UNIT 24  PRINCIPAL COMPONENT ANALYSIS

24.1 Introduction

Principal component analysis is a useful statistical technique that has found applications in diverse fields. It helps in reducing the dimensionality of the data set and identification of patterns in the data through graphical representation. In nutshell, we can say that principal component analysis is a data reduction technique.

The purpose of principal component analysis is to derive a small number of linear combinations (principal components) of a set of variables that retain as much information in the original variables as possible, which are discussed in Section 24.2. Often a small number of principal components can be used in place of the original variables for plotting, regression, clustering and so on. Principal component analysis can also be viewed as a technique to remove multicollinearity in the data. In this technique, we transform the original set of variables to a new set of uncorrelated random variables, called principal components. These new variables are linear combinations of the original variables and are derived in decreasing order of importance so that the first principal component accounts for as much as possible of the variation in the original data. The researcher often hopes that the first few components will account for most of the variation in the original data so that the effective dimensionality of the data can be reduced. The variance of principal components is computed in Section 24.3 In Section 24.4, contribution of variability through spectral decomposition of $\Sigma$ is discussed.

Principal component analysis often reveals relationships that were not previously suspected and thereby allows interpretations that would not ordinarily result. Geometrically, these linear combinations represent the selection of a new coordinate system obtained by rotating the original system with say, $x_1, x_2, \ldots, x_p$ as the coordinate axes. The new axes represent the directions with maximum variability and provide a simple and more parsimonious description of the covariance structure. The identification of important components are done in Section 24.5. Principal components from the correlation matrix are discussed in Section 24.6. In Section 24.7, estimation of principal components is discussed. The scaling convention, geometrical interpretation and applications of principal components are discussed in Section 24.8, 24.9 and 24.10 respectively.
Objectives
After studying this unit, one will be able to

- define principal components
- derive principal components from covariance matrix
- derive principal components from a correlation matrix
- identify the important principal components
- describe the geometrical interpretations of principal components
- apply the principal components in diverse fields

24.2 DERIVATION OF PRINCIPAL COMPONENTS

Let us start this section by defining and deriving the principal components. Suppose \(x' = \begin{bmatrix} x_1, \cdots, x_p \end{bmatrix}\) is a p-dimensional random variable with mean vector \(\mu\) and covariance matrix \(\Sigma = \begin{pmatrix} \sigma_{ij} \end{pmatrix}\). The problem is to find a new set of variables, say \(z_1, z_2, \cdots, z_p\), which are uncorrelated and whose variances decrease from first to last. Each \(z_j\) is taken to be a linear combination of the \(x\)'s, so that

\[
 z_j = a_{1j}x_1 + a_{2j}x_2 + \cdots + a_{pj}x_p \\
= a_j'x, \quad \text{where } j = 1, 2, \cdots, p,
\]

where \(a_j' = \begin{bmatrix} a_{1j}, a_{2j}, \cdots, a_{pj} \end{bmatrix}\) is a vector of constants. Equation (1) contains an arbitrary scale factor. We therefore impose the condition that \(a_j'a_j = \sum_{k=1}^{p} a_{kj}^2 = 1\). We shall see that this particular normalization procedure ensures that the overall transformation is orthogonal in other words, that distances in \(p\)-space are preserved.

The first principal component \(z_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1p}x_p\), is found by choosing \(a_1\) such that variance of \(z_1\) is as large as possible subject to the condition that \(a_{11}^2 + a_{12}^2 + \cdots + a_{1p}^2 = 1\) or \(a_1'a_1 = 1\).

This constraint is introduced because if this is not done, then \(\text{Var}(z_1)\) can be increased simply by multiplying each \(a_{ij}\)'s by a common constant factor.

The second principal component \(z_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2p}x_p\), is found by choosing \(a_2\) such that \(\text{Var}(z_2)\) is as large as possible (next to \(\text{Var}(z_1)\)) subject to the constraint that \(a_{21}^2 + a_{22}^2 + \cdots + a_{2p}^2 = 1\) or \(a_2'a_2 = 1\) and \(\text{Cov}(z_1, z_2) = 0\). Continuing this process, the remaining \(p-2\) principal components \(z_3, \cdots, z_p\) can be derived so that \(z_i\) is uncorrelated with the first \(i-1\) principal components and variance has maximum subject to \(a_j'a_j = 1\). Then \(z_i\)'s have variances in decreasing order.

We begin by finding the first principal component. We want to choose \(a_1\) so as to maximize the variance of \(z_1\) subject to the normalization constraint that \(a_1'a_1 = 1\).
As
\[ \text{Var}(z_i) = \text{Var}(a'_i X) = a'_i \Sigma a_i. \] (2)
so we take \( a'_i \Sigma a_i \) as our objective function.

The standard procedure for maximizing a function of several variables with one or more constraints is the method of Lagrange multipliers. With just one constraint, this method uses the fact that the stationary points of a differentiable function of \( p \) variables, say \( f(x_1, \ldots, x_p) \), subject to a constraint \( g(x_1, \ldots, x_p) = c \), are such that there exists a number \( \lambda \), called the Lagrange multiplier, so that
\[ \frac{\partial f}{\partial x_i} - \lambda \frac{\partial g}{\partial x_i} = 0 \quad i = 1, 2, \ldots, p \] (3)
at the stationary points. These \( p \) equations, together with the constraint, are sufficient to determine the co-ordinates of the stationary points (and the corresponding values of \( \lambda \), which, however, are usually of little interest). Further investigation is needed to see if a stationary point is a maximum, minimum or saddle point. It is helpful to form a new function, \( L(x) \), such that
\[ L(x) = f(x) - \lambda [g(x) - c] \]
where the term in the square brackets is of course zero. Then the set of equations in (3) may be written simply as
\[ \frac{\partial L}{\partial x} = 0. \]

Applying this method to maximization of \( a'_i \Sigma a_i \), we write
\[ L(a_i) = a'_i \Sigma a_i - \lambda(a'_i a_i - 1). \]
Then, we have
\[ \frac{\partial L}{\partial x} = 2\Sigma a_i - 2\lambda a_i. \]
Setting this equal to 0, we have
\[ (\Sigma - \lambda I) a_i = 0. \] (4)

We now come to the crucial step in the argument. If Equation (4) is to have a non trivial solution for \( a_i \), then \( (\Sigma - \lambda I) \) must be a singular matrix. Thus \( \lambda \) must be chosen that
\[ |\Sigma - \lambda I| = 0. \]

Thus a non-zero solution for Equation (4) exists if and only if \( \lambda \) is an eigenvalue of \( \Sigma \). But \( \Sigma \) will generally have \( p \) eigenvalues, all of which must be non-negative as \( \Sigma \) is positive semi-definite symmetric matrix. Let us denote the eigenvalue by \( \lambda_1, \lambda_2, \ldots, \lambda_p \), and assume for the moment that they are distinct, and that \( \lambda_1 > \lambda_2 > \cdots > \lambda_p \geq 0 \). Which one shall be choose to determine the first principal component? Now,
\[ \text{Var}(a'_ix) = (a'_i \Sigma a_i) = (a'_i \lambda I a_i) \quad (\text{using equation (4)}) = \lambda. \]
Since we are interested in maximization of this variance, therefore, we choose \( \lambda \) to be the largest eigenvalue, namely \( \lambda_1 \). Then, using equation (4), \( a_1 \), must be the eigenvector of \( \Sigma \) corresponding to the largest eigenvalue.

The second principal component, namely \( z_2 = a'_2 x \), is obtained by an extension of the above argument. In addition to the scaling constraint that \( a'_2 a_2 = 1 \), we now have a second constraint that \( z_2 \) should be uncorrelated with \( z_1 \). Now,

\[
\text{Cov}(z_2, z_1) = \text{Cov}(a'_2 x, a'_1 x) = E\left[a'_2 (x-\mu)(x-\mu)' a_1\right] = a'_2 \Sigma a_1.
\]

Since, the first two principal components should be uncorrelated, therefore it is required that \( a'_2 \Sigma a_1 \) is equal to zero. But since \( \Sigma a_1 = \lambda_1 a_1 \), an equivalent simpler condition is that \( a'_2 a_1 = 0 \). In other words, \( a_1 \) and \( a_2 \) should be orthogonal.

In order to maximize the variance of \( z_2 \), namely \( a'_2 \Sigma a_2 \), subject to the two constraints, viz. \( a'_2 a_2 = 1 \) and \( a'_2 a_1 = 0 \), we need to introduce two Lagrange multipliers, which may be denoted by \( \lambda \) and \( \Gamma \). Now consider the function

\[
L(a_2) = a'_2 \Sigma a_2 - \lambda (a'_2 a_2 - 1) - \Gamma a'_2 a_1.
\]

At the stationary point(s) we must have

\[
\frac{\partial L}{\partial a_2} = 2(\Sigma - \lambda I) a_2 - \Gamma a_1 = 0.
\]

If we pre-multiply this equation by \( a'_1 \) we get \( 2a'_1 \Sigma a_2 - \Gamma = 0 \), since \( a'_1 a_2 = 0 \). But from equation (5), \( a'_1 \Sigma a_2 \) required to be zero, so that \( \Gamma \) is zero at the stationary point(s). Thus equation (6) becomes \( (\Sigma - \lambda I)a_2 = 0 \).

This time we choose \( \lambda = \lambda_2 \), the second largest eigenvalue of \( \Sigma \), and \( a_2 \) to be the corresponding eigenvector.

Continuing as above, \( a_j \) turns out to be the eigenvector associated with the \( j \)th largest eigenvalue.

There is no difficulty in extending the above argument to the case where some of the eigenvalues of \( \Sigma \) are equal. In that case there is no unique way of choosing the corresponding eigenvectors, but as long as the eigenvectors associated with multiple roots are chosen to be orthogonal, the argument carries through.

### 24.3 VARIANCE OF PRINCIPAL COMPONENTS

Let \( A = [a_1, \cdots, a_p] \), denote the \((p \times p)\) matrix of the eigenvectors of \( \Sigma \) obtained in the last section and the \((p \times 1)\) vector of principal components by \( Z' = (z_1, z_2, \cdots, z_p) \).

Then

\[
Z = A' x,
\]

The \((p \times p)\) covariance matrix of \( Z \) is then
Note that the matrix is diagonal as the components have been chosen to be uncorrelated.

One can also express $\text{Var}(Z)$ in the form $A \Sigma A'$, so that

$$\Lambda = A \Sigma A$$

(7)

gives the important relation between the covariance matrix of original variables $x$ and the corresponding principal components. Note that the Equation (7) can be rewritten as

$$\Sigma = A \Lambda A'$$

(8)

since $A$ is an orthogonal matrix.

We have already noted that the eigenvalues can be interpreted as the respective variances of the different principal components. Now the sum of these variances is given by

$$\sum_{i=1}^{p} \text{Var}(z_i) = \sum_{i=1}^{p} \lambda_i = \text{trace}(\Lambda).$$

But

$$\text{trace}(\Lambda) = \text{trace}(A \Sigma A')$$

$$= \text{trace}(\Sigma A A')$$

$$= \text{trace}(\Sigma)$$

$$= \sum_{i=1}^{p} \text{Var}(x_i)$$

Thus we have the important result that the sums of the variances of the original variables and of their principal components are the same. It is, therefore, convenient to make statements such as the $i^{th}$ principal component accounts for a proportion $\frac{\lambda_i}{\sum_{j=1}^{p} \lambda_j}$ of the total variation in the original data, though it should be emphasized that this is not an analysis of variance in the usual sense of the expression. We will also say that the first $m$ principal components account for a proportion $\frac{\sum_{j=1}^{m} \lambda_j}{\sum_{j=1}^{p} \lambda_j}$ of the total variation.

### 24.4 CONTRIBUTION OF VARIABILITY THROUGH SPECTRAL DECOMPOSITION OF $\Sigma$

Spectral decomposition of $\Sigma$ is given by

$$\Sigma A A' = \lambda_1 a_1 a_1' + \lambda_2 a_2 a_2' + \cdots + \lambda_p a_p a_p'.$$
This clearly shows that the contribution of the first principal component to $\Sigma$ is $\lambda_1 a_1 a_1'$ when that is removed, the residual matrix of variances and covariances is

$$\Sigma - \lambda_1 a_1 a_1' = \sum_{j=2}^{p} \lambda_j a_j a_j'$$

The second principal component contributes $\lambda_2 a_2 a_2'$ and so on.

### 24.5 IDENTIFICATION OF IMPORTANT COMPONENTS

The usual procedure is to look at the first few components which, hopefully, account for a large proportion of the total variance. In order to do this, it is necessary to decide which eigenvalues are ‘large’ and which are ‘small’ so that components corresponding to the later may be disregarded. When analysing a correlation matrix where the sum of the eigenvalues is $p$, many researchers use the rule that eigenvalues less than 1 may be disregarded. This arbitrary policy is a useful rule of thumb but has no theoretical justification. It may be better to look at the pattern of eigenvalues and see if there is a natural breakpoint. The fact that there is no objective way of deciding how many components to retain is a serious drawback of the method.

Alternatively, the number of important principal components, (say $q < p$) can be known, if $\Sigma$ can be approximated by another matrix of order $p$, say $B$, of a smaller rank, say $q(< p)$. The goodness of approximation of the closeness of $B$ to $\Sigma$ can be measured by the norm of $\Sigma - B$ which is defined as

$$\|\Sigma - B\|_2 = \left\{ \sum_{i=1}^{p} \sum_{j=1}^{p} (\sigma_{ij} - b_{ij})^2 \right\}^{1/2}.$$ 

Here $\|\cdot\|_p$ denotes the entry wise p-norm of vectors. Entry wise p-norm of a matrix

$$A_{mon} = ((a_{ij})) \text{ is defined as } \|A\|_p = \left\{ \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^p \right\}^{1/p}.$$ 

Among all possible $p \times p$ matrices $B$ of rank $q$, we desire to find the one for which the norm is minimum. We see that

$$\|\Sigma - B\|^2_2 = \text{trace}(\Sigma - B) (\Sigma - B)'$$

$$= \text{trace}\{ AA' (\Sigma - B) AA' (\Sigma - B)' \}$$

$$= \text{trace}\{(A' \Sigma A - A'B A)(A' \Sigma A - A'BA)'\}$$

$$= \text{trace}(\Lambda - G)(\Lambda - G)'$$

where $G = A'B A$.

Since $A$ is an orthogonal matrix, $G$ is of the same rank as $B$, viz. $q$. If the elements of $G$ are $g_{ij}$, then

$$\|\Sigma - B\|^2_2 = \sum_{i=1}^{p} (\lambda_i - g_{ii})^2 + \sum_{i=1}^{p} \sum_{j=1}^{p} g_{ij}^2$$

To minimize this, we must obviously choose $g_{ij} = 0$ for all $i$, $j$ such that $i \neq j$, i.e., $G$ is a diagonal matrix. But $G$ is of rank $q$. Therefore, $p - q$ elements on the diagonal of $G$ must be zero and the remaining elements must be different from zero. In order to satisfy this and minimize the quantity $\sum_{i=1}^{p} (\lambda_i - g_{ii})^2$, it is obvious that the choice must
be $g_a = \lambda_i (i = 1, 2, \ldots, q)$. Therefore, it follows that the required matrix $B$ is

$$
\lambda_1 a_1a_1' + \lambda_2 a_2a_2' + \cdots + \lambda_q a_qa_q'.
$$

This result shows that, if $\Sigma$ is to be approximated by a matrix of order $p$ and rank $q$, choose the first $q$ terms in the spectral decomposition of $\Sigma$. These correspond to the first $q$ principal components.

### 24.6 PRINCIPAL COMPONENTS FROM THE CORRELATION MATRIX

It is quite common to calculate the principal components of a set of variables after they have been standardized to have unit variance. This means that one is effectively finding the principal components from the correlation matrix $R$ rather than from the covariance matrix $\Sigma$. The mathematical derivation is the same, so that the principal components turn out to be in terms of the eigenvectors of $R$. However, it is important to realize the eigenvalues and eigenvectors of $R$ will generally not be the same as those of $\Sigma$. Choosing to analyze $R$ rather than $\Sigma$ involves a definite but arbitrary decision to make the variables ‘equally important’.

For the correlation matrix, the diagonal terms are all unity. Thus the sum of the diagonal terms (or the sum of the variances of the standardized variables) will be equal to $p$. Thus the sum of the eigenvalues of $R$ will also be equal to $p$, so that the proportion of the total variation accounted for by the $j$th component is simply $\lambda_j / p$.

**Example 1 (Principal components from Covariance Matrix):**

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

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

It may be verified that eigenvalue-eigenvector pairs are

$$
\lambda_1 = 5.83; \quad a_1' = (0.383, -0.924, 0)
$$

$$
\lambda_2 = 2.00; \quad a_2' = (0, 0, 1)
$$

$$
\lambda_3 = 0.17; \quad a_3' = (-0.924, 0.383, 0).
$$

Therefore, the principal components become

$$
z_1 = a_1' x = 0.383 x_1 -0.924 x_2
$$

$$
z_2 = a_2' x = x_3
$$

$$
z_3 = a_3' x = -0.924 x_1 + 0.383 x_2.
$$

The variable $x_3$ is one of the principal components because it is uncorrelated with the other two variables.

$$
\text{Var}(z_1) = \text{Var}(0.383 x_1 -0.924 x_2)
$$

$$
= (0.383)^2 \text{Var}(x_1) +(-0.924)^2 \text{Var}(x_2) +2(0.383)(-0.924)
$$

$$
\text{Cov}(x_1, x_2)
$$

$$
= 0.147(1)+0.854(5) -0.708(-2)
$$

$$
= 5.83 = \lambda_1.
$$

$$
\text{Cov}(z_1, z_2) = \text{Cov}(0.383 x_1 -0.924 x_2, x_3)
$$
\[
\begin{align*}
&= 0.383 \text{ Cov}(x_1, x_3) - 0.924 \text{ Cov}(x_2, x_3) \\
&= 0.383 (0) - 0.924 (0) = 0.
\end{align*}
\]

It is also readily apparent that
\[
\sigma_{11} + \sigma_{22} + \sigma_{33} = 1 + 5 + 2 = \lambda_1 + \lambda_2 + \lambda_3 = 5.83 + 2.00 + 0.17.
\]

The proportion of total variance accounted for by the first principal component is
\[
\frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3} = \frac{5.83}{8} = 0.73.
\]
Continuing like this, the first two principal components account for a proportion \(\frac{5.83+2}{8} = 0.98\) of the population variance. In this case the components \(z_1\) and \(z_2\) could replace the three variables with little loss of information.

\[
B = \lambda_1 a_1 a_1' + \lambda_2 a_2 a_2' .
\]

\[
B = \begin{bmatrix}
0.85519687 & -2.0631787 & 0 \\
-2.0631787 & 4.97751408 & 0 \\
0 & 0 & 2
\end{bmatrix} = \Sigma \text{ (approx)}.
\]

That is, a few principal components will approximate covariance structure of \(x\)'s.

**Example 2 (Principal components from Covariance and Correlation Matrix):**

Consider the covariance matrix

\[
\Sigma = \begin{bmatrix}
1 & 4 \\
4 & 100
\end{bmatrix}
\]

and the derived correlation matrix as

\[
R = \begin{bmatrix}
1 & .4 \\
.4 & 1
\end{bmatrix}.
\]

The eigenvalue-eigenvector pairs from \(\Sigma\) are

\[
\lambda_1 = 100.16, \quad a_1' = [0.040, 0.999] \\
\lambda_2 = 0.84, \quad a_2' = [0.999, -0.040].
\]

Similarly, the eigenvalue-eigenvector pairs from \(R\) are

\[
\lambda_1 = 1 + \rho = 1.4, \quad a_1' = [0.707, 0.707] \\
\lambda_2 = 1 - \rho = 0.6, \quad a_2' = [0.707, -0.707].
\]

The respective principal components using \(\Sigma\) are

\[
z_1 = 0.040x_1 + 0.999x_2 \\
z_2 = 0.999x_1 - 0.040x_2
\]

and using \(R\), these are

\[
z_1 = 0.707[(x_1 - \mu_1)/1] + 0.707[(x_2 - \mu_2)/10] \\
= 0.707(x_1 - \mu_1) + 0.707(x_2 - \mu_2) \\
z_2 = 0.707(x_1 - \mu_1) + 0.707(x_2 - \mu_2).
\]
Because of its large variance, \( x_2 \) completely dominates the first principal component determined from \( \Sigma \). Moreover, this first principal component explains a proportion \( \frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{100.16}{101.00} = 0.992 \) of the total population variance.

When the variables \( x_1 \) and \( x_2 \) are standardized, however, the resulting variables contributed equally to the principal component determined from \( R \).

In this case, the first principal component explains a proportion \( \frac{\lambda_1}{p} = \frac{1.40}{2} = 0.70 \) of the total (standardized) population variance.

Most strikingly we see that the relative importance of the variables, for instance, the first principal component is greatly affected by the standardization. When the principal component obtained from \( R \) are expressed in terms of \( x_1 \) and \( x_2 \), the relative magnitudes of the weights 0.707 and 0.707 are in direct opposition to those of the weights 0.040 and 0.999 attached to these variables in the principal components obtained from \( \Sigma \).

Example 2 demonstrates that the principal components derived from \( \Sigma \) are different from those derived from \( R \). Furthermore, one set of principal components is not a simple function of the other. This suggests that the standardization has its own consequences. Variables should probably be standardized if they are measured on scales with widely differing ranges or if the measurement units are not commensurate.

### 24.7 ESTIMATING THE PRINCIPAL COMPONENTS

The above derivation of the principal components of \( x \) assumes that \( \Sigma \) is known. Generally we deal with sample data and \( \Sigma \) is unknown. In such situations, \( \Sigma \) is replaced by \( S \), the sample covariance matrix. The derivation of the principal components of \( x \) using the sample variances and covariances \( S \) is same as before. Let us denote the eigenvalues of \( S \) in descending order of size by \( \lambda_1, \lambda_2, \ldots, \lambda_p \) and the corresponding eigenvectors by \( a_1, a_2, \ldots, a_p \). Since \( S \) is positive semi-definite symmetric, the eigenvalues are all non-negative and represent the estimated variances of the different components.

If our sample of ‘individuals’ a random sample from a larger population, then \( \{\lambda_i\} \) and \( \{a_i\} \) may be regarded as estimates of the eigenvalues and vectors of \( \Sigma \), giving us estimates of the principal components of \( x \). But no assumptions have been made about the underlying population, and without such assumptions it is impossible to derive the sampling properties of the estimates. If we are prepared to assume that the observations are taken from a multivariate normal distribution, then some sampling theory is available. But this theory is of limited principal value, partly because many of the results are for the asymptotic case (as \( n \to \infty \)), and partly because the normality assumption is often questionable. In any case, the ‘sample’ may be observations for a complete population. Thus the modern tendency is to view Principal Component Analysis as a mathematical technique with no underlying statistical model. The principal components obtained from the sample covariance matrix \( S \) are seen as the principal components and not as estimates of the corresponding quantities obtained from \( \Sigma \). The ‘hats’ over \( \lambda_i \) and \( a_i \) are often omitted. Indeed, it is not even necessary to regard \( x \) and \( z \) as random variables.

Example 3 (Principal Components from sample data):
The data is same as given in Example 2 in Johnson and Wichern, 2002. Perspiration from 20 healthy females was analyzed. Three components, \( X_1 = \) sweat rate, \( X_2 = \) sodium content and \( X_3 = \) potassium content were measured and the results are presented in Table 1.
Table 1: Sweat Data

<table>
<thead>
<tr>
<th>Individual</th>
<th>(X_1) (sweat rate)</th>
<th>(X_2) (sodium content)</th>
<th>(X_3) (potassium content)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.7</td>
<td>48.5</td>
<td>9.3</td>
</tr>
<tr>
<td>2</td>
<td>5.7</td>
<td>65.1</td>
<td>8.0</td>
</tr>
<tr>
<td>3</td>
<td>3.8</td>
<td>47.2</td>
<td>10.9</td>
</tr>
<tr>
<td>4</td>
<td>3.2</td>
<td>53.2</td>
<td>12.0</td>
</tr>
<tr>
<td>5</td>
<td>3.1</td>
<td>55.5</td>
<td>9.7</td>
</tr>
<tr>
<td>6</td>
<td>4.6</td>
<td>36.1</td>
<td>7.9</td>
</tr>
<tr>
<td>7</td>
<td>2.4</td>
<td>24.8</td>
<td>14.0</td>
</tr>
<tr>
<td>8</td>
<td>7.2</td>
<td>33.1</td>
<td>7.6</td>
</tr>
<tr>
<td>9</td>
<td>6.7</td>
<td>47.4</td>
<td>8.5</td>
</tr>
<tr>
<td>10</td>
<td>5.4</td>
<td>54.1</td>
<td>11.3</td>
</tr>
<tr>
<td>11</td>
<td>3.9</td>
<td>36.9</td>
<td>12.7</td>
</tr>
<tr>
<td>12</td>
<td>4.5</td>
<td>58.8</td>
<td>12.3</td>
</tr>
<tr>
<td>13</td>
<td>3.5</td>
<td>27.8</td>
<td>9.8</td>
</tr>
<tr>
<td>14</td>
<td>4.5</td>
<td>40.2</td>
<td>8.4</td>
</tr>
<tr>
<td>15</td>
<td>1.5</td>
<td>13.5</td>
<td>10.1</td>
</tr>
<tr>
<td>16</td>
<td>8.5</td>
<td>56.4</td>
<td>7.1</td>
</tr>
<tr>
<td>17</td>
<td>4.5</td>
<td>71.6</td>
<td>8.2</td>
</tr>
<tr>
<td>18</td>
<td>6.5</td>
<td>52.8</td>
<td>10.9</td>
</tr>
<tr>
<td>19</td>
<td>4.1</td>
<td>44.1</td>
<td>11.2</td>
</tr>
<tr>
<td>20</td>
<td>5.5</td>
<td>40.9</td>
<td>9.4</td>
</tr>
</tbody>
</table>

The variance-covariance matrix of the above data is

\[
\Sigma = \begin{bmatrix} 2.879368 & 10.01 & -1.80905 \\ 10.01 & 199.7884 & -5.64 \\ -1.80905 & -5.64 & 3.627658 \end{bmatrix}
\]

Now find the eigenvalues and eigenvectors of the above matrix. Arrange the eigenvalues in decreasing order. Let the eigenvalues in decreasing order and corresponding eigenvectors are

\[
\begin{align*}
\lambda_1 &= 200.462 & a_1 &= (0.0508, 0.9983, -0.0291) \\
\lambda_2 &= 4.532 & a_2 &= (-0.5737, 0.0530, 0.8173) \\
\lambda_3 &= 1.301 & a_3 &= (0.8175, -0.0249, 0.5754) 
\end{align*}
\]

The principal components for this data are

\[
\begin{align*}
z_1 &= 0.0508x_1 + 0.9983x_2 - 0.0291x_3 \\
z_2 &= -0.5737x_1 + 0.0530x_2 + 0.8173x_3 \\
z_3 &= 0.8175x_1 - 0.0249x_2 + 0.5754x_3
\end{align*}
\]

The variance of principal components will be eigenvalues i.e.

\[
\text{Var}(z_1) = 200.462, \quad \text{Var}(z_2) = 4.532, \quad \text{Var}(z_3) = 1.301.
\]

The total variation explained by principal components is

\[
\lambda_1 + \lambda_2 + \lambda_3 = 200.462 + 4.532 + 1.301 = 206.295.
\]

As such, it can be seen that the total variation explained by principal components is same as that explained by original variables. It could also be proved mathematically as well as empirically that the principal components are uncorrelated.

The proportion of total variation accounted for by the first principal component is
\[
\frac{\lambda_i}{\lambda_1 + \lambda_2 + \lambda_3} = \frac{200.462}{206.295} = 0.9717
\]
of the total variance,

Where as, the first two components account for a proportion

\[
\frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3} = \frac{204.994}{206.295} = 0.9937
\]
of it.

Hence, in further analysis, the first or first two principal components \(z_1\) and \(z_2\) could replace three variables by sacrificing negligible information about the total variation in the system. The scores of principal components can be obtained by substituting the values of \(x_i\)’s in the equations of \(z_i\)’s. For above data, the scores from the first two principal components for first observation i.e. for first individual are

\[
z_1 = 0.0508 \times 3.7 + 0.9983 \times 48.5 - 0.0291 \times 9.3
\]
\[
z_2 = -0.5737 \times 3.7 + 0.0530 \times 48.5 + 0.8173 \times 9.3
\]

Similarly principal component scores for other individuals can be obtained. Thus the whole data with three variables can be converted to a new data set with two principal components.

Now try the following exercise.

---

**E1:** Weather data given below pertains to Raipur district (M.P.) from 1970 to 1986 for kharif crop season from 21st May to 7th October. The weather variables are average minimum temperature \((x_1)\), average relative humidity at 8 hrs \((x_2)\), average relative humidity at 14 hrs \((x_3)\) and total rainfall in cm \((x_4)\).

<table>
<thead>
<tr>
<th>S.No.</th>
<th>(x_1)</th>
<th>(x_2)</th>
<th>(x_3)</th>
<th>(x_4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>86</td>
<td>66</td>
<td>186.49</td>
</tr>
<tr>
<td>2</td>
<td>24.9</td>
<td>84</td>
<td>66</td>
<td>124.34</td>
</tr>
<tr>
<td>3</td>
<td>25.4</td>
<td>77</td>
<td>55</td>
<td>98.79</td>
</tr>
<tr>
<td>4</td>
<td>24.4</td>
<td>82</td>
<td>62</td>
<td>118.88</td>
</tr>
<tr>
<td>5</td>
<td>22.9</td>
<td>79</td>
<td>53</td>
<td>71.88</td>
</tr>
<tr>
<td>6</td>
<td>7.7</td>
<td>86</td>
<td>60</td>
<td>111.96</td>
</tr>
<tr>
<td>7</td>
<td>25.1</td>
<td>82</td>
<td>58</td>
<td>99.74</td>
</tr>
<tr>
<td>8</td>
<td>24.9</td>
<td>83</td>
<td>63</td>
<td>115.2</td>
</tr>
<tr>
<td>9</td>
<td>24.9</td>
<td>82</td>
<td>63</td>
<td>100.6</td>
</tr>
<tr>
<td>10</td>
<td>24.9</td>
<td>78</td>
<td>56</td>
<td>62.38</td>
</tr>
<tr>
<td>11</td>
<td>24.3</td>
<td>85</td>
<td>67</td>
<td>154.4</td>
</tr>
<tr>
<td>12</td>
<td>24.6</td>
<td>79</td>
<td>61</td>
<td>112.71</td>
</tr>
<tr>
<td>13</td>
<td>24.3</td>
<td>81</td>
<td>58</td>
<td>79.63</td>
</tr>
<tr>
<td>14</td>
<td>24.6</td>
<td>81</td>
<td>61</td>
<td>125.59</td>
</tr>
<tr>
<td>15</td>
<td>24.1</td>
<td>85</td>
<td>64</td>
<td>99.87</td>
</tr>
<tr>
<td>16</td>
<td>24.5</td>
<td>84</td>
<td>63</td>
<td>143.56</td>
</tr>
<tr>
<td>17</td>
<td>24</td>
<td>81</td>
<td>61</td>
<td>114.97</td>
</tr>
</tbody>
</table>

**Mean** | **23.56** | **82.06** | **61** | **112.97**

**S.D.** | **4.13** | **2.75** | **3.97** | **30.06**

Find the principal components using variance covariance matrix and correlation matrix.

**E2:** An experiment was conducted to compare the drying methods of microwave assisted convective drying with those of conventional methods such as microwave, convective and freeze drying of banana. The data obtained for
overall drying rate \( (x_1) \), rehydration ratio \( (x_2) \), total sugars \( (x_3) \), total carbohydrates \( (x_4) \) and energy use efficiency \( (x_5) \).
3.622 1.780 52.500 615.885 36.745
3.226 1.686 51.919 615.109 22.841
3.219 1.923 52.669 624.009 17.530
3.622 1.700 53.419 631.008 34.279
3.410 1.700 51.428 612.887 20.997
3.224 1.901 52.699 628.051 14.702
4.830 2.240 52.664 615.008 52.068
4.140 2.299 52.172 624.097 44.630
3.629 2.400 54.000 630.059 39.121
5.270 2.338 53.316 626.841 49.277
4.460 2.451 54.761 640.984 32.958
4.139 2.497 54.399 640.019 25.291
5.278 2.800 54.418 641.216 46.269
5.276 2.586 55.400 650.000 33.677
5.270 2.558 55.280 652.189 26.463
6.446 2.830 54.806 649.057 53.206
5.780 2.598 54.900 646.189 32.542
5.770 2.700 54.607 650.007 24.664

Obtain principal components using both covariance matrix and correlation matrix and interpret the results.

E3) Find the principal components, and the proportion of total population variance explained by each, when the covariance matrix is

\[
\sum = \begin{bmatrix}
\sigma^2 & \sigma^2 \rho & 0 \\
\sigma^2 \rho & \sigma^2 & \sigma^2 \rho \\
0 & \sigma^2 \rho & \sigma^2 \\
\end{bmatrix}, \quad -\frac{1}{\sqrt{2}} < \rho < \frac{1}{\sqrt{2}}.
\]

E4) Consider the census-tract data listed below. Suppose the observations on \(X_5 = \) median value home were recorded in thousands, rather than ten thousands of Rupees; that is, multiply all the numbers listed in the sixth column of the data Table by 10.

<table>
<thead>
<tr>
<th>Tract</th>
<th>Total Population (thousands)</th>
<th>Median School Years</th>
<th>Total employment (thousands)</th>
<th>Health services employment (hundreds)</th>
<th>Median Value home (Rs.10,000s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.935</td>
<td>14.2</td>
<td>2.265</td>
<td>2.27</td>
<td>2.91</td>
</tr>
<tr>
<td>2</td>
<td>1.52</td>
<td>13.1</td>
<td>.597</td>
<td>.75</td>
<td>2.62</td>
</tr>
<tr>
<td>3</td>
<td>32.599</td>
<td>12.7</td>
<td>1.237</td>
<td>1.11</td>
<td>1.72</td>
</tr>
<tr>
<td>4</td>
<td>4.009</td>
<td>15.2</td>
<td>1.649</td>
<td>.81</td>
<td>3.02</td>
</tr>
<tr>
<td>5</td>
<td>4.687</td>
<td>14.7</td>
<td>2.312</td>
<td>2.50</td>
<td>2.22</td>
</tr>
<tr>
<td>6</td>
<td>8.044</td>
<td>15.6</td>
<td>3.641</td>
<td>4.51</td>
<td>2.36</td>
</tr>
<tr>
<td>7</td>
<td>2.766</td>
<td>13.3</td>
<td>1.244</td>
<td>1.03</td>
<td>1.97</td>
</tr>
<tr>
<td>8</td>
<td>6.538</td>
<td>17.0</td>
<td>2.618</td>
<td>2.39</td>
<td>1.85</td>
</tr>
<tr>
<td>9</td>
<td>6.451</td>
<td>12.9</td>
<td>3.147</td>
<td>5.52</td>
<td>2.01</td>
</tr>
<tr>
<td>10</td>
<td>3.314</td>
<td>12.2</td>
<td>1.606</td>
<td>2.18</td>
<td>1.82</td>
</tr>
<tr>
<td>11</td>
<td>3.777</td>
<td>13.0</td>
<td>2.119</td>
<td>2.83</td>
<td>1.80</td>
</tr>
<tr>
<td>12</td>
<td>1.530</td>
<td>13.8</td>
<td>0.789</td>
<td>0.84</td>
<td>4.25</td>
</tr>
<tr>
<td>13</td>
<td>2.768</td>
<td>13.6</td>
<td>1.336</td>
<td>1.75</td>
<td>2.64</td>
</tr>
<tr>
<td>14</td>
<td>6.585</td>
<td>14.9</td>
<td>2.763</td>
<td>1.91</td>
<td>3.17</td>
</tr>
</tbody>
</table>

(a) Construct the sample covariance matrix, \(S\), for the census-tract data when \(X_5\) = median value home is recorded in thousands of Rupees.

(b) Obtain the eigenvalue-eigenvector pairs and take first two sample principal components for the covariance matrix in (a).
(c) Compute the proportion of total variance explained by the first two principal components obtained in (b).

E5) The following data table was published in Time Magazine in January, 1996. It gives the beer, wine and liquor consumption (in liters per year), life expectancy (in years) and heart disease rate (in cases per 100,000 per year) for the 10 countries listed.

<table>
<thead>
<tr>
<th>Country</th>
<th>Liquor</th>
<th>Wine</th>
<th>Beer</th>
<th>Life Expectancy in years</th>
<th>Heart disease rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>2.5</td>
<td>63.5</td>
<td>40.1</td>
<td>78</td>
<td>61.1</td>
</tr>
<tr>
<td>Italy</td>
<td>0.9</td>
<td>58.0</td>
<td>25.1</td>
<td>78</td>
<td>94.1</td>
</tr>
<tr>
<td>Switzerland</td>
<td>1.7</td>
<td>46.0</td>
<td>65.0</td>
<td>78</td>
<td>106.4</td>
</tr>
<tr>
<td>Australia</td>
<td>1.2</td>
<td>15.7</td>
<td>102.1</td>
<td>78</td>
<td>173.0</td>
</tr>
<tr>
<td>Great Britain</td>
<td>1.5</td>
<td>12.2</td>
<td>100.0</td>
<td>77</td>
<td>199.7</td>
</tr>
<tr>
<td>United States</td>
<td>2.0</td>
<td>8.9</td>
<td>87.8</td>
<td>76</td>
<td>176</td>
</tr>
<tr>
<td>Russia</td>
<td>3.8</td>
<td>2.7</td>
<td>17.1</td>
<td>69</td>
<td>373.6</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>1.0</td>
<td>1.7</td>
<td>140.0</td>
<td>73</td>
<td>283.7</td>
</tr>
<tr>
<td>Japan</td>
<td>2.1</td>
<td>1.0</td>
<td>55.0</td>
<td>79</td>
<td>34.7</td>
</tr>
<tr>
<td>Mexico</td>
<td>0.8</td>
<td>0.2</td>
<td>50.4</td>
<td>73</td>
<td>36.4</td>
</tr>
</tbody>
</table>

(a) Construct the sample covariance matrix, \( S \), for the census-tract data when \( X = \) median value home is recorded in thousands of dollars.

(b) Obtain the eigenvalue-eigenvector pairs and the first two sample principal components for the covariance matrix in (a).

(c) Compute the proportion of total variance explained by the first two principal components obtained in (b).

(d) Obtain the principal component score for the first two principal components.

(e) Standardize the data with mean 0 and variance 1 and repeat the steps (a) to (d) on the transformed data.

Now, let us discuss scaling convention in principal component analysis in the following section.

### 24.8 SCALING CONVENTION IN PCA

It is important to realize that the principal components of a set of variables depend critically upon the scales used to measure the variables. For example, suppose that for each of \( n \) individuals we measure their weight in pounds, their height in feet and their age in years to give a vector \( x \). Denote the resulting sample covariance matrix by \( S_x \), its eigenvalues by \( \{ \lambda_i \} \) and its eigenvectors by \( \{ a_i \} \).

If we transform to new co-ordinates so that \( u' = \) (weight in kilograms, height in metres, age in months). Then \( u = Kx \), where \( K \) is the diagonal matrix

\[
K = \begin{bmatrix}
1/2.2 & 0 & 0 \\
0 & 1/3.28 & 0 \\
0 & 0 & 12
\end{bmatrix}
\]

As there are, for example, 2.2 pounds in one kilogram, 3.28 feet in one meter and 12 months in a year. The covariance matrix of the new variables will be given by \( S_u = KS_x K' \).

since \( K' = K \). The eigenvalues and eigenvectors of \( S_u \) will generally be different from those of \( S_x \) and will be denoted by \( \{ \lambda_i' \} \) and \( \{ a_i' \} \). But will they give the same
principal components when transformed back to the original variables? The answer
generally is no. The principal components will change unless:

a) All the diagonal element of $K$ are the same, so that $K = cI$, where $c$ is a scalar
   constant. This would mean that all the variables are scaled in the same way.

b) Variables corresponding to unequal diagonal elements of $K$ are uncorrelated. In
   particular, if all the elements of $K$ are unequal and all variables are uncorrelated,
   then $S_x$ must be a diagonal matrix, in that case there is no point in carrying out
   PCA.

The practical outcome of the above result is that principal component are generally
changed by scaling and that they are, therefore, not a unique characteristic of the data.
If, for example, one variable has a much larger variance than all the other variables,
then this variable will dominate the first principal component of the covariance matrix
whatever be the correlation structure, whereas if the variables are all scaled to have
unit variance, then the first principal component will be quite different in kind.
Because of this, it is generally thought to be of little importance in carrying out a PCA
unless the variables have ‘roughly similar’ variance, as may be the case, for example, if
all the variables are percentage, or are measured in the same co-ordinates.

The conventional way of getting rid of the scaling problem is to analyze the
correlation matrix rather than the covariance matrix, so that each multivariate
observation, $x$, is transformed by

$$ u = Kx $$

where

$$ K = \begin{bmatrix}
\frac{1}{s_1} & \cdots & 0 & \cdots & 0 \\
0 & \cdots & \frac{1}{s_2} & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & 0 & \cdots & \frac{1}{s_p}
\end{bmatrix}. $$

and $s_i$ is the sample standard deviation for the $i$th variable. This ensures that all
variables are scaled to have unit variance and so in some sense have equal importance.
This scaling procedure is still arbitrary to some extent, is data-dependent and avoids
rather than solves the scaling problem. If the variables are not thought to be of equal
importance, then the analysis of the correlation matrix is not recommended.
Analysing the correlation matrix also makes it more difficult to compare the results of
Principal Component Analysis on two or more different samples.

Note that the principal components of the correlation matrix will not be orthogonal if
the variables are transformed back to their original co-ordinates. This is because a
linear transformation of two lines at right angles in Euclidean space will not generally
give two new lines at right angles, which is another way of explaining why the scaling
problem arise in the first place.

**Example 4:** Using the data given in E 1), discuss the scaling convention in PCA.

**Solution:** For the data given in E1), $x_1$ is measured in $^0$C, $x_2$ and $x_3$ in terms of
percentage and $x_4$ in cm. Then the sample covariance matrix $S_x = \Sigma$ and eigenvalue-
eigenvector pairs $\{\lambda_i, a_i\}$, $i = 1, 2, 3, 4$, have been obtained earlier.

If we transform $x$ to new coordinates $u$ by means of the transformation

$$ u = Kx $$

with $u' = (x_1$ in $F-32, x_2$ in $%, x_3$ in $%, x_4$ in mm), then
The covariance matrix of the new variables is 
\[ S_u = K S_x K \text{ as } K' = K. \]

We get the eigenvalue-eigenvector pairs of \( S_u \) as
\[ \lambda_1^* = 90400.807 \quad a_1^* = (0.001, 0.006, 0.010, 1) \]
\[ \lambda_2^* = 56.300 \quad a_2^* = (0.988, 0.150, 0.027, 0) \]
\[ \lambda_3^* = 8.061 \quad a_3^* = (0.055, 0.521, 0.852, -0.012) \]
\[ \lambda_4^* = 1.106 \quad a_4^* = (0.142, 0.840, -0.678, 0). \]

These are not the same as \( (\lambda_i, a'_i) \), \( i=1,2,3,4 \).

Seeing the matrix \( K \), we find that

1. At least one of the diagonal elements of \( K \) is different from other (in our case 3 out of 4 are different).
2. Variables corresponding, to unequal diagonal elements of \( K \) are correlated.

Hence, the eigenvalue-eigenvector pairs of \( S_u \) give different principal components when transformed back to original \( x \)'s Principal Components therefore do not depict the unique characteristic of the original data due to scaling problem.

The conventional way of getting round the scaling problem is to analyze the correlation matrix rather than the covariance matrix, so that \( x \) is transformed to \( u = K_1 x \), where
\[
K_1 = \begin{bmatrix}
1/4.13 & 0 & 0 & 0 \\
0 & 1/2.75 & 0 & 0 \\
0 & 0 & 1/3.97 & 0 \\
0 & 0 & 0 & 1/30.06
\end{bmatrix}.
\]

The diagonals being, inverses of the sample Standard deviation of the \( i^{th} \) variable and then carrying out Principal Component Analysis.

### 24.9 GEOMETRICAL INTERPRETATION OF PRINCIPAL COMPONENTS

The transformation \( Z = A' x \) is an orthogonal transformation and it transforms an ellipsoid \( x \Sigma^{-1} x = \text{Constant} \) into the ellipsoid \( Z' \Lambda^{-1} = \text{Constant} \), i.e.
\[
\sum_{i=1}^{p} \frac{Z_i^2}{\lambda_i} = \text{Constant}.
\]

This is an ellipsoid referred to its principal axes and \( \sqrt{\lambda_i} \) \( (i = 1, 2, ..., p) \) are the lengths of principal semi-axes. For \( p = 2 \), i.e., when \( x \)'s can be replaced by two
principal components, this is an ellipse in a coordinate system with axes \( z_1 \) and \( z_2 \) lying in the directions of \( a_1 \) and \( a_2 \) respectively as depicted in the figure below.

![Geometrical Interpretation of Principal Components](image)

If the eigenvalues of \( \Sigma \) are equal, the new ellipse is a circle and there is unique transformation. In fact, this only happens if the original variables are uncorrelated and have equal variance.

## 24.10 APPLICATIONS OF PRINCIPAL COMPONENTS

i) An important benefit of Principal Component Analysis is that it provides a quick way of assessing the effective dimensionality of a set of data. If the first few components account for most of the variation in the original data, it is often a good idea to use these first few components scores in subsequent analysis.

ii) Plotting of data becomes difficult with more than three variables. But if the first two components account for a large proportion of the total variation, then it will often be useful to plot the values of the first two components scores for each individual. In other words, Principal Component Analysis may enable us to plot the data in two dimensions. This procedure is known as Biplot. A biplot is a scatter plot that graphically displays both the row factors and the column factors of a two-way data. The concept of biplots were first developed by Gabriel (1971). Since then, biplot has been used in data visualization and pattern analysis in various research fields, from psychology to economics to agronomy. Currently there are over 50000 web pages containing the keyword ‘biplot’. This technique has extensively been used in the analysis of multi-environment trials. For generating a biplot, we take the matrix representing the effects of two factors. This matrix is then subjected to singular value decomposition.

iii) In particular, one can then look for outliers or for groups or ‘clusters’ of individuals. This is an important use of Principal Component Analysis and often groupings of variables which would not be found by other means.

iv) Reduction is dimensionality can help in discriminant analysis when a large number of correlated variables (p) are observed on n observation (n<p). As the number of observations is less than the number of variables, there will be unpleasant singularity problem unless the dimensionality is drastically reduced. Hopefully, Principal Component Analysis does that.

v) Multiple regression can be dangerous if the so-called independent variables are in fact highly correlated. Various techniques, such as ridge regression, have been developed to overcome this problem. An alternative approach, which may
well be more fruitful, is to regress, not on the original variables, but on those components which are most highly correlated with the dependent variable. This is called principal components regression.

It is quite likely that first few principal components account for most of the variability in the original data. If so, these few principal components can then replace the initial \( p \) variables in subsequent analysis, thus reducing the effective dimensionality of the problem. An analysis of principal components often reveals relationships that were not previously suspected and thereby allows interpretation that would not ordinarily result.

However, Principal Components Analysis is more of a mean to an end rather than end in itself because this frequently serves as intermediate steps in much larger investigations by reducing the dimensionality of the problem and providing easier interpretation. It is a mathematical technique, which does not require user to specify the statistical model or assumption about distribution of original variates. It may also be mentioned that principal components are artificial variables and often it is not possible to assign physical meaning to them. Further, since Principal Components Analysis transforms original set of variables to new set of uncorrelated variables. It is worth stressing that if the original variables are uncorrelated, then there is no point in carrying out the Principal Components Analysis. It is important to note here that the principal components depend on the scale of measurement. Conventional way of getting rid of this problem is to use the standardized variables with unit variances.

Now, let us summarize the unit.

### 24.11 SUMMARY

In this Unit, we covered the following points.

1. Principal Component Analysis is a dimensional reduction technique in which we derive a small number of linear combinations (principal components) of a set of variables that retain as much information in the original variables as possible.

2. Principal Components can be derived from covariance matrix or correlation matrix.

3. For obtaining principal components, one must know the eigenvalues of the sample covariance/correlation matrix.

4. Principal component scores are obtained by multiplying the row vector of variables with the column eigenvectors.

5. Principal components of a set of variables depend critically upon the scales used to measure the variables.

### 24.12 SOLUTIONS/ANSWERS

E1) The problem is to find first few principal components, which may account for most of the variation in the original data.

1. Identification of important principal components through covariance matrix of the original variables.

The variables \( \mathbf{X}' = (x_1, x_2, x_3, x_4) \) have the covariance matrix as
The eigenvalue-eigenvector pairs of $\Sigma$ are

\[
\begin{align*}
\lambda_1 &= 916.902 & \mathbf{a}_1' &= (0.006, 0.061, 0.103, 0.993) \\
\lambda_2 &= 18.375 & \mathbf{a}_2' &= (0.955, -0.296, 0.011, 0.012) \\
\lambda_3 &= 7.87 & \mathbf{a}_3' &= (0.141, 0.485, 0.855, -0.119) \\
\lambda_4 &= 1.056 & \mathbf{a}_4' &= (0.260, 0.820, -0.509, .001).
\end{align*}
\]

The principal component are $\mathbf{Z}_i = \mathbf{a}_i' \mathbf{x}, i = 1, 2, 3, 4$.

The proportion of total variance accounted for by the first principal component is

\[
\frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4} = \frac{916.902}{944.203} = 0.97.
\]

Continuing, the first two components account for proportion

\[
\frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4} = \frac{935.277}{944.203} = 0.99
\]

of the total variance. Hence the components $\mathbf{z}_1 = \mathbf{a}_1' \mathbf{x}$ and $\mathbf{z}_2 = \mathbf{a}_2' \mathbf{x}$ could replace the four variables with negligible loss of information

\[
\mathbf{B} = \lambda_1 \mathbf{a}_1 \mathbf{a}_1' + \lambda_2 \mathbf{a}_2 \mathbf{a}_2'.
\]

\[
\begin{bmatrix}
16.79 & -4.86 & 0.76 & 5.67 \\
5.02 & 5.70 & 55.47 \\
9.73 & 93.78 \\
904.11
\end{bmatrix}
= \Sigma \text{(approximately)}
\]

$\Rightarrow$ a few Principal Components will approximate the covariance structure of $\mathbf{x}, \mathbf{s}$

II. Principal components from the correlation matrix

The correlation matrix for $\mathbf{x}' = (x_1, x_2, x_3, x_4)$ is

\[
\mathbf{R} = \begin{bmatrix}
1 & -0.36 & 0.09 & 0.04 \\
1 & 0.78 & 0.66 & \cdot \\
1 & 0.78 & \cdot & \cdot \\
1 & \cdot & \cdot & \cdot
\end{bmatrix}.
\]

The eigenvalue-eigenvector pairs of $\mathbf{R}$ are

\[
\begin{align*}
\lambda_1 &= 2.493 & \mathbf{a}_1' &= (-0.088, 0.577, 0.588, 0.560) \\
\lambda_2 &= 1.137 & \mathbf{a}_2' &= (0.919, -0.269, 0.214, 0.195) \\
\lambda_3 &= 0.270 & \mathbf{a}_3' &= (0.200, 0.412, 0.388, -0.800).
\end{align*}
\]
\[ \lambda_4 = 0.100 \quad \mathbf{a}'_4 = (0.329, \quad 0.652, \quad -0.677, \quad 0.090). \]

The proportion of total variance accounted for by the first principal component is \( \lambda_1 / p = 0.62 \) and by the first two components is \( (\lambda_1 + \lambda_2) / p = 0.91 \).

Hence, the component \( \mathbf{z}_1 \) and \( \mathbf{z}_2 \) could replace the four \( \mathbf{x}_i \)'s with little loss of information. Here we get
\[
\mathbf{B} = \lambda_1 \mathbf{a}_1 \mathbf{a}'_1 + \lambda_2 \mathbf{a}_2 \mathbf{a}'_2 \quad \text{as}
\[
= \begin{bmatrix}
0.980 & -0.410 & 0.095 & 0.081 \\
0.912 & 0.780 & 0.750 \\
0.914 & 0.868 \\
0.825
\end{bmatrix}
= \mathbf{R} \quad \text{(approximately)}.
\]

Using \( \Sigma \), the important Principal Components are
\[
\mathbf{z}_1 = 0.006 \mathbf{x}_1 + 0.061 \mathbf{x}_2 + 0.103 \mathbf{x}_3 + 0.993 \mathbf{x}_4
\]
\[
\mathbf{z}_2 = 0.955 \mathbf{x}_1 + 0.296 \mathbf{x}_2 + 0.011 \mathbf{x}_3 + 0.012 \mathbf{x}_4.
\]

Using \( \mathbf{R} \), the important Principal Components are
\[
\mathbf{z}_1 = -0.088 \mathbf{x}_1 + 0.577 \mathbf{x}_2 + 0.588 \mathbf{x}_3 + 0.563 \mathbf{x}_4
\]
\[
\mathbf{z}_2 = 0.919 \mathbf{x}_1 + 0.269 \mathbf{x}_2 + 0.214 \mathbf{x}_3 + 0.195 \mathbf{x}_4.
\]

Because of its large variance of 903.87, \( \mathbf{x}_4 \) completely dominates the first Principal Component determined from \( \Sigma \). Moreover, this Principal Component explains a proportion 0.97 of the total variance.

When the variables are standardized, however, the resulting variables contribute equally to the Principal Components determined from \( \mathbf{R} \). In this case, the first principal component explains only a proportion 0.62 of the total (standardized) variance.

**E2) Compute Variance-Covariance matrix of the above data**

<table>
<thead>
<tr>
<th>( \text{odr} )</th>
<th>( \text{rr} )</th>
<th>( \text{ts} )</th>
<th>( \text{tc} )</th>
<th>( \text{ee} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{odr} )</td>
<td>2.61370216</td>
<td>0.46963160</td>
<td>3.22941054</td>
<td>40.17446370</td>
</tr>
<tr>
<td>( \text{rr} )</td>
<td>0.46963160</td>
<td>0.13532200</td>
<td>0.61445540</td>
<td>7.20586410</td>
</tr>
<tr>
<td>( \text{ts} )</td>
<td>3.22941054</td>
<td>0.61445540</td>
<td>4.94440810</td>
<td>61.06601780</td>
</tr>
<tr>
<td>( \text{tc} )</td>
<td>40.17446370</td>
<td>7.20586410</td>
<td>61.06601780</td>
<td>770.90290160</td>
</tr>
<tr>
<td>( \text{ee} )</td>
<td>16.00298700</td>
<td>2.89859030</td>
<td>16.53574530</td>
<td>190.35753710</td>
</tr>
</tbody>
</table>

Obtain eigenvalue and eigenvectors of the matrix obtained and arrange the eigenvalues and eigenvectors in descending order of eigenvalues.
Therefore, the principal components are

\[ \text{pc1} = 0.051913 \times \text{odr} + 0.009324 \times \text{rr} + 0.076287 \times \text{ts} + 0.956805 \times \text{tc} + 0.275537 \times \text{ee}; \]

\[ \text{pc2} = 0.039006 \times \text{odr} + 0.007244 \times \text{rr} - 0.009233 \times \text{ts} - 0.277839 \times \text{tc} + 0.959765 \times \text{ee}; \]

\[ \text{pc3} = 0.964708 \times \text{odr} + 0.184747 \times \text{rr} - 0.179117 \times \text{ts} - 0.025543 \times \text{tc} - 0.049719 \times \text{ee}; \]

\[ \text{pc4} = 0.048212 \times \text{odr} + 0.543387 \times \text{rr} + 0.835033 \times \text{ts} - 0.069282 \times \text{tc} - 0.018084 \times \text{ee}; \]

\[ \text{pc5} = -0.250595 \times \text{odr} + 0.818816 \times \text{rr} - 0.514521 \times \text{ts} + 0.043302 \times \text{tc} + 0.011590 \times \text{ee}; \]

Obtain Total variance explained by all the 5 principal components

Total variance = 832.839971+109.535336+0.215563+0.106023+0.027591

=942.7245

Obtain Proportion of Total variance explained by each principal component and cumulative variance explained

\[
\begin{array}{|c|c|c|c|c|}
\hline
\text{Eigenvalue} & \text{Proportion Explained by Principal components} & \text{Cumulative proportion of variance explained} \\
\hline
1 & 832.839971 & 0.8834 & 0.8834 \\
2 & 109.535336 & 0.1162 & 0.9996 \\
3 & 0.215563 & 0.0002 & 0.9999 \\
4 & 0.106023 & 0.0001 & 1.0000 \\
5 & 0.027591 & 0.0000 & 1.0000 \\
\hline
\end{array}
\]

One can easily see that first two principal components explain 99.96% of the total variance.
E3) Obtain eigenvalues of $\Sigma$. These are $\sigma^2(1 + \sqrt{2}\rho)$, $\sigma^2(1 - \sqrt{2}\rho)$ and 1. Now proportion of total population variance explained by each principal component can be obtained easily.

E4) (a) Multiply median value home by 10 to get median value home in thousand dollars. Now obtain variance-covariance matrix for Total Population (totalpop), median school years (schoolyear), total employment (totalemp), health services employment (healthemp) and median value home in thousand dollars (valhometh)

<table>
<thead>
<tr>
<th>Covariance Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>totalpop</td>
</tr>
<tr>
<td>totalpop</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>schoolyear</td>
</tr>
<tr>
<td>totalemp</td>
</tr>
<tr>
<td>healthemp</td>
</tr>
<tr>
<td>valhometh</td>
</tr>
</tbody>
</table>

(b) Obtain eigenvalues and eigenvectors of the variance covariance matrix obtained in (a)

<table>
<thead>
<tr>
<th>Eigenvalues</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Eigenvectors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prin1</td>
</tr>
<tr>
<td>totalpop</td>
</tr>
<tr>
<td>schoolyear</td>
</tr>
<tr>
<td>totalemp</td>
</tr>
<tr>
<td>healthemp</td>
</tr>
<tr>
<td>valhometh</td>
</tr>
</tbody>
</table>

(c) Compute the proportion of variance explained by first two components as:

a. Variance explained by first principal component = ratio of first eigenvalue to that of sum of all eigenvalues; 0.6557

b. Variance explained by first two principal components = ratio of sum of first two eigenvalues to that of sum of all eigenvalues; 0.9646

E5) (a) Now obtain variance-covariance matrix for liquor, wine, beer, life expectancy in years (LifeExpyears), heart disease rate (Heartdisrate)
(b) Obtain eigenvalues and eigenvectors of the variance covariance matrix obtained in (a)

<table>
<thead>
<tr>
<th>Eigenvalue</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Liquor</th>
<th>Wine</th>
<th>Beer</th>
<th>LifeExpyears</th>
<th>Heartdisrate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liquor</td>
<td>0.83389</td>
<td>-1.00167</td>
<td>-16.92111</td>
<td>-1.09444</td>
</tr>
<tr>
<td>Wine</td>
<td>-1.00167</td>
<td>621.35656</td>
<td>-1495.5071</td>
<td>-1090.38144</td>
</tr>
<tr>
<td>Beer</td>
<td>-16.92111</td>
<td>-375.77600</td>
<td>11.88444</td>
<td>1070.23089</td>
</tr>
<tr>
<td>LifeExpyears</td>
<td>-1.09444</td>
<td>41.99889</td>
<td>10.32222</td>
<td>-249.84778</td>
</tr>
<tr>
<td>Heartdisrate</td>
<td>43.34611</td>
<td>-1090.38144</td>
<td>12280.66678</td>
<td></td>
</tr>
</tbody>
</table>

(c) Compute the proportion of variance explained by first two components as:
c. Variance explained by first principal component = ratio of first eigenvalue to that of sum of all eigenvalues; 0.8673
d. Variance explained by first two principal components = ratio of sum of first two eigenvalues to that of sum of all eigenvalues; 0.9691

(d) Obtain the mean and arithmetic mean of each variable. Take the deviation of each observation from mean of the corresponding variable and divide the deviation by standard deviation of corresponding variable and repeat steps given above.

24.13 PRACTICAL ASSIGNMENTS

Session 9
Write a programme in ‘C’ language to find the principal components from a covariance matrix and test it for the data given in Example 1.

Session 10

Write a programme in ‘C’ language to find the principal components from a correlation matrix and test it for the data given in Example 2.