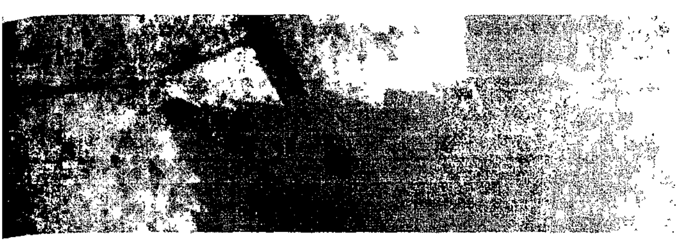
Chapter

9



FACTOR ANALYSIS AND INFERENCE FOR STRUCTURED COVARIANCE MATRICES

9.1 Introduction

Factor analysis has provoked rather turbulent controversy throughout its history. Its modern beginnings lie in the early-20th-century attempts of Karl Pearson, Charles Spearman, and others to define and measure intelligence. Because of this early association with constructs such as intelligence, factor analysis was nurtured and developed primarily by scientists interested in psychometrics. Arguments over the psychological interpretations of several early studies and the lack of powerful computing facilities impeded its initial development as a statistical method. The advent of high-speed computers has generated a renewed interest in the theoretical and computational aspects of factor analysis. Most of the original techniques have been abandoned and early controversies resolved in the wake of recent developments. It is still true, however, that each application of the technique must be examined on its own merits to determine its success.

The essential purpose of factor analysis is to describe, if possible, the covariance relationships among many variables in terms of a few underlying, but unobservable, random quantities called *factors*. Basically, the factor model is motivated by the following argument: Suppose variables can be grouped by their correlations. That is, suppose all variables within a particular group are highly correlated among themselves, but have relatively small correlations with variables in a different group. Then it is conceivable that each group of variables represents a single underlying construct, or factor, that is responsible for the observed correlations. For example, correlations from the group of test scores in classics, French, English, mathematics, and music collected by Spearman suggested an underlying "intelligence" factor. A second group of variables, representing physical-fitness scores, if available, might correspond to another factor. It is this type of structure that factor analysis seeks to confirm.

Factor analysis can be considered an extension of principal component analysis. Both can be viewed as attempts to approximate the covariance matrix $\mathbf{\Sigma}$. However, the approximation based on the factor analysis model is more elaborate. The primary question in factor analysis is whether the data are consistent with a prescribed structure.

9.2 The Orthogonal Factor Model

The observable random vector $\mathbf{X}$, with p components, has mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$. The factor model postulates that $\mathbf{X}$ is linearly dependent upon a few unobservable random variables $F_1, F_2, \ldots, F_m$, called *common factors*, and p additional sources of variation $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p$, called *errors* or, sometimes, *specific factors*.[1] In particular, the factor analysis model is

$$\begin{aligned}
X_1 - \mu_1 &= \ell_{11}F_1 + \ell_{12}F_2 + \cdots + \ell_{1m}F_m + \varepsilon_1 \\
X_2 - \mu_2 &= \ell_{21}F_1 + \ell_{22}F_2 + \cdots + \ell_{2m}F_m + \varepsilon_2 \\
&\;\;\vdots \\
X_p - \mu_p &= \ell_{p1}F_1 + \ell_{p2}F_2 + \cdots + \ell_{pm}F_m + \varepsilon_p
\end{aligned} \tag{9-1}$$

or, in matrix notation,

$$\underset{(p\times 1)}{\mathbf{X} - \boldsymbol{\mu}} = \underset{(p\times m)}{\mathbf{L}}\;\underset{(m\times 1)}{\mathbf{F}} + \underset{(p\times 1)}{\boldsymbol{\varepsilon}} \tag{9-2}$$

The coefficient ℓ_{ij} is called the *loading* of the ith variable on the jth factor, so the matrix $\mathbf{L}$ is the *matrix of factor loadings*. Note that the ith specific factor ε_i is associated only with the ith response X_i. The p deviations $X_1 - \mu_1, X_2 - \mu_2, \ldots, X_p - \mu_p$ are expressed in terms of $p + m$ random variables $F_1, F_2, \ldots, F_m, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p$ which are *unobservable*. This distinguishes the factor model of (9-2) from the multivariate regression model in (7-23), in which the independent variables [whose position is occupied by $\mathbf{F}$ in (9-2)] can be observed.

With so many unobservable quantities, a direct verification of the factor model from observations on $X_1, X_2, \ldots, X_p$ is hopeless. However, with some additional assumptions about the random vectors $\mathbf{F}$ and $\boldsymbol{\varepsilon}$, the model in (9-2) implies certain covariance relationships, which can be checked.

We assume that

$$E(\mathbf{F}) = \underset{(m\times 1)}{\mathbf{0}}, \qquad \text{Cov}(\mathbf{F}) = E[\mathbf{F}\mathbf{F}'] = \underset{(m\times m)}{\mathbf{I}}$$

$$E(\boldsymbol{\varepsilon}) = \underset{(p\times 1)}{\mathbf{0}}, \qquad \text{Cov}(\boldsymbol{\varepsilon}) = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \underset{(p\times p)}{\mathbf{\Psi}} = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix} \tag{9-3}$$

[1] As Maxwell [12] points out, in many investigations the ε_i tend to be combinations of measurement error and factors that are uniquely associated with the individual variables.

and that $\mathbf{F}$ and $\boldsymbol{\varepsilon}$ are independent, so

$$\operatorname{Cov}(\boldsymbol{\varepsilon}, \mathbf{F}) = E(\boldsymbol{\varepsilon}\mathbf{F}') = \underset{(p\times m)}{\mathbf{0}}$$

These assumptions and the relation in (9-2) constitute the *orthogonal factor model.*[2]

Orthogonal Factor Model with m Common Factors

$$\underset{(p\times 1)}{\mathbf{X}} = \underset{(p\times 1)}{\boldsymbol{\mu}} + \underset{(p\times m)}{\mathbf{L}}\underset{(m\times 1)}{\mathbf{F}} + \underset{(p\times 1)}{\boldsymbol{\varepsilon}}$$

$$\begin{aligned} \mu_i &= \textit{mean} \text{ of variable } i \\ \varepsilon_i &= i\text{th } \textit{specific factor} \\ F_j &= j\text{th } \textit{common factor} \\ \ell_{ij} &= \textit{loading} \text{ of the } i\text{th variable on the } j\text{th factor} \end{aligned} \tag{9-4}$$

The unobservable random vectors $\mathbf{F}$ and $\boldsymbol{\varepsilon}$ satisfy the following conditions:

$\mathbf{F}$ and $\boldsymbol{\varepsilon}$ are independent

$E(\mathbf{F}) = \mathbf{0}, \operatorname{Cov}(\mathbf{F}) = \mathbf{I}$

$E(\boldsymbol{\varepsilon}) = \mathbf{0}$, $\operatorname{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}$, where $\boldsymbol{\Psi}$ is a diagonal matrix

The orthogonal factor model implies a covariance structure for $\mathbf{X}$. From the model in (9-4),

$$\begin{aligned} (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' &= (\mathbf{LF} + \boldsymbol{\varepsilon})(\mathbf{LF} + \boldsymbol{\varepsilon})' \\ &= (\mathbf{LF} + \boldsymbol{\varepsilon})((\mathbf{LF})' + \boldsymbol{\varepsilon}') \\ &= \mathbf{LF}(\mathbf{LF})' + \boldsymbol{\varepsilon}(\mathbf{LF})' + \mathbf{LF}\boldsymbol{\varepsilon}' + \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \end{aligned}$$

so that

$$\begin{aligned} \boldsymbol{\Sigma} &= \operatorname{Cov}(\mathbf{X}) = E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' \\ &= \mathbf{L}E(\mathbf{FF}')\mathbf{L}' + E(\boldsymbol{\varepsilon}\mathbf{F}')\mathbf{L}' + \mathbf{L}E(\mathbf{F}\boldsymbol{\varepsilon}') + E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') \\ &= \mathbf{LL}' + \boldsymbol{\Psi} \end{aligned}$$

according to (9-3). Also by independence, $\operatorname{Cov}(\boldsymbol{\varepsilon}, \mathbf{F}) = E(\boldsymbol{\varepsilon}, \mathbf{F}') = \mathbf{0}$

Also, by the model in (9-4), $(\mathbf{X} - \boldsymbol{\mu})\mathbf{F}' = (\mathbf{LF} + \boldsymbol{\varepsilon})\mathbf{F}' = \mathbf{LFF}' + \boldsymbol{\varepsilon}\mathbf{F}'$. $\operatorname{Cov}(\mathbf{X}, \mathbf{F}) = E(\mathbf{X} - \boldsymbol{\mu})\mathbf{F}' = \mathbf{L}E(\mathbf{FF}') + E(\boldsymbol{\varepsilon}\mathbf{F}') = \mathbf{L}$.

[2] Allowing the factors $\mathbf{F}$ to be correlated so that $\operatorname{Cov}(\mathbf{F})$ is *not* diagonal gives the oblique factor model. The oblique model presents some additional estimation difficulties and will not be discussed in this book. (See [20].)

Covariance Structure for the Orthogonal Factor Model

1. $\text{Cov}(\mathbf{X}) = \mathbf{LL}' + \boldsymbol{\Psi}$

   or

$$\begin{aligned} \text{Var}(X_i) &= \ell_{i1}^2 + \cdots + \ell_{im}^2 + \psi_i \\ \text{Cov}(X_i, X_k) &= \ell_{i1}\ell_{k1} + \cdots + \ell_{im}\ell_{km} \end{aligned} \tag{9-5}$$

2. $\text{Cov}(\mathbf{X}, \mathbf{F}) = \mathbf{L}$

   or

$$\text{Cov}(X_i, F_j) = \ell_{ij}$$

The model $\mathbf{X} - \boldsymbol{\mu} = \mathbf{LF} + \boldsymbol{\varepsilon}$ is *linear* in the common factors. If the p responses $\mathbf{X}$ are, in fact, related to underlying factors, but the relationship is nonlinear, such as in $X_1 - \mu_1 = \ell_{11}F_1F_3 + \varepsilon_1$, $X_2 - \mu_2 = \ell_{21}F_2F_3 + \varepsilon_2$, and so forth, then the covariance structure $\mathbf{LL}' + \boldsymbol{\Psi}$ given by (9-5) may not be adequate. The very important assumption of linearity is inherent in the formulation of the traditional factor model.

That portion of the variance of the ith variable contributed by the m common factors is called the ith *communality*. That portion of $\text{Var}(X_i) = \sigma_{ii}$ due to the specific factor is often called the *uniqueness*, or *specific variance*. Denoting the ith communality by h_i^2, we see from (9-5) that

$$\underbrace{\sigma_{ii}}_{\text{Var}(X_i)} = \underbrace{\ell_{i1}^2 + \ell_{i2}^2 + \cdots + \ell_{im}^2}_{\text{communality}} + \underbrace{\psi_i}_{\text{specific variance}}$$

or

$$h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 + \cdots + \ell_{im}^2 \tag{9-6}$$

and

$$\sigma_{ii} = h_i^2 + \psi_i, \qquad i = 1, 2, \ldots, p$$

The ith communality is the sum of squares of the loadings of the ith variable on the m common factors.

Example 9.1 (Verifying the relation $\boldsymbol{\Sigma} = \mathbf{LL}' + \boldsymbol{\Psi}$ for two factors) Consider the covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} 19 & 30 & 2 & 12 \\ 30 & 57 & 5 & 23 \\ 2 & 5 & 38 & 47 \\ 12 & 23 & 47 & 68 \end{bmatrix}$$

The equality

$$\begin{bmatrix} 19 & 30 & 2 & 12 \\ 30 & 57 & 5 & 23 \\ 2 & 5 & 38 & 47 \\ 12 & 23 & 47 & 68 \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 7 & 2 \\ -1 & 6 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} 4 & 7 & -1 & 1 \\ 1 & 2 & 6 & 8 \end{bmatrix} + \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

or

$$\mathbf{\Sigma} = \mathbf{L}\mathbf{L}' + \mathbf{\Psi}$$

may be verified by matrix algebra. Therefore, $\mathbf{\Sigma}$ has the structure produced by an $m = 2$ orthogonal factor model. Since

$$\mathbf{L} = \begin{bmatrix} \ell_{11} & \ell_{12} \\ \ell_{21} & \ell_{22} \\ \ell_{31} & \ell_{32} \\ \ell_{41} & \ell_{42} \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 7 & 2 \\ -1 & 6 \\ 1 & 8 \end{bmatrix},$$

$$\mathbf{\Psi} = \begin{bmatrix} \psi_1 & 0 & 0 & 0 \\ 0 & \psi_2 & 0 & 0 \\ 0 & 0 & \psi_3 & 0 \\ 0 & 0 & 0 & \psi_4 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

the communality of X_1 is, from (9-6),

$$h_1^2 = \ell_{11}^2 + \ell_{12}^2 = 4^2 + 1^2 = 17$$

and the variance of X_1 can be decomposed as

$$\sigma_{11} = (\ell_{11}^2 + \ell_{12}^2) + \psi_1 = h_1^2 + \psi_1$$

or

$$\underbrace{19}_{\text{variance}} = \underbrace{4^2 + 1^2}_{\text{communality}} + \underbrace{2}_{\substack{\text{specific}\\\text{variance}}} = 17 + 2$$

A similar breakdown occurs for the other variables. ■

The factor model assumes that the $p + p(p-1)/2 = p(p+1)/2$ variances and covariances for $\mathbf{X}$ can be reproduced from the pm factor loadings ℓ_{ij} and the p specific variances ψ_i. When $m = p$, any covariance matrix $\mathbf{\Sigma}$ can be reproduced exactly as $\mathbf{LL}'$ [see (9-11)], so $\mathbf{\Psi}$ can be the zero matrix. However, it is when m is small relative to p that factor analysis is most useful. In this case, the factor model provides a "simple" explanation of the covariation in $\mathbf{X}$ with fewer parameters than the $p(p+1)/2$ parameters in $\mathbf{\Sigma}$. For example, if $\mathbf{X}$ contains $p = 12$ variables, and the factor model in (9-4) with $m = 2$ is appropriate, then the $p(p+1)/2 = 12(13)/2 = 78$ elements of $\mathbf{\Sigma}$ are described in terms of the $mp + p = 12(2) + 12 = 36$ parameters ℓ_{ij} and ψ_i of the factor model.

Unfortunately for the factor analyst, most covariance matrices cannot be factored as $\mathbf{LL}' + \boldsymbol{\Psi}$, where the number of factors m is much less than p. The following example demonstrates one of the problems that can arise when attempting to determine the parameters ℓ_{ij} and ψ_i from the variances and covariances of the observable variables.

Example 9.2 (Nonexistence of a proper solution) Let $p = 3$ and $m = 1$, and suppose the random variables X_1, X_2, and X_3 have the positive definite covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & .9 & .7 \\ .9 & 1 & .4 \\ .7 & .4 & 1 \end{bmatrix}$$

Using the factor model in (9-4), we obtain

$$\begin{aligned} X_1 - \mu_1 &= \ell_{11}F_1 + \varepsilon_1 \\ X_2 - \mu_2 &= \ell_{21}F_1 + \varepsilon_2 \\ X_3 - \mu_3 &= \ell_{31}F_1 + \varepsilon_3 \end{aligned}$$

The covariance structure in (9-5) implies that

$$\boldsymbol{\Sigma} = \mathbf{LL}' + \boldsymbol{\Psi}$$

or

$$\begin{array}{lll} 1 = \ell_{11}^2 + \psi_1 & .90 = \ell_{11}\ell_{21} & .70 = \ell_{11}\ell_{31} \\ & 1 = \ell_{21}^2 + \psi_2 & .40 = \ell_{21}\ell_{31} \\ & & 1 = \ell_{31}^2 + \psi_3 \end{array}$$

The pair of equations

$$\begin{aligned} .70 &= \ell_{11}\ell_{31} \\ .40 &= \ell_{21}\ell_{31} \end{aligned}$$

implies that

$$\ell_{21} = \left(\frac{.40}{.70}\right)\ell_{11}$$

Substituting this result for ℓ_{21} in the equation

$$.90 = \ell_{11}\ell_{21}$$

yields $\ell_{11}^2 = 1.575$, or $\ell_{11} = \pm 1.255$. Since $\text{Var}(F_1) = 1$ (by assumption) and $\text{Var}(X_1) = 1$, $\ell_{11} = \text{Cov}(X_1, F_1) = \text{Corr}(X_1, F_1)$. Now, a correlation coefficient cannot be greater than unity (in absolute value), so, from this point of view, $|\ell_{11}| = 1.255$ is too large. Also, the equation

$$1 = \ell_{11}^2 + \psi_1, \quad \text{or} \quad \psi_1 = 1 - \ell_{11}^2$$

gives

$$\psi_1 = 1 - 1.575 = -.575$$

which is unsatisfactory, since it gives a negative value for $\text{Var}(\varepsilon_1) = \psi_1$.

Thus, for this example with $m = 1$, it is possible to get a unique numerical solution to the equations $\mathbf{\Sigma} = \mathbf{LL}' + \mathbf{\Psi}$. However, the solution is not consistent with the statistical interpretation of the coefficients, so it is not a proper solution. ■

When $m > 1$, there is always some inherent ambiguity associated with the factor model. To see this, let $\mathbf{T}$ be any $m \times m$ orthogonal matrix, so that $\mathbf{TT}' = \mathbf{T}'\mathbf{T} = \mathbf{I}$. Then the expression in (9-2) can be written

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{LF} + \boldsymbol{\varepsilon} = \mathbf{LTT}'\mathbf{F} + \boldsymbol{\varepsilon} = \mathbf{L}^*\mathbf{F}^* + \boldsymbol{\varepsilon} \tag{9-7}$$

where

$$\mathbf{L}^* = \mathbf{LT} \quad \text{and} \quad \mathbf{F}^* = \mathbf{T}'\mathbf{F}$$

Since

$$E(\mathbf{F}^*) = \mathbf{T}'E(\mathbf{F}) = \mathbf{0}$$

and

$$\text{Cov}(\mathbf{F}^*) = \mathbf{T}'\text{Cov}(\mathbf{F})\mathbf{T} = \mathbf{T}'\mathbf{T} = \underset{(m\times m)}{\mathbf{I}}$$

it is impossible, on the basis of observations on $\mathbf{X}$, to distinguish the loadings $\mathbf{L}$ from the loadings $\mathbf{L}^*$. That is, the factors $\mathbf{F}$ and $\mathbf{F}^* = \mathbf{T}'\mathbf{F}$ have the same statistical properties, and even though the loadings $\mathbf{L}^*$ are, in general, different from the loadings $\mathbf{L}$, they both generate the same covariance matrix $\mathbf{\Sigma}$. That is,

$$\mathbf{\Sigma} = \mathbf{LL}' + \mathbf{\Psi} = \mathbf{LTT}'\mathbf{L}' + \mathbf{\Psi} = (\mathbf{L}^*)(\mathbf{L}^*)' + \mathbf{\Psi} \tag{9-8}$$

This ambiguity provides the rationale for "factor rotation," since orthogonal matrices correspond to rotations (and reflections) of the coordinate system for $\mathbf{X}$.

Factor loadings $\mathbf{L}$ are determined only up to an orthogonal matrix $\mathbf{T}$. Thus, the loadings

$$\mathbf{L}^* = \mathbf{LT} \quad \text{and} \quad \mathbf{L} \tag{9-9}$$

both give the same representation. The communalities, given by the diagonal elements of $\mathbf{LL}' = (\mathbf{L}^*)(\mathbf{L}^*)'$ are also unaffected by the choice of $\mathbf{T}$.

The analysis of the factor model proceeds by imposing conditions that allow one to uniquely estimate $\mathbf{L}$ and $\mathbf{\Psi}$. The loading matrix is then rotated (multiplied by an orthogonal matrix), where the rotation is determined by some "ease-of-interpretation" criterion. Once the loadings and specific variances are obtained, factors are identified, and estimated values for the factors themselves (called *factor scores*) are frequently constructed.

9.3 Methods of Estimation

Given observations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ on p generally correlated variables, factor analysis seeks to answer the question, Does the factor model of (9-4), with a small number of factors, adequately represent the data? In essence, we tackle this statistical model-building problem by trying to verify the covariance relationship in (9-5).

The sample covariance matrix $\mathbf{S}$ is an estimator of the unknown population covariance matrix $\boldsymbol{\Sigma}$. If the off-diagonal elements of $\mathbf{S}$ are small or those of the sample correlation matrix $\mathbf{R}$ essentially zero, the variables are not related, and a factor analysis will not prove useful. In these circumstances, the *specific* factors play the dominant role, whereas the major aim of factor analysis is to determine a few important common factors.

If $\boldsymbol{\Sigma}$ appears to deviate significantly from a diagonal matrix, then a factor model can be entertained, and the initial problem is one of estimating the factor loadings ℓ_{ij} and specific variances ψ_i. We shall consider two of the most popular methods of parameter estimation, the *principal component* (and the related *principal factor*) *method* and the *maximum likelihood method*. The solution from either method can be rotated in order to simplify the interpretation of factors, as described in Section 9.4. It is always prudent to try more than one method of solution; if the factor model is appropriate for the problem at hand, the solutions should be consistent with one another.

Current estimation and rotation methods require iterative calculations that must be done on a computer. Several computer programs are now available for this purpose.

The Principal Component (and Principal Factor) Method

The spectral decomposition of (2-16) provides us with one factoring of the covariance matrix $\boldsymbol{\Sigma}$. Let $\boldsymbol{\Sigma}$ have eigenvalue–eigenvector pairs $(\lambda_i, \mathbf{e}_i)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. Then

$$\begin{aligned}\boldsymbol{\Sigma} &= \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2' + \cdots + \lambda_p \mathbf{e}_p \mathbf{e}_p' \\ &= \left[\sqrt{\lambda_1}\, \mathbf{e}_1 \;\vdots\; \sqrt{\lambda_2}\, \mathbf{e}_2 \;\vdots\; \cdots \;\vdots\; \sqrt{\lambda_p}\, \mathbf{e}_p \right] \begin{bmatrix} \sqrt{\lambda_1}\, \mathbf{e}_1' \\ \hdashline \sqrt{\lambda_2}\, \mathbf{e}_2' \\ \hdashline \vdots \\ \hdashline \sqrt{\lambda_p}\, \mathbf{e}_p' \end{bmatrix}\end{aligned} \tag{9-10}$$

This fits the prescribed covariance structure for the factor analysis model having as many factors as variables $(m = p)$ and specific variances $\psi_i = 0$ for all i. The loading matrix has jth column given by $\sqrt{\lambda_j}\, \mathbf{e}_j$. That is, we can write

$$\underset{(p\times p)}{\boldsymbol{\Sigma}} = \underset{(p\times p)}{\mathbf{L}}\;\underset{(p\times p)}{\mathbf{L}'} + \underset{(p\times p)}{\mathbf{0}} = \mathbf{L}\mathbf{L}' \tag{9-11}$$

Apart from the scale factor $\sqrt{\lambda_j}$, the factor loadings on the jth factor are the coefficients for the jth principal component of the population.

Although the factor analysis representation of $\boldsymbol{\Sigma}$ in (9-11) is exact, it is not particularly useful: It employs as many common factors as there are variables and does not allow for any variation in the specific factors $\boldsymbol{\varepsilon}$ in (9-4). We prefer models that explain the covariance structure in terms of just a few common factors. One approach, when the last $p - m$ eigenvalues are small, is to neglect the contribution of $\lambda_{m+1}\mathbf{e}_{m+1}\mathbf{e}'_{m+1} + \cdots + \lambda_p\mathbf{e}_p\mathbf{e}'_p$ to $\boldsymbol{\Sigma}$ in (9-10). Neglecting this contribution, we obtain the approximation

$$\boldsymbol{\Sigma} \doteq \left[\sqrt{\lambda_1}\,\mathbf{e}_1 \;\vdots\; \sqrt{\lambda_2}\,\mathbf{e}_2 \;\vdots\; \cdots \;\vdots\; \sqrt{\lambda_m}\,\mathbf{e}_m\right] \begin{bmatrix} \sqrt{\lambda_1}\,\mathbf{e}'_1 \\ \cdots\cdots \\ \sqrt{\lambda_2}\,\mathbf{e}'_2 \\ \cdots\cdots \\ \vdots \\ \cdots\cdots \\ \sqrt{\lambda_m}\,\mathbf{e}'_m \end{bmatrix} = \underset{(p\times m)}{\mathbf{L}}\ \underset{(m\times p)}{\mathbf{L}'} \tag{9-12}$$

The approximate representation in (9-12) assumes that the specific factors $\boldsymbol{\varepsilon}$ in (9-4) are of minor importance and can also be ignored in the factoring of $\boldsymbol{\Sigma}$. If specific factors are included in the model, their variances may be taken to be the diagonal elements of $\boldsymbol{\Sigma} - \mathbf{LL}'$, where $\mathbf{LL}'$ is as defined in (9-12).

Allowing for specific factors, we find that the approximation becomes

$$\begin{aligned} \boldsymbol{\Sigma} &\doteq \mathbf{LL}' + \boldsymbol{\Psi} \\ &= \left[\sqrt{\lambda_1}\,\mathbf{e}_1 \;\vdots\; \sqrt{\lambda_2}\,\mathbf{e}_2 \;\vdots\; \cdots \;\vdots\; \sqrt{\lambda_m}\,\mathbf{e}_m\right] \begin{bmatrix} \sqrt{\lambda_1}\,\mathbf{e}'_1 \\ \cdots\cdots \\ \sqrt{\lambda_2}\,\mathbf{e}'_2 \\ \cdots\cdots \\ \vdots \\ \cdots\cdots \\ \sqrt{\lambda_m}\,\mathbf{e}'_m \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix} \end{aligned} \tag{9-13}$$

where $\psi_i = \sigma_{ii} - \sum_{j=1}^{m} \ell_{ij}^2$ for $i = 1, 2, \ldots, p$.

To apply this approach to a data set $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, it is customary first to center the observations by subtracting the sample mean $\bar{\mathbf{x}}$. The centered observations

$$\mathbf{x}_j - \bar{\mathbf{x}} = \begin{bmatrix} x_{j1} \\ x_{j2} \\ \vdots \\ x_{jp} \end{bmatrix} - \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix} = \begin{bmatrix} x_{j1} - \bar{x}_1 \\ x_{j2} - \bar{x}_2 \\ \vdots \\ x_{jp} - \bar{x}_p \end{bmatrix} \qquad j = 1, 2, \ldots, n \tag{9-14}$$

have the same sample covariance matrix $\mathbf{S}$ as the original observations.

In cases in which the units of the variables are not commensurate, it is usually desirable to work with the standardized variables

$$\mathbf{z}_j = \begin{bmatrix} \dfrac{(x_{j1} - \bar{x}_1)}{\sqrt{s_{11}}} \\ \dfrac{(x_{j2} - \bar{x}_2)}{\sqrt{s_{22}}} \\ \vdots \\ \dfrac{(x_{jp} - \bar{x}_p)}{\sqrt{s_{pp}}} \end{bmatrix} \qquad j = 1, 2, \ldots, n$$

whose sample covariance matrix is the sample correlation matrix $\mathbf{R}$ of the observations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. Standardization avoids the problems of having one variable with large variance unduly influencing the determination of factor loadings.

The representation in (9-13), when applied to the sample covariance matrix $\mathbf{S}$ or the sample correlation matrix $\mathbf{R}$, is known as the *principal component solution*. The name follows from the fact that the factor loadings are the scaled coefficients of the first few sample principal components. (See Chapter 8.)

Principal Component Solution of the Factor Model

The principal component factor analysis of the sample covariance matrix $\mathbf{S}$ is specified in terms of its eigenvalue-eigenvector pairs $(\hat{\lambda}_1, \hat{\mathbf{e}}_1)$, $(\hat{\lambda}_2, \hat{\mathbf{e}}_2), \ldots, (\hat{\lambda}_p, \hat{\mathbf{e}}_p)$, where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p$. Let $m < p$ be the number of common factors. Then the matrix of estimated factor loadings $\{\tilde{\ell}_{ij}\}$ is given by

$$\tilde{\mathbf{L}} = \left[\sqrt{\hat{\lambda}_1}\,\hat{\mathbf{e}}_1 \mid \sqrt{\hat{\lambda}_2}\,\hat{\mathbf{e}}_2 \mid \cdots \mid \sqrt{\hat{\lambda}_m}\,\hat{\mathbf{e}}_m\right] \tag{9-15}$$

The estimated specific variances are provided by the diagonal elements of the matrix $\mathbf{S} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}'$, so

$$\tilde{\boldsymbol{\Psi}} = \begin{bmatrix} \tilde{\psi}_1 & 0 & \cdots & 0 \\ 0 & \tilde{\psi}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tilde{\psi}_p \end{bmatrix} \quad \text{with} \quad \tilde{\psi}_i = s_{ii} - \sum_{j=1}^{m} \tilde{\ell}_{ij}^2 \tag{9-16}$$

Communalities are estimated as

$$\tilde{h}_i^2 = \tilde{\ell}_{i1}^2 + \tilde{\ell}_{i2}^2 + \cdots + \tilde{\ell}_{im}^2 \tag{9-17}$$

The principal component factor analysis of the sample correlation matrix is obtained by starting with $\mathbf{R}$ in place of $\mathbf{S}$.

For the principal component solution, the estimated loadings for a given factor do not change as the number of factors is increased. For example, if $m = 1$, $\tilde{\mathbf{L}} = \left[\sqrt{\hat{\lambda}_1}\,\hat{\mathbf{e}}_1\right]$, and if $m = 2$, $\tilde{\mathbf{L}} = \left[\sqrt{\hat{\lambda}_1}\,\hat{\mathbf{e}}_1 \mid \sqrt{\hat{\lambda}_2}\,\hat{\mathbf{e}}_2\right]$, where $(\hat{\lambda}_1, \hat{\mathbf{e}}_1)$ and $(\hat{\lambda}_2, \hat{\mathbf{e}}_2)$ are the first two eigenvalue-eigenvector pairs for $\mathbf{S}$ (or $\mathbf{R}$).

By the definition of $\tilde{\psi}_i$, the diagonal elements of $\mathbf{S}$ are equal to the diagonal elements of $\tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \tilde{\boldsymbol{\Psi}}$. However, the off-diagonal elements of $\mathbf{S}$ are not usually reproduced by $\tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \tilde{\boldsymbol{\Psi}}$. How, then, do we select the number of factors m?

If the number of common factors is not determined by a priori considerations, such as by theory or the work of other researchers, the choice of m can be based on the estimated eigenvalues in much the same manner as with principal components. Consider the *residual matrix*

$$\mathbf{S} - (\tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \tilde{\boldsymbol{\Psi}}) \tag{9-18}$$

resulting from the approximation of $\mathbf{S}$ by the principal component solution. The diagonal elements are zero, and if the other elements are also small, we may subjectively take the m factor model to be appropriate. Analytically, we have (see Exercise 9.5)

$$\text{Sum of squared entries of } (\mathbf{S} - (\tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \tilde{\boldsymbol{\Psi}})) \leq \hat{\lambda}_{m+1}^2 + \cdots + \hat{\lambda}_p^2 \tag{9-19}$$

Consequently, a small value for the sum of the squares of the neglected eigenvalues implies a small value for the sum of the squared errors of approximation.

Ideally, the contributions of the first few factors to the sample variances of the variables should be large. The contribution to the sample variance s_{ii} from the first common factor is $\tilde{\ell}_{i1}^2$. The contribution to the *total* sample variance, $s_{11} + s_{22} + \cdots + s_{pp} = \text{tr}(\mathbf{S})$, from the first common factor is then

$$\tilde{\ell}_{11}^2 + \tilde{\ell}_{21}^2 + \cdots + \tilde{\ell}_{p1}^2 = \left(\sqrt{\hat{\lambda}_1}\,\hat{\mathbf{e}}_1\right)'\left(\sqrt{\hat{\lambda}_1}\,\hat{\mathbf{e}}_1\right) = \hat{\lambda}_1$$

since the eigenvector $\hat{\mathbf{e}}_1$ has unit length. In general,

$$\begin{pmatrix}\text{Proportion of total}\\ \text{sample variance}\\ \text{due to } j\text{th factor}\end{pmatrix} = \begin{cases} \dfrac{\hat{\lambda}_j}{s_{11} + s_{22} + \cdots + s_{pp}} & \text{for a factor analysis of } \mathbf{S} \\[2ex] \dfrac{\hat{\lambda}_j}{p} & \text{for a factor analysis of } \mathbf{R} \end{cases} \tag{9-20}$$

Criterion (9-20) is frequently used as a heuristic device for determining the appropriate number of common factors. The number of common factors retained in the model is increased until a "suitable proportion" of the total sample variance has been explained.

Another convention, frequently encountered in packaged computer programs, is to set m equal to the number of eigenvalues of $\mathbf{R}$ greater than one if the sample correlation matrix is factored, or equal to the number of positive eigenvalues of $\mathbf{S}$ if the sample covariance matrix is factored. These rules of thumb should not be applied indiscriminately. For example, $m = p$ if the rule for $\mathbf{S}$ is obeyed, since all the eigenvalues are expected to be positive for large sample sizes. The best approach is to retain few rather than many factors, assuming that they provide a satisfactory interpretation of the data and yield a satisfactory fit to $\mathbf{S}$ or $\mathbf{R}$.

Example 9.3 (Factor analysis of consumer-preference data) In a consumer-preference study, a random sample of customers were asked to rate several attributes of a new product. The responses, on a 7-point semantic differential scale, were tabulated and the attribute correlation matrix constructed. The correlation matrix is presented next:

$$\begin{array}{lc} \textit{Attribute (Variable)} & \\ \text{Taste} & 1 \\ \text{Good buy for money} & 2 \\ \text{Flavor} & 3 \\ \text{Suitable for snack} & 4 \\ \text{Provides lots of energy} & 5 \end{array}
\begin{array}{c} \begin{array}{ccccc} \;\;1 & \;\;\;2 & \;\;\;3 & \;\;\;4 & \;\;\;5 \end{array} \\ \begin{bmatrix} 1.00 & .02 & .96 & .42 & .01 \\ .02 & 1.00 & .13 & .71 & .85 \\ .96 & .13 & 1.00 & .50 & .11 \\ .42 & .71 & .50 & 1.00 & .79 \\ .01 & .85 & .11 & .79 & 1.00 \end{bmatrix} \end{array}$$

It is clear from the circled entries in the correlation matrix that variables 1 and 3 and variables 2 and 5 form groups. Variable 4 is "closer" to the (2, 5) group than the (1, 3) group. Given these results and the small number of variables, we might expect that the apparent linear relationships between the variables can be explained in terms of, at most, two or three common factors.

The first two eigenvalues, $\hat{\lambda}_1 = 2.85$ and $\hat{\lambda}_2 = 1.81$, of $\mathbf{R}$ are the only eigenvalues greater than unity. Moreover, $m = 2$ common factors will account for a cumulative proportion

$$\frac{\hat{\lambda}_1 + \hat{\lambda}_2}{p} = \frac{2.85 + 1.81}{5} = .93$$

of the total (standardized) sample variance. The estimated factor loadings, communalities, and specific variances, obtained using (9-15), (9-16), and (9-17), are given in Table 9.1.

Table 9.1

Variable	Estimated factor loadings $\tilde{\ell}_{ij} = \sqrt{\hat{\lambda}_i}\,\hat{e}_{ij}$ F_1	F_2	Communalities $\tilde{h}_i^2$	Specific variances $\tilde{\psi}_i = 1 - \tilde{h}_i^2$
1. Taste	.56	.82	.98	.02
2. Good buy for money	.78	−.53	.88	.12
3. Flavor	.65	.75	.98	.02
4. Suitable for snack	.94	−.10	.89	.11
5. Provides lots of energy	.80	−.54	.93	.07
Eigenvalues	2.85	1.81		
Cumulative proportion of total (standardized) sample variance	.571	.932		

Now,

$$\tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \tilde{\mathbf{\Psi}} = \begin{bmatrix} .56 & .82 \\ .78 & -.53 \\ .65 & .75 \\ .94 & -.10 \\ .80 & -.54 \end{bmatrix} \begin{bmatrix} .56 & .78 & .65 & .94 & .80 \\ .82 & -.53 & .75 & -.10 & -.54 \end{bmatrix}$$

$$+ \begin{bmatrix} .02 & 0 & 0 & 0 & 0 \\ 0 & .12 & 0 & 0 & 0 \\ 0 & 0 & .02 & 0 & 0 \\ 0 & 0 & 0 & .11 & 0 \\ 0 & 0 & 0 & 0 & .07 \end{bmatrix} = \begin{bmatrix} 1.00 & .01 & .97 & .44 & .00 \\ & 1.00 & .11 & .79 & .91 \\ & & 1.00 & .53 & .11 \\ & & & 1.00 & .81 \\ & & & & 1.00 \end{bmatrix}$$

nearly reproduces the correlation matrix **R**. Thus, on a purely descriptive basis, we would judge a two-factor model with the factor loadings displayed in Table 9.1 as providing a good fit to the data. The communalities (.98, .88, .98, .89, .93) indicate that the two factors account for a large percentage of the sample variance of each variable.

We shall not interpret the factors at this point. As we noted in Section 9.2, the factors (and loadings) are unique up to an orthogonal rotation. A rotation of the factors often reveals a simple structure and aids interpretation. We shall consider this example again (see Example 9.9 and Panel 9.1) after factor rotation has been discussed. ■

Example 9.4 (Factor analysis of stock-price data) Stock-price data consisting of $n = 103$ weekly rates of return on $p = 5$ stocks were introduced in Example 8.5. In that example, the first two sample principal components were obtained from **R**. Taking $m = 1$ and $m = 2$, we can easily obtain principal component solutions to the orthogonal factor model. Specifically, the estimated factor loadings are the sample principal component coefficients (eigenvectors of **R**), scaled by the square root of the corresponding eigenvalues. The estimated factor loadings, communalities, specific variances, and proportion of total (standardized) sample variance explained by each factor for the $m = 1$ and $m = 2$ factor solutions are available in Table 9.2. The communalities are given by (9-17). So, for example, with $m = 2$, $\tilde{h}_1^2 = \tilde{\ell}_{11}^2 + \tilde{\ell}_{12}^2 = (.732)^2 + (-.437)^2 = .73$.

Table 9.2

	One-factor solution		Two-factor solution		
	Estimated factor loadings	Specific variances	Estimated factor loadings		Specific variances
Variable	F_1	$\tilde{\psi}_i = 1 - \tilde{h}_i^2$	F_1	F_2	$\tilde{\psi}_i = 1 - \tilde{h}_i^2$
1. J P Morgan	.732	.46	.732	−.437	.27
2. Citibank	.831	.31	.831	−.280	.23
3. Wells Fargo	.726	.47	.726	−.374	.33
4. Royal Dutch Shell	.605	.63	.605	.694	.15
5. ExxonMobil	.563	.68	.563	.719	.17
Cumulative proportion of total (standardized) sample variance explained	.487		.487	.769	

The residual matrix corresponding to the solution for $m = 2$ factors is

$$\mathbf{R} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}' - \tilde{\boldsymbol{\Psi}} = \begin{bmatrix} 0 & -.099 & -.185 & -.025 & .056 \\ -.099 & 0 & -.134 & .014 & -.054 \\ -.185 & -.134 & 0 & .003 & .006 \\ -.025 & .014 & .003 & 0 & -.156 \\ .056 & -.054 & .006 & -.156 & 0 \end{bmatrix}$$

The proportion of the total variance explained by the two-factor solution is appreciably larger than that for the one-factor solution. However, for $m = 2$, $\tilde{\mathbf{L}}\tilde{\mathbf{L}}'$ produces numbers that are, in general, larger than the sample correlations. This is particularly true for r_{13}.

It seems fairly clear that the first factor, F_1, represents general economic conditions and might be called a *market factor*. All of the stocks load highly on this factor, and the loadings are about equal. The second factor contrasts the banking stocks with the oil stocks. (The banks have relatively large negative loadings, and the oils have large positive loadings, on the factor.) Thus, F_2 seems to differentiate stocks in different industries and might be called an *industry factor*. To summarize, rates of return appear to be determined by general market conditions and activities that are unique to the different industries, as well as a residual or firm specific factor. This is essentially the conclusion reached by an examination of the sample principal components in Example 8.5. ■

A Modified Approach—the Principal Factor Solution

A modification of the principal component approach is sometimes considered. We describe the reasoning in terms of a factor analysis of $\mathbf{R}$, although the procedure is also appropriate for $\mathbf{S}$. If the factor model $\boldsymbol{\rho} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$ is correctly specified, the m *common* factors should account for the *off-diagonal* elements of $\boldsymbol{\rho}$, as well as the *communality portions* of the diagonal elements

$$\rho_{ii} = 1 = h_i^2 + \psi_i$$

If the specific factor contribution ψ_i is removed from the diagonal or, equivalently, the 1 replaced by h_i^2, the resulting matrix is $\boldsymbol{\rho} - \boldsymbol{\Psi} = \mathbf{L}\mathbf{L}'$.

Suppose, now, that initial estimates ψ_i^* of the specific variances are available. Then replacing the ith diagonal element of $\mathbf{R}$ by $h_i^{*2} = 1 - \psi_i^*$, we obtain a "reduced" sample correlation matrix

$$\mathbf{R}_r = \begin{bmatrix} h_1^{*2} & r_{12} & \cdots & r_{1p} \\ r_{12} & h_2^{*2} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1p} & r_{2p} & \cdots & h_p^{*2} \end{bmatrix}$$

Now, apart from sampling variation, all of the elements of the reduced sample correlation matrix $\mathbf{R}_r$ should be accounted for by the m common factors. In particular, $\mathbf{R}_r$ is factored as

$$\mathbf{R}_r \doteq \mathbf{L}_r^* \mathbf{L}_r^{*\prime} \tag{9-21}$$

where $\mathbf{L}_r^* = \{\ell_{ij}^*\}$ are the estimated loadings.

The *principal factor method* of factor analysis employs the estimates

$$\begin{aligned} \mathbf{L}_r^* &= \left[\sqrt{\hat{\lambda}_1^*}\,\hat{\mathbf{e}}_1^* \;\vdots\; \sqrt{\hat{\lambda}_2^*}\,\hat{\mathbf{e}}_2^* \;\vdots\; \cdots \;\vdots\; \sqrt{\hat{\lambda}_m^*}\,\hat{\mathbf{e}}_m^*\right] \\ \psi_i^* &= 1 - \sum_{j=1}^{m} \ell_{ij}^{*2} \end{aligned} \tag{9-22}$$

where $(\hat{\lambda}_i^*, \hat{\mathbf{e}}_i^*)$, $i = 1, 2, \ldots, m$ are the (largest) eigenvalue-eigenvector pairs determined from $\mathbf{R}_r$. In turn, the communalities would then be (re)estimated by

$$\tilde{h}_i^{*2} = \sum_{j=1}^{m} \ell_{ij}^{*2} \tag{9-23}$$

The principal factor solution can be obtained iteratively, with the communality estimates of (9-23) becoming the initial estimates for the next stage.

In the spirit of the principal component solution, consideration of the estimated eigenvalues $\hat{\lambda}_1^*, \hat{\lambda}_2^*, \ldots, \hat{\lambda}_p^*$ helps determine the number of common factors to retain. An added complication is that now some of the eigenvalues may be negative, due to the use of initial communality estimates. Ideally, we should take the number of common factors equal to the rank of the reduced *population* matrix. Unfortunately, this rank is not always well determined from $\mathbf{R}_r$, and some judgment is necessary.

Although there are many choices for initial estimates of specific variances, the most popular choice, when one is working with a correlation matrix, is $\psi_i^* = 1/r^{ii}$, where r^{ii} is the ith diagonal element of $\mathbf{R}^{-1}$. The initial communality estimates then become

$$h_i^{*2} = 1 - \psi_i^* = 1 - \frac{1}{r^{ii}} \tag{9-24}$$

which is equal to the square of the multiple correlation coefficient between X_i and the other $p - 1$ variables. The relation to the multiple correlation coefficient means that h_i^{*2} can be calculated even when $\mathbf{R}$ is not of full rank. For factoring $\mathbf{S}$, the initial specific variance estimates use s^{ii}, the diagonal elements of $\mathbf{S}^{-1}$. Further discussion of these and other initial estimates is contained in [6].

Although the principal component method for $\mathbf{R}$ can be regarded as a principal factor method with *initial* communality estimates of unity, or specific variances equal to zero, the two are philosophically and geometrically different. (See [6].) In practice, however, the two frequently produce comparable factor loadings if the number of variables is large and the number of common factors is small.

We do not pursue the principal factor solution, since, to our minds, the solution methods that have the most to recommend them are the principal component method and the maximum likelihood method, which we discuss next.

The Maximum Likelihood Method

If the common factors $\mathbf{F}$ and the specific factors $\boldsymbol{\varepsilon}$ can be assumed to be normally distributed, then maximum likelihood estimates of the factor loadings and specific variances may be obtained. When $\mathbf{F}_j$ and $\boldsymbol{\varepsilon}_j$ are jointly normal, the observations $\mathbf{X}_j - \boldsymbol{\mu} = \mathbf{L}\mathbf{F}_j + \boldsymbol{\varepsilon}_j$ are then normal, and from (4-16), the likelihood is

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= (2\pi)^{-\frac{np}{2}} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} e^{-\left(\frac{1}{2}\right)\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n} (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' + n(\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})'\right)\right]} \\ &= (2\pi)^{\frac{-(n-1)p}{2}} |\boldsymbol{\Sigma}|^{-\frac{(n-1)}{2}} e^{-\left(\frac{1}{2}\right)\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n} (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'\right)\right]} \\ &\quad \times (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\left(\frac{n}{2}\right)(\bar{\mathbf{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu})} \end{aligned} \tag{9-25}$$

which depends on $\mathbf{L}$ and $\boldsymbol{\Psi}$ through $\boldsymbol{\Sigma} = \mathbf{LL}' + \boldsymbol{\Psi}$. This model is still not well defined, because of the multiplicity of choices for $\mathbf{L}$ made possible by orthogonal transformations. It is desirable to make $\mathbf{L}$ well defined by imposing the computationally convenient *uniqueness condition*

$$\mathbf{L}'\boldsymbol{\Psi}^{-1}\mathbf{L} = \boldsymbol{\Delta} \qquad \text{a diagonal matrix} \tag{9-26}$$

The maximum likelihood estimates $\hat{\mathbf{L}}$ and $\hat{\boldsymbol{\Psi}}$ must be obtained by numerical maximization of (9-25). Fortunately, efficient computer programs now exist that enable one to get these estimates rather easily.

We summarize some facts about maximum likelihood estimators and, for now, rely on a computer to perform the numerical details.

Result 9.1. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a random sample from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \mathbf{LL}' + \boldsymbol{\Psi}$ is the covariance matrix for the m common factor model of (9-4). The maximum likelihood estimators $\hat{\mathbf{L}}$, $\hat{\boldsymbol{\Psi}}$, and $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ maximize (9-25) subject to $\hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}}$ being diagonal.

The maximum likelihood estimates of the communalities are

$$\hat{h}_i^2 = \hat{\ell}_{i1}^2 + \hat{\ell}_{i2}^2 + \cdots + \hat{\ell}_{im}^2 \qquad \text{for } i = 1, 2, \ldots, p \tag{9-27}$$

so

$$\begin{pmatrix} \text{Proportion of total sample} \\ \text{variance due to } j\text{th factor} \end{pmatrix} = \frac{\hat{\ell}_{1j}^2 + \hat{\ell}_{2j}^2 + \cdots + \hat{\ell}_{pj}^2}{s_{11} + s_{22} + \cdots + s_{pp}} \tag{9-28}$$

Proof. By the invariance property of maximum likelihood estimates (see Section 4.3), functions of $\mathbf{L}$ and $\boldsymbol{\Psi}$ are estimated by the same functions of $\hat{\mathbf{L}}$ and $\hat{\boldsymbol{\Psi}}$. In particular, the communalities $h_i^2 = \ell_{i1}^2 + \cdots + \ell_{im}^2$ have maximum likelihood estimates $\hat{h}_i^2 = \hat{\ell}_{i1}^2 + \cdots + \hat{\ell}_{im}^2$. ■

If, as in (8-10), the variables are standardized so that $\mathbf{Z} = \mathbf{V}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$, then the covariance matrix $\boldsymbol{\rho}$ of $\mathbf{Z}$ has the representation

$$\boldsymbol{\rho} = \mathbf{V}^{-1/2}\boldsymbol{\Sigma}\mathbf{V}^{-1/2} = (\mathbf{V}^{-1/2}\mathbf{L})(\mathbf{V}^{-1/2}\mathbf{L})' + \mathbf{V}^{-1/2}\boldsymbol{\Psi}\mathbf{V}^{-1/2} \tag{9-29}$$

Thus, $\boldsymbol{\rho}$ has a factorization analogous to (9-5) with loading matrix $\mathbf{L}_z = \mathbf{V}^{-1/2}\mathbf{L}$ and specific variance matrix $\boldsymbol{\Psi}_z = \mathbf{V}^{-1/2}\boldsymbol{\Psi}\mathbf{V}^{-1/2}$. By the invariance property of maximum likelihood estimators, the maximum likelihood estimator of $\boldsymbol{\rho}$ is

$$\begin{aligned} \hat{\boldsymbol{\rho}} &= (\hat{\mathbf{V}}^{-1/2}\hat{\mathbf{L}})(\hat{\mathbf{V}}^{-1/2}\hat{\mathbf{L}})' + \hat{\mathbf{V}}^{-1/2}\hat{\boldsymbol{\Psi}}\hat{\mathbf{V}}^{-1/2} \\ &= \hat{\mathbf{L}}_z\hat{\mathbf{L}}_z' + \hat{\boldsymbol{\Psi}}_z \end{aligned} \tag{9-30}$$

where $\hat{\mathbf{V}}^{-1/2}$ and $\hat{\mathbf{L}}$ are the maximum likelihood estimators of $\mathbf{V}^{-1/2}$ and $\mathbf{L}$, respectively. (See Supplement 9A.)

As a consequence of the factorization of (9-30), whenever the maximum likelihood analysis pertains to the correlation matrix, we call

$$\hat{h}_i^2 = \hat{\ell}_{i1}^2 + \hat{\ell}_{i2}^2 + \cdots + \hat{\ell}_{im}^2 \qquad i = 1, 2, \ldots, p \tag{9-31}$$

the maximum likelihood estimates of the communalities, and we evaluate the importance of the factors on the basis of

$$\begin{pmatrix}\text{Proportion of total (standardized)}\\ \text{sample variance due to } j\text{th factor}\end{pmatrix} = \frac{\hat{\ell}_{1j}^2 + \hat{\ell}_{2j}^2 + \cdots + \hat{\ell}_{pj}^2}{p} \tag{9-32}$$

To avoid more tedious notations, the preceding $\hat{\ell}_{ij}$'s denote the elements of $\hat{\mathbf{L}}_{\mathbf{z}}$.

Comment. Ordinarily, the observations are standardized, and a sample correlation matrix is factor analyzed. The sample correlation matrix $\mathbf{R}$ is inserted for $[(n-1)/n]\mathbf{S}$ in the likelihood function of (9-25), and the maximum likelihood estimates $\hat{\mathbf{L}}_{\mathbf{z}}$ and $\hat{\mathbf{\Psi}}_{\mathbf{z}}$ are obtained using a computer. Although the likelihood in (9-25) is appropriate for $\mathbf{S}$, not $\mathbf{R}$, surprisingly, this practice is equivalent to obtaining the maximum likelihood estimates $\hat{\mathbf{L}}$ and $\hat{\mathbf{\Psi}}$ based on the sample covariance matrix $\mathbf{S}$, setting $\hat{\mathbf{L}}_{\mathbf{z}} = \hat{\mathbf{V}}^{-1/2}\hat{\mathbf{L}}$ and $\hat{\mathbf{\Psi}}_{\mathbf{z}} = \hat{\mathbf{V}}^{-1/2}\hat{\mathbf{\Psi}}\hat{\mathbf{V}}^{-1/2}$. Here $\hat{\mathbf{V}}^{-1/2}$ is the diagonal matrix with the reciprocal of the sample standard deviations (computed with the divisor $\sqrt{n}$) on the main diagonal.

Going in the other direction, given the estimated loadings $\hat{\mathbf{L}}_{\mathbf{z}}$ and specific variances $\hat{\mathbf{\Psi}}_{\mathbf{z}}$ obtained from $\mathbf{R}$, we find that the resulting maximum likelihood estimates for a factor analysis of the covariance matrix $[(n-1)/n]\mathbf{S}$ are $\hat{\mathbf{L}} = \hat{\mathbf{V}}^{1/2}\hat{\mathbf{L}}_{\mathbf{z}}$ and $\hat{\mathbf{\Psi}} = \hat{\mathbf{V}}^{1/2}\hat{\mathbf{\Psi}}_{\mathbf{z}}\hat{\mathbf{V}}^{1/2}$, or

$$\hat{\ell}_{ij} = \hat{\ell}_{\mathbf{z},ij}\sqrt{\hat{\sigma}_{ii}} \quad \text{and} \quad \hat{\psi}_i = \hat{\psi}_{\mathbf{z},i}\hat{\sigma}_{ii}$$

where $\hat{\sigma}_{ii}$ is the sample variance computed with divisor n. The distinction between divisors can be ignored with principal component solutions. ■

The equivalency between factoring $\mathbf{S}$ and $\mathbf{R}$ has apparently been confused in many published discussions of factor analysis. (See Supplement 9A.)

Example 9.5 (Factor analysis of stock-price data using the maximum likelihood method) The stock-price data of Examples 8.5 and 9.4 were reanalyzed assuming an $m = 2$ factor model and using the *maximum likelihood method*. The estimated factor loadings, communalities, specific variances, and proportion of total (standardized) sample variance explained by each factor are in Table 9.3.[3] The corresponding figures for the $m = 2$ factor solution obtained by the *principal component method* (see Example 9.4) are also provided. The communalities corresponding to the maximum likelihood factoring of $\mathbf{R}$ are of the form [see (9-31)] $\hat{h}_i^2 = \hat{\ell}_{i1}^2 + \hat{\ell}_{i2}^2$. So, for example,

$$\hat{h}_1^2 = (.115)^2 + (.765)^2 = .58$$

[3] The maximum likelihood solution leads to a *Heywood case*. For this example, the solution of the likelihood equations give estimated loadings such that a specific variance is negative. The software program obtains a feasible solution by slightly adjusting the loadings so that all specific variance estimates are nonnegative. A Heywood case is suggested here by the .00 value for the specific variance of Royal Dutch Shell.

Table 9.3

	Maximum likelihood			Principal components		
	Estimated factor loadings		Specific variances	Estimated factor loadings		Specific variances
Variable	F_1	F_2	$\hat{\psi}_i = 1 - \hat{h}_i^2$	F_1	F_2	$\tilde{\psi}_i = 1 - \tilde{h}_i^2$
1. J P Morgan	.115	.755	.42	.732	−.437	.27
2. Citibank	.322	.788	.27	.831	−.280	.23
3. Wells Fargo	.182	.652	.54	.726	−.374	.33
4. Royal Dutch Shell	1.000	−.000	.00	.605	.694	.15
5. Texaco	.683	−.032	.53	.563	.719	.17
Cumulative proportion of total (standardized) sample variance explained	.323	.647		.487	.769	

The residual matrix is

$$\mathbf{R} - \hat{\mathbf{L}}\hat{\mathbf{L}}' - \hat{\mathbf{\Psi}} = \begin{bmatrix} 0 & .001 & -.002 & .000 & .052 \\ .001 & 0 & .002 & .000 & -.033 \\ -.002 & .002 & 0 & .000 & .001 \\ .000 & .000 & .000 & 0 & .000 \\ .052 & -.033 & .001 & .000 & 0 \end{bmatrix}$$

The elements of $\mathbf{R} - \hat{\mathbf{L}}\hat{\mathbf{L}}' - \hat{\mathbf{\Psi}}$ are much smaller than those of the residual matrix corresponding to the principal component factoring of $\mathbf{R}$ presented in Example 9.4. On this basis, we prefer the maximum likelihood approach and typically feature it in subsequent examples.

The cumulative proportion of the total sample variance explained by the factors is larger for principal component factoring than for maximum likelihood factoring. It is not surprising that this criterion typically favors principal component factoring. Loadings obtained by a principal component factor analysis are related to the principal components, which have, by design, a variance optimizing property. [See the discussion preceding (8-19).]

Focusing attention on the maximum likelihood solution, we see that all variables have positive loadings on F_1. We call this factor the *market factor*, as we did in the principal component solution. The interpretation of the second factor is not as clear as it appeared to be in the principal component solution. The bank stocks have large positive loadings and the oil stocks have negligible loadings on the second factor F_2. From this perspective, the second factor differentiaties the bank stocks from the oil stocks and might be called an *industry factor*. Alternatively, the second factor might be simply called a *banking factor*.

The patterns of the initial factor loadings for the maximum likelihood solution are constrained by the uniqueness condition that $\hat{\mathbf{L}}'\hat{\mathbf{\Psi}}^{-1}\hat{\mathbf{L}}$ be a diagonal matrix. Therefore, useful factor patterns are often not revealed until the factors are rotated (see Section 9.4). ■

Example 9.6 (Factor analysis of Olympic decathlon data) Linden [11] originally conducted a factor analytic study of Olympic decathlon results for all 160 complete starts from the end of World War II until the mid-seventies. Following his approach we examine the $n = 280$ complete starts from 1960 through 2004. The recorded values for each event were standardized and the signs of the timed events changed so that large scores are good for all events. We, too, analyze the correlation matrix, which based on all 280 cases, is

$$\mathbf{R} = \begin{bmatrix} 1.000 & .6386 & .4752 & .3227 & .5520 & .3262 & .3509 & .4008 & .1821 & -.0352 \\ .6386 & 1.0000 & .4953 & .5668 & .4706 & .3520 & .3998 & .5167 & .3102 & .1012 \\ .4752 & .4953 & 1.0000 & .4357 & .2539 & .2812 & .7926 & .4728 & .4682 & -.0120 \\ .3227 & .5668 & .4357 & 1.0000 & .3449 & .3503 & .3657 & .6040 & .2344 & .2380 \\ .5520 & .4706 & .2539 & .3449 & 1.0000 & .1546 & .2100 & .4213 & .2116 & .4125 \\ .3262 & .3520 & .2812 & .3503 & .1546 & 1.0000 & .2553 & .4163 & .1712 & .0002 \\ .3509 & .3998 & .7926 & .3657 & .2100 & .2553 & 1.0000 & .4036 & .4179 & .0109 \\ .4008 & .5167 & .4728 & .6040 & .4213 & .4163 & .4036 & 1.0000 & .3151 & .2395 \\ .1821 & .3102 & .4682 & .2344 & .2116 & .1712 & .4179 & .3151 & 1.0000 & .0983 \\ -.0352 & .1012 & -.0120 & .2380 & .4125 & .0002 & .0109 & .2395 & .0983 & 1.0000 \end{bmatrix}$$

From a principal component factor analysis perspective, the first four eigenvalues, 4.21, 1.39, 1.06, .92, of $\mathbf{R}$ suggest a factor solution with $m = 3$ or $m = 4$. A subsequent interpretation, much like Linden's original analysis, reinforces the choice $m = 4$.

In this case, the two solution methods produced very different results. For the principal component factorization, all events except the 1,500-meter run have large positive loading on the first factor. This factor might be labeled *general athletic ability*. Factor 2, which loads heavily on the 400-meter run and 1,500-meter run might be called a *running endurance* factor. The remaining factors cannot be easily interpreted to our minds.

For the maximum likelihood method, the first factor appears to be a *general athletic ability factor* but the loading pattern is not as strong as with principal component factor solution. The second factor is primarily a *strength* factor because shot put and discus load highly on this factor. The third factor is *running endurance* since the 400-meter run and 1,500-meter run have large loadings. Again, the fourth factor is not easily identified, although it may have something to do with jumping ability or *leg strength*. We shall return to an interpretation of the factors in Example 9.11 after a discussion of factor rotation.

The four-factor principal component solution accounts for much of the total (standardized) sample variance, although the estimated specific variances are large in some cases (for example, the javelin). This suggests that some events might require *unique* or specific attributes not required for the other events. The four-factor maximum likelihood solution accounts for less of the total sample

Table 9.4

	Principal component					Maximum likelihood				
	Estimated factor loadings				Specific variances	Estimated factor loadings				Specific variances
Variable	F_1	F_2	F_3	F_4	$\tilde{\psi}_i = 1 - \tilde{h}_i^2$	F_1	F_2	F_3	F_4	$\hat{\psi}_i = 1 - \hat{h}_i^2$
1. 100-m run	.696	.022	−.468	−.416	.12	.993	−.069	−.021	.002	.01
2. Long jump	.793	.075	−.255	−.115	.29	.665	.252	.239	.220	.39
3. Shot put	.771	−.434	.197	−.112	.17	.530	.777	−.141	−.079	.09
4. High jump	.711	.181	.005	.367	.33	.363	.428	.421	.424	.33
5. 400-m run	.605	.549	−.045	−.397	.17	.571	.019	.620	−.305	.20
6. 100 m hurdles	.513	−.083	−.372	.561	.28	.343	.189	.090	.323	.73
7. Discus	.690	−.456	.289	−.078	.23	.402	.718	−.102	−.095	.30
8. Pole vault	.761	.162	.018	.304	.30	.440	.407	.390	.263	.42
9. Javelin	.518	−.252	.519	−.074	.39	.218	.461	.084	−.085	.73
10. 1500-m run	.220	.746	.493	.085	.15	−.016	.091	.609	−.145	.60
Cumulative proportion of total variance explained	.42	.56	.67	.76		.27	.45	.57	.62	

variance, but, as the following residual matrices indicate, the maximum likelihood estimates $\hat{\mathbf{L}}$ and $\hat{\boldsymbol{\Psi}}$ do a better job of reproducing $\mathbf{R}$ than the principal component estimates $\tilde{\mathbf{L}}$ and $\tilde{\boldsymbol{\Psi}}$.

Principal component:

$$\mathbf{R} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}' - \tilde{\boldsymbol{\Psi}} =$$

$$\begin{bmatrix} 0 & -.082 & -.006 & -.021 & -.068 & .031 & -.016 & .003 & .039 & .062 \\ -.082 & 0 & -.046 & .033 & -.107 & -.078 & -.048 & -.059 & .042 & .006 \\ -.006 & -.046 & 0 & .006 & -.010 & -.014 & -.003 & -.013 & -.151 & .055 \\ -.021 & .033 & .006 & 0 & -.038 & -.204 & -.015 & -.078 & -.064 & -.086 \\ -.068 & -.107 & -.010 & -.038 & 0 & .096 & .025 & -.006 & .030 & -.074 \\ .031 & -.078 & -.014 & -.204 & .096 & 0 & .015 & -.124 & .119 & .085 \\ -.016 & -.048 & -.003 & -.015 & .025 & .015 & 0 & -.029 & -.210 & .064 \\ .003 & -.059 & -.013 & -.078 & -.006 & -.124 & -.029 & 0 & -.026 & -.084 \\ .039 & .042 & -.151 & -.064 & .030 & .119 & -.210 & -.026 & 0 & -.078 \\ .062 & .006 & .055 & -.086 & -.074 & .085 & .064 & -.084 & -.078 & 0 \end{bmatrix}$$

Maximum likelihood:

$$\mathbf{R} - \hat{\mathbf{L}}\hat{\mathbf{L}}' - \hat{\boldsymbol{\Psi}} =$$

$$\begin{bmatrix} 0 & .000 & .000 & -.000 & -.000 & .000 & -.000 & .000 & -.001 & 000 \\ .000 & 0 & -.002 & .023 & .005 & .017 & -.003 & -.030 & .047 & -.024 \\ .000 & -.002 & 0 & .004 & -.000 & -.009 & .000 & -.001 & -.001 & .000 \\ -.000 & .023 & .004 & 0 & -.002 & -.030 & -.004 & -.006 & -.042 & .010 \\ -.000 & .005 & -.001 & -.002 & 0 & -.002 & .001 & .001 & .000 & -.001 \\ .000 & -.017 & -.009 & -.030 & -.002 & 0 & .022 & .069 & .029 & -.019 \\ -.000 & -.003 & .000 & -.004 & .001 & .022 & 0 & -.000 & -.000 & .000 \\ .000 & -.030 & -.001 & -.006 & .001 & .069 & -.000 & 0 & .021 & .011 \\ -.001 & .047 & -.001 & -.042 & .001 & .029 & -.000 & .021 & 0 & -.003 \\ .000 & -.024 & .000 & .010 & -.001 & -.019 & .000 & .011 & -.003 & 0 \end{bmatrix}$$

■

A Large Sample Test for the Number of Common Factors

The assumption of a normal population leads directly to a test of the adequacy of the model. Suppose the m common factor model holds. In this case $\boldsymbol{\Sigma} = \mathbf{LL}' + \boldsymbol{\Psi}$, and testing the adequacy of the m common factor model is equivalent to testing

$$H_0: \underset{(p\times p)}{\boldsymbol{\Sigma}} = \underset{(p\times m)}{\mathbf{L}} \underset{(m\times p)}{\mathbf{L}'} + \underset{(p\times p)}{\boldsymbol{\Psi}} \tag{9-33}$$

versus H_1: $\boldsymbol{\Sigma}$ any other positive definite matrix. When $\boldsymbol{\Sigma}$ does not have any special form, the maximum of the likelihood function [see (4-18) and Result 4.11 with $\hat{\boldsymbol{\Sigma}} = ((n-1)/n)\mathbf{S} = \mathbf{S}_n$] is proportional to

$$|\mathbf{S}_n|^{-n/2} e^{-np/2} \tag{9-34}$$

Under H_0, $\boldsymbol{\Sigma}$ is restricted to have the form of (9-33). In this case, the maximum of the likelihood function [see (9-25) with $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ and $\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}$, where $\hat{\mathbf{L}}$ and $\hat{\boldsymbol{\Psi}}$ are the maximum likelihood estimates of $\mathbf{L}$ and $\boldsymbol{\Psi}$, respectively] is proportional to

$$|\hat{\boldsymbol{\Sigma}}|^{-n/2}\exp\left(-\tfrac{1}{2}\operatorname{tr}\left[\hat{\boldsymbol{\Sigma}}^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j-\bar{\mathbf{x}})(\mathbf{x}_j-\bar{\mathbf{x}})'\right)\right]\right)$$

$$= |\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}|^{-n/2}\exp\left(-\tfrac{1}{2}n\operatorname{tr}[(\hat{\mathbf{L}}\hat{\mathbf{L}} + \hat{\boldsymbol{\Psi}})^{-1}\mathbf{S}_n]\right) \tag{9-35}$$

Using Result 5.2, (9-34), and (9-35), we find that the likelihood ratio statistic for testing H_0 is

$$\begin{aligned} -2\ln\Lambda &= -2\ln\left[\frac{\text{maximized likelihood under } H_0}{\text{maximized likelihood}}\right] \\ &= -2\ln\left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\mathbf{S}_n|}\right)^{-n/2} + n\left[\operatorname{tr}(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}_n) - p\right] \end{aligned} \tag{9-36}$$

with degrees of freedom,

$$\begin{aligned} v - v_0 &= \tfrac{1}{2}p(p+1) - \left[p(m+1) - \tfrac{1}{2}m(m-1)\right] \\ &= \tfrac{1}{2}\left[(p-m)^2 - p - m\right] \end{aligned} \tag{9-37}$$

Supplement 9A indicates that $\operatorname{tr}(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}_n) - p = 0$ provided that $\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}$ is the maximum likelihood estimate of $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$. Thus, we have

$$-2\ln\Lambda = n\ln\left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\mathbf{S}_n|}\right) \tag{9-38}$$

Bartlett [3] has shown that the chi-square approximation to the sampling distribution of $-2\ln\Lambda$ can be improved by replacing n in (9-38) with the multiplicative factor $(n - 1 - (2p + 4m + 5)/6)$.

Using Bartlett's correction,[4] we reject H_0 at the α level of significance if

$$(n - 1 - (2p + 4m + 5)/6)\ln\frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}|}{|\mathbf{S}_n|} > \chi^2_{[(p-m)^2-p-m]/2}(\alpha) \tag{9-39}$$

provided that n and $n - p$ are large. Since the number of degrees of freedom, $\frac{1}{2}[(p-m)^2 - p - m]$, must be positive, it follows that

$$m < \tfrac{1}{2}\left(2p + 1 - \sqrt{8p+1}\right) \tag{9-40}$$

in order to apply the test (9-39).

[4] Many factor analysts obtain an approximate maximum likelihood estimate by replacing $\mathbf{S}_n$ with the unbiased estimate $\mathbf{S} = [n/(n-1)]\mathbf{S}_n$ and then minimizing $\ln|\boldsymbol{\Sigma}| + \operatorname{tr}[\boldsymbol{\Sigma}^{-1}\mathbf{S}]$. The dual substitution of $\mathbf{S}$ and the approximate maximum likelihood estimator into the test statistic of (9-39) does not affect its large sample properties.

Comment. In implementing the test in (9-39), we are testing for the adequacy of the m common factor model by comparing the generalized variances $|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}|$ and $|\mathbf{S}_n|$. If n is large and m is small relative to p, the hypothesis H_0 will usually be rejected, leading to a retention of more common factors. However, $\hat{\mathbf{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}$ may be close enough to $\mathbf{S}_n$ so that adding more factors does not provide additional insights, even though those factors are "significant." Some judgment must be exercised in the choice of m.

Example 9.7 (Testing for two common factors) The two-factor maximum likelihood analysis of the stock-price data was presented in Example 9.5. The residual matrix there suggests that a two-factor solution may be adequate. Test the hypothesis $H_0: \mathbf{\Sigma} = \mathbf{L}\mathbf{L}' + \mathbf{\Psi}$, with $m = 2$, at level $\alpha = .05$.

The test statistic in (9-39) is based on the ratio of generalized variances

$$\frac{|\hat{\mathbf{\Sigma}}|}{|\mathbf{S}_n|} = \frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}|}{|\mathbf{S}_n|}$$

Let $\hat{\mathbf{V}}^{-1/2}$ be the diagonal matrix such that $\hat{\mathbf{V}}^{-1/2}\mathbf{S}_n\hat{\mathbf{V}}^{-1/2} = \mathbf{R}$. By the properties of determinants (see Result 2A.11),

$$|\hat{\mathbf{V}}^{-1/2}||\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}||\hat{\mathbf{V}}^{-1/2}| = |\hat{\mathbf{V}}^{-1/2}\hat{\mathbf{L}}\hat{\mathbf{L}}'\hat{\mathbf{V}}^{-1/2} + \hat{\mathbf{V}}^{-1/2}\hat{\mathbf{\Psi}}\hat{\mathbf{V}}^{-1/2}|$$

and

$$|\hat{\mathbf{V}}^{-1/2}||\mathbf{S}_n||\hat{\mathbf{V}}^{-1/2}| = |\hat{\mathbf{V}}^{-1/2}\mathbf{S}_n\hat{\mathbf{V}}^{-1/2}|$$

Consequently,

$$\begin{aligned}\frac{|\hat{\mathbf{\Sigma}}|}{|\mathbf{S}_n|} &= \frac{|\hat{\mathbf{V}}^{-1/2}|}{|\hat{\mathbf{V}}^{-1/2}|}\frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}|}{|\mathbf{S}_n|}\frac{|\hat{\mathbf{V}}^{-1/2}|}{|\hat{\mathbf{V}}^{-1/2}|}\\ &= \frac{|\hat{\mathbf{V}}^{-1/2}\hat{\mathbf{L}}\hat{\mathbf{L}}'\hat{\mathbf{V}}^{-1/2} + \hat{\mathbf{V}}^{-1/2}\hat{\mathbf{\Psi}}\hat{\mathbf{V}}^{-1/2}|}{|\hat{\mathbf{V}}^{-1/2}\mathbf{S}_n\hat{\mathbf{V}}^{-1/2}|} \qquad (9\text{-}41)\\ &= \frac{|\hat{\mathbf{L}}_z\hat{\mathbf{L}}_z' + \hat{\mathbf{\Psi}}_z|}{|\mathbf{R}|}\end{aligned}$$

by (9-30). From Example 9.5, we determine

$$\frac{|\hat{\mathbf{L}}_z\hat{\mathbf{L}}_z' + \hat{\mathbf{\Psi}}_z|}{|\mathbf{R}|} = \frac{\begin{vmatrix} 1.000 & & & & \\ .632 & 1.000 & & & \\ .513 & .572 & 1.000 & & \\ .115 & .322 & .182 & 1.000 & \\ .103 & .246 & .146 & .683 & 1.000 \end{vmatrix}}{\begin{vmatrix} 1.000 & & & & \\ .632 & 1.000 & & & \\ .510 & .574 & 1.000 & & \\ .115 & .322 & .182 & 1.000 & \\ .154 & .213 & .146 & .683 & 1.000 \end{vmatrix}} = \frac{.17898}{.17519} = 1.0216$$

Using Bartlett's correction, we evaluate the test statistic in (9-39):

$$[n - 1 - (2p + 4m + 5)/6] \ln \frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}|}{|\mathbf{S}_n|}$$

$$= \left[103 - 1 - \frac{(10 + 8 + 5)}{6}\right] \ln(1.0216) = 2.10$$

Since $\frac{1}{2}[(p - m)^2 - p - m] = \frac{1}{2}[(5 - 2)^2 - 5 - 2] = 1$, the 5% critical value $\chi_1^2(.05) = 3.84$ is not exceeded, and we fail to reject H_0. We conclude that the data do not contradict a two-factor model. In fact, the observed significance level, or P-value, $P[\chi_1^2 > 2.10] \doteq .15$ implies that H_0 would not be rejected at *any* reasonable level. ■

Large sample variances and covariances for the maximum likelihood estimates $\hat{\ell}_{ij}, \hat{\psi}_i$ have been derived when these estimates have been determined from the sample covariance matrix $\mathbf{S}$. (See [10].) The expressions are, in general, quite complicated.

9.4 Factor Rotation

As we indicated in Section 9.2, all factor loadings obtained from the initial loadings by an orthogonal transformation have the same ability to reproduce the covariance (or correlation) matrix. [See (9-8).] From matrix algebra, we know that an orthogonal transformation corresponds to a rigid rotation (or reflection) of the coordinate axes. For this reason, an orthogonal transformation of the factor loadings, as well as the implied orthogonal transformation of the factors, is called *factor rotation*.

If $\hat{\mathbf{L}}$ is the $p \times m$ matrix of estimated factor loadings obtained by any method (principal component, maximum likelihood, and so forth) then

$$\hat{\mathbf{L}}^* = \hat{\mathbf{L}}\mathbf{T}, \qquad \text{where } \mathbf{T}\mathbf{T}' = \mathbf{T}'\mathbf{T} = \mathbf{I} \tag{9-42}$$

is a $p \times m$ matrix of "rotated" loadings. Moreover, the estimated covariance (or correlation) matrix remains unchanged, since

$$\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}} = \hat{\mathbf{L}}\mathbf{T}\mathbf{T}'\hat{\mathbf{L}} + \hat{\boldsymbol{\Psi}} = \hat{\mathbf{L}}^*\hat{\mathbf{L}}^{*\prime} + \hat{\boldsymbol{\Psi}} \tag{9-43}$$

Equation (9-43) indicates that the residual matrix, $\mathbf{S}_n - \hat{\mathbf{L}}\hat{\mathbf{L}}' - \hat{\boldsymbol{\Psi}} = \mathbf{S}_n - \hat{\mathbf{L}}^*\hat{\mathbf{L}}^{*\prime} - \hat{\boldsymbol{\Psi}}$, remains unchanged. Moreover, the specific variances $\hat{\psi}_i$, and hence the communalities $\hat{h}_i^2$, are unaltered. Thus, from a mathematical viewpoint, it is immaterial whether $\hat{\mathbf{L}}$ or $\hat{\mathbf{L}}^*$ is obtained.

Since the original loadings may not be readily interpretable, it is usual practice to rotate them until a "simpler structure" is achieved. The rationale is very much akin to sharpening the focus of a microscope in order to see the detail more clearly.

Ideally, we should like to see a pattern of loadings such that each variable loads highly on a single factor and has small to moderate loadings on the remaining factors. However, it is not always possible to get this simple structure, although the rotated loadings for the decathlon data discussed in Example 9.11 provide a nearly ideal pattern.

We shall concentrate on graphical and analytical methods for determining an orthogonal rotation to a simple structure. When $m = 2$, or the common factors are considered two at a time, the transformation to a simple structure can frequently be determined graphically. The uncorrelated common factors are regarded as unit vectors along perpendicular coordinate axes. A plot of the pairs of factor loadings $(\hat{\ell}_{i1}, \hat{\ell}_{i2})$ yields p points, each point corresponding to a variable. The coordinate axes can then be visually rotated through an angle—call it ϕ—and the new rotated loadings $\hat{\ell}^*_{ij}$ are determined from the relationships

$$\underset{(p\times 2)}{\hat{\mathbf{L}}^*} = \underset{(p\times 2)(2\times 2)}{\hat{\mathbf{L}}\ \mathbf{T}} \tag{9-44}$$

$$\text{where} \quad \begin{cases} \mathbf{T} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} & \text{clockwise rotation} \\[2ex] \mathbf{T} = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} & \text{counterclockwise rotation} \end{cases}$$

The relationship in (9-44) is rarely implemented in a two-dimensional graphical analysis. In this situation, clusters of variables are often apparent by eye, and these clusters enable one to identify the common factors without having to inspect the magnitudes of the rotated loadings. On the other hand, for $m > 2$, orientations are not easily visualized, and the magnitudes of the *rotated* loadings must be inspected to find a meaningful interpretation of the original data. The choice of an orthogonal matrix $\mathbf{T}$ that satisfies an *analytical* measure of simple structure will be considered shortly.

Example 9.8 (A first look at factor rotation) Lawley and Maxwell [10] present the sample correlation matrix of examination scores in $p = 6$ subject areas for $n = 220$ male students. The correlation matrix is

$$\begin{array}{c} \\ \mathbf{R} = \end{array} \begin{array}{c} \begin{array}{cccccc} \text{Gaelic} & \text{English} & \text{History} & \text{Arithmetic} & \text{Algebra} & \text{Geometry} \end{array} \\ \begin{bmatrix} 1.0 & .439 & .410 & .288 & .329 & .248 \\ & 1.0 & .351 & .354 & .320 & .329 \\ & & 1.0 & .164 & .190 & .181 \\ & & & 1.0 & .595 & .470 \\ & & & & 1.0 & .464 \\ & & & & & 1.0 \end{bmatrix} \end{array}$$

and a maximum likelihood solution for $m = 2$ common factors yields the estimates in Table 9.5.

Table 9.5

Variable	Estimated factor loadings F_1	F_2	Communalities $\hat{h}_i^2$
1. Gaelic	.553	.429	.490
2. English	.568	.288	.406
3. History	.392	.450	.356
4. Arithmetic	.740	−.273	.623
5. Algebra	.724	−.211	.569
6. Geometry	.595	−.132	.372

All the variables have positive loadings on the first factor. Lawley and Maxwell suggest that this factor reflects the overall response of the students to instruction and might be labeled a *general intelligence* factor. Half the loadings are positive and half are negative on the second factor. A factor with this pattern of loadings is called a *bipolar factor*. (The assignment of negative and positive poles is arbitrary, because the signs of the loadings on a factor can be reversed without affecting the analysis.) This factor is not easily identified, but is such that individuals who get above-average scores on the verbal tests get above-average scores on the factor. Individuals with above-average scores on the mathematical tests get below-average scores on the factor. Perhaps this factor can be classified as a "math-nonmath" factor.

The factor loading pairs $(\hat{\ell}_{i1}, \hat{\ell}_{i2})$ are plotted as points in Figure 9.1. The points are labeled with the numbers of the corresponding variables. Also shown is a clockwise orthogonal rotation of the coordinate axes through an angle of $\phi \doteq 20°$. This angle was chosen so that one of the new axes passes through $(\hat{\ell}_{41}, \hat{\ell}_{42})$. When this is done, all the points fall in the first quadrant (the factor loadings are all positive), and the two distinct clusters of variables are more clearly revealed.

The mathematical test variables load highly on F_1^* and have negligible loadings on F_2^*. The first factor might be called a *mathematical-ability* factor. Similarly, the three verbal test variables have high loadings on F_2^* and moderate to small loadings on F_1^*. The second factor might be labeled a *verbal-ability* factor. The *general-intelligence* factor identified initially is submerged in the factors F_1^* and F_2^*.

The rotated factor loadings obtained from (9-44) with $\phi \doteq 20°$ and the corresponding communality estimates are shown in Table 9.6. The magnitudes of the rotated factor loadings reinforce the interpretation of the factors suggested by Figure 9.1.

The communality estimates are unchanged by the orthogonal rotation, since $\hat{\mathbf{L}}\hat{\mathbf{L}}' = \hat{\mathbf{L}}\mathbf{T}\mathbf{T}'\hat{\mathbf{L}}' = \hat{\mathbf{L}}^*\hat{\mathbf{L}}^{*\prime}$, and the communalities are the diagonal elements of these matrices.

We point out that Figure 9.1 suggests an *oblique rotation* of the coordinates. One new axis would pass through the cluster $\{1,2,3\}$ and the other through the $\{4,5,6\}$ group. Oblique rotations are so named because they correspond to a *nonrigid* rotation of coordinate axes leading to new axes that are not perpendicular.

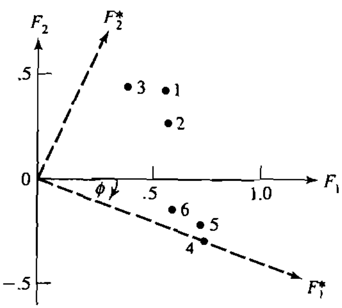


Figure 9.1 Factor rotation for test scores.

Table 9.6

Variable	Estimated rotated factor loadings F_1^*	F_2^*	Communalities $\hat{h}_i^{*2} = \hat{h}_i^2$
1. Gaelic	.369	.594	.490
2. English	.433	.467	.406
3. History	.211	.558	.356
4. Arithmetic	.789	.001	.623
5. Algebra	.752	.054	.568
6. Geometry	.604	.083	.372

It is apparent, however, that the interpretation of the oblique factors for this example would be much the same as that given previously for an orthogonal rotation. ■

Kaiser [9] has suggested an analytical measure of simple structure known as the *varimax* (or normal varimax) *criterion*. Define $\tilde{\ell}_{ij}^* = \hat{\ell}_{ij}^*/\hat{h}_i$ to be the rotated coefficients scaled by the square root of the communalities. Then the (normal) varimax procedure selects the orthogonal transformation $\mathbf{T}$ that makes

$$V = \frac{1}{p}\sum_{j=1}^{m}\left[\sum_{i=1}^{p}\tilde{\ell}_{ij}^{*4} - \left(\sum_{i=1}^{p}\tilde{\ell}_{ij}^{*2}\right)^2\Big/ p\right] \tag{9-45}$$

as large as possible.

Scaling the rotated coefficients $\hat{\ell}_{ij}^*$ has the effect of giving variables with small communalities relatively more weight in the determination of simple structure. After the transformation $\mathbf{T}$ is determined, the loadings $\tilde{\ell}_{ij}^*$ are multiplied by $\hat{h}_i$ so that the original communalities are preserved.

Although (9-45) looks rather forbidding, it has a simple interpretation. In words,

$$V \propto \sum_{j=1}^{m}\begin{pmatrix}\text{variance of squares of (scaled) loadings for} \\ j\text{th factor}\end{pmatrix} \tag{9-46}$$

Effectively, maximizing V corresponds to "spreading out" the squares of the loadings on each factor as much as possible. Therefore, we hope to find groups of large and negligible coefficients in any *column* of the rotated loadings matrix $\hat{\mathbf{L}}^*$.

Computing algorithms exist for maximizing V, and most popular factor analysis computer programs (for example, the statistical software packages SAS, SPSS, BMDP, and MINITAB) provide varimax rotations. As might be expected, varimax rotations of factor loadings obtained by different solution methods (principal components, maximum likelihood, and so forth) will not, in general, coincide. Also, the pattern of rotated loadings may change considerably if additional common factors are included in the rotation. If a dominant single factor exists, it will generally be obscured by any orthogonal rotation. By contrast, it can always be held fixed and the remaining factors rotated.

Example 9.9 (Rotated loadings for the consumer-preference data) Let us return to the marketing data discussed in Example 9.3. The original factor loadings (obtained by the principal component method), the communalities, and the (varimax) rotated factor loadings are shown in Table 9.7. (See the SAS statistical software output in Panel 9.1.)

Table 9.7

Variable	Estimated factor loadings F_1	F_2	Rotated estimated factor loadings F_1^*	F_2^*	Communalities $\tilde{h}_i^2$
1. Taste	.56	.82	.02	.99	.98
2. Good buy for money	.78	−.52	.94	−.01	.88
3. Flavor	.65	.75	.13	.98	.98
4. Suitable for snack	.94	−.10	.84	.43	.89
5. Provides lots of energy	.80	−.54	.97	−.02	.93
Cumulative proportion of total (standardized) sample variance explained	.571	.932	.507	.932	

It is clear that variables 2, 4, and 5 define factor 1 (high loadings on factor 1, small or negligible loadings on factor 2), while variables 1 and 3 define factor 2 (high loadings on factor 2, small or negligible loadings on factor 1). Variable 4 is most closely aligned with factor 1, although it has aspects of the trait represented by factor 2. We might call factor 1 a *nutritional* factor and factor 2 a *taste* factor.

The factor loadings for the variables are pictured with respect to the original and (varimax) rotated factor axes in Figure 9.2. ∎

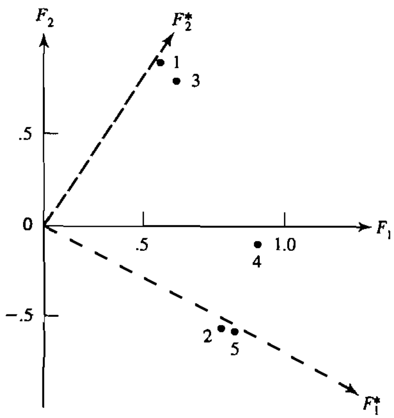


Figure 9.2 Factor rotation for hypothetical marketing data.

PANEL 9.1 SAS ANALYSIS FOR EXAMPLE 9.9 USING PROC FACTOR.

```
title 'Factor Analysis';
data consumer(type = corr);
_type_='CORR';
input _name_$ taste money flavor snack energy;
cards;
taste      1.00    .      .      .      .
money       .02   1.00    .      .      .
flavor      .96    .13   1.00    .      .
snack       .42    .71    .50   1.00    .
energy      .01    .85    .11    .79   1.00
;
proc factor res data=consumer
   method=prin nfact=2rotate=varimax preplot plot;
   var taste money flavor snack energy;
```

PROGRAM COMMANDS

OUTPUT

Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 5 Average = 1

	1	2	3	4	5
Eigenvalue	2.853090	1.806332	0.204490	0.102409	0.033677
Difference	1.046758	1.601842	0.102081	0.068732	
Proportion	0.5706	0.3613	0.0409	0.0205	0.0067
Cumulative	0.5706	0.9319	0.9728	0.9933	1.0000

2 factors will be retained by the NFACTOR criterion.

Factor Pattern

	FACTOR1	FACTOR2
TASTE	0.55986	0.81610
MONEY	0.77726	−0.52420
FLAVOR	0.64534	0.74795
SNACK	0.93911	−0.10492
ENERGY	0.79821	−0.54323

Final Communality Estimates: Total = 4.659423

TASTE	MONEY	FLAVOR	SNACK	ENERGY
0.97961	0.878920	0.975883	0.892928	0.932231

(continues on next page)

PANEL 9.1 *(continued)*

Rotation Method: Varimax

Rotated Factor Pattern

	FACTOR1	FACTOR2
TASTE	0.01970	0.98948
MONEY	0.93744	−0.01123
FLAVOR	0.12856	0.97947
SNACK	0.84244	0.42805
ENERGY	0.96539	−0.01563

Variance explained by each factor

FACTOR1	FACTOR2
2.537396	2.122027

Rotation of factor loadings is recommended particularly for loadings obtained by maximum likelihood, since the initial values are constrained to satisfy the uniqueness condition that $\hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}}$ be a diagonal matrix. This condition is convenient for computational purposes, but may not lead to factors that can easily be interpreted.

Example 9.10 (Rotated loadings for the stock-price data) Table 9.8 shows the initial and rotated maximum likelihood estimates of the factor loadings for the stock-price data of Examples 8.5 and 9.5. An $m = 2$ factor model is assumed. The estimated

Table 9.8

Variable	Maximum likelihood estimates of factor loadings F_1	F_2	Rotated estimated factor loadings F_1^*	F_2^*	Specific variances $\hat{\psi}_i^2 = 1 - \hat{h}_i^2$
J P Morgan	.115	.755	.763	.024	.42
Citibank	.322	.788	.821	.227	.27
Wells Fargo	.182	.652	.669	.104	.54
Royal Dutch Shell	1.000	−.000	.118	.993	.00
ExxonMobil	.683	.032	.113	.675	.53
Cumulative proportion of total sample variance explained	.323	.647	.346	.647	

specific variances and cumulative proportions of the total (standardized) sample variance explained by each factor are also given.

An interpretation of the factors suggested by the unrotated loadings was presented in Example 9.5. We identified *market* and *industry* factors.

The rotated loadings indicate that the bank stocks (JP Morgan, Citibank, and Wells Fargo) load highly on the first factor, while the oil stocks (Royal Dutch Shell and ExxonMobil) load highly on the second factor. (Although the rotated loadings obtained from the principal component solution are not displayed, the same phenomenon is observed for them.) The two rotated factors, together, differentiate the industries. It is difficult for us to label these factors intelligently. Factor 1 represents those unique economic forces that cause bank stocks to move together. Factor 2 appears to represent economic conditions affecting oil stocks.

As we have noted, a general factor (that is, one on which *all* the variables load highly) tends to be "destroyed after rotation." For this reason, in cases where a general factor is evident, an orthogonal rotation is sometimes performed with the general factor loadings fixed.[5] ■

Example 9.11 (Rotated loadings for the Olympic decathlon data) The estimated factor loadings and specific variances for the Olympic decathlon data were presented in Example 9.6. These quantities were derived for an $m = 4$ factor model, using both principal component and maximum likelihood solution methods. The interpretation of all the underlying factors was not immediately evident. A varimax rotation [see (9-45)] was performed to see whether the rotated factor loadings would provide additional insights. The varimax rotated loadings for the $m = 4$ factor solutions are displayed in Table 9.9, along with the specific variances. Apart from the estimated loadings, rotation will affect only the *distribution* of the proportions of the total sample variance explained by each factor. The cumulative proportion of the total sample variance explained for *all* factors does not change.

The rotated factor loadings for both methods of solution point to the same underlying attributes, although factors 1 and 2 are not in the same order. We see that shot put, discus, and javelin load highly on a factor, and, following Linden [11], this factor might be called *explosive arm strength*. Similarly, high jump, 110-meter hurdles, pole vault, and—to some extent—long jump load highly on another factor. Linden labeled this factor *explosive leg strength*. The 100-meter run, 400-meter run, and—again to some extent—the long jump load highly on a third factor. This factor could be called *running speed*. Finally, the 1500-meter run loads heavily and the 400-meter run loads heavily on the fourth factor. Linden called this factor *running endurance*. As he notes, "The basic functions indicated in this study are mainly consistent with the traditional classification of track and field athletics."

[5]Some general-purpose factor analysis programs allow one to fix loadings associated with certain factors and to rotate the remaining factors.

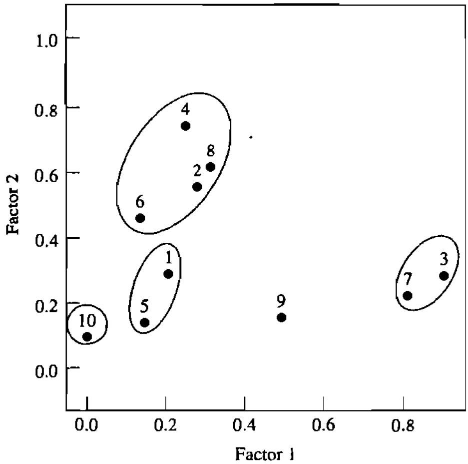
Table 9.9

	Principal component					Maximum likelihood				
	Estimated rotated factor loadings, $\tilde{\ell}^*_{ij}$				Specific variances	Estimated rotated factor loadings, $\hat{\ell}^*_{ij}$				Specific variances
Variable	F_1^*	F_2^*	F_3^*	F_4^*	$\tilde{\psi}_i = 1 - \tilde{h}_i^2$	F_1^*	F_2^*	F_3^*	F_4^*	$\hat{\psi}_i = 1 - \hat{h}_i^2$
100-m run	.182	$\boxed{.885}$	.205	−.139	.12	.204	.296	$\boxed{.928}$	−.005	.01
Long jump	.291	$\boxed{.664}$	$\boxed{.429}$	.055	.29	.280	$\boxed{.554}$	$\boxed{.451}$	.155	.39
Shot put	$\boxed{.819}$	.302	.252	−.097	.17	$\boxed{.883}$	.278	.228	−.045	.09
High jump	.267	.221	$\boxed{.683}$	.293	.33	.254	$\boxed{.739}$	.057	.242	.33
400-m run	.086	$\boxed{.747}$	.068	$\boxed{.507}$	.17	.142	.151	$\boxed{.519}$	$\boxed{.700}$	.20
110-m hurdles	.048	.108	$\boxed{.826}$	−.161	.28	.136	$\boxed{.465}$	.173	−.033	.73
Discus	$\boxed{.832}$	.185	.204	−.076	.23	$\boxed{.793}$	.220	.133	−.009	.30
Pole vault	.324	.278	$\boxed{.656}$	.293	.30	.314	$\boxed{.613}$	.169	.279	.42
Javelin	$\boxed{.754}$	.024	.054	.188	.39	$\boxed{.477}$	.160	.041	.139	.73
1500-m run	−.002	.019	.075	$\boxed{.921}$	.15	.001	.110	−.070	$\boxed{.619}$	.60
Cumulative proportion of total sample variance explained	.22	.43	.62	.76		.20	.37	.51	.62	

Plots of rotated maximum likelihood loadings for factors pairs $(1,2)$ and $(1,3)$ are displayed in Figure 9.3 on page 513. The points are generally grouped along the factor axes. Plots of rotated principal component loadings are very similar. ■

Oblique Rotations

Orthogonal rotations are appropriate for a factor model in which the common factors are assumed to be independent. Many investigators in social sciences consider *oblique* (nonorthogonal) rotations, as well as orthogonal rotations. The former are



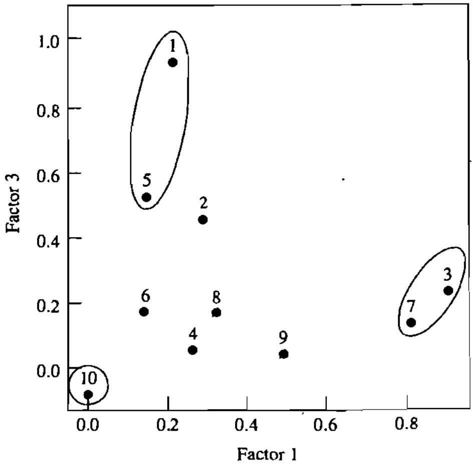


Figure 9.3 Rotated maximum likelihood loadings for factor pairs (1, 2) and (1, 3)—decathlon data. (The numbers in the figures correspond to variables.)

often suggested after one views the estimated factor loadings and do not follow from our postulated model. Nevertheless, an oblique rotation is frequently a useful aid in factor analysis.

If we regard the m common factors as coordinate axes, the point with the m coordinates $(\hat{\ell}_{i1}, \hat{\ell}_{i2}, \ldots, \hat{\ell}_{im})$ represents the position of the ith variable in the *factor space*. Assuming that the variables are grouped into nonoverlapping clusters, an orthogonal rotation to a simple structure corresponds to a *rigid* rotation of the coordinate axes such that the axes, after rotation, pass as closely to the clusters as possible. An oblique rotation to a simple structure corresponds to a *nonrigid* rotation of the coordinate system such that the rotated axes (no longer perpendicular) pass (nearly) through the clusters. An oblique rotation seeks to express each variable in terms of a minimum number of factors—preferably, a single factor. Oblique rotations are discussed in several sources (see, for example, [6] or [10]) and will not be pursued in this book.

9.5 Factor Scores

In factor analysis, interest is usually centered on the parameters in the factor model. However, the estimated values of the common factors, called *factor scores*, may also be required. These quantities are often used for diagnostic purposes, as well as inputs to a subsequent analysis.

Factor scores are not estimates of unknown parameters in the usual sense. Rather, they are estimates of values for the unobserved random factor vectors $\mathbf{F}_j$, $j = 1, 2, \ldots, n$. That is, factor scores

$$\hat{\mathbf{f}}_j = \text{estimate of the values } \mathbf{f}_j \text{ attained by } \mathbf{F}_j \ (j\text{th case})$$

The estimation situation is complicated by the fact that the unobserved quantities $\mathbf{f}_j$ and $\boldsymbol{\varepsilon}_j$ outnumber the observed $\mathbf{x}_j$. To overcome this difficulty, some rather heuristic, but reasoned, approaches to the problem of estimating factor values have been advanced. We describe two of these approaches.

Both of the factor score approaches have two elements in common:

1. They treat the estimated factor loadings $\hat{\ell}_{ij}$ and specific variances $\hat{\psi}_i$ as if they were the true values.

2. They involve linear transformations of the original data, perhaps centered or standardized. Typically, the estimated *rotated* loadings, rather than the original estimated loadings, are used to compute factor scores. The computational formulas, as given in this section, do not change when rotated loadings are substituted for unrotated loadings, so we will not differentiate between them.

The Weighted Least Squares Method

Suppose first that the mean vector $\boldsymbol{\mu}$, the factor loadings $\mathbf{L}$, and the specific variance $\boldsymbol{\Psi}$ are known for the factor model

$$\underset{(p\times 1)}{\mathbf{X}} - \underset{(p\times 1)}{\boldsymbol{\mu}} = \underset{(p\times m)}{\mathbf{L}}\underset{(m\times 1)}{\mathbf{F}} + \underset{(p\times 1)}{\boldsymbol{\varepsilon}}$$

Further, regard the specific factors $\boldsymbol{\varepsilon}' = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p]$ as errors. Since $\text{Var}(\varepsilon_i) = \psi_i$, $i = 1, 2, \ldots, p$, need not be equal, Bartlett [2] has suggested that weighted least squares be used to estimate the common factor values.

The sum of the squares of the errors, weighted by the reciprocal of their variances, is

$$\sum_{i=1}^{p} \frac{\varepsilon_i^2}{\psi_i} = \boldsymbol{\varepsilon}'\boldsymbol{\Psi}^{-1}\boldsymbol{\varepsilon} = (\mathbf{x} - \boldsymbol{\mu} - \mathbf{L}\mathbf{f})'\boldsymbol{\Psi}^{-1}(\mathbf{x} - \boldsymbol{\mu} - \mathbf{L}\mathbf{f}) \tag{9-47}$$

Bartlett proposed choosing the estimates $\hat{\mathbf{f}}$ of $\mathbf{f}$ to minimize (9-47). The solution (see Exercise 7.3) is

$$\hat{\mathbf{f}} = (\mathbf{L}'\boldsymbol{\Psi}^{-1}\mathbf{L})^{-1}\mathbf{L}'\boldsymbol{\Psi}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \tag{9-48}$$

Motivated by (9-48), we take the estimates $\hat{\mathbf{L}}$, $\hat{\boldsymbol{\Psi}}$, and $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ as the true values and obtain the factor scores for the jth case as

$$\hat{\mathbf{f}}_j = (\hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}})^{-1}\hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}) \tag{9-49}$$

When $\hat{\mathbf{L}}$ and $\hat{\boldsymbol{\Psi}}$ are determined by the maximum likelihood method, these estimates must satisfy the uniqueness condition, $\hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}} = \hat{\boldsymbol{\Delta}}$, a diagonal matrix. We then have the following:

Factor Scores Obtained by Weighted Least Squares from the Maximum Likelihood Estimates

$$\hat{\mathbf{f}}_j = (\hat{\mathbf{L}}'\hat{\mathbf{\Psi}}^{-1}\hat{\mathbf{L}})^{-1}\hat{\mathbf{L}}'\hat{\mathbf{\Psi}}^{-1}(\mathbf{x}_j - \hat{\boldsymbol{\mu}})$$
$$= \hat{\mathbf{\Delta}}^{-1}\hat{\mathbf{L}}'\hat{\mathbf{\Psi}}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}), \qquad j = 1,2,\dots,n$$

or, if the correlation matrix is factored (9-50)

$$\hat{\mathbf{f}}_j = (\hat{\mathbf{L}}_{\mathbf{z}}'\hat{\mathbf{\Psi}}_{\mathbf{z}}^{-1}\hat{\mathbf{L}}_{\mathbf{z}})^{-1}\hat{\mathbf{L}}_{\mathbf{z}}'\hat{\mathbf{\Psi}}_{\mathbf{z}}^{-1}\mathbf{z}_j$$
$$= \hat{\mathbf{\Delta}}_{\mathbf{z}}^{-1}\hat{\mathbf{L}}_{\mathbf{z}}'\hat{\mathbf{\Psi}}_{\mathbf{z}}^{-1}\mathbf{z}_j, \qquad j = 1,2,\dots,n$$

where $\mathbf{z}_j = \mathbf{D}^{-1/2}(\mathbf{x}_j - \bar{\mathbf{x}})$, as in (8-25), and $\hat{\boldsymbol{\rho}} = \hat{\mathbf{L}}_{\mathbf{z}}\hat{\mathbf{L}}_{\mathbf{z}}' + \hat{\mathbf{\Psi}}_{\mathbf{z}}$.

The factor scores generated by (9-50) have sample mean vector $\mathbf{0}$ and zero sample covariances. (See Exercise 9.16.)

If rotated loadings $\hat{\mathbf{L}}^* = \hat{\mathbf{L}}\mathbf{T}$ are used in place of the original loadings in (9-50), the subsequent factor scores, $\hat{\mathbf{f}}_j^*$, are related to $\hat{\mathbf{f}}_j$ by $\hat{\mathbf{f}}_j^* = \mathbf{T}'\hat{\mathbf{f}}_j$, $j = 1, 2, \dots, n$.

Comment. If the factor loadings are estimated by the principal component method, it is customary to generate factor scores using an unweighted (ordinary) least squares procedure. Implicitly, this amounts to assuming that the ψ_i are equal or nearly equal. The factor scores are then

$$\hat{\mathbf{f}}_j = (\tilde{\mathbf{L}}'\tilde{\mathbf{L}})^{-1}\tilde{\mathbf{L}}'(\mathbf{x}_j - \bar{\mathbf{x}})$$

or

$$\hat{\mathbf{f}}_j = (\tilde{\mathbf{L}}_{\mathbf{z}}'\tilde{\mathbf{L}}_{\mathbf{z}})^{-1}\tilde{\mathbf{L}}_{\mathbf{z}}'\mathbf{z}_j$$

for standardized data. Since $\tilde{\mathbf{L}} = \left[\sqrt{\hat{\lambda}_1}\,\hat{\mathbf{e}}_1 \mid \sqrt{\hat{\lambda}_2}\,\hat{\mathbf{e}}_2 \mid \cdots \mid \sqrt{\hat{\lambda}_m}\,\hat{\mathbf{e}}_m\right]$ [see (9-15)], we have

$$\hat{\mathbf{f}}_j = \begin{bmatrix} \dfrac{1}{\sqrt{\hat{\lambda}_1}}\hat{\mathbf{e}}_1'(\mathbf{x}_j - \bar{\mathbf{x}}) \\ \dfrac{1}{\sqrt{\hat{\lambda}_2}}\hat{\mathbf{e}}_2'(\mathbf{x}_j - \bar{\mathbf{x}}) \\ \vdots \\ \dfrac{1}{\sqrt{\hat{\lambda}_m}}\hat{\mathbf{e}}_m'(\mathbf{x}_j - \bar{\mathbf{x}}) \end{bmatrix} \tag{9-51}$$

For these factor scores,

$$\frac{1}{n}\sum_{j=1}^{n}\hat{\mathbf{f}}_j = \mathbf{0} \qquad \text{(sample mean)}$$

and

$$\frac{1}{n-1}\sum_{j=1}^{n}\hat{\mathbf{f}}_j\hat{\mathbf{f}}_j' = \mathbf{I} \qquad \text{(sample covariance)}$$

Comparing (9-51) with (8-21), we see that the $\hat{\mathbf{f}}_j$ are nothing more than the first m (scaled) principal components, evaluated at $\mathbf{x}_j$.

The Regression Method

Starting again with the original factor model $\mathbf{X} - \boldsymbol{\mu} = \mathbf{LF} + \boldsymbol{\varepsilon}$, we initially treat the loadings matrix $\mathbf{L}$ and specific variance matrix $\boldsymbol{\Psi}$ as known. When the common factors $\mathbf{F}$ and the specific factors (or errors) $\boldsymbol{\varepsilon}$ are jointly normally distributed with means and covariances given by (9-3), the linear combination $\mathbf{X} - \boldsymbol{\mu} = \mathbf{LF} + \boldsymbol{\varepsilon}$ has an $N_p(\mathbf{0}, \mathbf{LL}' + \boldsymbol{\Psi})$ distribution. (See Result 4.3.) Moreover, the joint distribution of $(\mathbf{X} - \boldsymbol{\mu})$ and $\mathbf{F}$ is $N_{m+p}(\mathbf{0}, \boldsymbol{\Sigma}^*)$, where

$$\underset{(m+p)\times(m+p)}{\boldsymbol{\Sigma}^*} = \left[\begin{array}{c|c} \boldsymbol{\Sigma} = \underset{(p\times p)}{\mathbf{LL}' + \boldsymbol{\Psi}} & \underset{(p\times m)}{\mathbf{L}} \\ \hline \underset{(m\times p)}{\mathbf{L}'} & \underset{(m\times m)}{\mathbf{I}} \end{array}\right] \tag{9-52}$$

and $\mathbf{0}$ is an $(m + p) \times 1$ vector of zeros. Using Result 4.6, we find that the conditional distribution of $\mathbf{F}|\mathbf{x}$ is multivariate normal with

$$\text{mean} = E(\mathbf{F}\,|\,\mathbf{x}) = \mathbf{L}'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \mathbf{L}'(\mathbf{LL}' + \boldsymbol{\Psi})^{-1}(\mathbf{x} - \boldsymbol{\mu}) \tag{9-53}$$

and

$$\text{covariance} = \text{Cov}(\mathbf{F}\,|\,\mathbf{x}) = \mathbf{I} - \mathbf{L}'\boldsymbol{\Sigma}^{-1}\mathbf{L} = \mathbf{I} - \mathbf{L}'(\mathbf{LL}' + \boldsymbol{\Psi})^{-1}\mathbf{L} \tag{9-54}$$

The quantities $\mathbf{L}'(\mathbf{LL}' + \boldsymbol{\Psi})^{-1}$ in (9-53) are the coefficients in a (multivariate) regression of the factors on the variables. Estimates of these coefficients produce factor scores that are analogous to the estimates of the conditional mean values in multivariate regression analysis. (See Chapter 7.) Consequently, given any vector of observations $\mathbf{x}_j$, and taking the maximum likelihood estimates $\hat{\mathbf{L}}$ and $\hat{\boldsymbol{\Psi}}$ as the true values, we see that the jth factor score vector is given by

$$\hat{\mathbf{f}}_j = \hat{\mathbf{L}}'\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}) = \hat{\mathbf{L}}'(\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}})^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}), \qquad j = 1, 2, \ldots, n \tag{9-55}$$

The calculation of $\hat{\mathbf{f}}_j$ in (9-55) can be simplified by using the matrix identity (see Exercise 9.6)

$$\underset{(m\times p)}{\hat{\mathbf{L}}'}\,(\hat{\mathbf{L}}\hat{\mathbf{L}}' + \underset{(p\times p)}{\hat{\boldsymbol{\Psi}}})^{-1} = \underset{(m\times m)}{(\mathbf{I} + \hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}})^{-1}}\,\underset{(m\times p)}{\hat{\mathbf{L}}'}\,\underset{(p\times p)}{\hat{\boldsymbol{\Psi}}^{-1}} \tag{9-56}$$

This identity allows us to compare the factor scores in (9-55), generated by the regression argument, with those generated by the weighted least squares procedure [see (9-50)]. Temporarily, we denote the former by $\hat{\mathbf{f}}_j^R$ and the latter by $\hat{\mathbf{f}}_j^{LS}$. Then, using (9-56), we obtain

$$\hat{\mathbf{f}}_j^{LS} = (\hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}})^{-1}(\mathbf{I} + \hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}})\hat{\mathbf{f}}_j^R = (\mathbf{I} + (\hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}})^{-1})\hat{\mathbf{f}}_j^R \tag{9-57}$$

For maximum likelihood estimates $(\hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}})^{-1} = \hat{\boldsymbol{\Delta}}^{-1}$ and if the elements of this diagonal matrix are close to zero, the regression and generalized least squares methods will give nearly the same factor scores.

In an attempt to reduce the effects of a (possibly) incorrect determination of the number of factors, practitioners tend to calculate the factor scores in (9-55) by using $\mathbf{S}$ (the original sample covariance matrix) instead of $\hat{\mathbf{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}$. We then have the following:

Factor Scores Obtained by Regression

$$\hat{\mathbf{f}}_j = \hat{\mathbf{L}}'\mathbf{S}^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}), \qquad j = 1, 2, \ldots, n$$

or, if a correlation matrix is factored, (9-58)

$$\hat{\mathbf{f}}_j = \hat{\mathbf{L}}_{\mathbf{z}}'\mathbf{R}^{-1}\mathbf{z}_j, \qquad j = 1, 2, \ldots, n$$

where, see (8-25),

$$\mathbf{z}_j = \mathbf{D}^{-1/2}(\mathbf{x}_j - \bar{\mathbf{x}}) \quad \text{and} \quad \hat{\boldsymbol{\rho}} = \hat{\mathbf{L}}_{\mathbf{z}}\hat{\mathbf{L}}_{\mathbf{z}}' + \hat{\mathbf{\Psi}}_{\mathbf{z}}$$

Again, if rotated loadings $\hat{\mathbf{L}}^* = \hat{\mathbf{L}}\mathbf{T}$ are used in place of the original loadings in (9-58), the subsequent factor scores $\hat{\mathbf{f}}_j^*$ are related to $\hat{\mathbf{f}}_j$ by

$$\hat{\mathbf{f}}_j^* = \mathbf{T}'\hat{\mathbf{f}}_j, \qquad j = 1, 2, \ldots, n$$

A numerical measure of agreement between the factor scores generated from two *different* calculation methods is provided by the sample correlation coefficient between scores on the same factor. Of the methods presented, none is recommended as uniformly superior.

Example 9.12 (Computing factor scores) We shall illustrate the computation of factor scores by the least squares and regression methods using the stock-price data discussed in Example 9.10. A maximum likelihood solution from $\mathbf{R}$ gave the estimated rotated loadings and specific variances

$$\hat{\mathbf{L}}_{\mathbf{z}}^* = \begin{bmatrix} .763 & .024 \\ .821 & .227 \\ .669 & .104 \\ .118 & .993 \\ .113 & .675 \end{bmatrix} \quad \text{and} \quad \hat{\mathbf{\Psi}}_{\mathbf{z}} = \begin{bmatrix} .42 & 0 & 0 & 0 & 0 \\ 0 & .27 & 0 & 0 & 0 \\ 0 & 0 & .54 & 0 & 0 \\ 0 & 0 & 0 & .00 & 0 \\ 0 & 0 & 0 & 0 & .53 \end{bmatrix}$$

The vector of standardized observations,

$$\mathbf{z}' = [.50, -1.40, -.20, -.70, 1.40]$$

yields the following scores on factors 1 and 2:

Weighted least squares (9-50):[6]

$$\hat{\mathbf{f}} = (\hat{\mathbf{L}}_z^{*\prime}\hat{\boldsymbol{\Psi}}_z^{-1}\hat{\mathbf{L}}_z^*)^{-1}\hat{\mathbf{L}}_z^{*\prime}\hat{\boldsymbol{\Psi}}_z^{-1}\mathbf{z} = \begin{bmatrix} -.61 \\ -.61 \end{bmatrix}$$

Regression (9-58):

$$\hat{\mathbf{f}} = \hat{\mathbf{L}}_z^{*\prime}\mathbf{R}^{-1}\mathbf{z} = \begin{bmatrix} .331 & .526 & .221 & -.137 & .011 \\ -.040 & -.063 & -.026 & 1.023 & -.001 \end{bmatrix} \begin{bmatrix} .50 \\ -1.40 \\ -.20 \\ -.70 \\ 1.40 \end{bmatrix} = \begin{bmatrix} -.50 \\ -.64 \end{bmatrix}$$

In this case, the two methods produce very similar results. All of the regression factor scores, obtained using (9-58), are plotted in Figure 9.4.

Comment. Factor scores with a rather pleasing intuitive property can be constructed very simply. Group the variables with high (say, greater than .40 in absolute value) loadings on a factor. The scores for factor 1 are then formed by summing the (standardized) observed values of the variables in the group, combined according to the sign of the loadings. The factor scores for factor 2 are the sums of the standardized observations corresponding to variables with high loadings

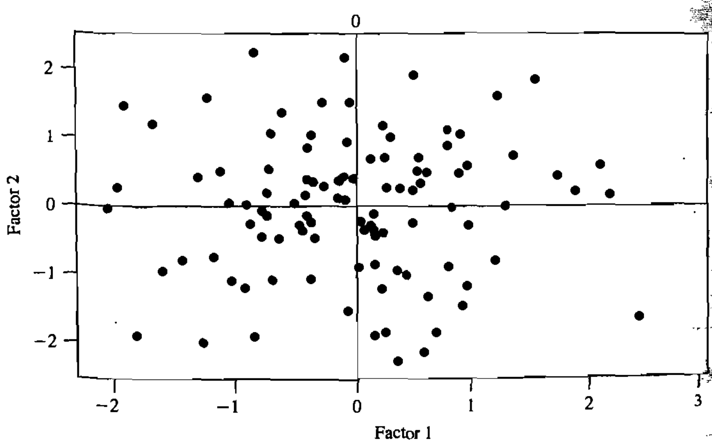


Figure 9.4 Factor scores using (9-58) for factors 1 and 2 of the stock-price data (maximum likelihood estimates of the factor loadings).

[6] In order to calculate the weighted least squares factor scores, .00 in the fourth diagonal position of $\hat{\boldsymbol{\Psi}}_z$ was set to .01 so that this matrix could be inverted.

on factor 2, and so forth. Data reduction is accomplished by replacing the standardized data by these simple factor scores. The simple factor scores are frequently highly correlated with the factor scores obtained by the more complex least squares and regression methods.

Example 9.13 (Creating simple summary scores from factor analysis groupings) The principal component factor analysis of the stock price data in Example 9.4 produced the estimated loadings

$$\tilde{\mathbf{L}} = \begin{bmatrix} .732 & -.437 \\ .831 & -.280 \\ .726 & -.374 \\ .605 & .694 \\ .563 & .719 \end{bmatrix} \quad \text{and} \quad \tilde{\mathbf{L}}^* = \tilde{\mathbf{L}}\mathbf{T} = \begin{bmatrix} .852 & .030 \\ .851 & .214 \\ .813 & .079 \\ .133 & .911 \\ .084 & .909 \end{bmatrix}$$

For each factor, take the loadings with largest absolute value in $\tilde{\mathbf{L}}$ as equal in magnitude, and neglect the smaller loadings. Thus, we create the linear combinations

$$\hat{f}_1 = x_1 + x_2 + x_3 + x_4 + x_5$$
$$\hat{f}_2 = x_4 + x_5 - x_1$$

as a summary. In practice, we would standardize these new variables.

If, instead of $\tilde{\mathbf{L}}$, we start with the varimax rotated loadings $\tilde{\mathbf{L}}^*$, the simple factor scores would be

$$\hat{f}_1 = x_1 + x_2 + x_3$$
$$\hat{f}_2 = x_4 + x_5$$

The identification of high loadings and negligible loadings is really quite subjective. Linear compounds that make subject-matter sense are preferable. ■

Although multivariate normality is often assumed for the variables in a factor analysis, it is very difficult to justify the assumption for a large number of variables. As we pointed out in Chapter 4, marginal transformations may help. Similarly, the factor scores may or may not be normally distributed. Bivariate scatter plots of factor scores can produce all sorts of nonelliptical shapes. Plots of factor scores should be examined prior to using these scores in other analyses. They can reveal outlying values and the extent of the (possible) nonnormality.

9.6 Perspectives and a Strategy for Factor Analysis

There are many decisions that must be made in any factor analytic study. Probably the most important decision is the choice of m, the number of common factors. Although a large sample test of the adequacy of a model is available for a given m, it is suitable only for data that are approximately normally distributed. Moreover, the test will most assuredly reject the model for small m if the number of variables and observations is large. Yet this is the situation when factor analysis provides a useful approximation. Most often, the final choice of m is based on some combination of

(1) the proportion of the sample variance explained, (2) subject-matter knowledge, and (3) the "reasonableness" of the results.

The choice of the solution method and type of rotation is a less crucial decision. In fact, the most satisfactory factor analyses are those in which rotations are tried with more than one method and all the results substantially confirm the same factor structure.

At the present time, factor analysis still maintains the flavor of an art, and no single strategy should yet be "chiseled into stone." We suggest and illustrate one reasonable option:

1. *Perform a principal component factor analysis.* This method is particularly appropriate for a first pass through the data. (It is not required that $\mathbf{R}$ or $\mathbf{S}$ be nonsingular.)
   - **(a)** Look for suspicious observations by plotting the factor scores. Also, calculate standardized scores for each observation and squared distances as described in Section 4.6.
   - **(b)** Try a varimax rotation.
2. *Perform a maximum likelihood factor analysis, including a varimax rotation.*
3. *Compare the solutions obtained from the two factor analyses.*
   - **(a)** Do the loadings group in the same manner?
   - **(b)** Plot factor scores obtained for principal components against scores from the maximum likelihood analysis.
4. *Repeat the first three steps for other numbers of common factors m.* Do extra factors necessarily contribute to the understanding and interpretation of the data?
5. *For large data sets, split them in half and perform a factor analysis on each part.* Compare the two results with each other and with that obtained from the complete data set to check the stability of the solution. (The data might be divided by placing the first half of the cases in one group and the second half of the cases in the other group. This would reveal changes over time.)

Example 9.14 (Factor analysis of chicken-bone data) We present the results of several factor analyses on bone and skull measurements of white leghorn fowl. The original data were taken from Dunn [5]. Factor analysis of Dunn's data was originally considered by Wright [15], who started his analysis from a different correlation matrix than the one we use.

The full data set consists of $n = 276$ measurements on bone dimensions:

$$\text{Head:}\quad \begin{cases} X_1 = \text{skull length} \\ X_2 = \text{skull breadth} \end{cases}$$

$$\text{Leg:}\quad \begin{cases} X_3 = \text{femur length} \\ X_4 = \text{tibia length} \end{cases}$$

$$\text{Wing:}\quad \begin{cases} X_5 = \text{humerus length} \\ X_6 = \text{ulna length} \end{cases}$$

The sample correlation matrix

$$\mathbf{R} = \begin{bmatrix} 1.000 & .505 & .569 & .602 & .621 & .603 \\ .505 & 1.000 & .422 & .467 & .482 & .450 \\ .569 & .422 & 1.000 & .926 & .877 & .878 \\ .602 & .467 & .926 & 1.000 & .874 & .894 \\ .621 & .482 & .877 & .874 & 1.000 & .937 \\ .603 & .450 & .878 & .894 & .937 & 1.000 \end{bmatrix}$$

was factor analyzed by the principal component and maximum likelihood methods for an $m = 3$ factor model. The results are given in Table 9.10.[7]

Table 9.10 Factor Analysis of Chicken-Bone Data

Principal Component

Variable	Estimated factor loadings			Rotated estimated loadings			
	F_1	F_2	F_3	F_1^*	F_2^*	F_3^*	$\tilde{\psi}_i$
1. Skull length	.741	.350	.573	.355	.244	.902	.00
2. Skull breadth	.604	.720	−.340	.235	.949	.211	.00
3. Femur length	.929	−.233	−.075	.921	.164	.218	.08
4. Tibia length	.943	−.175	−.067	.904	.212	.252	.08
5. Humerus length	.948	−.143	−.045	.888	.228	.283	.08
6. Ulna length	.945	−.189	−.047	.908	.192	.264	.07
Cumulative proportion of total (standardized) sample variance explained	.743	.873	.950	.576	.763	.950	

Maximum Likelihood

Variable	Estimated factor loadings			Rotated estimated loadings			
	F_1	F_2	F_3	F_1^*	F_2^*	F_3^*	$\hat{\psi}_i$
1. Skull length	.602	.214	.286	.467	.506	.128	.51
2. Skull breadth	.467	.177	.652	.211	.792	.050	.33
3. Femur length	.926	.145	−.057	.890	.289	.084	.12
4. Tibia length	1.000	.000	−.000	.936	.345	−.073	.00
5. Humerus length	.874	.463	−.012	.831	.362	.396	.02
6. Ulna length	.894	.336	−.039	.857	.325	.272	.09
Cumulative proportion of total (standardized) sample variance explained	.667	.738	.823	.559	.779	.823	

[7] Notice the estimated specific variance of .00 for tibia length in the maximum likelihood solution. This suggests that maximizing the likelihood function may produce a Heywood case. Readers attempting to replicate our results should try the Hey(wood) option if SAS or similar software is used.

After rotation, the two methods of solution appear to give somewhat different results. Focusing our attention on the principal component method and the cumulative proportion of the total sample variance explained, we see that a three-factor solution appears to be warranted. The third factor explains a "significant" amount of additional sample variation. The first factor appears to be a *body-size* factor dominated by wing and leg dimensions. The second and third factors, collectively, represent skull dimensions and might be given the same names as the variables, *skull breadth* and *skull length*, respectively.

The rotated maximum likelihood factor loadings are consistent with those generated by the principal component method for the first factor, but not for factors 2 and 3. For the maximum likelihood method, the second factor appears to represent head size. The meaning of the third factor is unclear, and it is probably not needed.

Further support for retaining three or fewer factors is provided by the residual matrix obtained from the maximum likelihood estimates:

$$\mathbf{R} - \hat{\mathbf{L}}_z\hat{\mathbf{L}}_z' - \hat{\mathbf{\Psi}}_z = \begin{bmatrix} .000 & & & & & \\ -.000 & .000 & & & & \\ -.003 & .001 & .000 & & & \\ .000 & .000 & .000 & .000 & & \\ -.001 & .000 & .000 & .000 & .000 & \\ .004 & -.001 & -.001 & .000 & -.000 & .000 \end{bmatrix}$$

All of the entries in this matrix are very small. We shall pursue the $m = 3$ factor model in this example. An $m = 2$ factor model is considered in Exercise 9.10.

Factor scores for factors 1 and 2 produced from (9-58) with the rotated maximum likelihood estimates are plotted in Figure 9.5. Plots of this kind allow us to identify observations that, for one reason or another, are not consistent with the remaining observations. Potential outliers are circled in the figure.

It is also of interest to plot pairs of factor scores obtained using the principal component and maximum likelihood estimates of factor loadings. For the chicken-bone data, plots of pairs of factor scores are given in Figure 9.6 on pages 524–526. If the loadings on a particular factor agree, the pairs of scores should cluster tightly about the 45° line through the origin. Sets of loadings that do not agree will produce factor scores that deviate from this pattern. If the latter occurs, it is usually associated with the last factors and may suggest that the number of factors is too large. That is, the last factors are not meaningful. This seems to be the case with the third factor in the chicken-bone data, as indicated by Plot (c) in Figure 9.6.

Plots of pairs of factor scores using estimated loadings from two solution methods are also good tools for detecting outliers. If the sets of loadings for a factor tend to agree, outliers will appear as points in the neighborhood of the 45° line, but far from the origin and the cluster of the remaining points. It is clear from Plot (b) in Figure 9.6 that one of the 276 observations is not consistent with the others. It has an unusually large F_2-score. When this point, $[39.1, 39.3, 75.7, 115, 73.4, 69.1]$, was removed and the analysis repeated, the loadings were not altered appreciably.

When the data set is large, it should be divided into two (roughly) equal sets, and a factor analysis should be performed on each half. The results of these analyses can be compared with each other and with the analysis for the full data set to

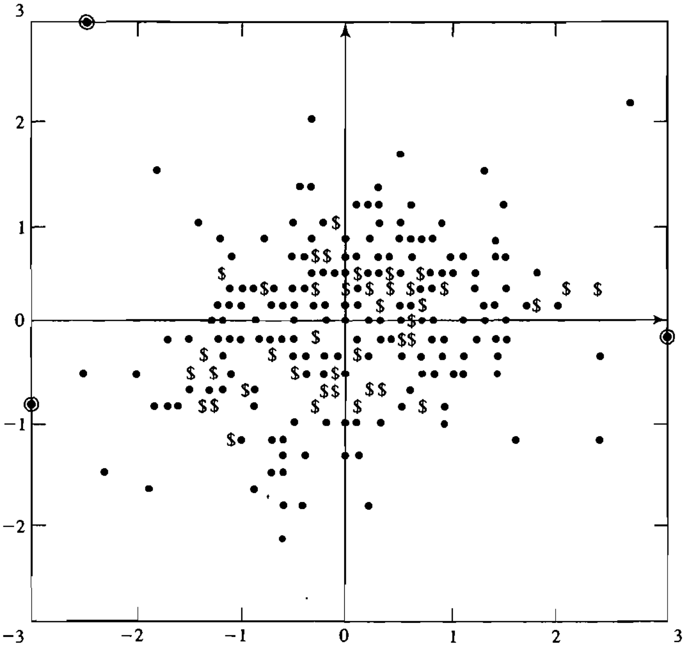


Figure 9.5 Factor scores for the first two factors of chicken-bone data.

test the stability of the solution. If the results are consistent with one another, confidence in the solution is increased.

The chicken-bone data were divided into two sets of $n_1 = 137$ and $n_2 = 139$ observations, respectively. The resulting sample correlation matrices were

$$\mathbf{R}_1 = \begin{bmatrix} 1.000 & & & & & \\ .696 & 1.000 & & & & \\ .588 & .540 & 1.000 & & & \\ .639 & .575 & .901 & 1.000 & & \\ .694 & .606 & .844 & .835 & 1.000 & \\ .660 & .584 & .866 & .863 & .931 & 1.000 \end{bmatrix}$$

and

$$\mathbf{R}_2 = \begin{bmatrix} 1.000 & & & & & \\ .366 & 1.000 & & & & \\ .572 & .352 & 1.000 & & & \\ .587 & .406 & .950 & 1.000 & & \\ .587 & .420 & .909 & .911 & 1.000 & \\ .598 & .386 & .894 & .927 & .940 & 1.000 \end{bmatrix}$$

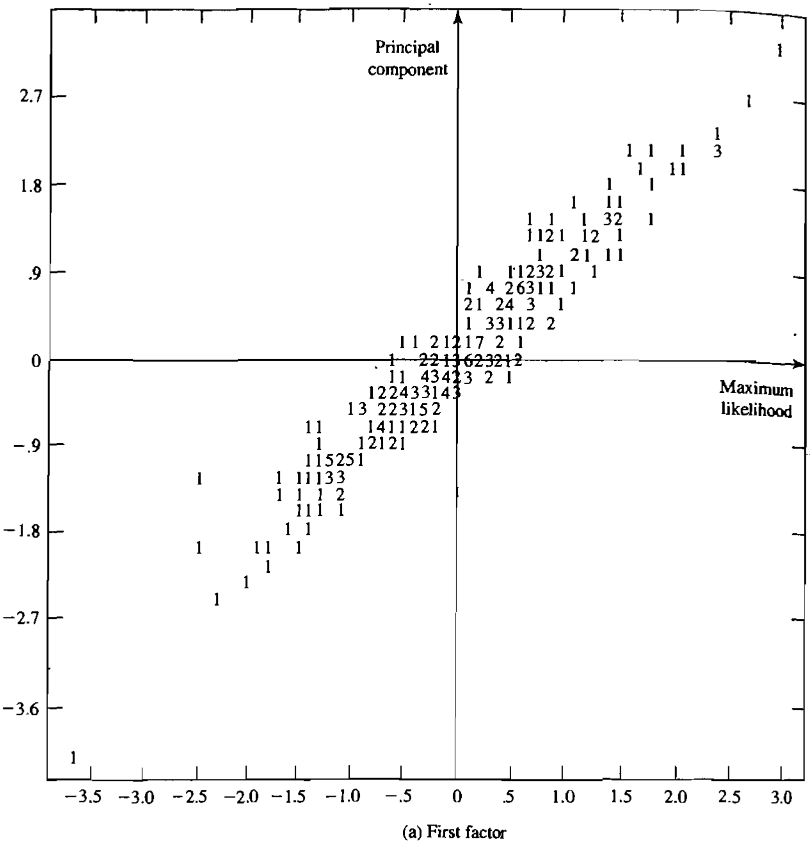


Figure 9.6 Pairs of factor scores for the chicken-bone data. (Loadings are estimated by principal component and maximum likelihood methods.)

The rotated estimated loadings, specific variances, and proportion of the total (standardized) sample variance explained for a principal component solution of an $m = 3$ factor model are given in Table 9.11 on page 525.

The results for the two halves of the chicken-bone measurements are very similar. Factors F_2^* and F_3^* interchange with respect to their labels, skull length and skull breadth, but they collectively seem to represent *head size*. The first factor, F_1^*, again appears to be a *body-size* factor dominated by leg and wing dimensions. These are the same interpretations we gave to the results from a principal component factor analysis of the entire set of data. The solution is remarkably stable, and we can be fairly confident that the large loadings are "real." As we have pointed out however, three factors are probably too many. A one- or two-factor model is surely sufficient for the chicken-bone data, and you are encouraged to repeat the analyses here with fewer factors and alternative solution methods. (See Exercise 9.10.) ■

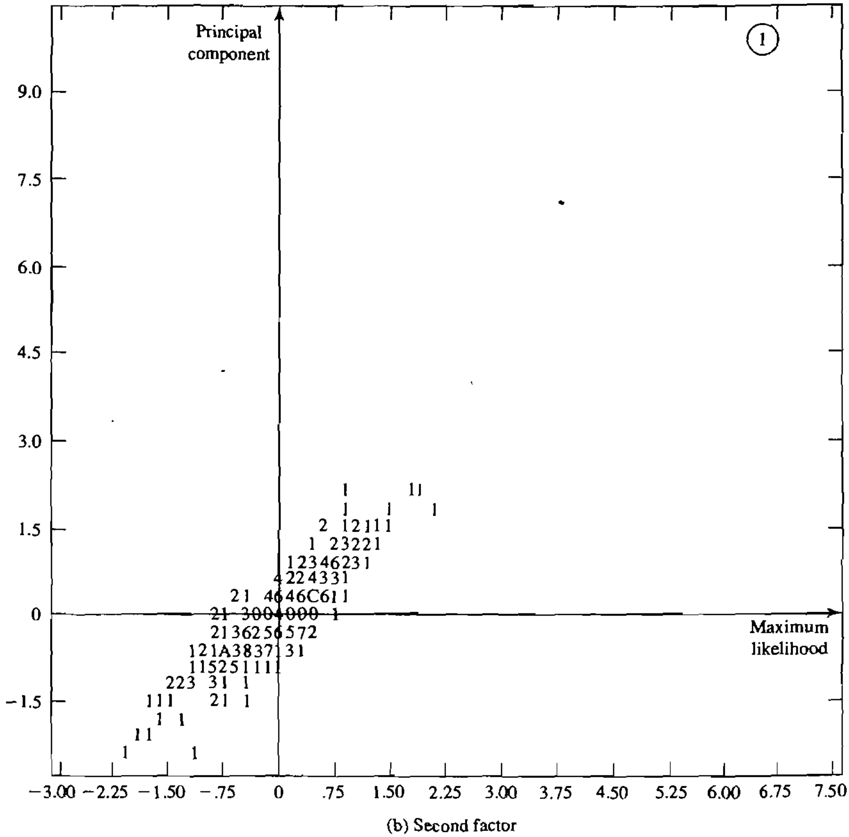
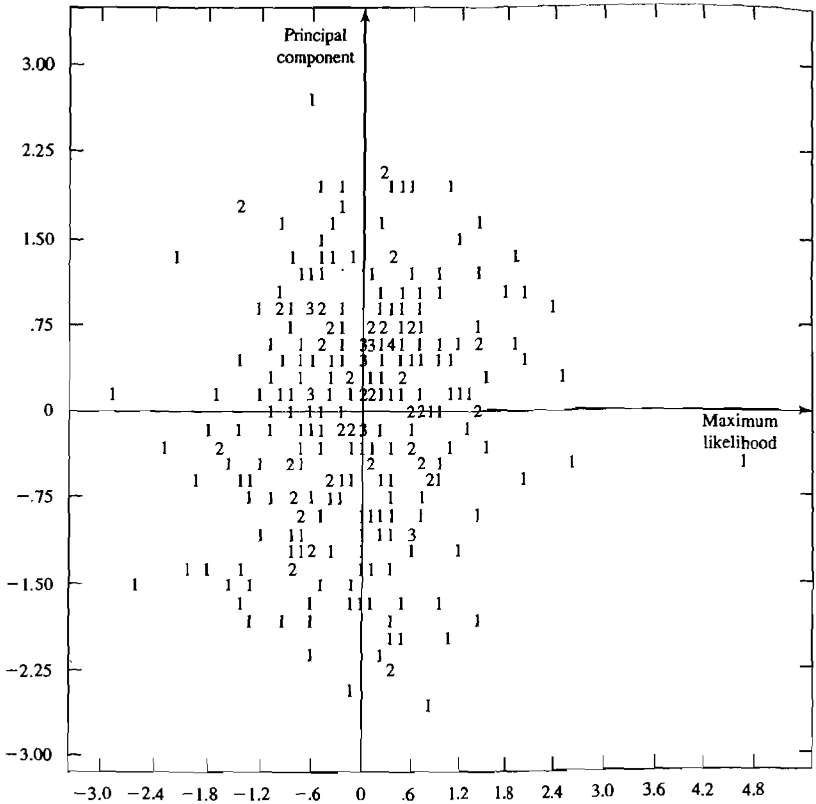


Figure 9.6 *(continued)*

Table 9.11

	First set (n_1 = 137 observations) Rotated estimated factor loadings				Second set (n_2 = 139 observations) Rotated estimated factor loadings			
Variable	F_1^*	F_2^*	F_3^*	$\tilde{\psi}_i$	F_1^*	F_2^*	F_3^*	$\tilde{\psi}_i$
1. Skull length	.360	.361	.853	.01	.352	.921	.167	.00
2. Skull breadth	.303	.899	.312	.00	.203	.145	.968	.00
3. Femur length	.914	.238	.175	.08	.930	.239	.130	.06
4. Tibia length	.877	.270	.242	.10	.925	.248	.187	.05
5. Humerus length	.830	.247	.395	.11	.912	.252	.208	.06
6. Ulna length	.871	.231	.332	.08	.914	.272	.168	.06
Cumulative proportion of total (standardized) sample variance explained	.546	.743	.940		.593	.780	.962	



(c) Third factor

Figure 9.6 *(continued)*

Factor analysis has a tremendous intuitive appeal for the behavioral and social sciences. In these areas, it is natural to regard multivariate observations on animal and human processes as manifestations of underlying unobservable "traits." Factor analysis provides a way of explaining the observed variability in behavior in terms of these traits.

Still, when all is said and done, factor analysis remains very subjective. Our examples, in common with most published sources, consist of situations in which the factor analysis model provides reasonable explanations in terms of a few interpretable factors. In practice, the vast majority of attempted factor analyses do not yield such clear-cut results. Unfortunately, the criterion for judging the quality of any factor analysis has not been well quantified. Rather, that quality seems to depend on a

WOW criterion

If, while scrutinizing the factor analysis, the investigator can shout "Wow, I understand these factors," the application is deemed successful.

Supplement 9A

SOME COMPUTATIONAL DETAILS FOR MAXIMUM LIKELIHOOD ESTIMATION

Although a simple analytical expression cannot be obtained for the maximum likelihood estimators $\hat{\mathbf{L}}$ and $\hat{\mathbf{\Psi}}$, they can be shown to satisfy certain equations. Not surprisingly, the conditions are stated in terms of the maximum likelihood estimator $\mathbf{S}_n = (1/n) \sum_{j=1}^{n} (\mathbf{X}_j - \overline{\mathbf{X}})(\mathbf{X}_j - \overline{\mathbf{X}})'$ of an unstructured covariance matrix. Some factor analysts employ the usual sample covariance $\mathbf{S}$, but still use the title *maximum likelihood* to refer to resulting estimates. This modification, referenced in Footnote 4 of this chapter, amounts to employing the likelihood obtained from the Wishart distribution of $\sum_{j=1}^{n} (\mathbf{X}_j - \overline{\mathbf{X}})(\mathbf{X}_j - \overline{\mathbf{X}})'$ and ignoring the minor contribution due to the normal density for $\overline{\mathbf{X}}$. The factor analysis of $\mathbf{R}$ is, of course, unaffected by the choice of $\mathbf{S}_n$ or $\mathbf{S}$, since they both produce the same correlation matrix.

Result 9A.1. Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ be a random sample from a normal population. The maximum likelihood estimates $\hat{\mathbf{L}}$ and $\hat{\mathbf{\Psi}}$ are obtained by maximizing (9-25) subject to the uniqueness condition in (9-26). They satisfy

$$(\hat{\mathbf{\Psi}}^{-1/2}\mathbf{S}_n\hat{\mathbf{\Psi}}^{-1/2})(\hat{\mathbf{\Psi}}^{-1/2}\hat{\mathbf{L}}) = (\hat{\mathbf{\Psi}}^{-1/2}\hat{\mathbf{L}})(\mathbf{I} + \hat{\mathbf{\Delta}}) \tag{9A-1}$$

so the jth column of $\hat{\mathbf{\Psi}}^{-1/2}\hat{\mathbf{L}}$ is the (nonnormalized) eigenvector of $\hat{\mathbf{\Psi}}^{-1/2}\mathbf{S}_n\hat{\mathbf{\Psi}}^{-1/2}$ corresponding to eigenvalue $1 + \hat{\Delta}_i$. Here

$$\mathbf{S}_n = n^{-1} \sum_{j=1}^{n} (\mathbf{x}_j - \overline{\mathbf{x}})(\mathbf{x}_j - \overline{\mathbf{x}})' = n^{-1}(n-1)\mathbf{S} \quad \text{and} \quad \hat{\Delta}_1 \geq \hat{\Delta}_2 \geq \cdots \geq \hat{\Delta}_m$$

Also, at convergence,

$$\hat{\psi}_i = i\text{th diagonal element of } \mathbf{S}_n - \hat{\mathbf{L}}\hat{\mathbf{L}}' \tag{9A-2}$$

and

$$\operatorname{tr}(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}_n) = p$$

We avoid the details of the proof. However, it is evident that $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ and a consideration of the log-likelihood leads to the maximization of $-(n/2)[\ln|\boldsymbol{\Sigma}| + \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}_n)]$ over $\mathbf{L}$ and $\boldsymbol{\Psi}$. Equivalently, since $\mathbf{S}_n$ and p are constant with respect to the maximization, we minimize

$$h(\hat{\boldsymbol{\mu}}, \boldsymbol{\Psi}, \mathbf{L}) = \ln|\boldsymbol{\Sigma}| - \ln|\mathbf{S}_n| + \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}_n) - p \tag{9A-3}$$

subject to $\mathbf{L}'\boldsymbol{\Psi}^{-1}\mathbf{L} = \boldsymbol{\Delta}$, a diagonal matrix. ■

Comment. Lawley and Maxwell [10], along with many others who do factor analysis, use the unbiased estimate $\mathbf{S}$ of the covariance matrix instead of the maximum likelihood estimate $\mathbf{S}_n$. Now, $(n - 1)\mathbf{S}$ has, for normal data, a Wishart distribution. [See (4-21) and (4-23).] If we ignore the contribution to the likelihood in (9-25) from the second term involving $(\boldsymbol{\mu} - \bar{\mathbf{x}})$, then maximizing the reduced likelihood over $\mathbf{L}$ and $\boldsymbol{\Psi}$ is equivalent to maximizing the Wishart likelihood

$$\text{Likelihood} \propto |\boldsymbol{\Sigma}|^{-(n-1)/2} e^{-[(n-1)/2]\operatorname{tr}[\boldsymbol{\Sigma}^{-1}\mathbf{S}]}$$

over $\mathbf{L}$ and $\boldsymbol{\Psi}$. Equivalently, we can minimize

$$\ln|\boldsymbol{\Sigma}| + \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S})$$

or, as in (9A-3),

$$\ln|\boldsymbol{\Sigma}| + \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) - \ln|\mathbf{S}| - p$$

Under these conditions, Result (9A-1) holds with $\mathbf{S}$ in place of $\mathbf{S}_n$. Also, for large n, $\mathbf{S}$ and $\mathbf{S}_n$ are almost identical, and the corresponding maximum likelihood estimates, $\hat{\mathbf{L}}$ and $\hat{\boldsymbol{\Psi}}$, would be similar. For testing the factor model [see (9-39)], $|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}|$ should be compared with $|\mathbf{S}_n|$ if the actual likelihood of (9-25) is employed, and $|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}|$ should be compared with $|\mathbf{S}|$ if the foregoing Wishart likelihood is used to derive $\hat{\mathbf{L}}$ and $\hat{\boldsymbol{\Psi}}$.

Recommended Computational Scheme

For $m > 1$, the condition $\mathbf{L}'\boldsymbol{\Psi}^{-1}\mathbf{L} = \boldsymbol{\Delta}$ effectively imposes $m(m - 1)/2$ constraints on the elements of $\mathbf{L}$ and $\boldsymbol{\Psi}$, and the likelihood equations are solved, subject to these contraints, in an iterative fashion. One procedure is the following:

1. Compute initial estimates of the specific variances $\psi_1, \psi_2, \ldots, \psi_p$. Jöreskog [8] suggests setting

$$\hat{\psi}_i = \left(1 - \frac{1}{2}\cdot\frac{m}{p}\right)\left(\frac{1}{s^{ii}}\right) \tag{9A-4}$$

where s^{ii} is the ith diagonal element of $\mathbf{S}^{-1}$.

2. Given $\hat{\Psi}$, compute the first m distinct eigenvalues, $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_m > 1$, and corresponding eigenvectors, $\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \ldots, \hat{\mathbf{e}}_m$, of the "uniqueness-rescaled" covariance matrix

$$\mathbf{S}^* = \hat{\Psi}^{-1/2}\mathbf{S}_n\hat{\Psi}^{-1/2} \tag{9A-5}$$

Let $\hat{\mathbf{E}} = [\hat{\mathbf{e}}_1 \mid \hat{\mathbf{e}}_2 \mid \cdots \mid \hat{\mathbf{e}}_m]$ be the $p \times m$ matrix of *normalized* eigenvectors and $\hat{\Lambda} = \text{diag}[\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_m]$ be the $m \times m$ diagonal matrix of eigenvalues. From (9A-1), $\hat{\Lambda} = \mathbf{I} + \hat{\Delta}$ and $\hat{\mathbf{E}} = \hat{\Psi}^{-1/2}\hat{\mathbf{L}}\hat{\Delta}^{-1/2}$. Thus, we obtain the estimates

$$\hat{\mathbf{L}} = \hat{\Psi}^{1/2}\hat{\mathbf{E}}\hat{\Delta}^{1/2} = \hat{\Psi}^{1/2}\hat{\mathbf{E}}(\hat{\Lambda} - \mathbf{I})^{1/2} \tag{9A-6}$$

3. Substitute $\hat{\mathbf{L}}$ obtained in (9A-6) into the likelihood function (9A-3), and minimize the result with respect to $\hat{\psi}_1, \hat{\psi}_2, \ldots, \hat{\psi}_p$. A numerical search routine must be used. The values $\hat{\psi}_1, \hat{\psi}_2, \ldots, \hat{\psi}_p$ obtained from this minimization are employed at Step (2) to create a new $\hat{\mathbf{L}}$. Steps (2) and (3) are repeated until convergence—that is, until the differences between successive values of $\hat{\ell}_{ij}$ and $\hat{\psi}_i$ are negligible.

Comment. It often happens that the objective function in (9A-3) has a relative minimum corresponding to *negative* values for some $\hat{\psi}_i$. This solution is clearly inadmissible and is said to be improper, or a *Heywood case*. For most packaged computer programs, negative $\hat{\psi}_i$, if they occur on a particular iteration, are changed to small positive numbers before proceeding with the next step.

Maximum Likelihood Estimators of $\boldsymbol{\rho} = \mathbf{L}_z\mathbf{L}_z' + \Psi_z$

When Σ has the factor analysis structure $\Sigma = \mathbf{L}\mathbf{L}' + \Psi$, $\boldsymbol{\rho}$ can be factored as $\boldsymbol{\rho} = \mathbf{V}^{-1/2}\Sigma\mathbf{V}^{-1/2} = (\mathbf{V}^{-1/2}\mathbf{L})(\mathbf{V}^{-1/2}\mathbf{L})' + \mathbf{V}^{-1/2}\Psi\mathbf{V}^{-1/2} = \mathbf{L}_z\mathbf{L}_z' + \Psi_z$. The loading matrix for the standardized variables is $\mathbf{L}_z = \mathbf{V}^{-1/2}\mathbf{L}$, and the corresponding specific variance matrix is $\Psi_z = \mathbf{V}^{-1/2}\Psi\mathbf{V}^{-1/2}$, where $\mathbf{V}^{-1/2}$ is the diagonal matrix with ith diagonal element $\sigma_{ii}^{-1/2}$. If $\mathbf{R}$ is substituted for $\mathbf{S}_n$ in the objective function of (9A-3), the investigator minimizes

$$\ln\left(\frac{|\mathbf{L}_z\mathbf{L}_z' + \Psi_z|}{|\mathbf{R}|}\right) + \text{tr}[(\mathbf{L}_z\mathbf{L}_z' + \Psi_z)^{-1}\mathbf{R}] - p \tag{9A-7}$$

Introducing the diagonal matrix $\hat{\mathbf{V}}^{1/2}$, whose ith diagonal element is the square root of the ith diagonal element of $\mathbf{S}_n$, we can write the objective function in (9A-7) as

$$\begin{aligned}
&\ln\left(\frac{|\hat{\mathbf{V}}^{1/2}||\mathbf{L}_z\mathbf{L}_z' + \Psi_z||\hat{\mathbf{V}}^{1/2}|}{|\hat{\mathbf{V}}^{1/2}||\mathbf{R}||\hat{\mathbf{V}}^{1/2}|}\right) + \text{tr}[(\mathbf{L}_z\mathbf{L}_z' + \Psi_z)^{-1}\hat{\mathbf{V}}^{-1/2}\hat{\mathbf{V}}^{1/2}\mathbf{R}\hat{\mathbf{V}}^{1/2}\hat{\mathbf{V}}^{-1/2}] - p \\
&\quad = \ln\left(\frac{|(\hat{\mathbf{V}}^{1/2}\mathbf{L}_z)(\hat{\mathbf{V}}^{1/2}\mathbf{L}_z)' + \hat{\mathbf{V}}^{1/2}\Psi_z\hat{\mathbf{V}}^{1/2}|}{|\mathbf{S}_n|}\right) \\
&\qquad + \text{tr}[((\hat{\mathbf{V}}^{1/2}\mathbf{L}_z)(\hat{\mathbf{V}}^{1/2}\mathbf{L}_z)' + \hat{\mathbf{V}}^{1/2}\Psi_z\hat{\mathbf{V}}^{1/2})^{-1}\mathbf{S}_n] - p \\
&\quad \geq \ln\left(\frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi}|}{|\mathbf{S}_n|}\right) + \text{tr}[(\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\Psi})^{-1}\mathbf{S}_n] - p
\end{aligned} \tag{9A-8}$$

The last inequality follows because the maximum likelihood estimates $\hat{\mathbf{L}}$ and $\hat{\boldsymbol{\Psi}}$ minimize the objective function (9A-3). [Equality holds in (9A-8) for $\hat{\mathbf{L}}_z = \hat{\mathbf{V}}^{-1/2}\hat{\mathbf{L}}$ and $\hat{\boldsymbol{\Psi}}_z = \hat{\mathbf{V}}^{-1/2}\hat{\boldsymbol{\Psi}}\hat{\mathbf{V}}^{-1/2}$.] Therefore, minimizing (9A-7) over $\mathbf{L}_z$ and $\boldsymbol{\Psi}_z$ is equivalent to obtaining $\hat{\mathbf{L}}$ and $\hat{\boldsymbol{\Psi}}$ from $\mathbf{S}_n$ and estimating $\mathbf{L}_z = \mathbf{V}^{-1/2}\mathbf{L}$ by $\hat{\mathbf{L}}_z = \hat{\mathbf{V}}^{-1/2}\hat{\mathbf{L}}$ and $\boldsymbol{\Psi}_z = \mathbf{V}^{-1/2}\boldsymbol{\Psi}\mathbf{V}^{-1/2}$ by $\hat{\boldsymbol{\Psi}}_z = \hat{\mathbf{V}}^{-1/2}\hat{\boldsymbol{\Psi}}\hat{\mathbf{V}}^{-1/2}$. The rationale for the latter procedure comes from the invariance property of maximum likelihood estimators. [See (4-20).]

Exercises

9.1. Show that the covariance matrix

$$\boldsymbol{\rho} = \begin{bmatrix} 1.0 & .63 & .45 \\ .63 & 1.0 & .35 \\ .45 & .35 & 1.0 \end{bmatrix}$$

for the $p = 3$ standardized random variables Z_1, Z_2, and Z_3 can be generated by the $m = 1$ factor model

$$\begin{aligned} Z_1 &= .9F_1 + \varepsilon_1 \\ Z_2 &= .7F_1 + \varepsilon_2 \\ Z_3 &= .5F_1 + \varepsilon_3 \end{aligned}$$

where $\text{Var}(F_1) = 1$, $\text{Cov}(\boldsymbol{\varepsilon}, F_1) = \mathbf{0}$, and

$$\boldsymbol{\Psi} = \text{Cov}(\boldsymbol{\varepsilon}) = \begin{bmatrix} .19 & 0 & 0 \\ 0 & .51 & 0 \\ 0 & 0 & .75 \end{bmatrix}$$

That is, write $\boldsymbol{\rho}$ in the form $\boldsymbol{\rho} = \mathbf{LL}' + \boldsymbol{\Psi}$.

9.2. Use the information in Exercise 9.1.

(a) Calculate communalities $h_i^2, i = 1, 2, 3$, and interpret these quantities.

(b) Calculate $\text{Corr}(Z_i, F_1)$ for $i = 1, 2, 3$. Which variable might carry the greatest weight in "naming" the common factor? Why?

9.3. The eigenvalues and eigenvectors of the correlation matrix $\boldsymbol{\rho}$ in Exercise 9.1 are

$$\begin{aligned} \lambda_1 &= 1.96, & \mathbf{e}_1' &= [.625, .593, .507] \\ \lambda_2 &= .68, & \mathbf{e}_2' &= [-.219, -.491, .843] \\ \lambda_3 &= .36, & \mathbf{e}_3' &= [.749, -.638, -.177] \end{aligned}$$

(a) Assuming an $m = 1$ factor model, calculate the loading matrix $\mathbf{L}$ and matrix of specific variances $\boldsymbol{\Psi}$ using the principal component solution method. Compare the results with those in Exercise 9.1.

(b) What proportion of the total population variance is explained by the first common factor?

9.4. Given $\boldsymbol{\rho}$ and $\boldsymbol{\Psi}$ in Exercise 9.1 and an $m = 1$ factor model, calculate the reduced correlation matrix $\tilde{\boldsymbol{\rho}} = \boldsymbol{\rho} - \boldsymbol{\Psi}$ and the principal factor solution for the loading matrix $\mathbf{L}$. Is the result consistent with the information in Exercise 9.1? Should it be?

9.5. Establish the inequality (9-19).

Hint: Since $\mathbf{S} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}' - \tilde{\boldsymbol{\Psi}}$ has zeros on the diagonal,

$$(\text{sum of squared entries of } \mathbf{S} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}' - \tilde{\boldsymbol{\Psi}}) \le (\text{sum of squared entries of } \mathbf{S} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}')$$

Now, $\mathbf{S}-\tilde{\mathbf{L}}\tilde{\mathbf{L}}' = \hat{\lambda}_{m+1}\hat{\mathbf{e}}_{m+1}\hat{\mathbf{e}}'_{m+1} + \cdots + \hat{\lambda}_p\hat{\mathbf{e}}_p\hat{\mathbf{e}}'_p = \hat{\mathbf{P}}_{(2)}\hat{\mathbf{\Lambda}}_{(2)}\hat{\mathbf{P}}'_{(2)}$, where $\hat{\mathbf{P}}_{(2)} = [\hat{\mathbf{e}}_{m+1} \vdots \cdots \vdots \hat{\mathbf{e}}_p]$ and $\hat{\mathbf{\Lambda}}_{(2)}$ is the diagonal matrix with elements $\hat{\lambda}_{m+1}, \ldots, \hat{\lambda}_p$.

Use (sum of squared entries of $\mathbf{A}$) = tr $\mathbf{AA}'$ and tr $[\hat{\mathbf{P}}_{(2)}\hat{\mathbf{\Lambda}}_{(2)}\hat{\mathbf{\Lambda}}_{(2)}\hat{\mathbf{P}}'_{(2)}] = $ tr $[\hat{\mathbf{\Lambda}}_{(2)}\hat{\mathbf{\Lambda}}_{(2)}]$.

9.6. Verify the following matrix identities.

(a) $(\mathbf{I} + \mathbf{L}'\mathbf{\Psi}^{-1}\mathbf{L})^{-1}\mathbf{L}'\mathbf{\Psi}^{-1}\mathbf{L} = \mathbf{I} - (\mathbf{I} + \mathbf{L}'\mathbf{\Psi}^{-1}\mathbf{L})^{-1}$

Hint: Premultiply both sides by $(\mathbf{I} + \mathbf{L}'\mathbf{\Psi}^{-1}\mathbf{L})$.

(b) $(\mathbf{LL}' + \mathbf{\Psi})^{-1} = \mathbf{\Psi}^{-1} - \mathbf{\Psi}^{-1}\mathbf{L}(\mathbf{I} + \mathbf{L}'\mathbf{\Psi}^{-1}\mathbf{L})^{-1}\mathbf{L}'\mathbf{\Psi}^{-1}$

Hint: Postmultiply both sides by $(\mathbf{LL}' + \mathbf{\Psi})$ and use (a).

(c) $\mathbf{L}'(\mathbf{LL}' + \mathbf{\Psi})^{-1} = (\mathbf{I} + \mathbf{L}'\mathbf{\Psi}^{-1}\mathbf{L})^{-1}\mathbf{L}'\mathbf{\Psi}^{-1}$

Hint: Postmultiply the result in (b) by $\mathbf{L}$, use (a), and take the transpose, noting that $(\mathbf{LL}' + \mathbf{\Psi})^{-1}$, $\mathbf{\Psi}^{-1}$, and $(\mathbf{I} + \mathbf{L}'\mathbf{\Psi}^{-1}\mathbf{L})^{-1}$ are symmetric matrices.

9.7. *(The factor model parameterization need not be unique.)* Let the factor model with $p = 2$ and $m = 1$ prevail. Show that

$$\sigma_{11} = \ell_{11}^2 + \psi_1, \qquad \sigma_{12} = \sigma_{21} = \ell_{11}\ell_{21}$$
$$\sigma_{22} = \ell_{21}^2 + \psi_2$$

and, for given σ_{11}, σ_{22}, and σ_{12}, there is an infinity of choices for $\mathbf{L}$ and $\mathbf{\Psi}$.

9.8. *(Unique but improper solution: Heywood case.)*
Consider an $m = 1$ factor model for the population with covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & .4 & .9 \\ .4 & 1 & .7 \\ .9 & .7 & 1 \end{bmatrix}$$

Show that there is a unique choice of $\mathbf{L}$ and $\mathbf{\Psi}$ with $\mathbf{\Sigma} = \mathbf{LL}' + \mathbf{\Psi}$, but that $\psi_3 < 0$, so the choice is not admissible.

9.9. In a study of liquor preference in France, Stoetzel [14] collected preference rankings of $p = 9$ liquor types from $n = 1442$ individuals. A factor analysis of the 9×9 sample correlation matrix of rank orderings gave the following estimated loadings:

	Estimated factor loadings		
Variable (X_1)	F_1	F_2	F_3
Liquors	.64	.02	.16
Kirsch	.50	−.06	−.10
Mirabelle	.46	−.24	−.19
Rum	.17	.74	.97*
Marc	−.29	.66	−.39
Whiskey	−.29	−.08	.09
Calvados	−.49	.20	−.04
Cognac	−.52	−.03	.42
Armagnac	−.60	−.17	.14

*This figure is too high. It exceeds the maximum value of .64, as a result of an approximation method for obtaining the estimated factor loadings used by Stoetzel.

Given these results, Stoetzel concluded the following: The major principle of liquor preference in France is the distinction between sweet and strong liquors. The second motivating element is price, which can be understood by remembering that liquor is both an expensive commodity and an item of conspicuous consumption. Except in the case of the two most popular and least expensive items (rum and marc), this second factor plays a much smaller role in producing preference judgments. The third factor concerns the sociological and primarily the regional, variability of the judgments. (See [14], p. 11.)

(a) Given what you know about the various liquors involved, does Stoetzel's interpretation seem reasonable?

(b) Plot the loading pairs for the first two factors. Conduct a graphical orthogonal rotation of the factor axes. Generate approximate rotated loadings. Interpret the rotated loadings for the first two factors. Does your interpretation agree with Stoetzel's interpretation of these factors from the unrotated loadings? Explain.

9.10. The correlation matrix for chicken-bone measurements (see Example 9.14) is

$$\begin{bmatrix} 1.000 & & & & & \\ .505 & 1.000 & & & & \\ .569 & .422 & 1.000 & & & \\ .602 & .467 & .926 & 1.000 & & \\ .621 & .482 & .877 & .874 & 1.000 & \\ .603 & .450 & .878 & .894 & .937 & 1.000 \end{bmatrix}$$

The following estimated factor loadings were extracted by the maximum likelihood procedure:

Variable	Estimated factor loadings F_1	F_2	Varimax rotated estimated factor loadings F_1^*	F_2^*
1. Skull length	.602	.200	.484	.411
2. Skull breadth	.467	.154	.375	.319
3. Femur length	.926	.143	.603	.717
4. Tibia length	1.000	.000	.519	.855
5. Humerus length	.874	.476	.861	.499
6. Ulna length	.894	.327	.744	.594

Using the *unrotated* estimated factor loadings, obtain the maximum likelihood estimates of the following.

(a) The specific variances.

(b) The communalities.

(c) The proportion of variance explained by each factor.

(d) The residual matrix $\mathbf{R} - \hat{\mathbf{L}}_z\hat{\mathbf{L}}_z' - \hat{\mathbf{\Psi}}_z$.

9.11. Refer to Exercise 9.10. Compute the value of the varimax criterion using both unrotated and rotated estimated factor loadings. Comment on the results.

9.12. The *covariance* matrix for the logarithms of turtle measurements (see Example 8.4) is

$$\mathbf{S} = 10^{-3}\begin{bmatrix} 11.072 & & \\ 8.019 & 6.417 & \\ 8.160 & 6.005 & 6.773 \end{bmatrix}$$

The following maximum likelihood estimates of the factor loadings for an $m = 1$ model were obtained:

Variable	Estimated factor loadings F_1
1. ln(length)	.1022
2. ln(width)	.0752
3. ln(height)	.0765

Using the estimated factor loadings, obtain the maximum likelihood estimates of each of the following.

(a) Specific variances.

(b) Communalities.

(c) Proportion of variance explained by the factor.

(d) The residual matrix $\mathbf{S}_n - \hat{\mathbf{L}}\hat{\mathbf{L}}' - \hat{\boldsymbol{\Psi}}$.

Hint: Convert $\mathbf{S}$ to $\mathbf{S}_n$.

9.13. Refer to Exercise 9.12. Compute the test statistic in (9-39). Indicate why a test of $H_0: \boldsymbol{\Sigma} = \mathbf{LL}' + \boldsymbol{\Psi}$ (with $m = 1$) versus $H_1: \boldsymbol{\Sigma}$ unrestricted cannot be carried out for this example. [See (9-40).]

9.14. The maximum likelihood factor loading estimates are given in (9A-6) by

$$\hat{\mathbf{L}} = \hat{\boldsymbol{\Psi}}^{1/2}\hat{\mathbf{E}}\hat{\boldsymbol{\Delta}}^{1/2}$$

Verify, for this choice, that

$$\hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}} = \hat{\boldsymbol{\Delta}}$$

where $\hat{\boldsymbol{\Delta}} = \hat{\boldsymbol{\Lambda}} - \mathbf{I}$ is a diagonal matrix.

9.15. Hirschey and Wichern [7] investigate the consistency, determinants, and uses of accounting and market-value measures of profitability. As part of their study, a factor analysis of accounting profit measures and market estimates of economic profits was conducted. The correlation matrix of accounting historical, accounting replacement, and market-value measures of profitability for a sample of firms operating in 1977 is as follows:

Variable	HRA	HRE	HRS	RRA	RRE	RRS	Q	REV
Historical return on assets, HRA	1.000							
Historical return on equity, HRE	.738	1.000						
Historical return on sales, HRS	.731	.520	1.000					
Replacement return on assets, RRA	.828	.688	.652	1.000				
Replacement return on equity, RRE	.681	.831	.513	.887	1.000			
Replacement return on sales, RRS	.712	.543	.826	.867	.692	1.000		
Market Q ratio, Q	.625	.322	.579	.639	.419	.608	1.000	
Market relative excess value, REV	.604	.303	.617	.563	.352	.610	.937	1.000

The following rotated principal component estimates of factor loadings for an $m = 3$ factor model were obtained:

Variable	Estimated factor loadings F_1	F_2	F_3
Historical return on assets	.433	.612	.499
Historical return on equity	.125	.892	.234
Historical return on sales	.296	.238	.887
Replacement return on assets	.406	.708	.483
Replacement return on equity	.198	.895	.283
Replacement return on sales	.331	.414	.789
Market Q ratio	.928	.160	.294
Market relative excess value	.910	.079	.355
Cumulative proportion of total variance explained	.287	.628	.908

(a) Using the estimated factor loadings, determine the specific variances and communalities.

(b) Determine the residual matrix, $\mathbf{R} - \hat{\mathbf{L}}_z\hat{\mathbf{L}}_z' - \hat{\boldsymbol{\Psi}}_z$. Given this information and the cumulative proportion of total variance explained in the preceding table, does an $m = 3$ factor model appear appropriate for these data?

(c) Assuming that estimated loadings less than .4 are small, interpret the three factors. Does it appear, for example, that market-value measures provide evidence of profitability distinct from that provided by accounting measures? Can you separate accounting historical measures of profitability from accounting replacement measures?

9.16. Verify that factor scores constructed according to (9-50) have sample mean vector $\mathbf{0}$ and zero sample covariances.

9.17. Refer to Example 9.12. Using the information in this example, evaluate $(\hat{\mathbf{L}}_z'\hat{\boldsymbol{\Psi}}_z^{-1}\hat{\mathbf{L}}_z)^{-1}$.

Note: Set the fourth diagonal element of $\hat{\boldsymbol{\Psi}}_z$ to .01 so that $\hat{\boldsymbol{\Psi}}_z^{-1}$ can be determined. Will the regression and generalized least squares methods for constructing factors scores for standardized stock price observations give nearly the same results? *Hint:* See equation (9-57) and the discussion following it.

The following exercises require the use of a computer.

9.18. Refer to Exercise 8.16 concerning the numbers of fish caught.

(a) Using only the measurements $x_1 - x_4$, obtain the principal component solution for factor models with $m = 1$ and $m = 2$.

(b) Using only the measurements $x_1 - x_4$, obtain the maximum likelihood solution for factor models with $m = 1$ and $m = 2$.

(c) Rotate your solutions in Parts (a) and (b). Compare the solutions and comment on them. Interpret each factor.

(d) Perform a factor analysis using the measurements $x_1 - x_6$. Determine a reasonable number of factors m, and compare the principal component and maximum likelihood solutions after rotation. Interpret the factors.

9.19. A firm is attempting to evaluate the quality of its sales staff and is trying to find an examination or series of tests that may reveal the potential for good performance in sales

The firm has selected a random sample of 50 sales people and has evaluated each on 3 measures of performance: growth of sales, profitability of sales, and new-account sales. These measures have been converted to a scale, on which 100 indicates "average" performance. Each of the 50 individuals took each of 4 tests, which purported to measure creativity, mechanical reasoning, abstract reasoning, and mathematical ability, respectively. The $n = 50$ observations on $p = 7$ variables are listed in Table 9.12 on page 536.

(a) Assume an orthogonal factor model for the *standardized variables* $Z_i = (X_i - \mu_i)/\sqrt{\sigma_{ii}}$, $i = 1, 2, \ldots, 7$. Obtain either the principal component solution or the maximum likelihood solution for $m = 2$ and $m = 3$ common factors.

(b) Given your solution in (a), obtain the rotated loadings for $m = 2$ and $m = 3$. Compare the two sets of rotated loadings. Interpret the $m = 2$ and $m = 3$ factor solutions.

(c) List the estimated communalities, specific variances, and $\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}$ for the $m = 2$ and $m = 3$ solutions. Compare the results. Which choice of m do you prefer at this point? Why?

(d) Conduct a test of $H_0: \boldsymbol{\Sigma} = \mathbf{LL}' + \boldsymbol{\Psi}$ versus $H_1: \boldsymbol{\Sigma} \neq \mathbf{LL}' + \boldsymbol{\Psi}$ for both $m = 2$ and $m = 3$ at the $\alpha = .01$ level. With these results and those in Parts b and c, which choice of m appears to be the best?

(e) Suppose a new salesperson, selected at random, obtains the test scores $\mathbf{x}' = [x_1, x_2, \ldots, x_7] = [110, 98, 105, 15, 18, 12, 35]$. Calculate the salesperson's factor score using the weighted least squares method and the regression method.

Note: The components of $\mathbf{x}$ must be standardized using the sample means and variances calculated from the original data.

9.20. Using the air-pollution variables X_1, X_2, X_5, and X_6 given in Table 1.5, generate the sample *covariance* matrix.

(a) Obtain the principal component solution to a factor model with $m = 1$ and $m = 2$.

(b) Find the maximum likelihood estimates of $\mathbf{L}$ and $\boldsymbol{\Psi}$ for $m = 1$ and $m = 2$.

(c) Compare the factorization obtained by the principal component and maximum likelihood methods.

9.21. Perform a varimax rotation of both $m = 2$ solutions in Exercise 9.20. Interpret the results. Are the principal component and maximum likelihood solutions consistent with each other?

9.22. Refer to Exercise 9.20.

(a) Calculate the factor scores from the $m = 2$ maximum likelihood estimates by (i) weighted least squares in (9-50) and (ii) the regression approach of (9-58).

(b) Find the factor scores from the principal component solution, using (9-51).

(c) Compare the three sets of factor scores.

9.23. Repeat Exercise 9.20, starting from the sample *correlation* matrix. Interpret the factors for the $m = 1$ and $m = 2$ solutions. Does it make a difference if $\mathbf{R}$, rather than $\mathbf{S}$, is factored? Explain.

9.24. Perform a factor analysis of the census-tract data in Table 8.5. Start with $\mathbf{R}$ and obtain both the maximum likelihood and principal component solutions. Comment on your choice of m. Your analysis should include factor rotation and the computation of factor scores.

9.25. Perform a factor analysis of the "stiffness" measurements given in Table 4.3 and discussed in Example 4.14. Compute factor scores, and check for outliers in the data. Use the sample covariance matrix $\mathbf{S}$.

Table 9.12 Salespeople Data

	Index of:			Score on:			
Salesperson	Sales growth (x_1)	Sales profitability (x_2)	New-account sales (x_3)	Creativity test (x_4)	Mechanical reasoning test (x_5)	Abstract reasoning test (x_6)	Mathematics test (x_7)
1	93.0	96.0	97.8	09	12	09	20
2	88.8	91.8	96.8	07	10	10	15
3	95.0	100.3	99.0	08	12	09	26
4	101.3	103.8	106.8	13	14	12	29
5	102.0	107.8	103.0	10	15	12	32
6	95.8	97.5	99.3	10	14	11	21
7	95.5	99.5	99.0	09	12	09	25
8	110.8	122.0	115.3	18	20	15	51
9	102.8	108.3	103.8	10	17	13	31
10	106.8	120.5	102.0	14	18	11	39
11	103.3	109.8	104.0	12	17	12	32
12	99.5	111.8	100.3	10	18	08	31
13	103.5	112.5	107.0	16	17	11	34
14	99.5	105.5	102.3	08	10	11	34
15	100.0	107.0	102.8	13	10	08	34
16	81.5	93.5	95.0	07	09	05	16
17	101.3	105.3	102.8	11	12	11	32
18	103.3	110.8	103.5	11	14	11	35
19	95.3	104.3	103.0	05	14	13	30
20	99.5	105.3	106.3	17	17	11	27
21	88.5	95.3	95.8	10	12	07	15
22	99.3	115.0	104.3	05	11	11	42
23	87.5	92.5	95.8	09	09	07	16
24	105.3	114.0	105.3	12	15	12	37
25	107.0	121.0	109.0	16	19	12	39
26	93.3	102.0	97.8	10	15	07	23
27	106.8	118.0	107.3	14	16	12	39
28	106.8	120.0	104.8	10	16	11	49
29	92.3	90.8	99.8	08	10	13	17
30	106.3	121.0	104.5	09	17	11	44
31	106.0	119.5	110.5	18	15	10	43
32	88.3	92.8	96.8	13	11	08	10
33	96.0	103.3	100.5	07	15	11	27
34	94.3	94.5	99.0	10	12	11	19
35	106.5	121.5	110.5	18	17	10	42
36	106.5	115.5	107.0	08	13	14	47
37	92.0	99.5	103.5	18	16	08	18
38	102.0	99.8	103.3	13	12	14	28
39	108.3	122.3	108.5	15	19	12	41
40	106.8	119.0	106.8	14	20	12	37
41	102.5	109.3	103.8	09	17	13	32
42	92.5	102.5	99.3	13	15	06	23
43	102.8	113.8	106.8	17	20	10	32
44	83.3	87.3	96.3	01	05	09	15
45	94.8	101.8	99.8	07	16	11	24
46	103.5	112.0	110.8	18	13	12	37
47	89.5	96.0	97.3	07	15	11	14
48	84.3	89.8	94.3	08	08	08	09
49	104.3	109.5	106.5	14	12	12	36
50	106.0	118.5	105.0	12	16	11	39

9.26. Consider the mice-weight data in Example 8.6. Start with the sample *covariance* matrix. (See Exercise 8.15 for $\sqrt{s_{ii}}$.)

(a) Obtain the principal component solution to the factor model with $m = 1$ and $m = 2$.

(b) Find the maximum likelihood estimates of the loadings and specific variances for $m = 1$ and $m = 2$.

(c) Perform a varimax rotation of the solutions in Parts a and b.

9.27. Repeat Exercise 9.26 by factoring $\mathbf{R}$ instead of the sample covariance matrix $\mathbf{S}$. Also, for the mouse with standardized weights $[.8, -.2, -.6, 1.5]$, obtain the factor scores using the maximum likelihood estimates of the loadings and Equation (9-58).

9.28. Perform a factor analysis of the national track records for women given in Table 1.9. Use the sample covariance matrix $\mathbf{S}$ and interpret the factors. Compute factor scores, and check for outliers in the data. Repeat the analysis with the sample correlation matrix $\mathbf{R}$. Does it make a difference if $\mathbf{R}$, rather than $\mathbf{S}$, is factored? Explain.

9.29. Refer to Exercise 9.28. Convert the national track records for women to speeds measured in meters per second. (See Exercise 8.19.) Perform a factor analysis of the speed data. Use the sample covariance matrix $\mathbf{S}$ and interpret the factors. Compute factor scores, and check for outliers in the data. Repeat the analysis with the sample correlation matrix $\mathbf{R}$. Does it make a difference if $\mathbf{R}$, rather than $\mathbf{S}$, is factored? Explain. Compare your results with the results in Exercise 9.28. Which analysis do you prefer? Why?

9.30. Perform a factor analysis of the national track records for men given in Table 8.6. Repeat the steps given in Exercise 9.28. Is the appropriate factor model for the men's data different from the one for the women's data? If not, are the interpretations of the factors roughly the same? If the models are different, explain the differences.

9.31. Refer to Exercise 9.30. Convert the national track records for men to speeds measured in meters per second. (See Exercise 8.21.) Perform a factor analysis of the speed data. Use the sample covariance matrix $\mathbf{S}$ and interpret the factors. Compute factor scores, and check for outliers in the data. Repeat the analysis with the sample correlation matrix $\mathbf{R}$. Does it make a difference if $\mathbf{R}$, rather than $\mathbf{S}$, is factored? Explain. Compare your results with the results in Exercise 9.30. Which analysis do you prefer? Why?

9.32. Perform a factor analysis of the data on bulls given in Table 1.10. Use the seven variables YrHgt, FtFrBody, PrctFFB, Frame, BkFat, SaleHt, and SaleWt. Factor the sample covariance matrix $\mathbf{S}$ and interpret the factors. Compute factor scores, and check for outliers. Repeat the analysis with the sample correlation matrix $\mathbf{R}$. Compare the results obtained from $\mathbf{S}$ with the results from $\mathbf{R}$. Does it make a difference if $\mathbf{R}$, rather than $\mathbf{S}$, is factored? Explain.

9.33. Perform a factor analysis of the psychological profile data in Table 4.6. Use the sample correlation matrix $\mathbf{R}$ constructed from measurements on the five variables, Indep, Supp, Benev, Conform and Leader. Obtain both the principal component and maximum likelihood solutions for $m = 2$ and $m = 3$ factors. Can you interpret the factors? Your analysis should include factor rotation and the computation of factor scores.

Note: Be aware that a maximum likelihood solution may result in a Heywood case.

9.34. The pulp and paper properties data are given in Table 7.7. Perform a factor analysis using observations on the four paper property variables, BL, EM, SF, and BS and the sample correlation matrix $\mathbf{R}$. Can the information in these data be summarized by a single factor? If so, can you interpret the factor? Try both the principal component and maximum likelihood solution methods. Repeat this analysis with the sample covariance matrix $\mathbf{S}$. Does your interpretation of the factor(s) change if $\mathbf{S}$ rather than $\mathbf{R}$ is factored?

9.35. Repeat Exercise 9.34 using observations on the pulp fiber characteristic variables AFL, LFF, FFF, and ZST. Can these data be summarized by a single factor? Explain.

9.36. Factor analyze the Mali family farm data in Table 8.7. Use the sample correlation matrix $\mathbf{R}$. Try both the principal component and maximum likelihood solution methods for $m = 3$, 4, and 5 factors. Can you interpret the factors? Justify your choice of m. Your analysis should include factor rotation and the computation of factor scores. Can you identify any outliers in these data?

References

1. Anderson, T. W. *An Introduction to Multivariate Statistical Analysis* (3rd ed.). New York: John Wiley, 2003.
2. Bartlett, M. S. "The Statistical Conception of Mental Factors." *British Journal of Psychology*, **28** (1937), 97–104.
3. Bartlett, M. S. "A Note on Multiplying Factors for Various Chi-Squared Approximations." *Journal of the Royal Statistical Society (B)* **16** (1954), 296–298.
4. Dixon, W. S. *Statistical Software Manual to Accompany BMDP Release 7/version 7.0* (paperback). Berkeley, CA: University of California Press, 1992.
5. Dunn, L. C. "The Effect of Inbreeding on the Bones of the Fowl." *Storrs Agricultural Experimental Station Bulletin*, **52** (1928), 1–112.
6. Harmon, H. H. *Modern Factor Analysis* (3rd ed.). Chicago: The University of Chicago Press, 1976.
7. Hirschey, M., and D. W. Wichern. "Accounting and Market-Value Measures of Profitability: Consistency, Determinants and Uses." *Journal of Business and Economic Statistics*, **2**, no. 4 (1984), 375–383.
8. Joreskog, K. G. "Factor Analysis by Least Squares and Maximum Likelihood." In *Statistical Methods for Digital Computers*, edited by K. Enslein, A. Ralston, and H. S. Wilf. New York: John Wiley, 1975.
9. Kaiser, H.F. "The Varimax Criterion for Analytic Rotation in Factor Analysis." *Psychometrika*, **23** (1958), 187–200.
10. Lawley, D. N., and A. E. Maxwell. *Factor Analysis as a Statistical Method* (2nd ed.). New York: American Elsevier Publishing Co., 1971.
11. Linden, M. "A Factor Analytic Study of Olympic Decathlon Data." *Research Quarterly*, **48**, no. 3 (1977), 562–568.
12. Maxwell, A. E. *Multivariate Analysis in Behavioral Research*. London: Chapman and Hall, 1977.
13. Morrison, D. F. *Multivariate Statistical Methods* (4th ed.). Belmont, CA: Brooks/Cole Thompson Learning, 2005.
14. Stoetzel, J. "A Factor Analysis of Liquor Preference." *Journal of Advertising Research*, **1** (1960), 7–11.
15. Wright, S. "The Interpretation of Multivariate Systems." In *Statistics and Mathematics in Biology*, edited by O. Kempthorne and others. Ames, IA: Iowa State University Press, 1954, 11–33.