be considered as theoretic communalities. The second and third rows contain communality type values used in factor extraction techniques. These will be discussed in subsequent paragraphs. The last four rows give results from four methods of factor extraction which should result in ideal communalities in the population. These methods will be discussed in detail in subsequent sections. All four of these techniques utilize iterative computing methods to arrive at stable determinations of communalities. Note that all four techniques are successful in arriving at the theoretic values given in the row for the major domain variances. If a correlation matrix were determined from a sample of individuals from a population characterized by a given population matrix which does not include discrepancies of fit, an objective of factor extraction from this sample correlation matrix would be to arrive at estimates of the major domain variances.

Consider the case in Table 8.3 when discrepancies of fit were included. As before, the first row in the bottom section contains the major domain variances which, as will be demonstrated, no longer can be considered as theoretic communalities. As before, the second and third rows contain communality type values used in factor extraction techniques and will be discussed subsequently. The last four rows give results for this case from four methods of factor extraction which will be discussed in detail in subsequent sections. The values in these rows not only vary from the values in Table 8.2 but also differ between methods of factor extraction. Note, especially the values for attribute 6 for which the communality type values for the four methods of factor extraction vary from .73 to .45 and are all greater than the .36 in Table 8.2. Inclusion of the discrepancies of fit has had an effect on these values which is different for different methods of factor extraction. There no longer is a single ideal solution. The population communalities differ by method of factor extraction and provide different objectives to be estimated from sample correlation matrices. This conclusion poses a considerable problem for statistical modeling which ignores discrepancies of fit. We conclude that ignoring discrepancies of fit is quite unrealistic and raises questions concerning several factoring techniques. This is an illustration of the discussion in Chapter 3 of the effects of lack of fit on results obtained by different methods of fitting the factor model.

8.2.1. Highest R Procedure

Thurstone developed the highest R procedure during the analyses of large experimental test batteries such as in his study of Primary Mental Abilities (1938). Computations were performed using mechanical desk calculators so that a simple procedure was needed which would yield values in the neighborhood of what might be considered as the most desirable values. The highest R technique is based on intuitive logic; there is no overall developmental justification.

Application of the highest R technique involves, for each attribute, j , finding the other attribute, i , which correlates most highly in absolute value with j . This correlation, in absolute value, is taken as a communality type value for attribute j . This procedure is illustrated in Table 8.4. The correlation matrix in this table is the same as in Table 8.3 with reversals of directions of attributes 3 and 5. In making these reversals of directions, algebraic signs are reversed in rows and columns for these two attributes. Note that there are double reversals for the diagonal entries and the correlation between the two attributes which results in no sign changes for these entries. Consider the column for attribute 1. Ignore the diagonal entry and find the largest other entry in absolute value. This value is .49 in row 2 and is recorded in the "Highest R" row. Note that for column 3 the highest value in absolute value is .37, this is the -.37 in row 1. The foregoing procedure is followed for each of the attributes. Note that the sign changes in the correlation matrix between Tables 8.3 and 8.4 did not change the values of the highest R's.

Justification of the highest R technique rests mostly on the idea that attributes for which communality type values should be high should correlate more highly with other attributes in a battery than would be true for attributes for which the communality type values should be low. This relation should be more nearly true for larger sized batteries of attributes than for smaller sized batteries such as used in the illustration.

There are a few algebraic relations to be considered. Consider two attributes, i and j . From equation (8.1) when there are no discrepancies of fit the correlation between these two attributes is:

$$r_{ij} = \sum_k a_{ik} a_{jk} \quad (8.24)$$

This correlation can be expressed in trigonometric terms as the scalar product between two vectors.

$$r_{ij} = h_i h_j \cos \theta_{ij} \quad (8.25)$$

where h_i and h_j are the lengths of these two vectors and $\cos \theta_{ij}$ is the cosine of the angle between these two vectors. When the absolute value of the correlation is considered, as required in the highest R procedure:

Table 8.4
Illustration of “Highest R”

	<u>Correlation Matrix</u>					
	1	2	3	4	5	6
1	1.00					
2	.49	1.00				
3	-.37	-.28	1.00			
4	.17	.16	-.15	1.00		
5	.04	.02	.03	-.02	1.00	
6	-.03	.02	-.07	.21	-.10	1.00
Highest R	.49	.49	.37	.21	.10	.21

$$|r_{ij}| = h_i h_j |\cos\theta_{ij}| \quad (8.26)$$

When the two attribute vectors are collinear

$$|\cos\theta_{ij}| = 1$$

so that

$$|r_{ij}| = h_i h_j$$

If the two attribute vectors have equal length,

$$h_i = h_j$$

and

$$|r_{ij}| = h_i^2 = h_j^2$$

Thus, in this special case $|r_{ij}|$ would yield the desired communality type values. In case the two vectors are not of equal length, such as:

$$h_i > h_j$$

then:

$$h_i^2 > |r_{ij}| > h_j^2$$

so that $|r_{ij}|$ is too low for one of the h^2 's and too high for the other. When the two attribute vectors are not collinear

$$|\cos\theta_{ij}| < 1$$

and there is a tendency for $|r_{ij}|$ to be less the desired communality type values. The selection of attribute 1 to have a high correlation in absolute value will tend toward $|\cos\theta_{ij}|$ having a value approaching unity.

8.2.2. Squared Multiple Correlations (SMC)'s

The "squared multiple correlation", or SMC, is the most commonly used communality type value used in factor extraction procedures. This coefficient is the squared multiple correlation for an attribute j in a multiple, linear regression of that attribute with all other attributes in a battery. Roff (1936) followed by Dwyer (1939) and Guttman (1940) showed that, in perfect cases, the communality of an attribute was equal to or greater than the SMC for that attribute. These developments presumed that the common factor model fit the correlation matrix precisely (there was no lack of fit) and that a population matrix was considered (there were no sampling discrepancies). Dwyer used a determinantal derivation while Guttman used super

matrices. We follow the Guttman form in our developments of this proposition. Guttman (1956) described the SMC's as the "best possible" systematic estimates of communalities; a conclusion justified in terms of a battery of attributes increasing indefinitely without increasing the number of common factors. Following this development, the SMC became widely adopted and incorporated into computer packages.

Standard procedures for computation of the SMC will be considered first. Initially, the case is to be considered when the correlation matrix is nonsingular. This case should cover the majority of factor analytic studies. Each attribute is considered in turn as a dependent variable with the remaining $(n - 1)$ attributes being considered as a battery of independent attributes. A linear regression is considered relating attribute j to the battery of independent attributes with the smc_j being the squared multiple correlation in this regression. The variance of the discrepancies between the observed values of j and regressed values is designated by s_{ej}^2 . From regression theory:

$$smc_j + s_{ej}^2 = 1 \quad (8.27)$$

a super matrix is constructed as illustrated in the following equations.

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{jj} & \mathbf{R}_{jI} \\ \mathbf{R}_{Ij} & \mathbf{R}_{II} \end{bmatrix} \quad \mathbf{R}^{-1} = \begin{bmatrix} \mathbf{R}^{jj} & \mathbf{R}^{jI} \\ \mathbf{R}^{Ij} & \mathbf{R}^{II} \end{bmatrix} \quad \mathbf{R}\mathbf{R}^{-1} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \quad (8.28)$$

The dependent attribute is indicated by the subscript j while the battery of independent attributes is indicated by the subscript I . Thus, \mathbf{R}_{jj} contains unity for the variance of the dependent attribute, \mathbf{R}_{jI} contains the correlations of the dependent attribute with the attributes in the independent battery. Similarly, \mathbf{R}_{Ij} contains the correlations of the independent attributes with j and \mathbf{R}_{II} contains the intercorrelations of the independent attributes. From regression theory:

$$smc_j = \mathbf{R}_{jI} \mathbf{R}_{II}^{-1} \mathbf{R}_{Ij} \quad (8.29)$$

The computational problem is to obtain smc_j .

The common computing procedure involves the inverse of the correlation matrix which is illustrated in equation (8.28) in super matrix form. Cells of the inverse matrix are indicated by superscripts. Only the \mathbf{R}^{jj} cell is important in the present context. Note, in equation (8.28) that a super identity matrix is indicated as the product of \mathbf{R} and its inverse. From this representation:

$$\mathbf{R}_{jj} \mathbf{R}^{jj} + \mathbf{R}_{jI} \mathbf{R}^{Ij} = \mathbf{I} \quad (8.30)$$

$$\mathbf{R}_{Ij} \mathbf{R}^{jj} + \mathbf{R}_{II} \mathbf{R}^{Ij} = \mathbf{0} \quad (8.31)$$

From equation (8.31):

$$\mathbf{R}^{Ij} = -\mathbf{R}_{II}^{-1} \mathbf{R}_{Ij} \mathbf{R}^{jj}$$

which yields \mathbf{R}^{Ij} which may be substituted into equation (8.30) to obtain:

$$\mathbf{R}_{jj} \mathbf{R}^{jj} - \mathbf{R}_{Ij} \mathbf{R}_{II}^{-1} \mathbf{R}^{jj} = \mathbf{I}$$

This equation may be solved to yield the desired equation for \mathbf{R}^{jj} :

$$\mathbf{R}^{jj} = (\mathbf{R}_{jj} - \mathbf{R}_{jI} \mathbf{R}_{II}^{-1} \mathbf{R}_{Ij})^{-1} \quad (8.32)$$

Equations for other cells of the inverse matrix may be obtained by similar solutions when desired. With equation (8.29) equation (8.32) becomes:

$$\mathbf{R}^{jj} = (\mathbf{R}_{jj} - smc_j)^{-1}$$

so that:

$$smc_j = \mathbf{R}_{jj} - (\mathbf{R}^{jj})^{-1} \quad (8.33)$$

An alternative involves s_{ej}^2 :

$$s_{ej}^2 = (\mathbf{R}_{jj})^{-1} \quad (8.34)$$

This equation is obtained from (8.27) and (8.34) with \mathbf{R}_{jj} being substituted for its equivalent unity in (8.27).

Equation (8.34) is extended to involving all attributes in the battery by defining diagonal matrix \mathbf{S}_e^2 containing the s_{ej}^2 as the diagonal elements. Equation (8.34) may be expanded to:

$$\mathbf{S}_e^2 = [\mathbf{Diag}(\mathbf{R}^{-1})]^{-1} \quad (8.35)$$

To adjust the diagonal elements of \mathbf{R} to having the SMC's, matrix \mathbf{S}_e^2 must be subtracted from \mathbf{R} in accordance to equation (8.27). Let \mathbf{R}_{smc} be the correlation matrix with the SMC's in the diagonal cells.

$$\mathbf{R}_{smc} = \mathbf{R} - \mathbf{S}_e^2 \quad (8.36)$$

The two preceding equation provide the basis for computations.

There is trouble in applying the preceding procedure when the correlation matrix is singular since, in this case, the inverse does not exist. A simple modification of this procedure was suggested by Finkbeiner and Tucker (1982). A small positive number, k , is to be added to each diagonal entry of \mathbf{R} to yield a matrix $\tilde{\mathbf{R}}$.

$$\tilde{\mathbf{R}} = \mathbf{R} + k\mathbf{I} \quad (8.37)$$

Define:

$$\tilde{S}_e^2 = [\text{diag}(\tilde{R}^{-1})]^{-1} \quad (8.38)$$

and compute \tilde{R}_{smc} by:

$$\tilde{R}_{smc} = \tilde{R} - \tilde{S}_e^2 \quad (8.39)$$

Two statements of approximation follow. As k approaches zero:

$$(\tilde{S}_e^2 - kI) \text{ approaches } S_e^2;$$

$$\tilde{R}_{smc} \text{ approaches } R_{smc}.$$

Finkbeiner and Tucker suggested a value of $k = .000001$ with which discrepancies in the approximations were in the fourth decimal place.

The advantage of the Finkbeiner and Tucker modification is that when R is a true Gramian matrix but is singular due to inclusion of dependent attributes in the battery, matrix \tilde{R} is not singular so that its inverse is possible. However, use of this modification does not remove the dependency; the procedure permits determination of which attributes form a dependent group. For example, if all scores of a multiple part test are included in the battery along with the total score, all part scores and the total score will be dependent so that their squared multiple correlations will be unity. Other attributes in the battery may not have unit multiple correlations. Use of the Finkbeiner and Tucker procedure will yield, within a close approximation, the unit squared multiple correlations and those that are not unity. Such dependencies should be eliminated from the battery by removing dependent measures such as the total score. These dependencies make what otherwise would be unique factors into common factors thus enlarging the common factor space as well as distorting this space.

There are several types of situations for which the correlation matrix may not be Gramian. One such type of situation is when the individual correlations are based on different samples. Missing scores can produce this situation. Another type situation is when tetrachoric correlations are used in the correlation matrix. The Finkbeiner and Tucker procedure usually will not correct for these situations. A possibility is to use the highest R technique.

The stated inequality between the communality and the squared multiple correlation is considered next. In the development here, the use of super matrices is continued. Assume that a population correlation matrix is being considered without discrepancies of fit so that the following equation holds:

$$R = AA' + U^2$$

where \mathbf{A} is an $n \times r$ factor matrix on orthogonal axes. This equation may be expressed in super matrix form for the correlation matrix in equation (8.28):

$$\begin{bmatrix} R_{jj} & R_{jI} \\ R_{Ij} & R_{II} \end{bmatrix} = \begin{bmatrix} A_j \\ A_I \end{bmatrix} \begin{bmatrix} A_j & A_I \\ A_j & A_I \end{bmatrix} + \begin{bmatrix} u_j^2 & 0 \\ 0 & U_I^2 \end{bmatrix} \quad (8.37)$$

A convenient transformation is considered next. For this transformation all uniquenesses for the attributes in battery I must be greater than zero. Let matrix \mathbf{T}' contain eigenvectors and diagonal matrix \mathbf{D} contain eigenvalues of the following matrix so that:

$$A_I U_I^{-2} A_I = T' D \quad (8.38)$$

Matrix \mathbf{T}' is square, orthonormal. Following is the transformation.

$$\begin{bmatrix} A_j \\ A_I \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & U_I^{-1} \end{bmatrix} \begin{bmatrix} A_j \\ A_I \end{bmatrix} T' \quad (8.39)$$

This transformation rescales the attributes in battery I to having unit uniquenesses and applies an orthogonal transformation on the common factor space. Then:

$$A_I' A_I = D \quad (8.40)$$

and:

$$\begin{bmatrix} R_{jj} & C_{jI} \\ C_{Ij} & C_{II} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & U_I^{-1} \end{bmatrix} \begin{bmatrix} R_{jj} & R_{jI} \\ R_{Ij} & R_{II} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & U_I^{-1} \end{bmatrix} \quad (8.41)$$

Note that the resealing of attributes in battery I results in covariances for these attributes; however, attribute j is not rescaled. Also, the orthogonal transformation leaves the formula for the communality of attribute j at:

$$h_j^2 = \sum_k a_{jk}^2 \quad (8.42)$$

The transformation is applied to the uniquenesses by:

$$\begin{bmatrix} 1 & 0 \\ 0 & U_I^{-1} \end{bmatrix} \begin{bmatrix} u_j^2 & 0 \\ 0 & U_I^{-1} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & U_I^{-1} \end{bmatrix} = \begin{bmatrix} u_j^2 & 0 \\ 0 & I \end{bmatrix} \quad (8.43)$$

The result of this transformation is that:

$$\begin{bmatrix} R_{jj} & C_{jI} \\ C_{Ij} & C_{II} \end{bmatrix} = \begin{bmatrix} A_j \\ A_I \end{bmatrix} \begin{bmatrix} A'_j & A'_I \end{bmatrix} + \begin{bmatrix} u_j^2 & 0 \\ 0 & U_I^2 \end{bmatrix} \quad (8.44)$$

This transformation leads to a simple form for the inverse of C_{II} which is used in the regression of attribute j on the battery I. Regression weights for the attributes in battery I have to be scaled accordingly; however, the squared multiple correlation is not altered by the scaling of battery I.

From multiple regression theory the normal equations relating attribute j to battery I are given by:

$$C_{II} B_{Ij} = C_{Ij} \quad (8.45)$$

where B_{Ij} is a column vector of regression weights. The inverse of C_{II} is given by the following equation:

$$C_{II}^{-1} = I - A_I(D + I)^{-1} A'_I \quad (8.46)$$

so that the solution for B_{Ij} is:

$$\begin{aligned} B_{Ij} &= C_{II}^{-1} C_{Ij} = C_{Ij} - A_I(D + I)^{-1} A'_I C_{Ij} \\ &= A'_I A_j - A_I(D + I)^{-1} A'_I A_I A'_j \\ &= A_I [I - (D + I)^{-1} D] A'_j \end{aligned} \quad (8.47)$$

The variance of the regressed values of j on battery I is designated by v_j and given by:

$$v_j = B'_{Ij} C_{II} B_{Ij} = B'_{Ij} C_{Ij}$$

With equations (8.44) and (8.47):

$$\begin{aligned} v_j &= A_j [I - D(D + I)^{-1}] A'_I A_I A'_j \\ &= A_j [I - D(D + I)^{-1}] D A'_j \\ &= A_j [D - D(D + I)^{-1} D] A'_j \end{aligned} \quad (8.48)$$

Let diagonal matrix W be defined in the present context by:

$$W = D - D(D + I)^{-1} D \quad (8.49)$$

Equation (8.48) becomes:

$$v_j = A_j W A'_j = \sum_k a_{jk}^2 w_k \quad (8.50)$$

where w_k is the k 'th diagonal entry in W . The value of this entry can be expressed as:

$$w_k = d_k - d_k(d_k + 1)^{-1}d_k$$

which becomes with algebraic manipulation:

$$w_k = d_k / (d_k + 1) \quad (8.51)$$

With the variance of standardized measures of attribute j being unity, R_{jj} equaling unity, the variance, v_j , of the regressed values of attribute j , equals the squared multiple correlation of attribute j on battery I.

$$smc_j = v_j \quad (8.52)$$

The important relation with which this development is concerned compares the communality of attribute j with the squared multiple correlation of this attribute. For this comparison a difference is taken:

$$h_j^2 - smc_j = \sum_k a_{jk}^2 - \sum_k a_{jk}^2 w_k = \sum_k a_{jk}^2 (1 - w_k) \quad (8.53)$$

From equation (8.51):

$$(1 - w_k) = 1 / (d_k + 1) \quad (8.54)$$

Note that the diagonal entries, d_k of matrix D are the sums of squares of the entries in columns of factor matrix A_I , as per equation (8.40), so that these d_k 's must be positive. The possibility of a zero value of a d_k is discarded since this would imply a column of zero factor loadings which is ruled out by the definition that the factor matrix have a rank equal to its column order. Then, for all $k = 1, r$:

$$d_k > 0$$

and from equation (8.54)

$$0 < (1 - w_k) < 1 \quad (8.55)$$

The possibility of $(1 - w_k)$ equaling 0 will be discussed later. Equations (8.53) and (8.55) lead to the following important inequality:

$$h_j^2 > smc_j \quad (8.56)$$

An illustration of the squared multiple correlations for the perfect case is given in Table 8.2. The smc's for all six attributes are markedly less than the major domain variances which are the theoretic communalities for this case. The relation between the communalities and the squared multiple correlations is dependent on the values of the d_k 's. An important effect is the relation of the d_k 's to the battery size. Since each d_k is the sum of squares of scaled factor weights, as the

battery size increases without increasing the number of factors, each d_k will increase which will lead to a decrease in the value of $(1 - w_k)$. This will lead to a reduction in the differences between the communalities and the squared multiple correlations so that the squared multiple correlations will become better estimates of the communalities as the battery size is increased. A limiting condition pointed out by Guttman (1956) is for the battery size to approach infinite; then the d_k 's will approach infinity and $(1 - w_k)$'s will approach zero so that the difference between the communalities and the squared multiple correlations also will approach zero.

Application of the foregoing inequality in practice is accompanied by some uncertainties. First, inclusion of lack of fit of the model has unknown effects on squared multiple correlations. This is in addition to the idea that no longer are there fixed ideal communalities. Analyses of correlation matrices obtained from samples introduce further uncertainties. A well known effect is that the squared multiple correlation obtained from a sample of observations is biased upwards. This might lead to a violation of the inequality in case the communality is not biased similarly, the possible bias of the communality not being well known. However, use of squared multiple correlations has yielded satisfactory to good results in many practical applications.

8.2.3. Iterated Communalities

A procedure followed sometimes is to iterate the communality values. Such a procedure starts with trial communality values, extracts a factor matrix by one of the factor extraction techniques and computes output communalities from this matrix which are substituted into the diagonal entries of the correlation matrix as the next trial communalities. Each iteration starts with trial communalities and ends with output communalities which become next trial communalities. This procedure is continued until there are minimal changes from trial communalities to output communalities. Results of this type procedure for several methods of factor extraction are illustrated in Tables 8.2 and 8.3. The general idea is that these iterations lead to communality values with which the extracted factor matrix better fits the input correlation matrix.

As seen in Table 8.2 for the perfect case in the population, the iterated communalities settle to equaling the theoretic values of the major domain variances. The scheme of iterating the communalities works in this case. However, consider Table 8.3 which presents illustrations of iterated communalities when discrepancies of fit have been included in the correlation matrix. The output communalities do not equal the major domain variances nor do the communalities obtained by different methods of factor extraction equal each other. Introduction of sampling discrepancies will produce even more differences between obtained communality values and any theoretic values and among the obtained values from different methods of factor extraction. Considerable problems are raised for practical applications. Information concerning these problems might be obtained from extensive simulation, Monte Carlo type studies. Use of the

Tucker, Koopman, Linn simulation procedure (1966 and described in Chapter 3) is recommended so that obtained results may be compared with input major domain matrices. Preliminary results of such studies indicate that the iterated communalities procedure works better than more approximate procedures only for large samples and few factors compared with the battery sizes.

Convergence of the iteration procedure may be a problem. Often convergence is slow so that some techniques to speed convergence could be advisable. For methods of factor extraction, including the principal factors technique and the maximum likelihood factor analysis method, alternate computing routines have been developed for which the computing time is greatly reduced. Of the factor extraction methods illustrated in Tables 8.2 and 8.3, proof of convergence exists only for the principal factors technique and the maximum likelihood method. However, experience indicates that convergence does occur for the other techniques.

Another area of problems with iterated communalities consists of generalized Heywood cases. Initial work in this area was by H. B. Heywood (1931) who indicated that the rank of a correlation matrix using a limited number of common factors may imply either a communality greater than unity or a negative communality. The concern for iterated communalities is that one or more of the communalities becomes greater than unity which is not permissible since such a communality implies a negative uniqueness. Special procedures, to be discussed subsequently, are required to avoid this situation. Unfortunately, this case is ignored some times.

8.3. Centroid Method of Factor Extraction

The centroid method of factor extraction is presented partly for its historical value and partly for some very useful techniques used in special situations. Thurstone (1935, 1947) developed the centroid method of factor extraction in the 1930's for his major factor analytic studies. This time was prior to the age of electronic computers and used mechanical desk calculators to perform the needed computations. There was a need for simple procedures and the centroid method filled this need. Today, with the availability of electronic computers, much more computationally complex procedures are readily available. However, the sign change procedure and criterion L, to be described later, are very useful in obtaining a measure of complexity of relations in a covariance or correlation matrix.

Thurstone developed the centroid method using a geometric view involving a configuration of vectors to represent the attributes. The centroid vector is the mean vector through which a centroid axis is passed. This is a centroid factor with the orthogonal projections of attribute vectors on it being the factor weights. One factor is extracted at a time and a residual correlation matrix is computed. The next factor is extracted from the residual matrix and a further residual matrix is computed. Thus, there is a series of residual correlation matrices with a

centroid factor extracted from each residual matrix. One problem is that the configuration of vectors frequently has vectors splayed out in many directions. This is true of many correlation matrices among personality traits and is almost always true of residual correlation matrices. As a consequence there is a need to reverse some of vectors to obtain a more stable centroid. This is the sign change procedure. Factor extraction is continued until the residual correlations become small enough to be ignored and resulting factors have only quite small factor weights (in absolute value).

Rather than Thurstone's geometric approach, an algebraic approach is used here involving the general theory of matrix factoring presented earlier in this chapter. Let matrix \mathbf{G} be any of the original correlation matrix and residual correlation matrices. One centroid factor is to be extracted from \mathbf{G} and a new matrix of residual correlations is to be computed. A major restriction is that weights, w_j , in the single column of matrix \mathbf{W} are to be either +1 or -1. The sign change procedure is used to reverse signs of weights for selected attributes. The signs are changed so as to maximize coefficient \mathbf{P} defined in equation (8.5). After the sign change procedure, the sum of the absolute values of factor weights equals \mathbf{F} , the square root of \mathbf{P} . Thus, the centroid method combined with the sign change procedure tends to maximize the sum of absolute values of the factor weights. However, there may be several maxima and there is no guarantee that an absolute maximum is obtained. If an absolute maximum is not obtained in one factor extracted, a factor related to the absolute maximum is likely to be obtained from the next matrix of residual correlations.

There are several matters to be considered when the weights are restricted to +1 or -1. First, it will be seen subsequently that the sign of the weight for an attribute has no effect on the contribution of the diagonal entry on coefficient \mathbf{P} . As a result, the diagonal entries in \mathbf{G} are eliminated from the sign change computations. Table 8.5 presents a correlation matrix with zeros in the diagonal. Such a matrix may be symbolized by $\tilde{\mathbf{G}}$ and defined by:

$$\tilde{\mathbf{G}} = \mathbf{G} - \mathbf{Diag}(\mathbf{G}) \quad (8.57)$$

so that

$$\tilde{g}_{jk} = g_{jk} \text{ for } j \neq k \text{ and}$$

$$\tilde{g}_{jj} = 0$$

Then:

$$\mathbf{G} = \tilde{\mathbf{G}} + \mathbf{Diag}(\mathbf{G}) \quad (8.58)$$

Table 8.5
Illustration of Determination of Centroid Factor with Sign Change

Correlation Matrix with Zeros in Diagonal							
	1	2	3	4	5	6	
1	--	.49	-.37	.17	.04	-.03	
2	.49	--	-.28	.16	.02	.02	
3	-.37	-.28	--	-.15	.03	-.07	
4	.17	.16	-.15	--	-.02	.21	
5	.04	.02	.03	-.02	--	-.10	
6	-.03	.02	-.07	.21	-.10	--	
D(R)	.49	.49	.37	.21	.10	.21	sum=1.87
W ₁ '	+1	+1	+1	+1	+1	+1	
\tilde{Q}_1'	.30	.41	-.84	.37	-.03	.03	$\tilde{p}_1=.24$
c ₁			-2				
W ₂ '	+1	+1	-1	+1	+1	+1	
\tilde{Q}_2'	1.04	.97	-.84	.67	-.09	.17	$\tilde{P}_2=3.60$
c ₂					-2		
W ₃ '	+1	+1	-1	+1	-1	+1	
\tilde{Q}_3'	.96	.93	-.90	.71	-.09	.37	$\tilde{p}_3=3.96$
Q'	1.45	1.42	-1.27	.92	-.19	.58	p=5.83
A'	.60	.59	-.53	.38	-.08	.24	

$$F = \sqrt{P} = 2.4145.$$

$$\sum_K a_k w_k = 2.42.$$

There will be several trials so that a trial number subscript will be used. Let W_t be the weight matrix for trial t with entries w_{tk} . Matrix Q_t is obtained from equation (8.4).

$$Q_t = GW_t$$

and using equation (8.58):

$$Q_t = \tilde{G} W_t + [Diag(G)]W_t$$

Define \tilde{Q}_t by:

$$\tilde{Q}_t = \tilde{G} W_t . \quad (8.59)$$

Then:

$$Q_t = \tilde{Q}_t + [Diag(G)]W_t . \quad (8.60)$$

By equations (8.5) and (8.60):

$$P_t = W_t' Q_t = W_t' \tilde{Q}_t + W_t' [Diag(G)]W_t . \quad (8.61)$$

Define \tilde{P}_t by:

$$\tilde{P}_t = W_t' \tilde{Q}_t . \quad (8.62)$$

Equation (8.63) gives an interesting relation:

$$W_t' [Diag(G)]W_t = \sum_j w_{tj}^2 g_{jj} = \sum_j g_{jj} . \quad (8.63)$$

Remember that the square of either +1 or -1 is +1. Then:

$$P_t = \tilde{P}_t + \sum_j g_{jj} . \quad (8.64)$$

This results supports the statement that the contribution of the diagonal entries of G is independent of the signs of the weights. From equation (8.6):

$$F_t = \sqrt{P_t} . \quad (8.65)$$

From equation (8.7)

$$A_t = Q_t(1/F_t) . \quad (8.66)$$

Then from equation (8.23)

$$F_t = W_t' A_t = \sum_k w_{tk} a_{tk} . \quad (8.67)$$

Changes from trial t to trial $(t+1)$ are considered next. At each step, the sign of the weight for only one attribute, i , is to be changed. When $w_{ti} = +1$, let $c_{ti} = -1$, and when $w_{ti} = -1$, let $c_{ti} = +1$; that is, the sign of c_{ti} is opposite to the sign of w_{ti} . Then:

$$w_{(t+1)i} = w_{ti} + c_{ti} ; \quad (8.68)$$

and, for $k \neq i$:

$$w_{(t+1)k} = w_{tk} . \quad (8.69)$$

For all attributes, j :

$$\tilde{q}_{(t+1)j} = \sum_h \tilde{g}_{jh} w_{(t+1)h} = \sum_{k \neq i} \tilde{g}_{jk} w_{(t+1)k} + \tilde{g}_{ji} w_{(t+1)i} . \quad (8.70)$$

Substitution for equations (8.68) and (8.69) yields:

$$\begin{aligned} \tilde{q}_{(t+1)j} &= \sum_{k \neq i} \tilde{g}_{jk} w_{tk} + \tilde{g}_{ji} w_{ti} + \tilde{g}_{ji} c_{ti} \\ &= \sum_h \tilde{g}_{jh} w_{th} + \tilde{g}_{ji} c_{ti} . \end{aligned}$$

This yields the important result:

$$\tilde{q}_{(t+1)j} = \tilde{q}_{tj} + \tilde{q}_{ji} c_{ti} ; \quad (8.71)$$

and, since \tilde{g}_{ii} equals zero:

$$\tilde{q}_{(t+1)i} = \tilde{q}_{ti} . \quad (8.72)$$

Interpretation of equation (8.62) in terms of trial $(t+1)$ yields:

$$\tilde{P}_{(t+1)} = \sum_h w_{(t+1)h} \tilde{q}_{(t+1)h} = \sum_{k \neq i} w_{(t+1)k} \tilde{q}_{(t+1)k} + w_{(t+1)i} \tilde{q}_{(t+1)i} .$$

Substitution from equations (8.68) and (8.69) yields:

$$\tilde{P}_{(t+1)} = \sum_{k \neq i} w_{tk} \tilde{q}_{tk} + c_{ti} \sum_{k \neq i} \tilde{g}_{ki} w_{tk} + w_{ti} \tilde{q}_{ti} + c_{ti} \tilde{q}_{ti}$$

which reduces to:

$$\tilde{P}_{(t+1)} = \tilde{P}_t + 2c_{ti} \tilde{q}_{ti} \quad (8.73)$$

The sign change procedure utilizes relations developed in the preceding paragraphs. This procedure is illustrated in Table 8.5 which gives a correlation matrix with zeros in the diagonal elements. This is matrix \tilde{G} . One general requirement is that the diagonal entries to be used subsequently must be all positive which is not necessarily true for residual correlation matrices when SMC's are used as communality like values. A procedure followed by Thurstone appears to

work very satisfactorily; that is, to use the highest R values for every correlation matrix and residual correlation matrix. Row D(R) of Table 8.5 contains the highest R (in absolute value) for the given correlation matrix. The sum of the entries in this row is given at the right. At trial 1, the weights in row W'_1 are all taken to be +1 and the first trial \tilde{Q}'_1 contains the column sums of matrix \tilde{G} . Coefficient \tilde{P}_1 is the sum of the entries in row \tilde{Q}'_1 . These are the preliminary steps before starting the sign change procedure.

Since, by equation (8.64), coefficient P_t is \tilde{P}_t plus the sum of the entries to be inserted in the diagonal of G and this sum is necessarily positive, increasing \tilde{P}_t necessarily increases P_t . The objective of the sign change procedure is to increase \tilde{P}_t as much as possible. In each step, or trial, the attribute is selected whereby \tilde{P}_t is increased most to $\tilde{P}_{(t+1)}$. By equation (8.73), the change from \tilde{P}_t to $\tilde{P}_{(t+1)}$ equals $2c_{ti}\tilde{q}_{ti}$. For this change to be an increase, c_{ti} and \tilde{q}_{ti} must have the same algebraic sign. Note that c_{ti} is defined to have the opposite sign to w_{ti} . Therefore, w_{ti} and \tilde{q}_{ti} must have opposite signs. The strategy is to select that attribute for which w_{ti} and \tilde{q}_{ti} have opposite signs and \tilde{q}_{ti} is the largest in absolute value satisfying the signs condition. In the example in Table 8.5, since all weights in row W'_1 are positive, the attribute with the most negative value in row \tilde{Q}'_1 is attribute 3 with an entry of -.84. Consequently, attribute 3 was chosen and a change coefficient c_{13} was set at -2, the sign being opposite to the sign of weight w_{13} .

After having established rows W'_1 , \tilde{Q}'_1 and selected the attribute for the sign change, the next series of steps is to make the changes accompanying this first sign change. In row W'_2 the signs of the weights for all attributes except attribute 3 remain unchanged at +1. The weight for selected attribute 3 is changed to -1. These weights in row W'_2 are the results of application of equations (8.68) and (8.69). Next, row \tilde{Q}'_2 is established as per equations (8.71) and (8.72). For an example consider the first entry in row \tilde{Q}'_2 , \tilde{q}_{21} . The value of 1.04 equals \tilde{q}_{21} plus c_{13} times \tilde{g}_{31} ; that is:

$$1.04 = .30 + (-2)(-.37)$$

These computations are continued for all attributes except the selected attribute 3 for which \tilde{q}_{23} equals \tilde{q}_{13} , a value of -.84. Coefficient \tilde{P}_2 can be computed two ways: one by obtaining the sum of products of entries in rows W'_2 and \tilde{Q}'_2 , as per equation (8.62); and by equation (8.73). For the second method:

$$3.60 = .24 + 2(-2)(-.84)$$

When using hand computing with the aid of a desk calculator, this value should be computed both ways to provide a check.

A selection is made next of the second attribute to have the sign of its weight changed. Rows W'_2 and \tilde{Q}'_2 are inspected for those attributes having entries with opposite signs and that attribute is selected for which the absolute value of the entry in row \tilde{Q}'_2 is largest. In the example, only for attribute 5 are the entries in rows W'_2 and \tilde{Q}'_2 opposite in sign; consequently, this attribute is selected to have its sign changed and a -2 is inserted into line c_2 for attribute 5.

Computations for trial 3 from the results in trial 2 are similar to those carried out in going from trial 1 to trial 2. The signs of weights in row W'_3 are the same as those in row W'_2 with exception of w_{35} , this being the weight having its sign changed. Entries in row \tilde{Q}'_3 are obtained from row 5 of G , the weights in row W'_2 , and the entries in row \tilde{Q}'_2 . Coefficient \tilde{P}_3 is obtained from \tilde{P}_2 , c_{25} , \tilde{q}_{25} as well as from the sum of products between entries in rows W'_3 and \tilde{Q}'_3 .

Inspection of rows W'_3 and \tilde{Q}'_3 reveals that the signs of the entries in these two rows are in agreement. There are no more signs to be changed. Row W'_3 is the final weight vector. A final row Q' is to be computed by adding in the diagonal entries of G with the proper signs, see equation (8.60). The entry in row Q' of Table 8.5 for the first attribute is:

$$1.45 = .96 + (.49)(+1) .$$

For attribute 3:

$$-1.27 = -.90 + (.37)(-1) .$$

For the final coefficient P see equation (8.64). The final P equals the final \tilde{P}_t plus the sum of the diagonal entries as well as equaling the sum of the products of the entries in the final W and Q . The value for the example is:

$$5.85 = 3.96 + 1.87 .$$

Factor weights in row A' are obtained by dividing the entries in row Q' by F which equals the square root of P . For the example the factor weight for attribute 1 is:

$$.60 = 1.45/2.4145 .$$

As indicated in equation (8.67), F equals the sum of products between the weights in the final row W' of the example and the factor weights in row A' .

The example in Table 8.5 is too small to illustrate one kind of difficulty encountered with larger correlation matrices. Sometimes when the sign has been changed for one attribute and after a number of further changes the \tilde{q} for that attribute becomes positive which is, now, opposite to the sign of the weight. The sign of the weight has to be changed back to a +1. In this case the change coefficient, c , is a +2. Then equations (8.62), (8.71), (8.72), and (8.73) provide the means for making this reverse change.

It is of interest to note now that the use of vectors \tilde{Q} without the diagonal entries provides a more effective sign change than would the use of vectors Q which includes the diagonal entries. For example, in trial 2 of the example in Table 8.5, adding in the diagonal entry of .10 to \tilde{q}_{25} of -.09 yields a value of .01 which agrees in sign with the weight w_{25} . If this value of .01 is compared with the weight of +1, then the sign for attribute 5 would not be changed and the increase in P would have been missed.

We return to the maximization proposition for coefficient P . With each trial resulting in an increase in P , a maximum should be reached since the value can not exceed the sum of absolute values of entries in the correlation matrix. However, there is no guarantee that there is only one maximum nor that an obtained result yields the largest maximum. There is one statement possible using the vectors \tilde{Q} without the diagonal entries: it is not possible to increase P further after reaching a solution by changing the sign of the weight for only one attribute. To go from one maximum to another must involve the changing of the weight signs for two or more attributes. Each solution, thus, involves at least a local maximum. We observe that when a major maximum has been missed, a solution involving the factor for this maximum is likely to appear in the results for the next factor extracted from the ensuing residual matrix.

Extraction of centroid factors from the correlation matrix among nine mental tests is given in Table 8.6. This correlation matrix was given in Chapter 1, Table 1.1, which includes the names of tests. For the original correlation matrix given at the top of the first page of this table, row D(R) contains the highest correlation for each attribute. These values will be substituted for the unities in the diagonal. Since all correlations are positive, all sign change weights are +1 and the entries in row Q' are the column sums of the correlation matrix, the diagonal unity having been replaced by the entry in row D(R). Coefficients P and F are given along with the factor weights in row A'_1 for the first centroid factor. Coefficient L is a very useful criterion to measure the structure indicated in the correlation matrix. In general:

$$L = \tilde{P} / [\sum_j \sum_{k \neq j} |g_{jk}|] \quad (8.74)$$

This criterion may be used in decisions on the number of factors to be extracted from a correlation matrix, more about this later. Since all original correlations are positive in the example, L for this matrix is unity.

The first factor residual matrix is given at the bottom of the first page of Table 8.6. The diagonal entries are residuals from the substituted diagonals of the correlation matrix. The column sums including these residual diagonals are zero within rounding error. These sums provide an excellent check on hand computations. Revised diagonal entries are given in row D(R). The signs were changed for attributes 4, 5, 6, and 8 after which row Q' was obtained

Table 8.6
Extraction of Centroid Factors from
Correlation Matrix among Nine Mental Tests

<u>Original Correlation Matrix</u>									
	1	2	3	4	5	6	7	8	9
1	<u>1.000</u>								
2	.499	<u>1.000</u>							
3	.394	.436	<u>1.000</u>						
4	.097	.007	.292	<u>1.000</u>					
5	.126	.023	.307	.621	<u>1.000</u>				
6	.085	.083	.328	.510	.623	<u>1.000</u>			
7	.284	.467	.291	.044	.114	.086	<u>1.000</u>		
8	.152	.235	.309	.319	.376	.337	.393	<u>1.000</u>	
9	.232	.307	.364	.213	.276	.271	.431	.489	<u>1.000</u>
D(R)*	.499	.499	.436	.621	.623	.623	.467	.489	.489
Q'	2.368	2.556	3.157	2.724	3.089	2.946	2.577	3.099	3.072
			P=25.588	F=5.058	L=1.000				
A ₁ '	.468	.505	.624	.539	.611	.582	.509	.613	.607

<u>First Residual Correlation Matrix</u>									
	1	2	3	4	5	6	7	8	9
1	<u>.290</u>								
2	.262	<u>.244</u>							
3	.102	.121	<u>.046</u>						
4	-.155	-.265	-.044	<u>.331</u>					
5	-.160	-.286	-.074	.292	<u>.250</u>				
6	-.188	-.211	-.035	.196	.267	<u>.284</u>			
7	.046	.210	-.027	-.230	-.197	-.211	<u>.207</u>		
8	-.135	-.075	-.073	-.011	.002	-.020	.081	<u>.114</u>	
9	-.052	.000	-.015	-.114	-.095	-.083	.122	.117	<u>.120</u>
Sum	.000	.000	.001	.000	-.001	-.001	.001	.000	.000
D(R)*	.262	.286	.121	.292	.292	.267	.230	.135	.122
Q'	1.257	1.715	.528	-1.578	-1.665	-1.439	1.137	-.191	.351
			P=9.862	F=3.140	L=.859				
A ₂ '	.400	.546	.168	-.503	-.530	-.458	.362	-.061	.112

* Highest R.

Table 8.6 (Continued)
Extraction of Centroid Factors from
Correlation Matrix among Nine Mental Tests

<u>Second Residual Correlation Matrix</u>									
	1	2	3	4	5	6	7	8	9
1	<u>.102</u>								
2	.044	<u>-.013</u>							
3	.034	.029	<u>.092</u>						
4	.046	.009	.040	<u>.040</u>					
5	.052	.004	.015	.026	<u>.011</u>				
6	-.004	.039	.042	-.034	.024	<u>.057</u>			
7	-.099	.012	-.088	-.048	-.005	-.045	<u>.099</u>		
8	-.110	-.041	-.063	-.041	-.030	-.048	.103	<u>.131</u>	
9	-.097	-.061	-.034	-.058	-.036	-.032	.081	.124	<u>.109</u>
<hr/>									
Weighted									
Sum	.000	.000	-.001	-.002	-.001	.001	.000	.001	.000
D(R)*	.110	.061	.088	.058	.052	.048	.103	.124	.124
Q'	-.590	-.276	-.433	-.293	-.245	-.238	.561	.685	.645
			P=3.967	F=1.992	L=.941				
A ₁ '	-.296	-.139	-.217	-.147	-.123	-.120	.281	.344	.324
<hr/>									
<u>Third Residual Correlation Matrix</u>									
	1	2	3	4	5	6	7	8	9
1	<u>.023</u>								
2	.003	<u>.042</u>							
3	-.030	-.001	<u>.041</u>						
4	.003	-.011	.008	<u>.036</u>					
5	.016	-.013	-.012	.008	<u>.037</u>				
6	-.040	.022	.016	-.052	.010	<u>.033</u>			
7	-.016	.051	-.027	-.007	.030	-.011	<u>.024</u>		
8	-.009	.006	.012	.009	.012	-.006	.006	<u>.006</u>	
9	-.001	-.016	.037	-.010	.004	.007	-.010	.012	<u>.019</u>
<hr/>									
Weighted									
Sum	-.001	-.001	.000	.000	.000	.001	.000	.000	.000
D(R)*	.040	.051	.037	.052	.030	.052	.051	.012	.037
Q'	-.119	.108	.082	-.109	.037	.180	.113	.054	.082
			P=.883	F=.940	L=.483				
A ₁ '	-.126	.115	.088	-.116	.039	.192	.120	.057	.088

* Highest R.

along with coefficients P , F , and L . For this matrix the sign change did not result in all positive contributions by the off-diagonal entries so that L is less than unity. Factor weights for the second factor in row A'_2 are obtained from row Q' and coefficient F .

The second factor residual correlation matrix is given at the top of the second page of Table 8.6. This matrix is obtained from the first factor residual correlation matrix with substituted diagonal entries and the second factor weights. The row of weighted sums uses the just preceding sign change weights as multipliers of the residual correlations. Again, these sums should equal zero within rounding error which provides a check on hand computations. See equation (8.22) for the basis for these zero weighted sums. Computation of the third factor weights progresses in the same manner as the computations for preceding factors.

The third factor residual correlation matrix is given at the bottom of the second page of Table 8.6. Computations for this matrix and the fourth factor weights are similar to the computations for preceding factors.

Decisions as to the number of factors to extract by the centroid method had only sketchy bases to support these decisions. Residual matrices were inspected for the magnitudes of the entries and factor extraction was stopped when these residuals were small so that they might be ignored. Table 8.7 gives three coefficients which might be used including the largest residual correlation in each residual matrix. For the example of nine mental tests the largest third factor residual was .052 and a decision might be made that this was small enough to be ignored. By this reason, the three factor solution would be accepted. Another coefficient which might be considered is the criterion L . Note in the example this coefficient is relatively high for the first three matrices and factors but drops substantially for the third factor residuals. In this example the low value of L for the third factor residuals could be taken as an indication to accept the three factor solution. A third criterion used by some analysts was the magnitude of the factor weights obtained. Frequently, when the largest factor weight was less than a value such as .2, a factor was not accepted. In the example this criterion would, again, indicate a three factor solution. Beyond such criteria as the foregoing, trial transformations of factors was considered and that number of factors accepted which led to the most meaningful solution. Some individuals advocated using an extra "residual" factor to help clean up transformed factors.

The centroid method of factor extraction has been presented partly for its historic value and partly to provide some useful techniques. For example, the sign change technique with criterion L has been found useful in testing the simplicity of special covariance matrices in research on the dimensionality of binary data such as item responses. Undoubtedly, there may be other cases of special covariance matrices for which a simple criterion related to complexity of the matrix would be helpful.

Table 8.7
 Illustration of Indices used for Number of Factors
 in Centroid Factor Extraction

	Largest Correlation	Criterion L	Largest Factor Loading*
Original Correlation matrix, Factor 1	.623	1.000	.624
First Residual Matrix, Factor 2	.292	.859	.546
Second Residual Matrix, Factor 3	.124	.941	.344
Third residual Matrix, Factor 4	.052	.483	.192

* In absolute value

8.4. Group Centroid Method of Factor Extraction

The group centroid method of factor extraction provides a simple technique which may be applied in special situations. In particular, a partitioning of the attributes into clusters with high within cluster correlations should be possible. The correlations between clusters should be low. For an example consider the correlation matrix in Table 8.8. Attributes 1, 2, and 4 intercorrelate relatively highly and have low correlations with the remaining attributes. These attributes are listed first in Table 8.9. Ignore the diagonal entries for the present. Attributes 5 and 6 have a high correlation while attribute 3 correlates negatively with them. A sign reversal of the correlations of attribute 3 produces moderately high, positive correlations with attributes 5 and 6. In making such a sign reversal, the signs of the correlations in both the rows and the columns are reversed for the attribute. Note that the diagonal entry remains positive since its sign is reversed twice. This operation yields the second cluster in Table 8.9. As seen in this table, there are two clusters of attributes with relatively high intercorrelations within the clusters and relatively low correlations between clusters. This is the type of situation for which the group centroid method of factor extraction could be appropriate.

For the operation of the group centroid method of factor extraction return to the original correlation matrix in Table 8.8; the clustered correlation matrix provided a guiding step but will not be used in the computations. A first consideration is the diagonal entries. The extracted factors will depend on the intercorrelations of the attributes in the clusters, these intercorrelations forming relatively small matrices. There is a problem in using SMC's as diagonal entries. As shown earlier, the SMC's tend to be smaller than desired for small matrices. For example, the SMC for attribute 1 in our example is .274 which appears small for the intercorrelations of the attributes in the example. In contrast, the "highest R" of .44 appears appropriate. Further, for computations using a desk calculator, the "highest R" technique is much more convenient. Thus, in general with the group centroid method of factor extraction, the "highest R" would be the preferred value to be inserted into the diagonal of the correlation matrix. This has been done for the middle matrix of Table 8.8 and were given in Table 8.9. Computations of the factor matrix will progress from the middle matrix of Table 8.8.

Equations used in the group centroid method of factor analysis method are repeated here for convenience. Matrix G_1 is the correlation matrix with desired diagonal entries such as the middle matrix of Table 8.8.

$$Q = G_1 W . \quad (8.4)$$

$$P = W' Q . \quad (8.5)$$

Table 8.8
Correlation matrices for
Illustration of Group Centroid Method of Factor Extraction

Correlation Matrix with Unities in Diagonal

	1	2	3	4	5	6
1	<u>1.00</u>	.44	-.04	.43	.04	.05
2	.44	<u>1.00</u>	-.06	.38	.07	.09
3	-.04	-.06	<u>1.00</u>	-.06	-.33	-.35
4	.43	.38	-.06	<u>1.00</u>	.09	.08
5	.04	.07	-.33	.09	<u>1.00</u>	.40
6	.05	.09	-.35	.08	.40	<u>1.00</u>

Correlation Matrix with Highest R's in Diagonal

	1	2	3	4	5	6
1	<u>.44</u>	.44	-.04	.43	.04	.05
2	.44	<u>.44</u>	-.06	.38	.07	.09
3	-.04	-.06	<u>.35</u>	-.06	-.33	-.35
4	.43	.38	-.06	<u>.43</u>	.09	.08
5	.04	.07	-.33	.09	<u>.40</u>	.40
6	.05	.09	-.35	.08	.40	<u>.40</u>

Residual Correlation Matrix

	1	2	3	4	5	6
1	<u>-.01</u>	.01	-.01	.01	-.01	.00
2	.01	<u>.02</u>	.00	-.03	-.01	.01
3	-.01	.00	<u>.03</u>	.01	.02	.01
4	.01	-.03	.01	<u>.03</u>	.01	-.01
5	-.01	-.01	.02	.01	<u>.01</u>	.01
6	.00	.01	.01	-.01	.01	<u>.00</u>

Table 8.9
 Clustered Correlation Matrix for
 Illustration of Group Centroid Method of Factor Extraction

Correlation Matrix with Highest R's in Diagonal

	1	2	4	5	6	-3
1	<u>.44</u>	.44	.43	.04	.05	.04
2	.44	<u>.44</u>	.38	.07	.09	.06
4	.43	.38	<u>.43</u>	.09	.08	.06
5	.04	.07	.09	<u>.40</u>	.40	.33
6	.05	.09	.08	.40	<u>.40</u>	.35
-3	.04	.06	.06	.33	.35	<u>.35</u>

$$FF' = P . \quad (8.6)$$

$$A = Q(F^{-1})' . \quad (8.7)$$

$$G_2 = G_1 - AA' \quad (8.8)$$

where G_2 is the matrix of residual correlations.

$$W'A = F . \quad (8.23)$$

Reference will be made to these equations during the discussion of the computing procedures.

Computations for the example are given in Table 8.10. Weight matrix W is the starting point. This matrix reflects the clusters which have been determined during inspection of the correlation matrix. There is a column of W for each cluster, or group, and contains weights of +1 or -1 for attributes in the cluster and weights of 0 for attributes not in the cluster. Weights of +1 are assigned to attributes which are not reflected in sign and weights of -1 for attributes reflected in sign. In the example, the first cluster was composed of attributes 1, 2, and 4 without any reflections in sign. Consequently, the first column of weight matrix in Table 8.10 has +1's for these three attributes and 0's for the other attributes. The second column of the weight matrix is for the second cluster with +1's for attributes 5 and 6 and a -1 for attribute 3 since this attribute was reflected in sign to form the cluster. The weight of -1 performs this reflection. In this second column of weights, 0's are recorded for attributes 1, 2, and 4 which are not in the second cluster. In general, the weight matrix reflects the clusters found during inspection of the correlation matrix.

Once the weight matrix has been established, computations follow the given equations. Matrix Q is computed by equation (8.4). Since the weight matrix contains only +1's, -1's, and 0's, this matrix multiplication involves only addition, or subtraction, of entries in the correlation matrix and may be accomplished quite readily with a desk calculator. Matrix P is obtained by equation (8.5) which, again, involves only addition or subtraction of entries in Q . Matrix F is a decomposition of P as indicated by equation (8.6). A Cholesky decomposition of P to triangular matrix F is most convenient. At this point note a requirement that matrix F must be of full rank. This reflects on the composition of the weight matrix and the correlation matrix. Having matrix F its inverse is obtained, this being a simple solution when F is triangular. The matrix of factor weights is obtained by equation (8.7). Equation (8.23) gives an interesting relation.

Once the factor matrix has been determined, a matrix of residual correlations should be computed by equation (8.8). For the example, the matrix of residual correlations is given at the bottom of Table 8.8. For our example, these residuals are all quite tiny indicating that the

Table 8.10
 Computation of Factor Matrix for
 Illustration of Group Centroid Method of Factor Analysis

<u>Weight Matrix</u>			<u>Matrix Q</u>			<u>Factor Matrix</u>		
	1	2		1	2		1	2
1	+1	0	1	1.31	.13	1	.67	-.04
2	+1	0	2	1.26	.22	2	.65	.02
3	0	-1	3	-.16	-1.03	3	-.08	-.56
4	+1	0	4	1.24	.23	4	.64	.02
5	0	+1	5	.20	1.13	5	.10	.61
6	0	+1	6	.22	1.15	6	.11	.62

<u>Matrix P</u>			<u>Factor Matrix</u>		
	1	2		1	2
1	3.81	.58	1	1.952	.000
2	.58	3.31	2	.000	1.795

<u>Matrix (F⁻¹)'</u>		
	1	2
5	.512	-.085
6	.000	.557

obtained factor matrix provides an excellent fit to the input correlation matrix. When the fit is not as good, further factors could be extracted from the residual matrix and added to the factor matrix by adjoining these new factors to those already obtained. New attribute clusters could be determined in the residual matrix and the group centroid method used to establish these new factors. An alternative is to apply the centroid method to the matrix of residual correlations.

The group centroid method of factor extraction appears to be a simple technique which could be useful in less formal analyses such as for pilot studies and analyses. For more formal studies more precise methods would be advisable.

8.5. Principal Factors Method of Factor Extraction

With the development of digital computers the principal factors method has become the most popular method for factor extraction. Prior to the large computer the calculation labor was prohibitively extensive to use this method on any but the most trivial sized matrices. The key to use of principal factors is the availability of solutions for eigenvalues and eigenvectors of real, symmetric matrices. Now, these solutions may be obtained quite readily for all but very large correlation and covariance matrices. The principal factors method has a number of desirable properties including a maximization of the sum of squares of factor weights on the extracted factors. Minimization of the sum of squares of residual correlations will be discussed in detail in the next chapter.

A numerical example is discussed before a presentation of mathematical properties of principal factors. Table 8.11 gives the correlation matrix for the nine mental tests example with SMC's in the diagonal. The eigenvalues and eigenvectors were computed for this matrix and are presented in Table 8.12. Note that all eigenvalues after the first three are negative. Use of SMC's in the diagonal of a correlation matrix must result in a number of negative eigenvalues. At the bottom of Table 8.12 is the principal factors matrix, each column of which being obtained by multiplying the entries in an eigenvector by the square root of corresponding eigenvalue. There are only three columns in the principal factors matrix since the square roots of the eigenvalues beyond the first three present problems with imaginary numbers. However, in some studies, the factor extraction may stop short of number of positive eigenvalues, this being a problem as to the number of factors which will be discussed subsequently.

Mathematical relations for principal factors are considered next. Let C be a square, symmetric matrix with real numbers. There is no restriction that C be Gramian as was stated in the general theory of matrix factoring; however, a number of the relations given in the general theory will be used. There must be considerable care in this usage to avoid violating several restrictions. For example, matrix C could be a correlation matrix with SMC's in the diagonal such as given for the nine mental tests in Table 8.11. As seen in Table 8.12, there are a number of

Table 8.11
Correlation and Residual Matrices for
Principal Factors for Nine Mental tests

Correlation Matrix with SMC's in Diagonal

	1	2	3	4	5	6	7	8	9
1	<u>.297</u>	.499	.394	.097	.126	.085	.284	.152	.232
2	.499	<u>.424</u>	.436	.007	.023	.083	.467	.235	.307
3	.394	.436	<u>.356</u>	.292	.307	.328	.291	.309	.364
4	.097	.007	.292	<u>.428</u>	.621	.510	.044	.319	.213
5	.126	.023	.307	.621	<u>.535</u>	.623	.114	.376	.276
6	.085	.083	.328	.510	.623	<u>.440</u>	.086	.337	.271
7	.284	.467	.291	.044	.114	.086	<u>.350</u>	.393	.431
8	.152	.235	.309	.319	.376	.337	.393	<u>.361</u>	.489
9	.232	.307	.364	.213	.276	.271	.431	.489	<u>.349</u>

First Residual Correlation Matrix

	1	2	3	4	5	6	7	8	9
1	<u>.123</u>	.303	.140	-.128	-.137	-.159	.083	-.101	-.014
2	.303	<u>.204</u>	.150	-.247	-.273	-.191	.241	-.050	.030
3	.140	.150	<u>-.014</u>	-.037	-.078	-.028	-.002	-.060	.005
4	-.128	-.247	-.037	<u>.136</u>	.280	.194	-.216	-.009	-.106
5	-.137	-.273	-.078	.280	<u>.136</u>	.254	-.190	-.007	-.097
6	-.159	-.191	-.028	.194	.254	<u>.098</u>	-.196	-.018	-.074
7	.083	.241	-.002	-.216	-.190	-.196	<u>.118</u>	.101	.147
8	-.101	-.050	-.060	-.009	-.007	-.018	.101	<u>-.007</u>	.131
9	-.014	.030	.005	-.106	-.097	-.074	.147	.131	<u>.001</u>

Second Residual Correlation Matrix

	1	2	3	4	5	6	7	8	9
1	<u>-.009</u>	.108	.083	.037	.038	-.011	-.064	-.102	-.072
2	.108	<u>-.085</u>	.066	-.001	-.014	.028	.024	-.051	-.055
3	.083	.066	<u>-.039</u>	.035	-.002	.036	-.066	-.061	-.020
4	.037	-.001	.035	<u>-.071</u>	.060	.008	-.032	-.008	-.034
5	.038	-.014	-.002	.060	<u>-.096</u>	.057	.004	-.006	-.020
6	-.011	.028	.036	.008	.057	<u>-.069</u>	-.031	-.017	-.010
7	-.064	.024	-.066	-.032	.004	-.031	<u>-.045</u>	.100	.083
8	-.102	-.051	-.061	-.008	-.006	-.017	.100	<u>-.007</u>	.131
9	-.072	-.055	-.020	-.034	-.020	-.010	.083	.131	<u>-.024</u>

Table 8.11(Continued)

Third Residual Correlation Matrix

	1	2	3	4	5	6	7	8	9
1	<u>-.086</u>	.064	.029	.012	.023	-.028	-.006	-.020	-.002
2	.064	<u>-.110</u>	.035	-.015	-.022	.019	.056	-.005	-.016
3	.029	.035	<u>-.077</u>	.017	-.012	.024	-.025	-.004	.029
4	.012	-.015	.017	<u>-.079</u>	.056	.002	-.013	.019	-.011
5	.023	-.022	-.012	.056	<u>-.098</u>	.054	.015	.009	-.007
6	-.028	.019	.024	.002	.054	<u>-.073</u>	-.018	.001	.006
7	-.006	.056	-.025	-.013	.015	-.018	<u>-.088</u>	.039	.031
8	-.020	-.005	-.004	.019	.009	.001	.039	<u>-.093</u>	.057
9	-.002	-.016	.029	-.011	-.007	.006	.031	.057	-.087

Table 8.12
 Computation of Principal Factors Matrix for Nine Mental Tests Example from
 Correlation Matrix with SMC's in Diagonal

<u>Eigenvalues</u>								
1	2	3	4	5	6	7	8	9
2.746	1.241	.346	-.048	-.064	-.125	-.153	-.175	-.255

<u>Eigenvectors</u>									
	1	2	3	4	5	6	7	8	9
1	.252	.326	.472	.265	-.213	.511	.259	.052	.401
2	.283	.483	.266	.094	.305	-.267	.190	.228	-.595
3	.367	.141	.331	-.557	-.258	-.279	-.272	-.447	.084
4	.326	-.409	.154	.311	-.444	-.437	-.082	.456	.057
5	.381	-.433	.087	.349	.146	.349	-.213	-.440	-.394
6	.353	-.366	.106	-.355	.629	.035	.178	.263	.326
7	.291	.362	-.353	.393	.292	-.214	-.448	-.084	.407
8	.366	.002	-.499	.019	-.184	-.146	.684	-.305	.030
9	.356	.14	-.425	-.334	-.249	.456	-.265	.418	-.214

<u>Principal Factor Matrix</u>			
	1	2	3
1	.417	.363	.278
2	.469	.538	.156
3	.609	.157	.195
4	.540	-.456	.090
5	.632	-.482	.051
6	.585	-.408	.062
7	.481	.404	-.208
8	.607	.003	-.293
9	.590	.159	-.250

negative eigenvalues for this matrix so that this matrix is not Gramian. Further, residual correlation matrices must have as many eigenvalues equal to zero as the number of factors that have been extracted. Allowance must be made for zero and negative eigenvalues in the development.

At the present, a single factor is considered so that there is a single column of factor weights. Also, matrix \mathbf{W} has a single column which will be designated as the vector $\underline{\mathbf{w}}$. Analogous to equation (8.4), vector $\underline{\mathbf{q}}$ is defined by:

$$\underline{\mathbf{q}} = \mathbf{C}\underline{\mathbf{w}}. \quad (8.24)$$

Matrix \mathbf{P} of equation (8.5) is replaced by a scalar p :

$$p = \underline{\mathbf{w}}'\mathbf{C}\underline{\mathbf{w}}. \quad (8.25)$$

Since p is a scalar, f is a scalar also and equals the square root of p as from equation (8.6). The vector of factor weights, $\underline{\mathbf{a}}$, is obtained analogously to equation (8.7):

$$\underline{\mathbf{a}} = q[1/p^{\frac{1}{2}}] = \mathbf{C}\underline{\mathbf{w}}/[(\underline{\mathbf{w}}'\mathbf{C}\underline{\mathbf{w}})^{\frac{1}{2}}]. \quad (8.26)$$

The major criterion for principal factors is that the sum of squares of the factor weights is to be a maximum. Let θ designate the sum of squares of the factor weights in vector $\underline{\mathbf{a}}$. Then:

$$\theta = \underline{\mathbf{a}}'\underline{\mathbf{a}} \quad (8.27)$$

with $\underline{\mathbf{a}}$ determined so that θ is a maximum.

A major restriction on the solution for vector $\underline{\mathbf{w}}$ is that $\underline{\mathbf{w}}'\mathbf{C}\underline{\mathbf{w}}$ be greater than zero so that its square root is possible with a real value so as to satisfy equation (8.26). All possible solutions for which this is not true are to be rejected.

The solution for maximum θ can be simplified with a transformation using eigenvalues and eigenvectors of \mathbf{C} :

$$\mathbf{C} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}' \quad (8.28)$$

where $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues in descending algebraic order and \mathbf{V} is an orthonormal matrix containing the corresponding eigenvectors. Since \mathbf{V} is orthonormal:

$$\mathbf{V}\mathbf{V}' = \mathbf{V}'\mathbf{V} = \mathbf{I}. \quad (8.29)$$

Substitution from equation (8.28) into equation (8.26) yields:

$$\underline{\mathbf{a}} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}'\underline{\mathbf{w}}/[(\underline{\mathbf{w}}'\mathbf{V}\mathbf{\Lambda}\mathbf{V}'\underline{\mathbf{w}})^{\frac{1}{2}}]$$

or

$$V'\underline{a} = \Lambda V'\underline{w}/[(\underline{w}'V\Lambda V'\underline{w})^{\frac{1}{2}}]. \quad (8.30)$$

Define vectors $\underline{\tilde{a}}$ and $\underline{\tilde{w}}$ by:

$$\underline{\tilde{a}} = V'\underline{a}; \quad (8.31)$$

$$\underline{\tilde{w}} = V'\underline{w}. \quad (8.32)$$

Then:

$$\underline{\tilde{a}} = \Lambda\underline{\tilde{w}} / [(\underline{\tilde{w}}'\Lambda\underline{\tilde{w}})^{\frac{1}{2}}], \quad (8.33)$$

and from equation (8.27)

$$\theta = \underline{\tilde{a}}'V'V\underline{\tilde{a}} = \underline{\tilde{a}}'\underline{\tilde{a}}. \quad (8.34)$$

With equation (8.33)

$$\theta = \underline{\tilde{w}}'\Lambda^2\underline{\tilde{w}} / (\underline{\tilde{w}}'\Lambda\underline{\tilde{w}}). \quad (8.35)$$

To obtain a maximum θ a partial derivative with respect to the elements of vector $\underline{\tilde{w}}$ is set equal to zero. There may be several optima with the largest solution being chosen.

$$\frac{\partial\theta}{\partial\underline{\tilde{w}}} = [2\Lambda^2\underline{\tilde{w}} (\underline{\tilde{w}}'\Lambda\underline{\tilde{w}}) - 2\Lambda\underline{\tilde{w}} (\underline{\tilde{w}}'\Lambda^2\underline{\tilde{w}})] / (\underline{\tilde{w}}'\Lambda\underline{\tilde{w}})^2 = \mathbf{0}. \quad (8.36)$$

With the restriction that $(\underline{\tilde{w}}'\Lambda\underline{\tilde{w}})$ be finite and not equal to zero:

$$\Lambda^2\underline{\tilde{w}} - [(\underline{\tilde{w}}'\Lambda\underline{\tilde{w}}) / (\underline{\tilde{w}}'\Lambda\underline{\tilde{w}})]\Lambda\underline{\tilde{w}} = \mathbf{0}. \quad (8.37)$$

Using equation (8.35) with algebraic manipulations yields:

$$(\Lambda - \theta I)(\Lambda\underline{\tilde{w}}) = \mathbf{0} \quad (8.38)$$

which is the equation for an eigen problem with eigenvalues θ and eigenvectors $(\Lambda\underline{\tilde{w}})$. Note that $(\Lambda\underline{\tilde{w}})$ is a vector with entries $(\lambda_i\tilde{w}_i)$ for $i = 1, 2, \dots, n$.

Since Λ is a diagonal matrix, the eigenvalues θ of equation (8.38) equal the diagonal values of Λ . Also, since the eigenvalues λ of C are in descending algebraic order, the maximum θ equals the first λ . Thus:

$$\theta_1 = \lambda_1. \quad (8.39)$$

The first eigenvector has an entry of unity for the first element and entries of zero for remaining entries.

$$(\lambda_1\tilde{w}_1) = 1 \quad (8.40)$$

$$(\lambda_i \tilde{w}_i) = 0 \text{ for } i = 2, \dots, n. \quad (8.41)$$

From (8.40)

$$\tilde{w}_1 = 1/\lambda_1 \quad (8.42)$$

so that

$$(\tilde{w}' \Lambda \tilde{w}) = 1/\lambda_1 \quad (8.43)$$

Unless all eigenvalues of \mathbf{C} are zero or negative, the obtained solution is acceptable.

The first factor weights are considered next. A subscript 1 is used with vectors \tilde{w} , \tilde{a} , and \underline{a} to designate this first factor. From equations (8.33) and (8.43):

$$\tilde{a}_1 = (\Lambda \tilde{w}_1)/(1/\lambda_1)^{\frac{1}{2}} = (\Lambda \tilde{w}_1) \lambda_1^{\frac{1}{2}}. \quad (8.44)$$

From equation (8.31):

$$\underline{a}_1 = \mathbf{V} \tilde{a}_1 = \mathbf{V} (\Lambda \tilde{w}_1) \lambda_1^{\frac{1}{2}}$$

From equations (8.40) and (8.41) vector $(\Lambda \tilde{w}_1)$ has a first entry of unity with all other entries equal to zero. Let matrix \mathbf{V} be partitioned as below:

$$\mathbf{V} = [\mathbf{V}_1, \mathbf{V}_2]$$

where \mathbf{V}_1 is $n \times 1$ containing the first column of \mathbf{V} and \mathbf{V}_2 is $n \times (n-1)$ containing the remaining columns of \mathbf{V} . With this construction and nature of vector $(\Lambda \tilde{w}_1)$:

$$\underline{a}_1 = \mathbf{V}_1 \lambda_1^{\frac{1}{2}} \quad (8.45)$$

which is the equation for the first factor weights.

The preceding paragraph concerned the first principal factor. Each of the eigenvalues λ_j greater than zero yields a principal factor. The sum of squares of the factor loadings on each of these factors is θ_j :

$$\theta_j = \lambda_j \quad (8.46)$$

The j 'th element of the eigenvector $(\Lambda \tilde{w}_j)$ equals unity while the remaining elements equal zero. Following similar steps which led to equation (8.45), the vector of factor weights is:

$$\underline{a}_j = \mathbf{V}_j \lambda_j^{\frac{1}{2}}. \quad (8.47)$$

As will be presented subsequently, each of these principal factors will be the first principal factor for a matrix of residual covariances.

The preceding results may be combined to yield a factor matrix for r principal factors. Let A_r be a factor matrix for the first r principal factors. Also, let V_r be an $n \times r$ matrix containing the first r eigenvectors and Λ_r be an $r \times r$ diagonal matrix containing the first r eigenvalues of C . Then, equations (8.45) and (8.47) may be combined to:

$$A_r = V_r \Lambda_r^{\frac{1}{2}} . \quad (8.48)$$

Since V_r is a vertical section of an orthonormal matrix:

$$V_r' V_r = I .$$

Then:

$$A_r' A_r = \Lambda_r . \quad (8.49)$$

The columns of A_r are orthogonal and their sums of squares equal the corresponding eigenvalues of C . For the nine mental tests example, three principal factors were extracted; thus, $r = 3$. The principal factors matrix is given at the bottom of Table 8.12. Multipliers for the three columns of eigenvectors are the square roots of the first three eigenvalues, these square roots being: 1.657, 1.114, and .588. The three columns of principal factors matrix are obtained by multiplying the corresponding columns of eigenvectors by these multipliers.

Residual correlation matrices are considered next. When one principal factor is extracted at a time there is a sequence of residual matrices with one factor being obtained from one such matrix and a residual matrix being determined from the matrix used in determining the factor. For the nine mental tests example, this series of residual correlation matrices is given in Table 8.11. The operations will be similar for each of these steps so that only the first factor residual matrix will be considered explicitly. Let $C_{\cdot 1}$ designate the first factor residual matrix. From equation (8.8):

$$C_{\cdot 1} = C - \underline{a}_1 \underline{a}'_1 . \quad (8.50)$$

Equations are written in this section with expanded matrices involving partitions of 1 factor and $(n-1)$ factors similar to the expansion used previously for matrix V . The eigenvalues and eigenvectors of C in equation (8.28) are written as:

$$C = [V_1, V_2] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} \begin{bmatrix} V_1' \\ V_2' \end{bmatrix} \quad (8.51)$$

where matrix Λ_2 , in the present context, is an $(n-1) \times (n-1)$ diagonal matrix containing the eigenvalues after the first one. The matrix product $\underline{a}_1 \underline{a}'_1$ is expressed as:

$$\underline{a}_1 \underline{a}'_1 = [\underline{a}_1, 0] \begin{bmatrix} \underline{a}'_1 \\ 0 \end{bmatrix}.$$

with:

$$\begin{aligned} [\underline{a}_1, 0] &= [V_1, V_2] \begin{bmatrix} \lambda_1^{\frac{1}{2}}, & 0 \\ 0 & 0 \end{bmatrix} \\ \underline{a}_1 \underline{a}'_1 &= [V_1, V_2] \begin{bmatrix} \lambda_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V'_1 \\ V'_2 \end{bmatrix} \end{aligned} \quad (8.52)$$

In the subtraction of $\underline{a}_1 \underline{a}'_1$ from C , as per equation (8.50), the eigenvectors on the left and right are factored out so that:

$$C_{\cdot 1} = [V_1, V_2] \begin{bmatrix} 0, & 0 \\ 0, & \Lambda_2 \end{bmatrix} \begin{bmatrix} V'_1 \\ V'_2 \end{bmatrix} \quad (8.53)$$

Thus, the first eigenvalue of C has been replaced by a zero for $C_{\cdot 1}$ so that the largest eigenvalue of $C_{\cdot 1}$ is the second largest eigenvalue of C . Note that the eigenvectors have not been changed. As a consequence of the foregoing, the first principal factor of $C_{\cdot 1}$ is the second principal factor of C . In a similar manner going from the first factor residual matrix to the second factor residual matrix sets the second eigenvalue of C to zero and the third principal factor of C is the first principal factor of the second residual matrix. These relations continue through as many factors as are extracted.

Several types of information are used in judging the number of factors to be extracted. However, not one criterion can be trusted completely so that an analyst must consider several indications available before making a judgment as to the number of factors. This is in contrast to a common procedure in computer packages which use a single criterion to automate this judgment so that each analysis can be completed automatically in a single run. Several of the types of information for number of factors will be discussed in following paragraphs.

An important point is that factor extraction is only the first part of a complete analysis. After an original factor matrix has been established, there is factor transformation. The transformed factors do not correspond directly, one to one, to the extracted factors but are mixtures of the extracted factors. A final criterion for number of factors extracted is the validity and interpretability of the transformed factor structure.

Major indicators for the number of factors are derived from the series of eigenvalues of the correlation matrix with unities in the diagonal cells and with SMC's in the diagonal cells. A procedure, called by some individuals as "root staring", involves inspection of the series of

eigenvalues, especially of the correlation matrix with SMC's in the diagonal cells. Table 8.13 lists the eigenvalues of the correlation matrix with unities in the diagonal cells and with SMC's in the diagonal cells for the nine mental tests example. Figure 8.1 presents a graph of the eigenvalues of the correlation matrix with SMC's in the diagonal. In this graph, the eigenvalues are plotted against the number of the eigenvalue. This graph illustrates results frequently obtained for well edited test batteries, the series of eigenvalues after a few large ones form an almost straight line. This phenomena may be interpreted as indicating that there are two influences in the formation of the data: first, a relevant factor structure and a second influence of random noise. Cattell (1966) described his "Scree Test" for the number of common factors based on the foregoing observation. The points in an eigenvalue graph are not to be interpreted as goodness of fit measures. If they were so interpreted, factor extraction would continue until a satisfactory goodness of fit is obtained. In contrast, factor extraction should be continued as long as eigenvalues are above the random noise line. Thus, for the nine mental tests example a three factor solution would be accepted. However, Thurstone as well as Cattell advocated extracting one or more extra factors which might be used in the factor transformation process to "clean up" the meaningful transformed factors. Such an operation should be followed only with great care.

An alternative to making an eigenvalue graph is illustrated in Table 8.13. On the right of the section for the eigenvalues of the correlation matrix with SMC's in the diagonal is a column of differences. These values are the differences between consecutive eigenvalues. Geometrically, for a straight line of points, such differences would be equal. Since the eigenvalues are ordered in descending algebraic value, all of the differences between consecutive eigenvalues must be equal to or greater than zero. Then, a series of points which approximate a straight line would have almost equal, positive differences. For the nine mental tests series of differences in Table 8.13, starting with the difference between the fourth and fifth eigenvalue, the values of the differences are quite small with little variation when compared with preceding differences. The last large difference is between the third and fourth eigenvalue indicating that the third eigenvalue is the last one above the random noise line.

Guttman (1964) developed three lower bounds for the number of common factors for a correlation matrix. In this development he considered only cases for population correlation matrices for which the common factor model fitted exactly. Guttman's strongest lower bound for the number of common factors is the number of non-negative eigenvalues of the correlation matrix with SMC's in the diagonal cells. That is, the number of common factors is equal to or greater than the number of these eigenvalues which are positive, including those equal to zero. As can be seen from Table 8.13 for the nine mental tests example, by this criterion there are at least 3 common factors for this matrix. However, as shown by Kaiser and Hunka (1973) from analyses of 64 correlation matrices found in the literature, this criterion leads to the extraction of

Table 8.13
 Information for Number of Factors from
 Eigenvalues of Correlation Matrix for Nine Mental Tests Example

Correlation Matrix with

<u>Unities in Diagonal</u>		<u>SMC's in Diagonal</u>		
	<u>Eigenvalues</u>		<u>Eigenvalues</u>	<u>Differences</u>
1	3.347	1	2.746	1.505
2	1.820	2	1.241	.894
3	.997	3	.346	.395
4	.580	4	-.048	.016
5	.549	5	-.064	.061
6	.497	6	-.125	.028
7	.476	7	-.153	.022
8	.412	8	-.175	.050
9	.322	9	-.225	

Parallel Analysis for Eigenvalues

	<u>Real Data</u>	<u>Approximate Random Data</u>	<u>Differences</u>
1	2.746	.189	2.557
2	1.241	.127	1.114
3	.346	.087	.259
4	-.048	.049	-.097

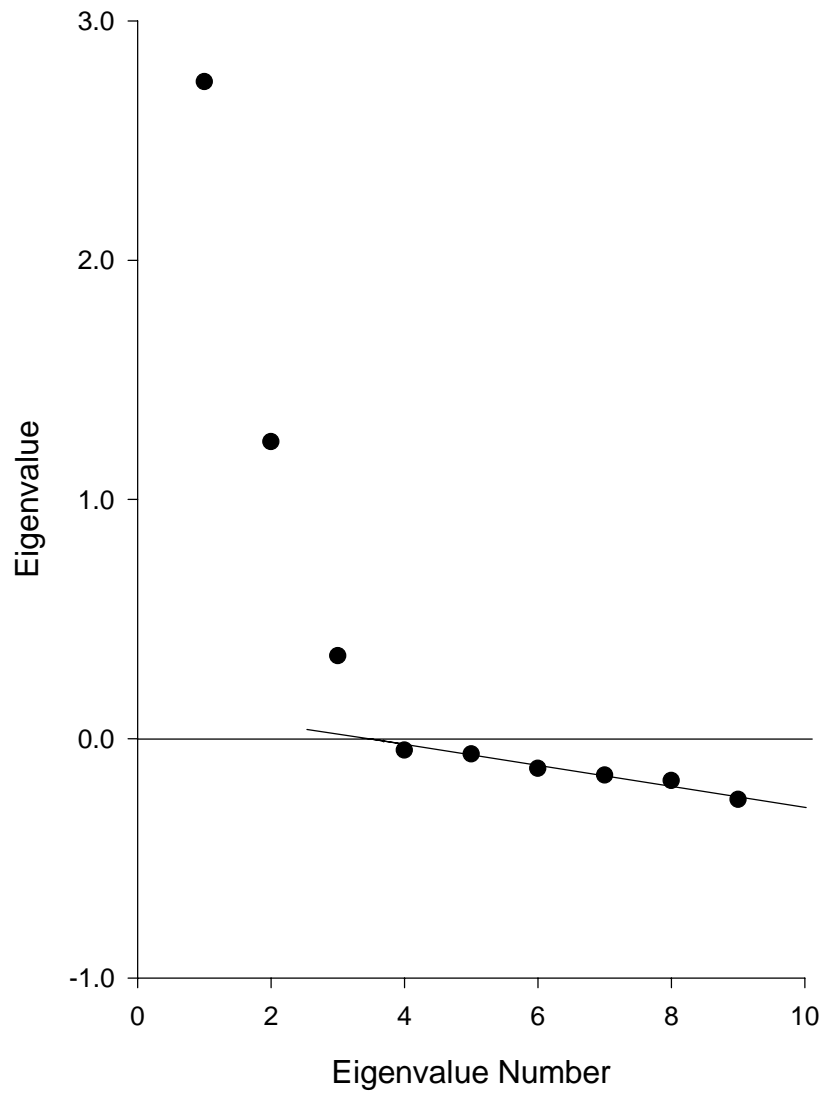


Figure 8.1: Eigenvalue graph for correlation matrix with SMC's in the diagonal, nine mental tests example

a large number of factors. They conclude that this "lower bound is not of practical use in determining the effective number of common factors." An illustration of this stronger lower bound leading to an undesirably large number of extracted factors is given in a second example presented subsequently.

Guttman's weaker lower bound for the number of common factors is the number of eigenvalues of the correlation matrix with unities in the diagonal cells equal to or greater than one. This is a very commonly used value in computer packages for the number of factors to be extracted. Following considerable experience in analyzing a variety of correlation matrices, Kaiser (1960) suggested a simple, approximate procedure using principal components analysis and extracting the number of components equal to Guttman's weaker lower bound. Later, this procedure became called "Little Jiffy" after a remark by Chester Harris. Use of this weaker lower bound has been carried over to computer packages. Analysts, however, must remember that this is a lower bound and may lead to extracting too few factors. See the first column of values in Table 8.13 for the nine mental tests example, these are the eigenvalues of the correlation matrix with unities in the diagonal cells. There are two eigenvalues greater than 1.000 so that Guttman's weaker lower bound for the number of factors would indicate that there are at least two common factors. The third eigenvalue is just less than 1.000; however, a computer output from a number of computer packages would extract only two factors.

Table 8.14 presents results from transformations of the two factor solution and the three factor solution for the nine mental tests example. The two factor solution was indicated by a blind following of the procedure based on Guttman's weaker lower bound. The three factor solution was indicated by the series of eigenvalues of the correlation matrix with SMC's in the diagonal. These results illustrate the types of difficulties which may be encountered when too few factors are extracted. For the three factor solution the three transformed factors are the previously identified numerical operations factor, the spatial factor, and the perceptual speed factor. For the two factor solution, the perceptual speed factor has been collapsed into the numerical operations factor. The spatial factor representation in the two factor solution was OK. Limiting the common factor space by extracting too few factors causes a loss of some weaker factors with the attributes then having improper weights on other factors. As a result, the transformed factor solution is defective. As Kaiser has put it (personal communication) "it's a sin to extract too few factors." Remember that eigenvalues (roots) greater than one of the correlation matrix with unities in the diagonal is a lower bound to the number of factors. Analysts should inspect computer outputs to see if too few factors were extracted. If more factors are indicated, a computer parameter should be set for a rerun to use a larger number of factors.

Humphreys with Ilgen and Montanelli (see Humphreys and Ilgen (1959), Humphreys and Montanelli (1975), Montanelli and Humphreys (1976)) developed a different type of information

Table 8.14
 Transformed Factor Weight Matrices from
 Principal Factors for Nine Mental Tests Example

Test	<u>Factor Weights</u>						
	<u>Two Factor Solution</u>			<u>Three Factor Solution</u>			
		1	2		1	2	3
Addition	1	.55	.02	1	.66	.05	-.09
Multiplication	2	.72	-.07	2	.67	-.05	.11
Three-Higher	3	.50	.31	3	.52	.32	.01
Figures	4	-.03	.71	4	.00	.72	-.04
Cards	5	.00	.79	5	-.02	.80	.03
Flags	6	.03	.71	6	.02	.71	.02
Identical Numbers	7	.62	.04	7	.24	.03	.49
Faces	8	.38	.42	8	-.06	.40	.53
Mirror Reading	9	.49	.30	9	.08	.28	.52

<u>Factor Intercorrelations</u>						
	<u>Two Factor Solution</u>		<u>Three Factor Solution</u>			
	1	2	1	2	3	
1	1.00	.16	1	1.00	.15	.57
2	.16	1.00	2	.15	1.00	.11
			3	.57	.11	1.00

relevant to the number of factors to be extracted. Their suggestion was to compare results obtained from the real data with results obtained from random data. Paralleling the real data score matrix, they drew a matrix of random normal deviates (mean = 0, SD = 1) having the same number of rows and columns as the real data matrix. They, then obtained a correlation matrix for the random data and found the eigenvalues of this matrix with SMC's in the diagonal. Their idea was to continue factor extraction until the eigenvalues for the real data were not larger than the eigenvalues for the random data. See the bottom section of Table 8.13 for an example. The first column for "Real Data" is a copy of the first four eigenvalues given above for the correlation matrix having SMC's in the diagonal. The middle column for "Approximate Random Data" was computed by a procedure to be given later. The third column gives the differences between the real data eigenvalues and the approximate random data eigenvalues. Note that the first three real data eigenvalues are materially greater than the approximate random data eigenvalues while there is a switch at the fourth pair of eigenvalues. The parallel analysis criterion indicates that three factors should be extracted from this correlation matrix. In general, Humphreys and associates suggest that when the real data eigenvalues are not greater than the random data eigenvalues the real data eigenvectors and factors contain no more real information than exists for the random data. Consequently, factor extraction can be stopped.

To implement the parallel analysis procedure Montanelli and Humphreys (1976) ran a large Monte Carlo study involving replications of random data analyses for a selection of matrix sizes. The numbers of statistical individuals were 25 , 96 , 384 . or 1533 . They used a total of 21 battery sizes ranging from 6 to 90 . The number of replications per cell varied from 10 to 40 with the cells having fewer replications being those for the larger sample sizes. For each matrix order, $N \times n$, a series of mean eigenvalues across replications was computed, each of these mean values being for the i 'th eigenvalue. Montanelli and Humphreys provided a system to approximate such series of eigenvalues using tabled weights. The data of means was provided to Tucker who developed a system for use with a computer to approximate the mean eigenvalues. A first point is that the eigenvalues for random data become negative after m eigenvalues where $m = \frac{1}{2}n$ for n even and $\frac{1}{2}(n - 1)$ for n odd. As a consequence, the series of eigenvalues is truncated after the m 'th eigenvalue. Following are the equations for Tucker's system.

$$n' = n(N + 1)/N \quad (8.54)$$

$$x_{Ni} = \ln\{N - 1 + 100(3N - 4n)/N(N + 4n)\} \quad (8.55)$$

$$x_{ni} = \ln\left\{\frac{1}{2}n'(n' + 1) - ni - 10n/\{1 + 2[\ln(50i/n)]^2\}\right\} \quad (8.56)$$

$$a_i = .81848 - .39125 \cdot \ln(i - .54151) \quad (8.57)$$

$$b_{Ni} = -.53365 - .03711 \cdot \ln(i + 4.66402) \quad (8.58)$$

$$b_{ni} = .40202 + .03169 \cdot \ln(i - .71657) \quad (8.59)$$

where N is the sample size, n is the battery size, and i is the eigenvalue number. Then, the approximation to the random data eigenvalue is $\hat{\rho}_{Nni}$:

$$\hat{\rho}_{Nni} = \exp(a_i + b_{Ni}x_{Ni} + b_{ni}x_{ni}) \quad (8.60)$$

The approximate random data eigenvalues in Table 8.13 were computed by this system. Tucker used an empirical, trial and error procedure. Measures of goodness of fit to the Montanelli and Humphreys data indicated a satisfactory level for practical use.

A second example of principal factor extraction is presented in Tables 8.15 through 8.17 and Figure 8.2. This example uses 18 verbal and numerical tests selected from a 66 test battery by Thurstone and Thurstone (1941). Table 8.15 gives the correlation matrix with the test names being given in Table 8.17. First consideration is given to the number of factors to be extracted. For the roots greater than one from the correlation matrix with unities in the diagonal, the fourth eigenvalue is 1.122 while the fifth eigenvalue is .741. By this criterion there are four common factors. Figure 8.2 presents the eigenvalue plot for the correlation matrix with SMC's in the diagonal. There are four eigenvalues above a well defined line for random noise eigenvalues. By this criterion there appear to be four common factors to be extracted. The parallel analysis is given in Table 8.16. The first four real data eigenvalues are distinctly greater than the corresponding approximate random data eigenvalues while there is a switch at eigenvalue 5 for which the real data eigenvalue is less than the approximate random data eigenvalue. By this criterion, also, there appears to be four common factors. From the convergence of these three criteria for the number of factors to be extracted, a decision to consider a four factor solution is well justified.

A second look at the Guttman stronger lower bound for the number of factors is provided in Table 8.16. The first 9 real data eigenvalues are listed for the correlation matrix with SMC's in the diagonal. Guttman's stronger lower bound states that there are at least as many common factors as the number of these eigenvalues that are non-negative. For the 18 verbal and numerical test example, this criterion indicates that there are at least 8 common factors. This is an undesirable answer when compared with the number of factors indicated by the criteria discussed in the preceding paragraph. The Kaiser and Hunka (1973) conclusion appears to be upheld that the Guttman stronger lower bound is not usable for real world data. Table 8.17 contains the principal factors matrix for the 18 verbal and numerical tests example.

Table 8.15

Correlation Matrix among 18 Verbal and Numerical Tests Example

	1	2	3	4	5	6	7	8	9
1	<u>1.000</u>								
2	.264	<u>1.000</u>							
3	.232	.231	<u>1.000</u>						
4	.276	.465	.311	<u>1.000</u>					
5	.221	.248	.378	.460	<u>1.000</u>				
6	.284	.034	.431	.221	.237	<u>1.000</u>			
7	.234	.243	.377	.367	.357	.286	<u>1.000</u>		
8	.207	.395	.378	.439	.400	.154	.437	<u>1.000</u>	
9	.499	.208	.307	.328	.281	.467	.295	.203	<u>1.000</u>
10	.160	.092	.294	.236	.254	.299	.294	.321	.242
11	.195	.444	.220	.476	.384	.054	.237	.308	.183
12	.126	.403	.307	.420	.349	.116	.341	.473	.126
13	.159	.430	.209	.497	.454	-.008	.226	.335	.133
14	.184	.420	.206	.511	.455	.026	.187	.342	.172
15	.168	.365	.240	.486	.497	.126	.258	.326	.210
16	.165	.329	.331	.342	.256	.238	.396	.473	.163
17	.394	.454	.364	.411	.341	.291	.334	.376	.436
18	.217	.230	.385	.490	.541	.407	.291	.303	.356

	10	11	12	13	14	15	16	17	18
1									
2									
3									
4									
5									
6									
7									
8									
9									
10	<u>1.000</u>								
11	.112	<u>1.000</u>							
12	.218	.385	<u>1.000</u>						
13	.048	.624	.396	<u>1.000</u>					
14	.090	.623	.418	.769	<u>1.000</u>				
15	.112	.550	.396	.730	.661	<u>1.000</u>			
16	.296	.239	.459	.212	.241	.223	<u>1.000</u>		
17	.274	.324	.313	.281	.331	.310	.330	<u>1.000</u>	
18	.224	.418	.387	.439	.479	.500	.277	.364	<u>1.000</u>

* Selected from 66 test study by Thurstone and Thurstone (1941), Factorial studies of intelligence; samole size=710.

Table 8.16

Eigenvalues and Parallel Analysis for 18 Verbal and Numerical Tests Example

	Real Data	Approximate Random Data	Differences
1	6.031	.316	5.715
2	1.649	.254	1.395
3	.764	.209	.554
4	.532	.175	.357
5	.112	.146	-.035
6	.054	.120	-.066
7	.020	.094	-.074
8	.011	.065	-.054
9	-.013	.017	-.030

Table 8.17

Principal Factors Matrix for 18 Verbal and Numerical Tests Example

Test		1	2	3	4
Addition	1	.39	.28	.22	.31
Arithmetic (Word Problems)	2	.57	-.12	-.17	.36
Mirror Reading	3	.52	.33	-.03	-.14
Directions	4	.71	-.04	-.01	.06
Disarranged Sentences	5	.65	.01	.09	-.22
Identical Numbers	6	.35	.53	.24	-.16
Letter Grouping	7	.52	.27	-.18	-.06
Letter Series	8	.61	.12	-.36	.00
Multiplication	9	.45	.41	.34	.19
Number Patterns	10	.34	.34	-.12	-.11
Paragraph Recall	11	.64	-.32	.05	.07
Pedigrees	12	.60	-.05	-.31	-.06
Vocabulary	13	.70	-.52	.10	-.02
Sentences (Completion)	14	.70	-.45	.11	.00
Same or Opposite	15	.69	-.35	.16	-.11
Secret Writing	16	.51	.20	-.37	-.03
Three-Higher	17	.60	.23	.03	.26
Verbal Enumeration	18	.67	.05	.23	-.26

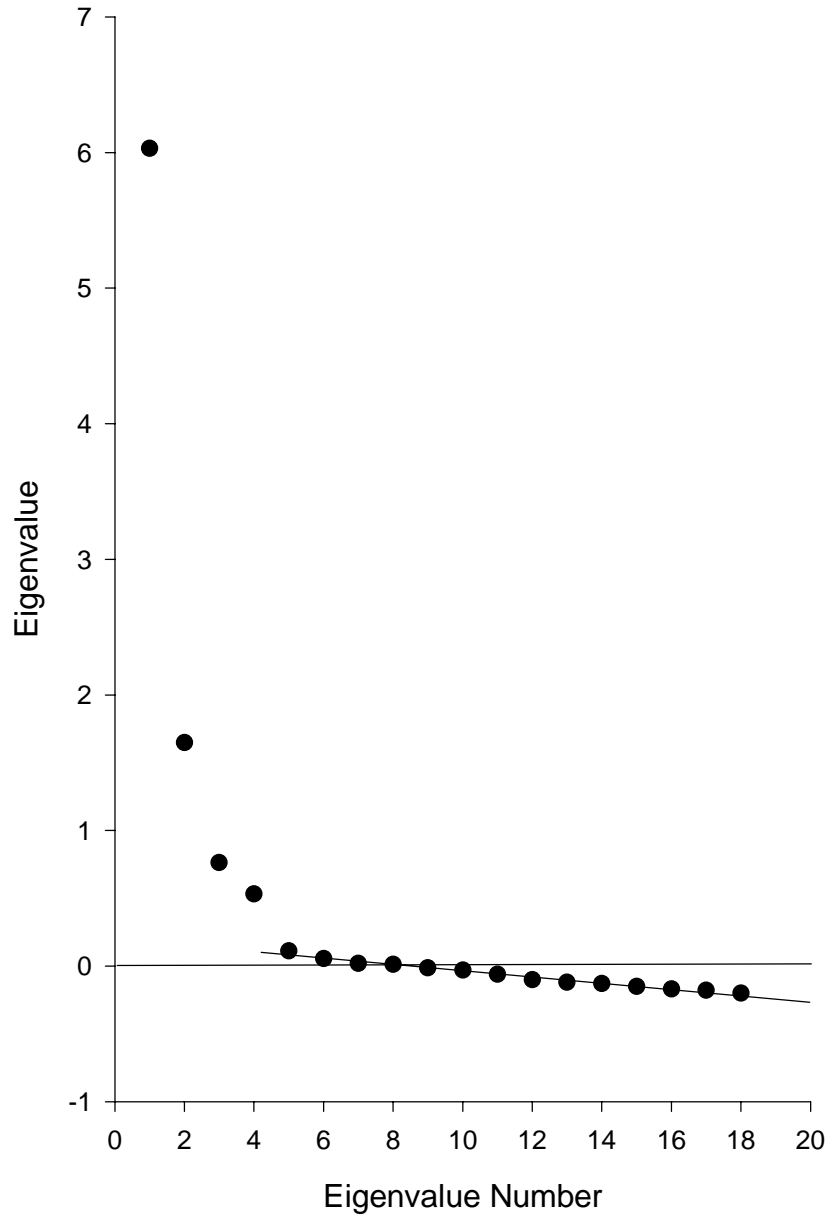


Figure 8.2: Eigenvalue graph for correlation matrix with SMC's in diagonal, 18 verbal & numerical test example.

8.6. Alpha Factor Analysis

Kaiser and Caffrey (1965) presented alpha factor analysis as a psychometric approach to factor analysis in contrast to what they termed statistical factor analysis. They emphasized, in a population of individuals, the generalization of results to a universe of content from observations on a battery of attributes which they considered to be a sample (usually nonrandom) from the universe of content. They contrasted this conception with statistical analysis generalization to a population of individuals from an observed sample (usually random) of individuals. Alpha factor analysis considers relations in a population of individuals and does not consider the sampling of individuals problem. The argument is that this sampling of attributes and generalization to a universe of content is an extremely important psychometric problem not considered in statistical factor analysis.

Alpha factor analysis considers the common parts of observed attribute measures. In accordance with the factor analysis model presented in Chapter 3 and outlined in Chapter 7, the vector of common parts scores is related to a vector of common factor scores for uncorrelated factors by:

$$\mathbf{z}_\alpha = \mathbf{x}_\alpha \mathbf{A}' \quad (8.61)$$

This equation is an interpretation of equation (7.11) in terms of uncorrelated factors (Remember that score vectors are taken as row vectors.). In the present context, reference to uncorrelated common factors is a matter of convenience. Problems with transformations to correlated common factors are to be discussed subsequently. A basic relation in alpha factor analysis involves an inverse from equation (8.61) which expresses common factor scores as weighted linear combinations of the common part scores. For a given battery, the common factor score for an individual on factor k is expressed as:

$$x_{\alpha k} = \sum_{j=1}^n w_{kj} z_{\alpha j} \quad (8.62)$$

where j refers to attributes in the battery. This relation pertains to a given battery since the common factor scores and common part scores may, and probably will change with a change in the battery. In comparison, the common factor score for the individual as related to the universe of content is related to the common part scores by:

$$\xi_{\alpha k} = \sum_{p=1}^{\infty} w_{kp} z_{\alpha p} \quad (8.63)$$

where p refers to attributes in the universe of content. The correlation between $x_{\alpha k}$ and $\xi_{\alpha k}$ is used as a measure of generalizability from the given battery to the universe of content. Following

the work by Kuder and Richardson (1937) on test reliability which was extended by Cronbach (1951) and with Rajaratnam and Glesser (1963), an adaptation of coefficient alpha was developed for generalizability in the present situation.

$$\alpha = \left[\frac{n}{n-1} \right] \left[1 - \frac{w' H^2 w}{w' (R - U^2) w} \right] \quad (8.64)$$

where \mathbf{w} is a column vector of weights for a given factor. In factor extraction this measure is to be maximized for successive factors. Solution for the maximum is facilitated by defining:

$$\lambda = \frac{w' (R - U^2) w}{w' H^2 w} \quad (8.65)$$

Setting the partial derivative of λ with respect to \mathbf{w} equal to zero leads, with algebraic operations, to:

$$[(R - U^2) - \lambda H^2] \mathbf{w} = \mathbf{0} \quad (8.66)$$

With column vector \mathbf{v} defined by:

$$v = Hw \quad (8.67)$$

Equation (8.66) can be written as follows.

$$[H^{-1}(R - U^2)H^{-1} - \lambda I] \mathbf{v} = \mathbf{0} \quad (8.68)$$

λ is an eigenvalue of $H^{-1}(R - U^2)H^{-1}$ and \mathbf{v} is the corresponding unit eigenvector. From equations (8.64) and (8.65) the k 'th value of α is:

$$\alpha_k = \left[\frac{n}{n-1} \right] \left[1 - \frac{1}{\lambda_k} \right] \quad (8.69)$$

Note that the maximum α corresponds to the maximum λ .

Matrix $(R - U^2)$ has the communalities, diagonal entries of \mathbf{H}^2 , in its diagonal and, thus, is the correlation matrix of the common parts of the observed attributes. Premultiplication and postmultiplication of this matrix by H^{-1} scales the attribute measures to unit communalities, thus, yielding the correlation matrix among the common parts. The eigensolution of equation (8.68) may be used to yield principal factors of this correlation matrix. For r factors (the number of factors to be extracted will be discussed in subsequent paragraphs.) the principal factors matrix of $H^{-1}(R - U^2)H^{-1}$ is:

$$\tilde{\mathbf{A}} = \mathbf{V}_r \mathbf{\Lambda}_r^{\frac{1}{2}} \quad (8.70)$$

To obtain the factor matrix for the attributes in terms of the original scale, it is necessary to perform an inverse scaling:

$$\mathring{\mathbf{A}} = H\tilde{\mathbf{A}}. \quad (8.71)$$

Combining equations (8.70) and (8.71):

$$\mathring{\mathbf{A}} = H\mathbf{V}_r\Lambda_r^{\frac{1}{2}}. \quad (8.72))$$

To this point the communalities in \mathbf{H}^2 have been taken to be known. However, these values are not known so that a solution for them is necessary. For convenience, repeating equation (3.49):

$$\mathbf{H}^2 = \text{Diag}(\mathring{\mathbf{A}}\mathring{\mathbf{A}}') \quad (3.49)$$

Also, from equation (3.34):

$$\mathbf{U}^2 = \mathbf{I} - \mathbf{H}^2 \quad (8.73)$$

An iterative solution involves starting with an initial \mathbf{H}^2 and performing the following steps.

1. Obtain \mathbf{U}^2 as per equation (8.73).
2. Form matrix $H^{-1}(R - \mathbf{U}^2)H^{-1}$.
3. Obtain Λ_r and \mathbf{V}_r from an eigensolution of $H^{-1}(R - \mathbf{U}^2)H^{-1}$.
4. Obtain matrix $\mathring{\mathbf{A}}$ as per equation (8.72).
5. Obtain \mathbf{H}^2 as per equation (3.49).
6. Return to step 1 until there is a minimal change of \mathbf{H}^2 .

Kaiser and Caffrey (1965) outlined an efficient program to implement this solution incorporating some short cuts to speed up the computations. While there is no proof that this system will converge, experience indicates that a converged solution will be obtained in almost all cases.

The ALPHA factor analysis of the correlation matrix in Table 1.1 for the nine mental tests example is given in Table 8.18 for two factors. The eigenvalues of $H^{-1}(R - \mathbf{U}^2)H^{-1}$ are given in the first row and the corresponding values of α are given in the second row. Only the first three α 's are given for reasons to be described later. The iterated factor matrix and communalities are presented in the lower section of the table.

Kaiser and Caffrey (1965) suggest that the number, r , of factors to be extracted from a correlation matrix equals the number of positive α 's ; that is, all factors, and only those factors, are to be extracted for which the generalizability is positive. Thus, the number of ALPHA factors to be extracted for the nine mental tests example is two, this being the number of positive α 's . Note from equation (8.69) that for a positive α_k the eigenvalue λ_k must be greater than one. This relation supports a commonly used principal that the number of factors to be extracted from a

Table 8.18

Alpha Factor Analysis for Nine Mental Tests Example
Two Factors Extracted

	Dimension								
	1	2	3	4	5	6	7	8	9
Eigenvalue	6.38	2.62	.99	.06	.03	-.07	-.27	-.34	-.41
Alpha	.95	.70	-.01						

Factor Matrix			
	1	2	Communality
1	.44	-.29	.28
2	.58	-.55	.64
3	.63	-.09	.40
4	.48	.53	.51
5	.58	.60	.70
6	.53	.49	.52
7	.54	-.32	.39
8	.60	.09	.37
9	.62	-.07	.38

correlation matrix equals the number of eigenvalues of the correlation matrix which are greater than one. To support the transition from eigenvalues of $H^{-1}(R - U^2)H^{-1}$ to eigenvalues of the correlation matrix R consider the following development. From equation (8.73):

$$H^{-1}(R - U^2)H^{-1} = H^{-1}(R - I + H^2)H^{-1} = H^{-1}(R - I)H^{-1} + I .$$

Equation (8.68) becomes:

$$[H^{-1}(R - I)H^{-1} - (\lambda - 1)I]V = 0 \quad (8.74)$$

Several logical steps follow.

1. For any α to be positive, the corresponding $(\lambda - 1)$ must be positive.
2. Since $(\lambda - 1)$ is an eigenvalue of $H^{-1}(R - I)H^{-1}$, the number of positive α 's equals the number of positive eigenvalues of $H^{-1}(R - I)H^{-1}$.
3. By Sylvester's Law of Inertia (see: Guttman, 1954 and Bôcher, 1907) the number of positive eigenvalues of value of $H^{-1}(R - I)H^{-1}$ is independent of the value of H .
4. A possible H to be considered is an identity matrix so that the number of positive eigenvalues of $(R - I)$ equals the number of positive eigenvalues of $H^{-1}(R - I)H^{-1}$ for any other value of H .
5. Each eigenvalue of $(R - I)$ equals the corresponding eigenvalue of R decreased by 1.
6. The number of positive eigenvalues of $(R - I)$ equals the number of eigenvalues of R which are greater than one.
7. The number of positive α 's equals the number of eigenvalues of R greater than one.

From the preceding, the number of factors, r , may be set from the eigenvalues of the correlation matrix and not changed with different approximations of the communalities.

Kaiser and Caffrey (1965) did not consider the topic of transformations of factors from the obtained ALPHA factors. For the nine mental tests example, Table 8.19 presents a transformation to correlated factors of the two factor ALPHA factor matrix. The Alpha coefficients for the transformed factors were obtained by entering the transformed factor weights into equation (8.64). The results in Table 8.19 differ from the solution given in Table 1.2 which included three factors. The first transformed ALPHA factor in Table 8.19 appears to be a combination of a numbers factor and a perceptual speed factor, that is, a combination of factors 1 and 3 of Table 1.2. This inspection suggests that too few factors had been extracted using the principle of positive generalizability.

Consider the series of eigenvalues in Table 8.18, the third eigenvalue is only slightly less than one and the corresponding α is only slightly negative. There is a distinct break in the eigenvalue series between the third and fourth eigenvalue which suggests that a three factor

Table 8.19

Transformed Alpha factors for Nine Mental Tests Example
Two Factor Solution

Factor Weights		
Attribute	1	2
1. Addition	.52	.04
2. Multiplication	.80	-.08
3. Three-Higher	.51	.32
4. Figures	-.03	.71
5. Cards	.00	.84
6. Flags	.03	.72
7. Ident. Numbers	.61	.07
8. Faces	.37	.44
9. Mirror Reading	.49	.33

Factor Correlations		
	1	2
1	1	
2	.52	1

Alpha for Transformed Factors		
	1	2
	.80	.86

solution might be appropriate. Table 8.20 presents a three factor ALPHA solution and Table 8.21 presents the transformed solution. Note that the transformed factors correspond to the three factors in Table 1.2 . A point of interest is that, while the third ALPHA factor had slightly negative generalizability, all three transformed factors had positive α 's .

The preceding material suggests a distinction between two views: that of factors being determined from the observed scores and the view of factor analytic studies being conducted to obtain indications of major internal attributes, or latent variables which are characteristics in a domain of mental behavior. Use of the generalizability of factors provides no mechanism to distinguish between major dimensions and possible trivial dimensions due to lack of fit. As noted by Kaiser and Caffrey, enlarging a battery of measures will lead to increasing numbers of factors. A small battery such as the nine mental tests example may lead to extraction of too few factors some of which are combinations of major factors which might be obtained with a larger battery. An argument may be made that this small battery is not adequate to determine the common factor space. This is a question of battery adequacy. However, extraction of three factors, ignoring the negative generalizability of the third factor, does lead to a quite interpretable transformed solution. This battery appears to be adequate to provide indications of three major internal attributes. There appears to be a contrast between two opinions as to the purpose of factor analytic studies. One opinion is that factor analytic studies are conducted to provide information about the structure of the dynamics of mental behavior by identifying internal attributes. The other opinion appears to be that factor analytic studies are conducted to determine factors, including factor scores, from the observed attributes. ALPHA factor analysis appears to be a representative of a procedure to determine factors.

8.7. Image Factor Analysis

Jöreskog (1963) described a most interesting but little used factor analysis model and analytic procedure which, subsequently, became known as Image Factor Analysis due to mathematical relations to Guttman's (1953) image theory. However, Jöreskog (1969) commented that Image Factor Analysis was a model in its own right. In Jöreskog's notation, his model for uncorrelated factors is:

$$\Sigma = \Lambda\Lambda' + \Psi \quad (8.75)$$

where Σ is the population dispersion (covariance) matrix among the observed attributes, Λ is the factor weights matrix, and Ψ is the diagonal matrix of unique variances. Note that Jöreskog does not incorporate a term for lack of fit of the model. For Image Factor Analysis, Jöreskog specializes this model by assuming that:

Table 8.20
Alpha Factor Analysis for Nine Mental Tests Example
Three Factors Extracted

	Dimension								
	1	2	3	4	5	6	7	8	9
Eigenvalue	5.53	2.48	.99	.14	.10	.03	-.03	-.10	-.14
Alpha	.92	.67	-.01						

	Factor Matrix			
	1	2	3	Communality
1	.46	-.35	.32	.43
2	.54	-.56	.20	.64
3	.63	-.11	.21	.45
4	.51	.51	.12	.53
5	.61	.57	.09	.71
6	.55	.46	.08	.52
7	.52	-.37	-.25	.47
8	.63	.06	-.36	.53
9	.62	-.11	-.29	.47

Table 8.21
 Transformed Alpha factors for Nine Mental Tests Example
 Three Factor Solution

Factor Weights			
Attribute	1	2	3
1. Addition	.70	.06	-.12
2. Multiplication	.76	-.06	.08
3. Three-Higher	.53	.34	.02
4. Figures	.01	.73	-.04
5. Cards	-.01	.84	.02
6. Flags	.02	.72	.03
7. Ident. Numbers	.24	.02	.53
8. Faces	-.09	.41	.62
9. Mirror Reading	.08	.28	.56

Factor Correlations			
	1	2	3
1	1.00	.12	.52
2	.12	1.00	.08
3	.52	.08	1.00

Alpha for Transformed Factors		
1	2	3
.24	.85	.72

$$\Psi = \theta[diag(\Sigma^{-1})]^{-1} \quad (8.76)$$

where θ is a constant parameter of the model.

A translation from Jöreskog's notation to that used in this book is needed. Our matrix $\mathring{\mathbf{A}}$ replaces Jöreskog's matrix $\mathbf{\Lambda}$ and \mathbf{U}^2 replaces Ψ . Our Δ_{Σ} is taken to be zero so that:

$$\Sigma = \Sigma_{yy} = \Sigma_{zz} = \mathring{\mathbf{A}}\mathring{\mathbf{A}}' + \mathbf{U}^2. \quad (8.77)$$

Equation (8.35) gives the squares of the standard errors of estimating the attributes from the remaining attributes in a battery in a sample. The equivalent relation for the population and involving the covariance matrix Σ is:

$$\Sigma_e = [diag(\Sigma^{-1})]^{-1} \quad (8.78)$$

with diagonal entries σ_{ej}^2 which are the error variances in estimating scores on attribute j from the remaining attributes in the battery. With these translations of notation and equations (8.76) and (8.78), Jöreskog's model becomes:

$$\Sigma = \mathring{\mathbf{A}}\mathring{\mathbf{A}}' + \theta\Sigma_e. \quad (8.79)$$

Understanding of the Image Factor Analysis model is facilitated by consideration of the inequality of equation (8.56). This equation may be revised to apply to covariance matrices instead of correlation matrices so as to yield with algebraic operations an inequality between uniqueness and error of estimate variances:

$$u_j^2 < \sigma_{ej}^2 \quad (8.80)$$

Usually this inequality includes possible equality; however, this equality occurs only for very special situations and is ignored here. With θ a positive constant between zero and one, u_j^2 might be approximated by $\theta\sigma_{ej}^2$. This would be especially true when both u_j^2 and $\theta\sigma_{ej}^2$ are nearly constant over the attributes in the battery. With this approximation, the Image Factor Analysis model may be written as:

$$(\Sigma - \theta\Sigma_e) = \mathring{\mathbf{A}}\mathring{\mathbf{A}}' \quad (8.81)$$

In this form, use of $\theta\sigma_{ej}^2$ appears to provide a solution to the communality problem. There remains to be the solution for an appropriate value for θ . This will be discussed in terms of analysis for a sample. The factor matrix could be determined by the principal factors procedure.

In a sample: Σ is replaced by \mathbf{C} , $\mathring{\mathbf{A}}$ by \mathbf{A} and Σ_e by S_e^2 with, per equation (8.35),

$$S_e^2 = [diag(\mathbf{C}^{-1})]^{-1}$$

With the addition of a residual term, the model of equation (8.81) may be written for the sample as:

$$(\mathbf{C} - \theta \mathbf{S}_e^2) = \mathbf{A}\mathbf{A}' + \mathbf{\Delta} \quad (8.82)$$

Jöreskog considers $(\theta \mathbf{S}_e^2 + \mathbf{\Delta})$ to be the residual in representing \mathbf{C} by $\mathbf{A}\mathbf{A}'$. In the computing procedure Jöreskog described in 1963 he performed a scaling transformation instead of directly factoring $(\mathbf{C} - \theta \mathbf{S}_e^2)$ to obtain:

$$S_e^{-1}(\mathbf{C} - \theta \mathbf{S}_e^2)S_e^{-1} = S_e^{-1}\mathbf{A}\mathbf{A}'S_e^{-1} + S_e^{-1}\mathbf{\Delta}S_e^{-1}$$

so that

$$(S_e^{-1}\mathbf{C}S_e^{-1} - \theta \mathbf{I}) = S_e^{-1}\mathbf{A}\mathbf{A}'S_e^{-1} + S_e^{-1}\mathbf{\Delta}S_e^{-1} \quad (8.83)$$

With the following definitions:

$$\tilde{\mathbf{C}} = S_e^{-1}\mathbf{C}S_e^{-1} \quad (8.84)$$

$$\tilde{\mathbf{A}} = S_e^{-1}\mathbf{A} \quad (8.85)$$

$$\tilde{\mathbf{\Delta}} = S_e^{-1}\mathbf{\Delta}S_e^{-1} \quad (8.86)$$

equation (8.83) yields:

$$(\tilde{\mathbf{C}} - \theta \mathbf{I}) = \tilde{\mathbf{A}}\tilde{\mathbf{A}}' + \tilde{\mathbf{\Delta}} \quad (8.87)$$

An important point is that $\tilde{\mathbf{C}}$ can be shown to be invariant within reflection of the attributes for any rescaling of the attribute scores. The form of equation (8.87) is very convenient: the eigenvectors of $(\tilde{\mathbf{C}} - \theta \mathbf{I})$ are invariant with changes in θ and the eigenvalues change by an additive procedure. The eigensolution for \mathbf{C} is (the eigenvalues are in descending algebraic order):

$$\tilde{\mathbf{C}} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}' \quad (8.88)$$

so that the eigensolution for $(\tilde{\mathbf{C}} - \theta \mathbf{I})$ is:

$$(\tilde{\mathbf{C}} - \theta \mathbf{I}) = \mathbf{V}(\mathbf{\Lambda} - \theta \mathbf{I})\mathbf{V}' \quad (8.89)$$

This is in accord with the general theory of eigensolutions. A principal factors solution to r factors of $(\tilde{\mathbf{C}} - \theta \mathbf{I})$ yields:

$$\tilde{\mathbf{A}} = \mathbf{V}_r(\mathbf{\Lambda}_r - \theta \mathbf{I}_r)^{\frac{1}{2}} \quad (8.90)$$

where \mathbf{V}_r contains the first r eigenvectors, $\mathbf{\Lambda}_r$ is a diagonal matrix containing the first r eigenvalues, and I_r is an $r \times r$ identity matrix. Then the factor matrix \mathbf{A} in terms of the original scaling may be obtained from equation (8.85) as:

$$\mathbf{A} = \mathbf{S}_e \tilde{\mathbf{A}}. \quad (8.91)$$

Jöreskog's suggested solution for θ was to set it equal to the mean of the $(n - r)$ discarded eigenvalues of $\tilde{\mathbf{C}}$

$$\theta = \sum_{k=r+1}^n \lambda_k / (n - r) \quad (8.92)$$

With this value of θ the sum of the discarded eigenvalues of $(\tilde{\mathbf{C}} - I)$ equals zero, that is:

$$\sum_{k=r+1}^n (\lambda_k - \theta) / (n - r) = 0 \quad (8.93)$$

A major problem is the choice of the number of factors to be extracted. Fortunately, the solution for $\tilde{\mathbf{C}}$ and the eigensolution do not depend upon the choice of r so that several choices may be considered. Jöreskog suggested a statistical coefficient for testing for significant departure of a chosen model (number of factors) from the number for the population. Subsequently, he found that this coefficient did not follow an expected chi-square distribution so that he advised (personal communication) that this coefficient would not be useful. Consequently, this coefficient will not be presented here. Choice of the number of factors remains a matter of judgment. Information from the series of eigenvalues might be useful; also, the size of residuals could be used.

Jöreskog (1969) described a maximum likelihood method for estimating the parameters in the Image Factor Analysis model and presented a measure of goodness of fit. The method described in the preceding paragraphs sacrifices some efficiency in the parameter estimates for the sake of speed. The maximum likelihood estimates are fully efficient in large samples. However, this procedure is relatively slow involving an iterative solution. In this respect it is very similar to the maximum likelihood factor analysis to be discussed in Chapter 9. While the maximum likelihood solution for image factor analysis is of theoretic interest it is very seldom used.

Reference notes for Sylvester's "Law of Inertia":

Guttman, L. Some necessary conditions for common-factor analysis. Psychometrika. 1954, 19, 149-161.

Bôcher, Maxime. Introduction to higher algebra. New York, Macmillan, 1907. (Sixteenth printing, 1952).

Kaiser, Henry F. & Caffrey. Alpha factor analysis. Psychometrika. 1965, 30, 1-14.

References for Image Factor Analysis

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

Jöreskog, K. G. Statistical estimation in factor analysis. Stockholm: Almqvist & Wiksell, 1963.

Jöreskog, K. G. Efficient estimation in image factor analysis. Psychometrika, 1969, 34, 51-75.