5. Multivariate Methods by Projection

5.1 Principal Component Analysis

A principle component analysis is concerned with explaining the variance-covariance structure of a set of variables through a few *linear combinations* of these variables. It general objectives are

(1) ***data reduction***
Although p components are required to reproduce the total system variability, often much of this variability can be accounted for by a small number k of the principle components.

(2) ***interpretation***
An analysis of principle components often reveals relationships that were not previously suspected and thereby allows interpretations that would not ordinarily results.

5.1.1 Population Principle Components

- Algebraically, principal components are particular linear combinations of the p random variables $X_1, X_2, \ldots, X_p$.

- Geometrically, these linear combination represent the selection of a new coordinate system obtained by rotating the original system with $X_1, X_2, \ldots, X_p$ as the coordinate axes. The new axes represents the directions with maximum variability and provide a simpler and more parsimonious description of the covariance structure.

- Principal components depend solely on the covariance matrix $\mathbf{\Sigma}$ (or the correlation matrix $\boldsymbol{\rho}$) of $X_1, X_2, \ldots, X_p$. Their development does not require a multivariate normal assumption. On the other hand, principal components derived for multivariate normal populations have useful interpretations in terms of the constant density ellipsoids.

let the random vector $\boldsymbol{X}' = [X_1, X_2, \ldots, X_p]$ have the covariance matrix $\boldsymbol{\Sigma}$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$.

Consider the linear combinations

$$\begin{aligned}
Y_1 &= \mathbf{a}_1'\boldsymbol{X} = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \\
Y_2 &= \mathbf{a}_2'\boldsymbol{X} = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p \\
&\vdots \\
Y_p &= \mathbf{a}_p'\boldsymbol{X} = a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p
\end{aligned}$$

Then

$$\begin{aligned}
\mathrm{Var}(Y_i) &= \mathbf{a}_i'\boldsymbol{\Sigma}\mathbf{a}_i \quad i = 1, 2, \ldots, p \\
\mathrm{Cov}(Y_i, Y_k) &= \mathbf{a}_i'\boldsymbol{\Sigma}\mathbf{a}_k \quad i, k = 1, 2, \ldots, p
\end{aligned}$$

Define

$$
\begin{aligned}
\text{First principle component} &= \text{linear combination } \mathbf{a}_1'\boldsymbol{X} \text{ that maximizes} \\
&\quad \mathrm{Var}(\mathbf{a}_1'\boldsymbol{X}) \text{ subject to } \mathbf{a}_1'\mathbf{a}_1 = 1 \\
\text{Second principle component} &= \text{linear combination } \mathbf{a}_2'\boldsymbol{X} \text{ that maximizes} \\
&\quad \mathrm{Var}(\mathbf{a}_2'\boldsymbol{X}) \text{ subject to } \mathbf{a}_2'\mathbf{a}_2 = 1 \text{ and} \\
&\quad \mathrm{Cov}(\mathbf{a}_1'\boldsymbol{X}, \mathbf{a}_2'\boldsymbol{X}) = 0
\end{aligned}
$$

At the ith step,

$$
\begin{aligned}
i\text{th principle component} &= \text{linear combination } \mathbf{a}_i'\boldsymbol{X} \text{ that maximizes} \\
&\quad \mathrm{Var}(\mathbf{a}_i'\boldsymbol{X}) \text{ subject to } \mathbf{a}_i'\mathbf{a}_i = 1 \text{ and} \\
&\quad \mathrm{Cov}(\mathbf{a}_i'\boldsymbol{X}, \mathbf{a}_k'\boldsymbol{X}) = 0 \quad \text{for} \quad k < i
\end{aligned}
$$

Results 5.1 Let $\boldsymbol{\Sigma}$ be the covariance matrix associated with the random vector $\boldsymbol{X}' = [X_1, X_2, \ldots, X_p]$. Let $\boldsymbol{\Sigma}$ have the eigenvalue-eigenvector pair $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \ldots, (\lambda_p, \mathbf{e}_p)$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. Then the ***ith principal component*** is given by

$$Y_i = \mathbf{e}_i'\boldsymbol{X} = e_{i1}X_1 + e_{i2}X_2 + \cdots + e_{ip}X_p, i = 1, 2, \ldots, p$$

With these choices,

$$\begin{aligned} \mathrm{Var}(Y_i) &= \mathbf{e}_i'\boldsymbol{\Sigma}\mathbf{e}_i = \lambda_i, i = 1, 2, \ldots, p \\ \mathrm{Cov}(Y_i, Y_k) &= \mathbf{e}_i'\boldsymbol{\Sigma}\mathbf{e}_k = 0, i \neq k \end{aligned}$$

If some λ_i are equal, the choices of corresponding coefficients vectors, $\mathbf{e}_i$, and hence Y_i are not unique.

Results 5.2 Let $\boldsymbol{X}' = [X_1, X_2, \ldots, X_p]$ have covariance matrix $\boldsymbol{\Sigma}$, with eigenvalue-eigenvector pairs $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \ldots, (\lambda_p, \mathbf{e}_p)$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. Let $Y_1 = \mathbf{e}_1'\boldsymbol{X}, Y_2 = \mathbf{e}_2'\boldsymbol{X}, \ldots, Y_p = \mathbf{e}_p'\boldsymbol{X}$ be the principal components. Then

$$\sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp} = \sum_{i=1}^{p} \mathrm{Var}(X_i) = \lambda_1 + \lambda_2 + \cdots + \lambda_p = \sum_{i=1}^{p} \mathrm{Var}(Y_i)$$

Results 5.3 If $Y_1 = \mathbf{e}_1'\boldsymbol{X}, Y_2 = \mathbf{e}_2'\boldsymbol{X}, \ldots, Y_p = \mathbf{e}_p'\boldsymbol{X}$ are the principal components obtained from the covariance matrix $\mathbf{\Sigma}$, then

$$\rho_{Y_i, X_k} = \frac{e_{ik}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}, i, k = 1, 2, \ldots, p$$

are the correlation coefficients between the components Y_i and the variables X_k. Here $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \ldots, (\lambda_p, \mathbf{e}_p)$ are the eigenvalue-eigenvector pair for $\mathbf{\Sigma}$.

Example 5.1 Suppose the random variables X_1, X_2 and X_3 have the covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Calculating the population principal components

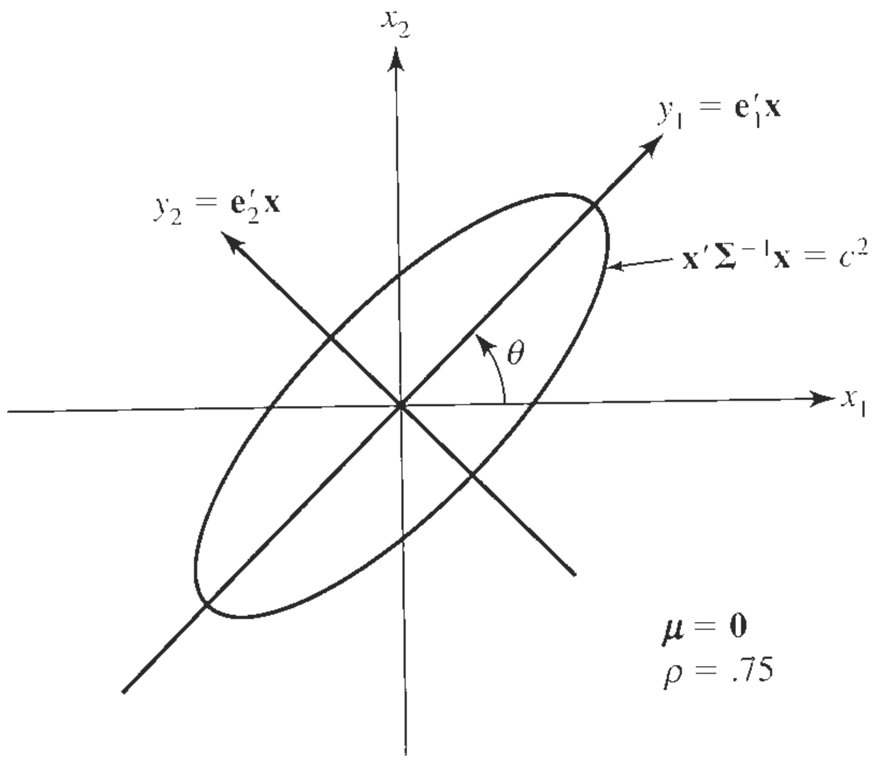

Figure 8.1 The constant density ellipse $\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x} = c^2$ and the principal components y_1, y_2 for a bivariate normal random vector $\mathbf{X}$ having mean $\mathbf{0}$.

Suppose $\boldsymbol{X}$ is distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We know that the density of $\boldsymbol{X}$ is constant on the $\boldsymbol{\mu}$ centered ellipsoids

$$(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = c^2$$

which have axes $\pm\sqrt{\lambda_i}\mathbf{e}_i, i = 1, 2, \ldots, p$, where the $(\lambda_i, \mathbf{e}_i)$ are the eigenvalue-eigenvector pairs of $\boldsymbol{\Sigma}$. Assume $\boldsymbol{\mu} = 0$, the equation above can be rewritten as

$$\begin{aligned} c^2 &= \mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x} = \frac{1}{\lambda_1}(\mathbf{e}'\mathbf{x})^2 + \frac{1}{\lambda_2}(\mathbf{e}_2'\mathbf{x})^2 + \cdots + \frac{1}{\lambda_p}(\mathbf{e}_p'\mathbf{x})^2 \\ &= \frac{1}{\lambda_1}y_1^2 + \frac{1}{\lambda_2}y_2^2 + \cdots + \frac{1}{\lambda_p}y_p^2 \end{aligned}$$

where $\mathbf{e}_1'\mathbf{x}, \mathbf{e}_2'\mathbf{x}, \ldots, \mathbf{e}_p'\mathbf{x}$ are recognized as the principal components of $\mathbf{x}$. The equation above defines in a coordinate system with axes $y_1, y_2, \ldots, y_p$ lying in the direction $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p$, respectively.

Principal Components Obtained from Standardized Variables

Principal components may also be obtained from the standardized variables

$$Z_i = \frac{X_i - \mu_i}{\sqrt{\sigma_{ii}}}, i = 1, 2, \dots, p$$

Or in matrix notation $\mathbf{Z} = (\mathbf{V}^{1/2})^{-1}(\boldsymbol{X} - \boldsymbol{\mu})$. Clearly $\mathrm{E}(\mathbf{Z}) = 0$ and $\mathrm{Cov}(\mathbf{Z}) = (\mathbf{V}^{1/2})^{-1}\Sigma(\mathbf{V}^{1/2})^{-1} = \boldsymbol{\rho}$

Results 5.4 The ith principal component of the standardized variables $\mathbf{Z}' = [Z_1, Z_2, \dots, Z_p]$ with $\mathrm{Cov}(\mathbf{Z}) = \boldsymbol{\rho}$, is given by

$$Y_i = \mathbf{e}_i'\mathbf{Z} = \mathbf{e}_i(\mathbf{V}^{1/2})^{-1}(\boldsymbol{X} - \boldsymbol{\mu}), i = 1, 2, \dots, p$$

Moveover,

$$\sum_{i=1}^{p} \mathrm{Var}(Y_i) = \sum_{i=1}^{p} \mathrm{Var}(Z_i) = p$$

and

$$\rho_{Y_i, Z_k} = e_{ik}\sqrt{\lambda_i}, i, k = 1, 2, \dots, p$$

In this case $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \dots, (\lambda_p, \mathbf{e}_p)$ are the eigenvalue-eigenvector pairs for $\boldsymbol{\rho}$, with $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_p \geq 0$.

Example 5.2 Consider the covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & 4 \\ 4 & 100 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\rho} = \begin{bmatrix} 1 & .4 \\ .4 & 1 \end{bmatrix}$$

Obtain the principle components by the covariance matrix $\mathbf{\Sigma}$ and correlation matrix $\boldsymbol{\rho}$.

Principal Components for Covariance Matrices with Special Structures

1.

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{pp} \end{bmatrix}$$

2.

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \cdots & \rho\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho\sigma^2 & \rho\sigma^2 & \cdots & \sigma^2 \end{bmatrix} \quad \text{or} \quad \boldsymbol{\rho} = \mathbf{\Sigma} = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}$$

Summarizing Sample Variation by Principle Components

Suppose the data $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ represent n independent drawings from some p-dimensional population withe mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. These data yield the sample mean vector $\bar{\mathbf{x}}$, the sample covariance matrix $\mathbf{S}$, and the sample correlation matrix $\mathbf{R}$.

If $\mathbf{S} = s_{ik}$ be $p \times p$ sample covariance matrix with eigenvalue-eigenvector pairs $(\hat{\lambda}_1, \hat{\mathbf{e}}_1), (\hat{\lambda}_2, \hat{\mathbf{e}}_2), \ldots, (\hat{\lambda}_p, \hat{\mathbf{e}}_p)$, the ith sample principal component is given by

$$\hat{y}_i = \hat{\mathbf{e}}_i'\mathbf{x} = \hat{e}_{i1}x_1 + \hat{e}_{i2}x_2 + \cdots + \hat{e}_{ip}x_p, i = 1, 2, \ldots, p$$

where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p \geq 0$ and $\mathbf{x}$ is any observation on the variables $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p$. Also

$$\begin{aligned}
\text{Sample variance}(\hat{y}_k &= \hat{\lambda}_k, k = 1, 2, \ldots, p \\
\text{Sample covariance}(\hat{y}_i, \hat{y}_k) &= 0, i \neq k \\
\text{Total sample variance} &= \sum_{i=1}^{n} s_{ii} = \hat{\lambda}_1 + \hat{\lambda}_2 + \cdots + \hat{\lambda}_p \\
r_{\hat{y}_i, x_k} &= \frac{\hat{e}_{ik}\sqrt{\hat{\lambda}_i}}{\sqrt{s_{kk}}}, i, k = 1, 2, \ldots, p.
\end{aligned}$$

Example 5.3 (Summarizing sample variability with two sample principal components) A census provided information, by tract, on five socioeconomic variables for Madison, Wisconsin, area. The data from 61 tracts are list in Table 8.5. These data produced the following summary statistics

$\bar{\mathbf{x}}' = [4.47,$	$3.96,$	$71.42,$	$26.91,$	$1.64]$
total population (thousands)	professional degree (percent)	employed age over 16 (percent)	government employment (percent)	median home value \$100,000

and

$$\mathbf{S} = \begin{bmatrix} 3.397 & -1.102 & 4.306 & -2.078 & 0.027 \\ -1.102 & 9.673 & -1.513 & 10.953 & 1.203 \\ 4.306 & -1.513 & 55.626 & -28.937 & -0.044 \\ -2.078 & 10.953 & -28.937 & 89.067 & 0.957 \\ 0.027 & 1.203 & -0.044 & 0.957 & 0.319 \end{bmatrix}$$

Can the sample variation be summarized by one or two principal components ?

Coefficients for the Principal Components
(Correlation Coefficients in Parentheses)

Variable	$\hat{\mathbf{e}}_1\ (r_{\hat{y}_1, x_k})$	$\hat{\mathbf{e}}_2\ (r_{\hat{y}_2, x_k})$	$\hat{\mathbf{e}}_3$	$\hat{\mathbf{e}}_4$	$\hat{\mathbf{e}}_5$
Total population	−0.039(−.22)	0.071(.24)	0.188	0.977	−0.058
Profession	0.105(.35)	0.130(.26)	−0.961	0.171	−0.139
Employment (%)	−0.492(−.68)	0.864(.73)	0.046	−0.091	0.005
Government employment (%)	0.863(.95)	0.480(.32)	0.153	−0.030	0.007
Medium home value	0.009(.16)	0.015(.17)	−0.125	0.082	0.989
Variance ($\hat{\lambda}_i$):	107.02	39.67	8.37	2.87	0.15
Cumulative percentage of total variance	67.7	92.8	98.1	99.9	1.000

The number of Principal Components

There is always the question of how many components to retain. There is no definitive answer to this question. Things to consider include

- the amount of total sample variance explained,
- the relative sizes of the eigenvalues (the variances of the sample components,)
- the subject-matter interpretation of the components.
- In addition, a component associated with an eigenvalue near zero and, hence deemed unimportant, may indicate an unsuspected linear dependency in the data.

A useful visual aid to determining an appropriate number of principal components is a ***scree plot***. With the eigenvalues ordered from largest to smallest, a scree plot is a plot of $\hat{\lambda}_i$ versus i—the magnitude of an eigenvalue versus its number.

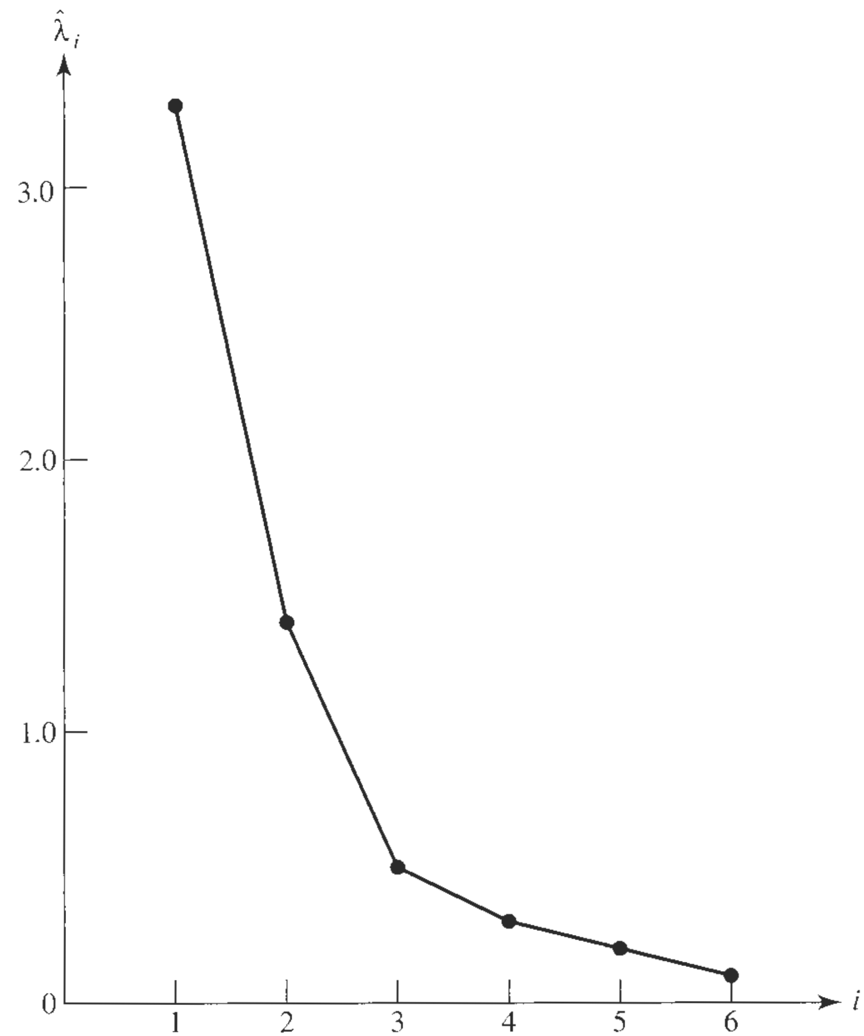

Figure 8.2 A scree plot.

Example 5.4 (Summarizing sample variability with one sample principal component) In a study of size and shape relationships for painted turtles, Jolicoeur and Mosimann measured carapace length, width, and height. Their data, reproduced in Table 6.9 suggest an analysis in term s of logarithms (Jolicoeur generally suggests a logarithmic transformation in studies of size-and-shape relationships). Perform a principal component analysis.

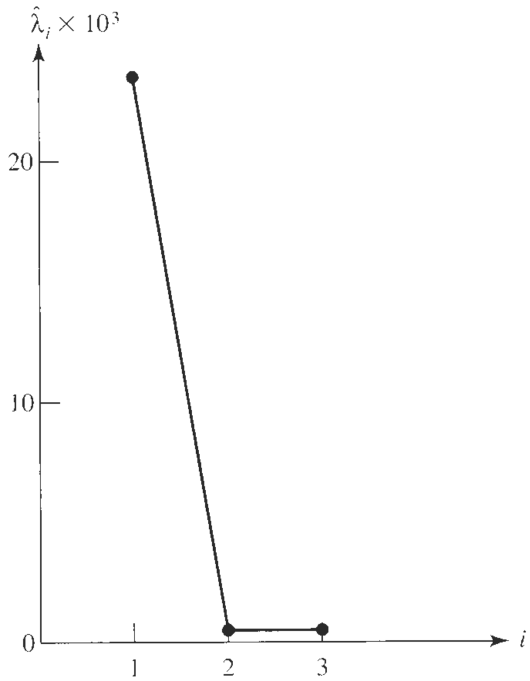

Figure 8.3 A scree plot for the turtle data.

Coefficients for the Principal Components
(Correlation Coefficients in Parentheses)

Variable	$\hat{\mathbf{e}}_1 (r_{\hat{y}_1, x_k})$	$\hat{\mathbf{e}}_2$	$\hat{\mathbf{e}}_3$
ln (length)	.683 (.99)	−.159	−.713
ln (width)	.510 (.97)	−.594	.622
ln (height)	.523 (.97)	.788	.324
Variance ($\hat{\lambda}_i$):	23.30×10^{-3}	$.60 \times 10^{-3}$	$.36 \times 10^{-3}$
Cumulative percentage of total variance	96.1	98.5	100

PANEL 8.1 SAS ANALYSIS FOR EXAMPLE 8.4 USING PROC PRINCOMP.

PROGRAM COMMANDS

```
title 'Principal Component Analysis';
data turtle;
infile 'E8-4.dat';
input length width height;
x1 = log(length); x2 =log(width); x3 =log(height);
proc princomp cov data = turtle out = result;
var x1 x2 x3;
```

OUTPUT

Principal Components Analysis

24 Observations
3 Variables

Simple Statistics

	X1	X2	X3
Mean	4.725443647	4.477573765	3.703185794
StD	0.105223590	0.080104466	0.082296771

Covariance Matrix

	X1	X2	X3
X1	0.0110720040	0.0080191419	0.0081596480
X2	0.0080191419	0.0064167255	0.0060052707
X3	0.0081596480	0.0060052707	0.0067727585

Total Variance = 0.024261488

Eigenvalues of the Covariance Matrix

	Eigenvalue	Difference	Proportion	Cumulative
PRIN1	0.023303	0.022705	0.960508	0.96051
PRIN2	0.000598	0.000238	0.024661	0.98517
PRIN3	0.000360		0.014832	1.00000

Eigenvectors

	PRIN1	PRIN2	PRIN3
X1	0.683102	−.159479	−.712697
X2	0.510220	−.594012	0.621953
X3	0.522539	0.788490	0.324401

Interpretation the Sample Principal Components

The sample principal components have serval interpretations

- Suppose the underlying distribution of $\boldsymbol{X}$ is nearly $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, Then the sample principal components $\hat{y}_i = \hat{\mathbf{e}}_i'(\mathbf{x} - \bar{\mathbf{x}})$ are realizations of population principal components $Y_i = \mathbf{e}_i(\boldsymbol{X} - \boldsymbol{\mu})$, which have an $N_p(0, \Lambda)$ distribution. The diagonal matrix Λ has entries $\lambda_1, \lambda_2, \ldots, \lambda_p$ and $(\lambda_i, \mathbf{e}_i)$ are the eigenvalue-eigenvector pairs of $\boldsymbol{\Sigma}$.

- From the sample value $\mathbf{x}_j$, we can approximate $\boldsymbol{\mu}$ by $\bar{\mathbf{x}}$ and $\boldsymbol{\Sigma}$ by $\mathbf{S}$. If $\mathbf{S}$ is positive definite, the contour consisting of all $p \times 1$ vector $\mathbf{x}$ satisfying

$$(\mathbf{x} - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x} - \bar{\mathbf{x}}) = c^2$$

  estimates the constant density contour $(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = c^2$ of the underlying normal density.

- Even when the normal assumption is suspect and the scatter plot may depart somewhat from an elliptical pattern, we can still extract eigenvalues from $\mathbf{S}$ and obtain the sample principal components.

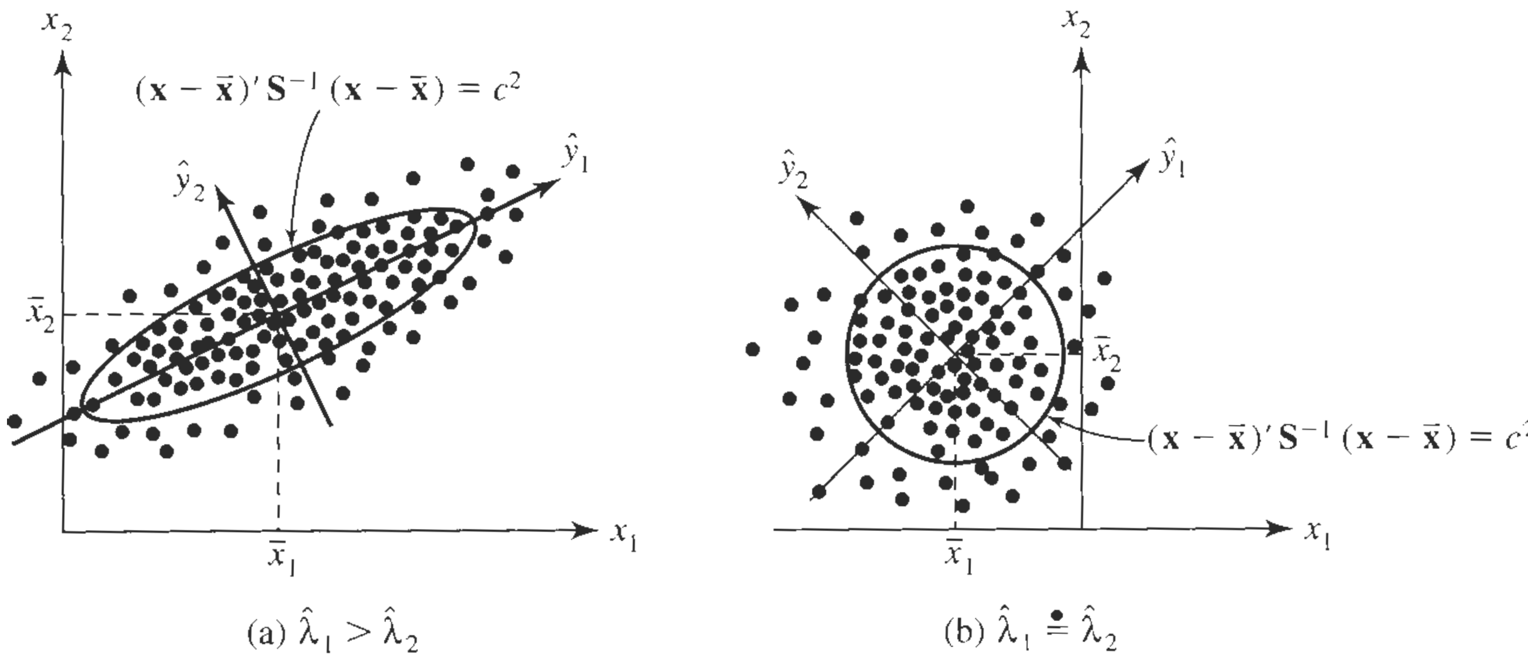

Figure 8.4 Sample principal components and ellipses of constant distance.

Standardizing the Sample Principal Components

- Sample principal components are, in general, not invariant with respect to changes in scale.

- Variables measured on different scales or on a common scale with widely differing ranges are often standardized. For example, standardization is accomplished by constructing

$$\mathbf{z}_j = \mathbf{D}^{-1/2}(\mathbf{x}_j - \bar{\mathbf{x}}) = \left[\frac{x_{j1} - \bar{x}_1}{\sqrt{s_{11}}}, \frac{x_{j2} - \bar{x}_2}{\sqrt{s_{22}}}, \ldots, \frac{x_{j1} - \bar{x}_p}{\sqrt{s_{pp}}}\right]', j = 1, 2, \ldots, n.$$

If $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n$ are standardized observations with covariance matrix $\mathbf{R}$, the ith sample principal component is

$$\hat{y}_i = \hat{\mathbf{e}}_i'\mathbf{z} = \hat{e}_{i1}z_1 + \hat{e}_{i2}z_2 + \cdots + \hat{e}_{ip}z_p, i = 1, 2, \ldots, p$$

where $(\hat{\lambda}_i, \hat{\mathbf{e}}_i)$ is the ith eigenvalue-eigenvector pair of $\mathbf{R}$ with

$$\text{Sample variance}(\hat{y}_i) = \hat{\lambda}_i, i = 1, 2, \ldots, p$$
$$\text{Sample covariance}(\hat{y}_i, \hat{y}_k) = 0, i \neq k$$

In addition,

$$\text{Total(standardized) sample variance} = \text{tr}(\mathbf{R}) = p = \hat{\lambda}_1 + \hat{\lambda}_2 + \cdots + \hat{\lambda}_p$$

and

$$r_{\hat{y}_i, z_k} = \hat{e}_{ik}\sqrt{\hat{\lambda}_i}, i, k = 1, 2, \ldots, p$$

Example 5.5 (Sample principal components from standardized data) The weekly rates return for five stocks (JP Morgen, Citibank, Wells Fargo, Royal Dutch Shell, and ExxonMobil) list on the New York Stock Exchange were determined for the period January 2004 through December 2005. The weekly rates of return are defined as (current week closing price-previous week closing price)/(previous week closing price), adjusted for stock splits and dividends, The data are listed in Table 8.4. The observations in 103 successive weeks appear to be independently distributed, but the rates of return across stocks are correlated, because as one expects, stocks tend to move together in response to general economic conditions. Standardizing this data set and find sample principal components data set after standardized.

Example 5.6 (Components from a correlation matrix with a special structure) Geneticists are often concerned with the inheritance of characteristics that can be measured several times during an animal's lifetime. Body weight (in grams) for $n = 150$ female mice were obtained immediately after the birth of their first four litters. The sample mean vector and sample correlation matrix were, respectively,

$$\bar{\mathbf{x}}' = [39.88, 45.08, 48.11, 49.95]$$

and

$$\mathbf{R} = \begin{bmatrix} 1.000 & .7501 & .6329 & .6363 \\ .7501 & 1.000 & .6925 & 7386 \\ .6329 & .6925 & 1.000 & .6625 \\ .6363 & .7386 & .6625 & 1.000 \end{bmatrix}$$

Find sample principal components by $\mathbf{R}$.

5.2 Factor Analysis and Inference for Structured Covariance Matrices

- 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***

- 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.

5.2.1 The Orthogonal Factor Model

- The observable random vector $\boldsymbol{X}$, with p components, has mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.
- The factor model postulates that $\boldsymbol{X}$ 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, sometimes, specific factors.
- In particular, the factor analysis model is

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

or in matrix notation

$$\boldsymbol{X} - \boldsymbol{\mu} = \mathbf{LF} + \boldsymbol{\varepsilon}$$

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***.

- The unobservable random vectors **F** and ε satisfy the following conditions:

  F and ε are independent
  $\mathrm{E}(\mathbf{F}) = 0, \mathrm{Cov}(\mathbf{F}) = \mathbf{I}$
  $\mathrm{E}(\varepsilon) = 0, \mathrm{Cov}(\varepsilon) = \mathbf{\Psi}$, where $\mathbf{\Psi}$ is diagonal matrix.

- Covariance structure for the Orthogonal Factor Model

  1. $\mathrm{Cov}(\boldsymbol{X}) = \mathbf{L}\mathbf{L}' + \mathbf{\Psi}$ or

$$\begin{aligned} \mathrm{Var}(X_i) &= \ell_{i1}^2 + \cdots + \ell_{im}^2 + \psi_i \hat{=} h_i^2 + \psi_i \\ \mathrm{Cov}(X_i, X_k) &= \ell_{i1}\ell_{k1} + \cdots + \ell_{im}\ell_{km} \end{aligned}$$

  2. $\mathrm{Cov}(\boldsymbol{X}, \mathbf{F}) = \mathbf{L}$ or

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

Example 5.7 Consider the covariance matrix

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

Verifying the relation $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}' + \mathbf{\Psi}$ for two factors

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.

Example 5.8 Let $p = 3$ and $m = 1$, and suppose the random variables X_1, X_2 and X_3 have the positive definite covariance matrix

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

Show $\boldsymbol{\Sigma}$ can not be factored by a factor analysis model with $m = 1$.

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

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

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

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

$$\boldsymbol{\Sigma} = \mathbf{LL}' + \boldsymbol{\Psi} = \mathbf{LTT}'\mathbf{L}' + \boldsymbol{\Psi} = (\mathbf{L}^*)(\mathbf{L}^*)' + \boldsymbol{\Psi}$$

Methods of Estimation

The Principal Component Solution of the Factor Model

The principal component analysis of the sample covariance matrix **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 estimate factor loading $\{\tilde{\ell}_{ij}\}$ is give by

$$\tilde{\mathbf{L}} = \left[\sqrt{\hat{\lambda}_1}\hat{\mathbf{e}}_1 \vdots \sqrt{\hat{\lambda}_2}\hat{\mathbf{e}}_2 \vdots \cdots \vdots \sqrt{\hat{\lambda}_m}\hat{\mathbf{e}}_m\right]$$

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

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

Communalities are estimated as

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

The principal component factor analysis of the sample correlation matrix is obtained by starting with **R** in place of **S**.

- For the principal component solution, the estimated loading for a given factor do not changes as the number of factors is increased.

- The choice of m can be based on the estimated eigenvalues in much the same manner as with principal components.

- Analytically, we have

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

- Ideally, the contributions of the first few factors to the sample variance of the variables should be large.

$$\begin{pmatrix} \text{Proportion of total} \\ \text{sample variance} \\ \text{due to } j\text{th factor} \end{pmatrix} = \begin{cases} \frac{\hat{\lambda}_j}{s_{11}+s_{22}+\cdots+s_{pp}} & \text{for a factor analysis of } \mathbf{S} \\ \frac{\lambda_j}{p} & \text{for a factor analysis of } \mathbf{R} \end{cases}$$

Example 5.9 In a consumer-preference study, a random sample of customers were asked to rate several attributions of a new product. The response, on a 7-point semantic differential scale, were tabulated and the attribute correlation matrix constructed. The correlation matrix is presented next:

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

do factor analysis for this consumer-preference data

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		

Example 5.10 Stock-price data consisting of $n = 103$ weekly rates of return on $p = 5$ stocks were introduced in Example 5.5. Do factor analysis for this data.

Table 9.2

Variable	One-factor solution: Estimated factor loadings F_1	One-factor solution: Specific variances $\tilde{\psi}_i = 1 - \tilde{h}_i^2$	Two-factor solution: Estimated factor loadings F_1	Two-factor solution: Estimated factor loadings F_2	Two-factor solution: Specific variances $\tilde{\psi}_i = 1 - \tilde{h}_i^2$
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 Maximum Likelihood Method

Results 5.5 Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be a random sample from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$ is the covariance matrix for the m common factor model. The maximum likelihood estimator $\hat{\mathbf{L}}$ and $\hat{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ maximize the likelihood function of $\boldsymbol{X}_j - \boldsymbol{\mu} = \mathbf{L}\mathbf{F}_j + \boldsymbol{\varepsilon}_j, j = 1, 2, \ldots, n$

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{np}{2}} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} e^{-\frac{1}{2}\text{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' + n(\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})'\right)\right]}$$

subject to $\hat{\mathbf{L}}\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}}$ be diagonal.

The maximum likelihood estimates of the communalities are

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

so

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

Although the likelihood in Results 5.5 is appropriate for **S**, not **R**, surprisingly, this practice is equivalent to obtaining the maximum likelihood estimate $\hat{\mathbf{L}}$ and $\hat{\boldsymbol{\Psi}}$ based on the sample covariance matrix **S**, setting $\hat{\mathbf{L}}_z = \hat{\mathbf{V}}^{-1/2}\hat{\boldsymbol{\Psi}}\hat{\mathbf{V}}^{-1/2}$.

Here $\hat{\mathbf{V}}^{-1/2}$ is the diagonal matrix with reciprocal of the sample standard deviation (computed with the divisor $\sqrt{n}$) on the main diagonal, and **Z** is the standardized observation with sample mean 0 and sample standard deviation 1.

Example 5.11 Using the maximum likelihood method do factor analysis for the stock-price data.

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	

Example 5.12 (Factor analysis of Olympic decathlon data) Linden 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

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

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	

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:

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

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

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

$$\hat{\mathbf{L}}^* = \mathbf{L}\mathbf{T}, \text{where } \mathbf{T}\mathbf{T}' = \mathbf{T}'\mathbf{T} = \mathbf{I}$$

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

$$\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}} = \hat{\mathbf{L}}\mathbf{T}\mathbf{T}'\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}} = \hat{\mathbf{L}}^*\hat{\mathbf{L}}^{*'} + \hat{\mathbf{\Psi}}$$

Example 5.13 (A first look at factor rotation) Lawley and Maxwell present the sample correlation matrix of examination scores in $p = 6$ subject areas for $n = 220$ male students. The correlation matrix is

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

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

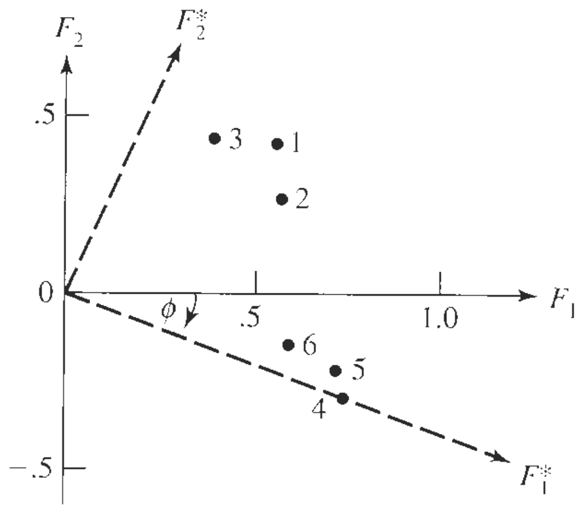

Figure 9.1 Factor rotation for test scores.

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

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

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

as large as possible.

Scaling the rotated coefficient $\hat{\ell}^*_{ij}$ 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.

Example 5.14 (Rotated Loading for the consumer-preference data)

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	

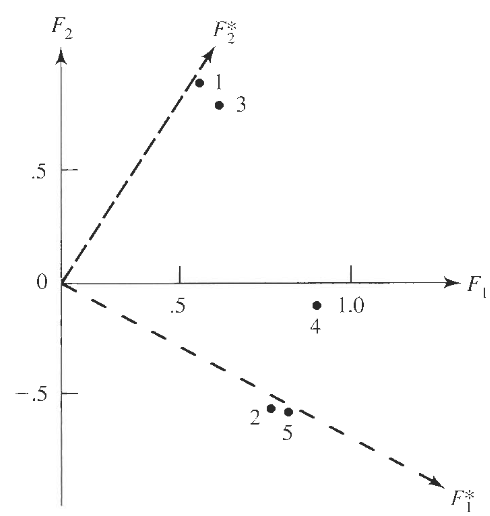

Figure 9.2 Factor rotation for hypothetical marketing data.

PANEL 9.1 SAS ANALYSIS FOR EXAMPLE 9.9 USING PROC FACTOR.

PROGRAM COMMANDS

```
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;
```

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

Example 5.15 (Rotated loading for the stock-price data)

Table 9.8

Variable	Maximum likelihood estimates of factor loadings F_1	F_2	Rotated estimated factor loadings F_1^*	F_2^*	Specific variances $\tilde{\psi}_i^2 = 1 - \hat{h}_i^2$
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	

Example 5.15 (Rotated loadings for the Olympic decathlon data)

Table 9.9

Variable	Principal component: Estimated rotated factor loadings, $\tilde{\ell}^*_{ij}$ F^*_1	F^*_2	F^*_3	F^*_4	Specific variances $\tilde{\psi}_i = 1 - \tilde{h}^2_i$	Maximum likelihood: Estimated rotated factor loadings, $\hat{\ell}^*_{ij}$ F^*_1	F^*_2	F^*_3	F^*_4	Specific variances $\hat{\psi}_i = 1 - \hat{h}^2_i$
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	

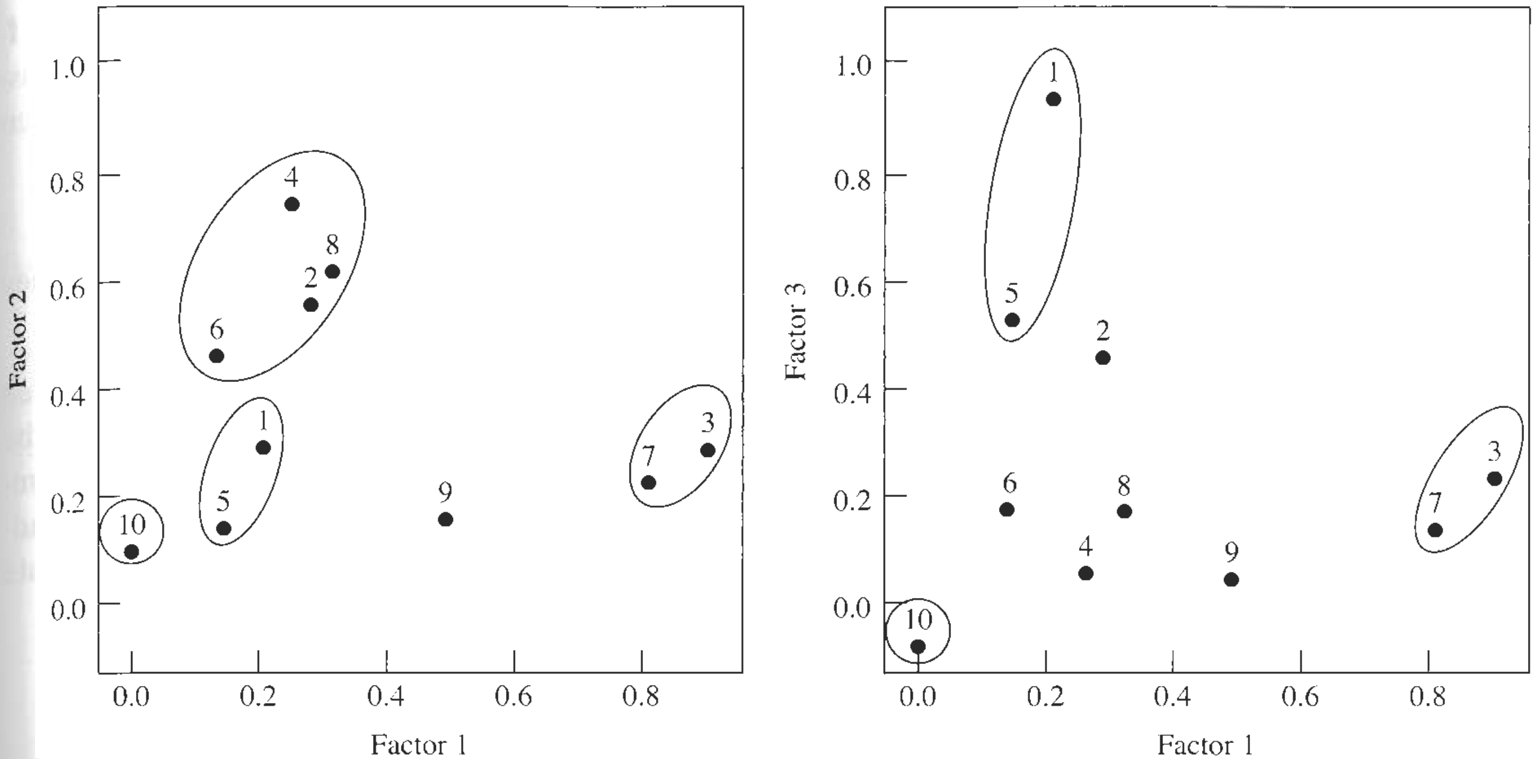

Figure 9.3 Rotated maximum likelihood loadings for factor pairs (1, 2) and (1, 3)—decathlon data. (The numbers in the figures correspond to variables.)

Factor Scores

- The estimate values of the common factors, called ***factor scores*** may also 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

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

- Normally the factor score approaches have two elements in common:

  1. They treat the estimate factor loadings $\hat{\ell}_{ij}$ and specific variance $\hat{\psi}_i$ as if they were 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.

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

$$
\begin{aligned}
\hat{\mathbf{f}}_j &= (\hat{\mathbf{L}}'\hat{\mathbf{\Psi}}^{-1}\hat{\mathbf{L}})^{-1}\hat{\mathbf{L}}'\hat{\mathbf{\Psi}}^{-1}(\mathbf{x}_j - \hat{\boldsymbol{\mu}}) \\
&= \hat{\mathbf{\Delta}}^{-1}\hat{\mathbf{L}}'\hat{\mathbf{\Psi}}^{-1}(\mathbf{x}_j - \hat{\mathbf{x}}), j = 1, 2, \ldots, n
\end{aligned}
$$

or, if the correlation matrix is factored

$$
\begin{aligned}
\hat{\mathbf{f}}_j &= (\hat{\mathbf{L}}_z'\hat{\mathbf{\Psi}}_z^{-1}\hat{\mathbf{L}}_z)^{-1}\hat{\mathbf{L}}_z'\hat{\mathbf{\Psi}}_z^{-1}\mathbf{z}_j \\
&= \hat{\mathbf{\Delta}}_z^{-1}\hat{\mathbf{L}}_z'\hat{\mathbf{\Psi}}_z^{-1}\mathbf{z}_j, j = 1, 2, \ldots, n
\end{aligned}
$$

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

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

- 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 amount to assuming that the ψ_i are equal or nearly equal. The factor scores are then

$$\hat{\mathbf{f}}_j = (\hat{\mathbf{L}}'\hat{\mathbf{L}})^{-1}\hat{\mathbf{L}}'(\mathbf{x}_j - \hat{\boldsymbol{\mu}}) \quad \text{or} \quad \hat{\mathbf{f}}_j = (\hat{\mathbf{L}}_z'\hat{\mathbf{L}}_z)^{-1}\hat{\mathbf{L}}_z'\mathbf{z}_j$$

  for standardized data.

Factor Scores Obtained by Regression

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

or, if a correlation matrix is factored

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

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

Example 5.16 (Computing factor scores) Compute factor scores by the least squares and regression methods using the stock-price data discussed in 5.11.

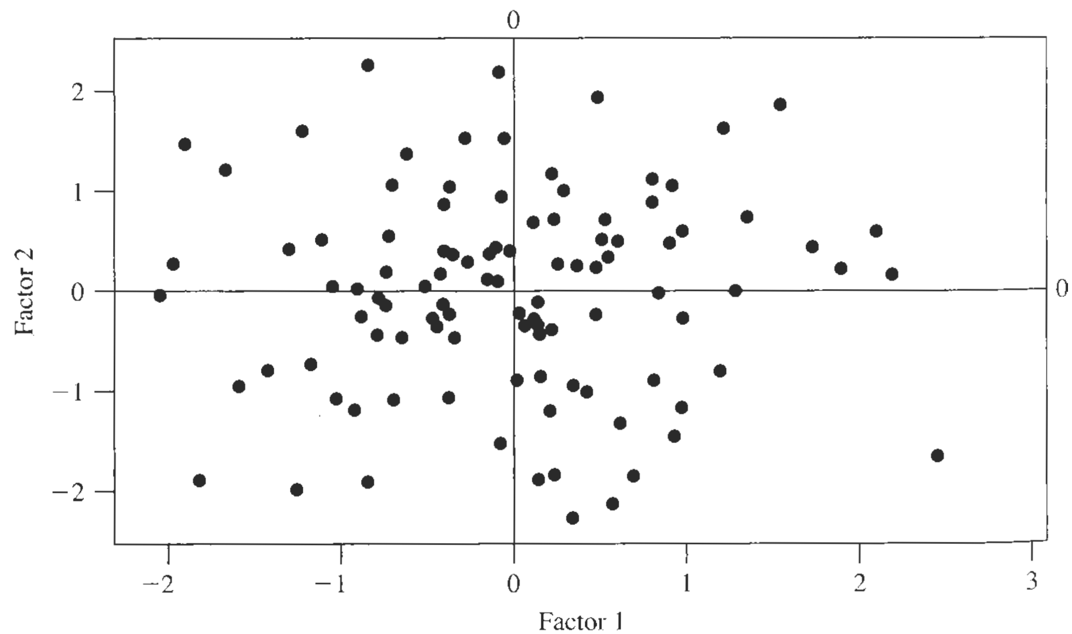

Figure 9.4 Factor scores using (9-58) for factors 1 and 2 of the stock-price data (maximum likelihood estimates of the factor loadings).

Perspectives and a Stragegy for Factor Analysis

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 **R** or **S** be nonsingular)

   (a) Look for suspicious observations by plotting the factor scores. Also, calculate standardized scores for each observation and squared distances.
   (b) Try a varimax rotation.

2. ***Perform a maximum likelihood factor analysis, including a varimax rotation.***

3. ***Compare the solution obtained from the two factor analysis.***

   (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.