Political Science 239

Principal Components and Factor Analysis

Rocio Titiunik

Spring 2008

1 Review of eigenvalues and eigenvectors

1.1 Eigenvalues

Let A be a square matrix. The scalar r is an eigenvalue of A if and only if $A - rI$ is a *singular* matrix. Thus r is an eigenvalue of A if and only if

$$det(A - rI) = 0 \tag{1}$$

In words, an *eigenvalue* of A is a number r such that when subtracted from each of the diagonal elements of A converts A into a singular matrix.

Note that for an $n \times n$ square matrix A, the left hand side of equation (1) defines a polynomial of order n in the variable r. This polynomial is referred to as the *characteristic polynomial* of A, and you can see from equation (1) that r is an eigenvalue of A if and only if it is a root of the characteristic polynomial of A. Also, since an nth order polynomial has at most n roots, then equation (1) implies that an $n \times n$ matrix A has *at most* n eigenvalues.

1.2 Eigenvectors

Recall that a matrix M is called *nonsingular* if the system of equations $Mx = b$ has a unique solution for every b (this is a definition). Remember also that an $n \times n$ matrix is invertible *if and only if* it is nonsingular (this is a theorem).

Consider the following system of equations, where M is an $n \times n$ matrix and x and 0 are $n \times 1$ vectors:

$$Mx = 0 \tag{2}$$

The matrix M is nonsingular if and only if the only solution to (2) is $x = 0$. (This is because if M is nonsingular, we have $x = M^{-1} \times 0 = 0$ and this is the only solution.) Conversely, M is singular if and only if the system defined in equation (2) has a nonzero solution.

Now, back to eigenvalues, the fact that the matrix $(A - rI)$ is singular means that the system of equations $(A - rI)\, v = 0$ has a nonzero solution.

An eigenvector of A *corresponding to eigenvalue* r is a *nonzero* vector v such that

$$(A - rI)\, v = 0 \tag{3}$$

where r is an eigenvector of A. Note that there is one eigenvector for every eigenvalue of A.

Note that if v is an eigenvector corresponding to eigenvalue r we have

$$\begin{aligned} (A - rI)\, v &= 0 && (4) \\ Av - rv &= 0 && (5) \\ Av &= rv && (6) \end{aligned}$$

The space spanned by eigenvector v corresponding to eigenvalue r is called *the eigenspace of A with respect to r.*

1.3 A very important theorem

Let $r_1, r_2, \cdots, r_k$ be eigenvalues of the $k \times k$ matrix A, and let $v_1, v_2, \cdots, v_k$ be the corresponding eigenvectors. Let the matrix P be formed by binding together all the eigenvectors of A, so that the jth column of P is eigenvector v_j. Then we have:

$$\begin{aligned}
AP &= A\,[v_1 \cdots v_k] \\
&= [Av_1 \cdots Av_k] \\
&= [Ar_1 \cdots Ar_k] \\
&= [v_1 \cdots v_k] \begin{pmatrix} r_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & r_k \end{pmatrix} \\
&= P \begin{pmatrix} r_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & r_k \end{pmatrix} \qquad (7)
\end{aligned}$$

Multiplying both sides of equation (7) by P^{-1}, we get

$$P^{-1}AP = \begin{pmatrix} r_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & r_k \end{pmatrix} \qquad (8)$$

Thus, we have the following theorem: Let $r_1, r_2, \cdots, r_k$ be eigenvalues of the $k \times k$ matrix A, and let $v_1, v_2, \cdots, v_k$ be the corresponding eigenvectors. Form the matrix

$$P = [v_1 \cdots v_k]$$

whose columns are A's k eigenvectors.

If P is invertible, then

$$P^{-1}AP = \begin{pmatrix} r_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & r_k \end{pmatrix} \tag{9}$$

Conversely, if $P^{-1}AP$ is a diagonal matrix D, then the columns of P must be the eigenvectors of A and the diagonal entries of D must be the eigenvalues of A.

Finally, note that it follows that A can be diagonalized in the following way:

$$A = PDP^{-1} \tag{10}$$

1.4 Example

Let A be

$$A = \begin{pmatrix} -1 & 3 \\ 2 & 0 \end{pmatrix} \tag{11}$$

Its characteristic polynomial is

$$\begin{aligned} det(A - rI) &= det \begin{pmatrix} -1-r & 3 \\ 2 & 0-r \end{pmatrix} \\ &= r\,(r+1) - 6 \\ &= (r+3)\,(r-2) \end{aligned}$$

The two eigenvalues of A are therefore $r_1 = -3$ and $r_2 = 2$.

The eigenvector v_1 associated with eigenvalue r_1 solves the system:

$$(A - (-3)I)v_1 = \begin{pmatrix} 2 & 3 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} v_1^1 \\ v_1^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

So we have

$$\begin{aligned} 2v_1^1 + 3v_1^2 &= 0 \\ v_1^1 &= \frac{-3}{2}v_1^2 \end{aligned} \tag{12}$$

and hence one possible eigenvector is

$$v_1 = \begin{pmatrix} -3 \\ 2 \end{pmatrix}$$

Of course, since equation (12) has infinitely many solutions, we have infinitely many v_1's, but all possible v_1's span the *same* eigenspace. In other words, the eigenspace with respect to r_1 is the one-dimensional space of all solutions to the linear equation (12).

1.5 Important properties to remember

Finally, remember the following properties of eigenvalues. Let A be an $n \times n$ matrix with eigenvalues $r_1, r_2, \cdots, r_n$. Then,

$$\begin{aligned} r_1 + r_2 + \cdots + r_n &= trace(A) \\ r_1 \cdot r_2 \cdot \cdots r_n &= det(A) \end{aligned}$$

2 Latent models: When the explanatory variables are "latent"

Now that we've reviewed eigenvalues and eigenvectors, we can get to the topic of this section, which is latent models, and in particular principal component analysis and factor analysis.

In the textbook account of statistical modeling, one has a theory of how a certain outcome variable Y is affected by an "independent" variable X, which is assumed to be observed or *manifest*.

The idea behind principal component and factor analysis (and behind latent models more generally) is that sometimes the explanatory variables cannot be directly observed, not because you don't have data on them but because they couldn't possibly be observed. These variables are called *latent*. As put by Everitt (1984), "they are essentially hypothetical constructs invented by a scientist for the purpose of understanding some research area of interest, and for which there exists no operational method for direct measurement". Everitt cites social class, public opinion or extrovert personality as examples of such concepts. You may wonder why one would want pursue a theory whose main explanatory variable cannot be clearly operationalized nor measured. That's a fair question, but let's put it aside for now.

Factor analysis is a method for studying the dependence of a set of manifest variables on a number of latent variables. Factor analysis assumes that both the manifest and latent variables are continuous. The idea is somehow to reduce the full set of manifest observations to a few latent values, so it is a sort of data reduction method.

2.1 The general latent variable model

Let $x = [x_1, x_2, \cdots x_p]$ be the manifest variables and $y = [y_1, y_2, \cdots y_m]$ be the latent variables. Usually, we have $p > m$.

Latent variable models assume that $x_1, x_2, \cdots x_p$ have a joint probability distribution conditional on $y_1, y_2, \cdots y_m$. Let this conditional distribution be $\phi(x|y)$. Let's assume that the manifest variables are continuous so that ϕ is a density function. Also, let the marginal distribution of y be $h(y)$. The marginal distribution of x is therefore:

$$f(x) = \int \phi(x|y)\, h(y)\, dy \tag{13}$$

The crucial assumption of latent models is *conditional independence*, this is, that manifest variables are independent of each other *given* the values of the latent variables. This is

$$\phi\left(x|y\right) = \phi_1\left(x_1|y\right), \phi_2\left(x_2|y\right), \cdots, \phi_p\left(x_p|y\right) \tag{14}$$

2.2 The principal components model

In one word, principal components analysis reduces to finding the eigenstructure of the covariance matrix of the original data of manifest variables. Let's see how this works.

The principal component model starts by defining a scalar ξ which is a linear combination of the original p manifest variables x:

$$\xi = \gamma_1 x_1 + \gamma_2 x_2 + \cdots + \gamma_p x_p \tag{15}$$

Which can be rewritten in matrix form as :

$$\xi = \gamma' x \tag{16}$$

Let Σ be the covariance matrix of x:

$$\Sigma \equiv E\left(xx'\right) \tag{17}$$

The covariance of the transformed variable ξ is

$$E\left(\xi\xi'\right) = E\left(\gamma' x x' \gamma\right) = \gamma' \Sigma \gamma \tag{18}$$

Now, the principal components analysis is to find the weight vector γ such that the variance $\gamma'\Sigma\gamma$ of the transformed variable is maximum over the class of linear combinations that can be formed subject to the constraint that

$$\gamma'\gamma = \sum_{i=1}^{p} \gamma_i^2 = 1 \tag{19}$$

The solution to the maximization problem is found as follows:

Let the Lagrangian be

$$L = \gamma' \Sigma \gamma + \lambda \left(1 - \gamma' \gamma\right) \tag{20}$$

The first-order conditions are:

$$\frac{\partial L}{\partial \gamma} = 2\Sigma\gamma - 2\lambda\gamma = 0 \tag{21}$$

$$\Sigma\gamma - \lambda\gamma = 0$$

$$(\Sigma - \lambda I)\,\gamma = 0 \tag{22}$$

Now, equation (22) looks a lot like equation (3) above. If we want this equation to have a solution other than $\gamma = 0$, the matrix $(\Sigma - \lambda I)$ must be singular. This is, we must have:

$$det\,(\Sigma - \lambda I) = 0 \tag{23}$$

which is precisely the definition of eigenvalue. This is, the solution to this problem requires that the Lagrange multiplier λ be an eigenvalue of the *original* covariance matrix Σ.

More precisely, equation (23) defines the characteristic polynomial of the matrix Σ, which is a polynomial of order p. This polynomial will have p (not necessarily distinct) roots, each of which will be an eigenvalue of the matrix Σ.

For each of these eigenvalues, we will have a different value of the weight vector γ defined by the following equations:

$$(\Sigma - \lambda I)\,\gamma = 0 \tag{24}$$

$$\gamma'\gamma = 1 \tag{25}$$

Thus, eigenvector γ_1 corresponding to the eigenvalue λ_1 is obtained by solving the equations

$$(\Sigma - \lambda_1 I)\,\gamma_1 = 0 \tag{26}$$

$$\gamma_1'\gamma_1 = 1 \tag{27}$$

Premultiplying equation (26) by γ_1' yields:

$$\begin{aligned} \gamma_1' \left(\Sigma - \lambda_1 I \right) \gamma_1 &= 0 \\ \gamma_1' \Sigma \gamma_1 - \lambda_1 \gamma_1' I \gamma_1 &= 0 \\ \gamma_1' \Sigma \gamma_1 - \lambda_1 \gamma_1' \gamma_1 &= 0 \\ \gamma_1' \Sigma \gamma_1 - \lambda_1 \cdot 1 &= 0 \\ \gamma_1' \Sigma \gamma_1 &= \lambda_1 \end{aligned} \tag{28}$$

But $\gamma_1' \Sigma \gamma_1 \equiv E\left(\xi \xi'\right)$, this is, the right-hand side of equation (28) is the variance of the linear transformation of x. Thus, we have:

$$E\left(\xi \xi'\right) = \lambda_1 \tag{29}$$

But since the vector γ_1 was chosen to maximize the variance of ξ, the eigenvalue λ_1 must be the *largest* eigenvalue of Σ.

So, let us summarize all these steps in a definition for the first principal component:

*The **first principal component** of the manifest random variables x is the linear combination $\gamma_{11}x_1 + \gamma_{12}x_2 + \cdots + \gamma_{1p}x_p \equiv \xi_1$ such that the coefficients $\gamma_{11}, \gamma_{12}, \cdots, \gamma_{1p}$ are the elements of the eigenvector γ_1 corresponding to the largest eigenvalue λ_1 of the covariance matrix Σ of the responses x.*

The second principal component, γ_2, given by

$$\xi_2 \equiv \gamma_{21}x_1 + \gamma_{22}x_2 + \cdots + \gamma_{2p}x_p \equiv \gamma_2' x \tag{30}$$

is found by maximizing $\gamma_2' \Sigma \gamma_2$, this is, the variance of ξ_2, subject to the constraints

$$\gamma_2' \gamma_2 = 1 \tag{31}$$

$$\gamma_1' \gamma_2 = 0 \tag{32}$$

Just like in the case of the first principal component, the constraint given by equation (31) is a scaling that ensures the uniqueness of the coefficients. The constraint given by equation (32) requires that γ_1 and γ_2 be orthogonal, which implies that the sum of the variances of the successive components (the first, second, third,..., and pth principal components) is equal to the total variance Σ of the x responses. This is, the orthogonality requirement ensures that Σ be decomposed into p orthogonal components.

The Lagrangian of this problem is

$$L_2 = \gamma_2' \Sigma \gamma_2 + \lambda_2 \left(1 - \gamma_2' \gamma_2\right) + \mu \left(0 - \gamma_1' \gamma_2\right) \tag{33}$$

The first-order conditions are:

$$\begin{aligned} \frac{\partial L_2}{\partial \gamma_2} = 2\Sigma\gamma_2 - 2\lambda_2\gamma_2 - \mu\gamma_1 &= 0 \\ 2\Sigma\gamma_2 - 2\lambda_2\gamma_2 &= \mu\gamma_1 \end{aligned} \tag{34}$$

Premultiplying equation (34) by γ_2' and using the fact that $\gamma_1'\gamma_1 = 1$ and $\gamma_1'\gamma_2 = 0$ (equations (31) and (32)), we have

$$\begin{aligned} \gamma_2'\Sigma\gamma_2 - \lambda_2\gamma_2'\gamma_2 &= \mu\gamma_2'\gamma_1 \\ \gamma_2'\Sigma\gamma_2 &= \lambda_2 \end{aligned} \tag{35}$$

Premultiplying equation (34) by γ_1' and using the fact that $\gamma_1'\gamma_1 = 1$ and $\gamma_1'\gamma_2 = 0$ (equations (31) and (32)), we have

$$\begin{aligned} 2\gamma_1'\Sigma\gamma_2 - 2\lambda_2\gamma_1'\gamma_2 &= \mu\gamma_1'\gamma_1 \\ 2\gamma_1'\Sigma\gamma_2 &= \mu \end{aligned} \tag{36}$$

Now, premultiplying equation (26) by γ_2' yields

$$\begin{aligned}\gamma_2'\Sigma\gamma_1 - \lambda_1\gamma_2'\gamma_1 &= 0 \\ \gamma_2'\Sigma\gamma_1 &= 0 \qquad (37)\end{aligned}$$

Equation (37) together with equation (36) imply that $\mu = 0$, and therefore equation (34) becomes

$$\begin{aligned}2\Sigma\gamma_2 - 2\lambda_2\gamma_2 &= 0 \\ (\Sigma - \lambda_2 I)\,\gamma_2 &= 0 \qquad (38)\end{aligned}$$

which means that the coefficients of the second principal component, $\gamma_2 \equiv \gamma_{21}, \gamma_{22}, \cdots \gamma_{2p}$, are the elements of the eigenvector corresponding to Σ's second largest eigenvalue, λ_2.

So, the variance of the *ith* principal component is Σ's *ith* largest eigenvalue, λ_i, and hence

$$Var\,(\xi_1) + Var\,(\xi_2) + \cdots + Var\,(\xi_p) = \lambda_1 + \lambda_2 + \cdots + \lambda_p = trace\,(\Sigma) \qquad (39)$$

which shows that the total variance of the linearly transformed variables (the ξ's) is equal to the total variance of the x responses.

As you can see, the problem of finding the weights γ reduces to finding the eigenstructure of the covariance matrix Σ. The eigenvectors give the vectors of weights and the eigenvalues give the variance of the transformed variable ξ, which are also referred to as the *principal component scores.*

2.3 The factor analysis model

Factor analysis aims to study whether the correlations between a set of manifest variables can be explained in terms of a small number of latent variables (the "*factors*", which we'll denote by f). The model is the following

$$\begin{aligned} x_1 &= \lambda_{11} f_1 + \lambda_{12} f_2 + \cdots + \lambda_{1k} f_k + u_1 && (40) \\ x_2 &= \lambda_{21} f_1 + \lambda_{22} f_2 + \cdots + \lambda_{2k} f_k + u_2 \\ &\cdots && (41) \\ x_p &= \lambda_{p1} f_1 + \lambda_{p2} f_2 + \cdots + \lambda_{pk} f_k + u_p \end{aligned}$$

Which can be rewritten in matrix form as :

$$x = \Lambda f + u \quad (42)$$

The $p \times k$ matrix Λ is called the matrix of *factor loadings*, $f_1, f_2, \cdots f_k$ are referred to as the *common factors* and $u_1, u_2, \cdots u_p$ represent the combined effect of *specific factors* and random error.

Note that this model models the p manifest variables in terms of $k + p$ hypothetical variables and hence is clearly not testable at the level of the raw data. But the following assumption is made: given the k common factors, the manifest variables are independent of each other (in other words, the residuals u are uncorrelated with each other, which is the conditional independence assumption mentioned above). We also assume that the latent variables f have zero means and unit variances.

Thus, the variance of the manifest variables is given by

$$\Sigma \equiv E\left(xx'\right) \quad (43)$$

Given the assumptions made above about u and f, we also have

$$E\left(fu'\right) = 0 \quad (44)$$

Now, let

$$\Psi \equiv E\left(uu'\right) \quad (45)$$

and

$$\Phi \equiv E\left(ff'\right) \tag{46}$$

This is, Ψ the variance-covariance matrix of the residual terms u, and Φ is the $k \times k$ variance-covariance matrix of f, which given the assumption of unit variance has ones down the main diagonal (the off-diagonals, of course, have the covariance between the factors or latent variables).

Given these definitions, we see that Σ is equal to the following expression:

$$\begin{aligned} \Sigma \equiv E\left(xx'\right) &= E\left(\left(\Lambda f + u\right)\left(\Lambda f + u\right)'\right) \\ &= E\left(\Lambda f f' \Lambda' + \Lambda f u' + u f' \Lambda' + u u'\right) \\ &= \Lambda E\left(ff'\right)\Lambda' + E\left(uu'\right) \end{aligned} \tag{47}$$

where the cross products become zero by equation (44).

So, the general expression for Σ is

$$\Sigma = \Lambda \Phi \Lambda' + \Psi \tag{48}$$

Also note that the correlation between factors and the manifest variables is

$$E\left(xf'\right) = E\left(\left(\Lambda f + u\right) f'\right) = E\left(\Lambda f f' + u f'\right) = \Lambda E\left(ff'\right) = \Lambda \Phi \tag{49}$$

And if we assume that the factors are orthogonal to each other (ie,$\Phi = I$), this simplifies to

$$\Sigma = \Lambda \Lambda' + \Psi \tag{50}$$

so that the variance of the manifest variable x_i is given by

$$\sigma_{ii} = \sum_{j=1}^{k} \lambda_{ij}^2 + \psi_{ii} \tag{51}$$

and the covariance between the manifest variable x_i and the manifest variable x_j is given by

$$\sigma_{ij} = \sum_{r=1}^{k} \lambda_{ir}\lambda_{jr} \tag{52}$$

In equation (51), the term $\sum_{j=1}^{k} \lambda_{ij}^2$ is known as the *communality* and represents the part of the variance of variable x_i which is shared with the other manifest variables via the common factors. The term ψ_{ii} is the variance of the residual term and is known as the *specific* or *unique* variance, and represents the part of the variance of variable x_i which is not shared with the other manifest variables.

In what follows, we will assume that $\Phi = I$, so that Σ is given by equation (50). The factor analysis problem seeks to answer the following question: given the matrix Σ, is it possible to define a unique matrix Ψ with positive diagonal elements for a specified value of k less than p and a unique matrix Λ satisfying equation (50)?

Note that the $p \times k$ matrix $\Sigma - \Psi$ is equal to

$$\Sigma - \Psi = \Lambda\Lambda' \tag{53}$$

This is, $\Sigma - \Psi$ is a covariance matrix, in which each diagonal element represents not the total variance of the corresponding manifest variable, but rather only that part of the manifest variable's variance which is due to the k common factors. This is what we called above the *communality* of the manifest variable.

When $k = 1$, Λ reduces to a $p \times 1$ column vector and it is unique, apart from a possible change of sign in all its elements.

When $k > 1$ the model is not identified because there are an infinity of choices for Λ which lead to the same prediction for Σ. To see this, note that equations (42) and (50) are still satisfied if we replace f by Mf and Λ by $M\Lambda$, where M is a $k \times k$ orthogonal matrix. In the factor analysis lingo, this is referred to as "a rotation of the factors". Bottom line is this: when $k > 1$ we *must* impose restrictions upon the elements of Λ, otherwise the model is not identified.

Let us now rescale each manifest variable so that the variance that is not due to the common factors, Ψ, is unity.

This is achieved by defining:

$$\Sigma^{\star} \equiv \Psi^{-1/2}\Sigma\Psi^{-1/2} \tag{54}$$

Thus, we rescale the $p \times k$ matrix $\Sigma - \Psi$ as follows:

$$\Psi^{-1/2}\left(\Sigma - \Psi\right)\Psi^{-1/2} \equiv \Sigma^{\star} - I \tag{55}$$

Using the theorem of section (1.3) and in particular equation (10), we can decompose the matrix $\Sigma^{\star} - I$ as follows (since this matrix is of rank k):

$$\Sigma^{\star} - I = \Omega\Delta\Omega' \tag{56}$$

where Δ is a $k \times k$ diagonal matrix whose diagonal entries are the k distinct eigenvalues of $\Sigma^{\star} - I$, and Ω is a $p \times k$ matrix whose columns are the eigenvectors corresponding to the eigenvalues of $\Sigma^{\star} - I$, and has the property that $\Omega\Omega' = I$ (ie, $\Omega' = \Omega^{-1}$).

If we assume that the k eigenvalues are all positive and distinct and that they are arranged from largest to smallest, then Ω is uniquely determined and we may define Λ by the following equation:

$$\Lambda = \Psi^{1/2}\Omega\Delta^{1/2} \tag{57}$$

To see why, note the it follows from equation (55) and equation (56) that :

$$\begin{aligned}
\Sigma^{\star} - I \equiv \Psi^{-1/2}\left(\Lambda\Lambda'\right)\Psi^{-1/2} &= \Omega\Delta\Omega' \\
\left(\Psi^{-1/2}\right)^{-1}\Psi^{-1/2}\Lambda\Lambda'\Psi^{-1/2}\left(\Psi^{-1/2}\right)^{-1} &= \left(\Psi^{-1/2}\right)^{-1}\Omega\Delta\Omega'\left(\Psi^{-1/2}\right)^{-1} \\
\Lambda\Lambda' &= \Psi^{1/2}\Omega\Delta\Omega'\Psi^{1/2} \\
\Lambda\Lambda' &= \Psi^{1/2}\Omega\Delta^{1/2}\left(\Delta^{1/2}\right)'\Omega'\Psi^{1/2} \qquad (58) \\
\Lambda\Lambda' &= \Psi^{1/2}\Omega\Delta^{1/2}\Delta^{1/2}\Omega'\Psi^{1/2} \qquad (59)
\end{aligned}$$

where we've used the fact that Δ is symmetric. So we see how equation (57) follows from the definitions and decompositions above.

And, of course, we have as desired

$$
\begin{aligned}
\Lambda\Lambda' &= \Psi^{1/2}\Omega\Delta^{1/2}\Delta^{1/2}\Omega'\Psi^{1/2} \\
&= \Psi^{1/2}\Omega\Delta\Omega'\Psi^{1/2} \\
&= \Psi^{1/2}\left(\Sigma^\star - I\right)\Psi^{1/2} \\
&= (\Sigma - \Psi)
\end{aligned}
$$

In order to interpret the results, note that the matrix $\Sigma^\star$ has the same eigenvectors as $\Sigma^\star - I$ and its eigenvalues are those of $\Sigma^\star - I$ increased by unity. So, what exactly have we done? Well, assuming that there is a unique diagonal matrix Ψ with positive elements such that the k largest eigenvalues of $\Sigma^\star = \Psi^{-1/2}\Sigma\Psi^{-1/2}$ are distinct and greater than unity and the remaining $p-k$ eigenvalues are each unity, then Λ may be defined by equation (57) and the factors are completely identified. This method is in fact a *principal component analysis* of the matrix $\Sigma^\star - I$.

Finally, note that since

$$\Psi^{-1/2}\Lambda = \Omega\Delta^{1/2} \tag{60}$$

we have

$$\Lambda'\Psi^{-1}\Lambda = \Delta^{1/2}\Omega'\Omega\Delta^{1/2} = \Delta^{1/2}I\Delta^{1/2} = \Delta \tag{61}$$

Thus, the constraint that we have imposed on Λ for identification is that $\Lambda'\Psi^{-1}\Lambda$ is a diagonal matrix whose positive and distinct diagonal entries (the eigenvalues of $\Sigma^\star - I$) are arranged in increasing order of magnitude. This identification condition thus amounts to **choosing the factor loadings in such a way that the first factor makes a maximum contribution to the variance of the manifest variables, the second makes a maximum contribution subject to being uncorrelated with the first factor, and so on**.

But, of course, this identification condition may be replaced by any other you want. And also, note that in practice we don't know the matrix Ψ, so we have to estimate it before we can successfully rescale the matrix Σ.

Finally, to see what we did in this section in one shot, note the following. The principal components solution of the factor analysis problem is

$$\xi = Mx \tag{62}$$

where ξ is a $1 \times p$ vector of principal components, and M is a $p \times p$ matrix of weights. The matrix M is such that $MM' = M'M = I$ (i.e., it is orthonormal). We can premultiply equation (62) by M' to obtain:

$$\begin{aligned} M'\xi &= M'Mx \\ &= Ix \end{aligned} \tag{63}$$

thus

$$x = M'\xi \tag{64}$$

Equation (64) means that the manifest variables x can be written as functions of the principal components of the ξ matrix. And, as we've seen in this section, this method of extracting factors assumes that the first k principal components of the ξ matrix represent the k common factors and the remaining $p - k$ principal components are used to determine the unique variance.

2.4 Principal components vs. Factor analysis: what was the difference?

In principal component analysis, we create an auxiliary variable ξ which is a linear combination of the manifest variables x, and then find the linear combination that explains most of the variance of the manifest variables. In factor analysis, on the other hand, we model the manifest variables as a linear combination of some latent factors which we don't observe plus some disturbance. In performing a principal component analysis of the scaled covariance matrix of the manifest variables,

$\Sigma^\star - I$, we are expressing each manifest variable as a linear combination of the same factors, and we are finding those factors.

According to Sharma(1996), “the objective of principal components analysis is to reduce the number of variables to a few components such that each component forms a new variable and the number of retained components explains the maximum amount of variance in the data. The objective of factor analysis, on the other hand, is to search or identify the underlying factor(s) or latent constructs that can explain the intercorrelation among the variables”. So, a major difference between the two techniques is that the goal of principal components is to explain the variance in the data, while the goal of factor analysis is to explain the correlation among the manifest variables. As Lawley and Maxwell (1971) put it, “whereas principal component is variance-oriented, factor analysis is covariance- or correlation-oriented”. [1]

Another important difference between the two methods is this. In a principal components analysis, the components are linear functions of the original manifest variables from which they are derived. Thus, there is really no difficulty in estimating the scores of any individual on the components. On the other hand, in factor analysis the common factors do not fully account for the total variance of the manifest variables, so the problem is more difficult. Since the factors are not linear functions of the manifest variables alone, the scores of an individual cannot be found exactly. They cannot be estimated in the usual statistical sense, and some criterion must be imposed in order to obtain reasonable estimates. Actually, in factor analysis estimation is in a sense a two-stage process: first, parameters in the model are estimated, then these parameters are used to obtain individual factor scores. See Lawley and Maxwell (1971), chapter one, for more details of this discussion.

References

- Sharma, S., 1996. *Applied Multivalued Techniques.* John Wiley and Sons, Inc.

[1] Note that the covariance matrix $E\left(xx'\right)$ is equal to the correlation matrix when the data have been standardized.

- Everitt, B. S., 1984. *An Introduction to Latent Variable Models.* Monographs on Statistics and Applied Probability. Chapman and Hall.
- Lawley, D. N. and A. E. Maxwell, 1971. *Factor Analysis as a Statistical Method.* American Elsevier Publishing Company, Inc.
- Simon, C. P. and L. Blume, 1994. *Mathematics for Economists.* Norton.
- Morrison, D. F, 2005. *Multivariate Statistical Methods.* Duxbury Advanced Series.