Chapter

2

MATRIX ALGEBRA AND RANDOM VECTORS

2.1 Introduction

We saw in Chapter 1 that multivariate data can be conveniently displayed as an array of numbers. In general, a rectangular array of numbers with, for instance, n rows and p columns is called a *matrix* of dimension $n \times p$. The study of multivariate methods is greatly facilitated by the use of matrix algebra.

The matrix algebra results presented in this chapter will enable us to concisely state statistical models. Moreover, the formal relations expressed in matrix terms are easily programmed on computers to allow the routine calculation of important statistical quantities.

We begin by introducing some very basic concepts that are essential to both our geometrical interpretations and algebraic explanations of subsequent statistical techniques. If you have not been previously exposed to the rudiments of matrix algebra, you may prefer to follow the brief refresher in the next section by the more detailed review provided in Supplement 2A.

2.2 Some Basics of Matrix and Vector Algebra

Vectors

An array $\mathbf{x}$ of n real numbers $x_1, x_2, \ldots, x_n$ is called a *vector*, and it is written as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{or} \quad \mathbf{x}' = [x_1, x_2, \ldots, x_n]$$

where the prime denotes the operation of *transposing* a column to a row.

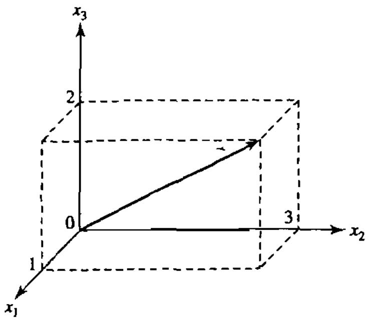

Figure 2.1 The vector $x' = [1, 3, 2]$.

A vector $\mathbf{x}$ can be represented geometrically as a directed line in n dimensions with component x_1 along the first axis, x_2 along the second axis, ..., and x_n along the nth axis. This is illustrated in Figure 2.1 for $n = 3$.

A vector can be *expanded* or *contracted* by multiplying it by a constant c. In particular, we define the vector $c\mathbf{x}$ as

$$c\mathbf{x} = \begin{bmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_n \end{bmatrix}$$

That is, $c\mathbf{x}$ is the vector obtained by multiplying each element of $\mathbf{x}$ by c. [See Figure 2.2(a).]

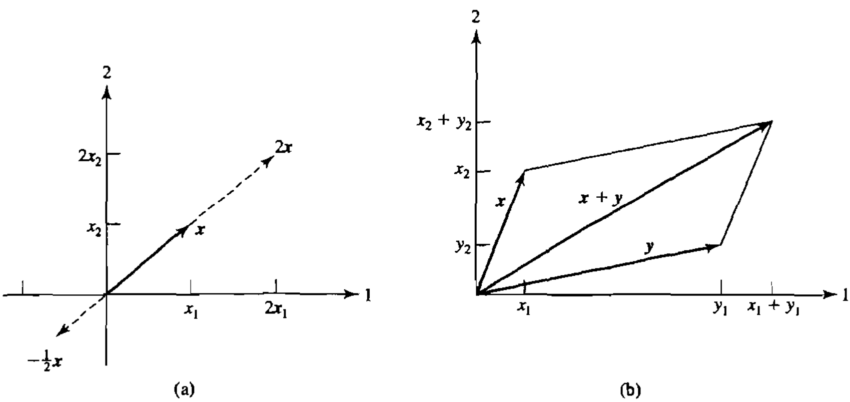

Figure 2.2 Scalar multiplication and vector addition.

Two vectors may be added. *Addition* of $\mathbf{x}$ and $\mathbf{y}$ is defined as

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

so that $\mathbf{x} + \mathbf{y}$ is the vector with ith element $x_i + y_i$.

The sum of two vectors emanating from the origin is the diagonal of the parallelogram formed with the two original vectors as adjacent sides. This geometrical interpretation is illustrated in Figure 2.2(b).

A vector has both direction and length. In $n = 2$ dimensions, we consider the vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

The length of $\mathbf{x}$, written $L_{\mathbf{x}}$, is defined to be

$$L_{\mathbf{x}} = \sqrt{x_1^2 + x_2^2}$$

Geometrically, the length of a vector in two dimensions can be viewed as the hypotenuse of a right triangle. This is demonstrated schematically in Figure 2.3.

The *length* of a vector $\mathbf{x}' = [x_1, x_2, \ldots, x_n]$, with n components, is defined by

$$L_{\mathbf{x}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \tag{2-1}$$

Multiplication of a vector $\mathbf{x}$ by a scalar c changes the length. From Equation (2-1),

$$\begin{aligned} L_{c\mathbf{x}} &= \sqrt{c^2x_1^2 + c^2x_2^2 + \cdots + c^2x_n^2} \\ &= |c|\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = |c|L_{\mathbf{x}} \end{aligned}$$

Multiplication by c does not change the direction of the vector $\mathbf{x}$ if $c > 0$. However, a negative value of c creates a vector with a direction opposite that of $\mathbf{x}$. From

$$L_{c\mathbf{x}} = |c|L_{\mathbf{x}} \tag{2-2}$$

it is clear that $\mathbf{x}$ is expanded if $|c| > 1$ and contracted if $0 < |c| < 1$. [Recall Figure 2.2(a).] Choosing $c = L_{\mathbf{x}}^{-1}$, we obtain the *unit vector* $L_{\mathbf{x}}^{-1}\mathbf{x}$, which has length 1 and lies in the direction of $\mathbf{x}$.

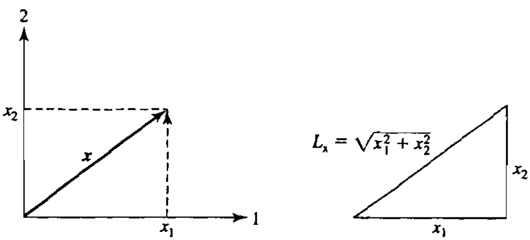

Figure 2.3
Length of $\mathbf{x} = \sqrt{x_1^2 + x_2^2}$.

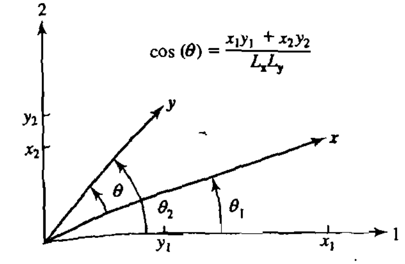

Figure 2.4 The angle θ between $\mathbf{x}' = [x_1, x_2]$ and $\mathbf{y}' = [y_1, y_2]$.

A second geometrical concept is *angle*. Consider two vectors in a plane and the angle θ between them, as in Figure 2.4. From the figure, θ can be represented as the difference between the angles θ_1 and θ_2 formed by the two vectors and the first coordinate axis. Since, by definition,

$$\cos(\theta_1) = \frac{x_1}{L_\mathbf{x}} \qquad \cos(\theta_2) = \frac{y_1}{L_\mathbf{y}}$$

$$\sin(\theta_1) = \frac{x_2}{L_\mathbf{x}} \qquad \sin(\theta_2) = \frac{y_2}{L_\mathbf{y}}$$

and

$$\cos(\theta) = \cos(\theta_2 - \theta_1) = \cos(\theta_2)\cos(\theta_1) + \sin(\theta_2)\sin(\theta_1)$$

the angle θ between the two vectors $\mathbf{x}' = [x_1, x_2]$ and $\mathbf{y}' = [y_1, y_2]$ is specified by

$$\cos(\theta) = \cos(\theta_2 - \theta_1) = \left(\frac{y_1}{L_\mathbf{y}}\right)\left(\frac{x_1}{L_\mathbf{x}}\right) + \left(\frac{y_2}{L_\mathbf{y}}\right)\left(\frac{x_2}{L_\mathbf{x}}\right) = \frac{x_1y_1 + x_2y_2}{L_\mathbf{x}L_\mathbf{y}} \tag{2-3}$$

We find it convenient to introduce the *inner product* of two vectors. For $n = 2$ dimensions, the inner product of $\mathbf{x}$ and $\mathbf{y}$ is

$$\mathbf{x}'\mathbf{y} = x_1y_1 + x_2y_2$$

With this definition and Equation (2-3),

$$L_\mathbf{x} = \sqrt{\mathbf{x}'\mathbf{x}} \qquad \cos(\theta) = \frac{\mathbf{x}'\mathbf{y}}{L_\mathbf{x}L_\mathbf{y}} = \frac{\mathbf{x}'\mathbf{y}}{\sqrt{\mathbf{x}'\mathbf{x}}\sqrt{\mathbf{y}'\mathbf{y}}}$$

Since $\cos(90°) = \cos(270°) = 0$ and $\cos(\theta) = 0$ only if $\mathbf{x}'\mathbf{y} = 0$, $\mathbf{x}$ and $\mathbf{y}$ are perpendicular when $\mathbf{x}'\mathbf{y} = 0$.

For an arbitrary number of dimensions n, we define the inner product of $\mathbf{x}$ and $\mathbf{y}$ as

$$\mathbf{x}'\mathbf{y} = x_1y_1 + x_2y_2 + \cdots + x_ny_n \tag{2-4}$$

The inner product is denoted by either $\mathbf{x}'\mathbf{y}$ or $\mathbf{y}'\mathbf{x}$.

Using the inner product, we have the natural extension of length and angle to vectors of n components:

$$L_{\mathbf{x}} = \text{length of } \mathbf{x} = \sqrt{\mathbf{x}'\mathbf{x}} \tag{2-5}$$

$$\cos(\theta) = \frac{\mathbf{x}'\mathbf{y}}{L_{\mathbf{x}}L_{\mathbf{y}}} = \frac{\mathbf{x}'\mathbf{y}}{\sqrt{\mathbf{x}'\mathbf{x}}\sqrt{\mathbf{y}'\mathbf{y}}} \tag{2-6}$$

Since, again, $\cos(\theta) = 0$ only if $\mathbf{x}'\mathbf{y} = 0$, we say that $\mathbf{x}$ and $\mathbf{y}$ are *perpendicular* when $\mathbf{x}'\mathbf{y} = 0$.

Example 2.1 (Calculating lengths of vectors and the angle between them) Given the vectors $\mathbf{x}' = [1, 3, 2]$ and $\mathbf{y}' = [-2, 1, -1]$, find $3\mathbf{x}$ and $\mathbf{x} + \mathbf{y}$. Next, determine the length of $\mathbf{x}$, the length of $\mathbf{y}$, and the angle between $\mathbf{x}$ and $\mathbf{y}$. Also, check that the length of $3\mathbf{x}$ is three times the length of $\mathbf{x}$.

First,

$$3\mathbf{x} = 3\begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \\ 6 \end{bmatrix}$$

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} + \begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1-2 \\ 3+1 \\ 2-1 \end{bmatrix} = \begin{bmatrix} -1 \\ 4 \\ 1 \end{bmatrix}$$

Next, $\mathbf{x}'\mathbf{x} = 1^2 + 3^2 + 2^2 = 14$, $\mathbf{y}'\mathbf{y} = (-2)^2 + 1^2 + (-1)^2 = 6$, and $\mathbf{x}'\mathbf{y} = 1(-2) + 3(1) + 2(-1) = -1$. Therefore,

$$L_{\mathbf{x}} = \sqrt{\mathbf{x}'\mathbf{x}} = \sqrt{14} = 3.742 \qquad L_{\mathbf{y}} = \sqrt{\mathbf{y}'\mathbf{y}} = \sqrt{6} = 2.449$$

and

$$\cos(\theta) = \frac{\mathbf{x}'\mathbf{y}}{L_{\mathbf{x}}L_{\mathbf{y}}} = \frac{-1}{3.742 \times 2.449} = -.109$$

so $\theta = 96.3°$. Finally,

$$L_{3\mathbf{x}} = \sqrt{3^2 + 9^2 + 6^2} = \sqrt{126} \quad \text{and} \quad 3L_{\mathbf{x}} = 3\sqrt{14} = \sqrt{126}$$

showing $L_{3\mathbf{x}} = 3L_{\mathbf{x}}$. ■

A pair of vectors $\mathbf{x}$ and $\mathbf{y}$ of the same dimension is said to be *linearly dependent* if there exist constants c_1 and c_2, both not zero, such that

$$c_1\mathbf{x} + c_2\mathbf{y} = \mathbf{0}$$

A set of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ is said to be *linearly dependent* if there exist constants $c_1, c_2, \ldots, c_k$, not all zero, such that

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_k\mathbf{x}_k = \mathbf{0} \tag{2-7}$$

Linear dependence implies that at least one vector in the set can be written as a linear combination of the other vectors. Vectors of the same dimension that are not linearly dependent are said to be *linearly independent*.

Example 2.2 (Identifying linearly independent vectors) Consider the set of vectors

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \quad \mathbf{x}_3 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

Setting

$$c_1 \mathbf{x}_1 + c_2 \mathbf{x}_2 + c_3 \mathbf{x}_3 = \mathbf{0}$$

implies that

$$\begin{aligned} c_1 + c_2 + c_3 &= 0 \\ 2c_1 - 2c_3 &= 0 \\ c_1 - c_2 + c_3 &= 0 \end{aligned}$$

with the unique solution $c_1 = c_2 = c_3 = 0$. As we cannot find three constants c_1, c_2, and c_3, *not all zero*, such that $c_1 \mathbf{x}_1 + c_2 \mathbf{x}_2 + c_3 \mathbf{x}_3 = \mathbf{0}$, the vectors $\mathbf{x}_1, \mathbf{x}_2$, and $\mathbf{x}_3$ are *linearly independent*. ■

The *projection* (or shadow) of a vector $\mathbf{x}$ on a vector $\mathbf{y}$ is

$$\text{Projection of } \mathbf{x} \text{ on } \mathbf{y} = \frac{(\mathbf{x}'\mathbf{y})}{\mathbf{y}'\mathbf{y}}\mathbf{y} = \frac{(\mathbf{x}'\mathbf{y})}{L_{\mathbf{y}}}\frac{1}{L_{\mathbf{y}}}\mathbf{y} \tag{2-8}$$

where the vector $L_{\mathbf{y}}^{-1}\mathbf{y}$ has unit length. The *length* of the projection is

$$\text{Length of projection} = \frac{|\mathbf{x}'\mathbf{y}|}{L_{\mathbf{y}}} = L_{\mathbf{x}}\left|\frac{\mathbf{x}'\mathbf{y}}{L_{\mathbf{x}}L_{\mathbf{y}}}\right| = L_{\mathbf{x}}|\cos(\theta)| \tag{2-9}$$

where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$. (See Figure 2.5.)

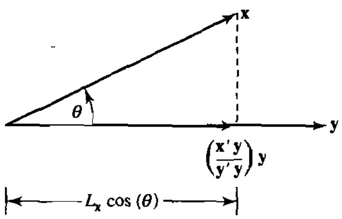

Figure 2.5 The projection of $\mathbf{x}$ on $\mathbf{y}$.

Matrices

A *matrix* is any rectangular array of real numbers. We denote an arbitrary array of n rows and p columns by

$$\underset{(n \times p)}{\mathbf{A}} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix}$$

Many of the vector concepts just introduced have direct generalizations to matrices.

The *transpose* operation $\mathbf{A}'$ of a matrix changes the columns into rows, so that the first column of $\mathbf{A}$ becomes the first row of $\mathbf{A}'$, the second column becomes the second row, and so forth.

Example 2.3 (The transpose of a matrix) If

$$\underset{(2\times3)}{\mathbf{A}} = \begin{bmatrix} 3 & -1 & 2 \\ 1 & 5 & 4 \end{bmatrix}$$

then

$$\underset{(3\times2)}{\mathbf{A}'} = \begin{bmatrix} 3 & 1 \\ -1 & 5 \\ 2 & 4 \end{bmatrix}$$

■

A matrix may also be multiplied by a constant c. The product $c\mathbf{A}$ is the matrix that results from multiplying each element of $\mathbf{A}$ by c. Thus

$$\underset{(n\times p)}{c\mathbf{A}} = \begin{bmatrix} ca_{11} & ca_{12} & \cdots & ca_{1p} \\ ca_{21} & ca_{22} & \cdots & ca_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{n1} & ca_{n2} & \cdots & ca_{np} \end{bmatrix}$$

Two matrices $\mathbf{A}$ and $\mathbf{B}$ of the same dimensions can be added. The sum $\mathbf{A} + \mathbf{B}$ has (i, j)th entry $a_{ij} + b_{ij}$.

Example 2.4 (The sum of two matrices and multiplication of a matrix by a constant) If

$$\underset{(2\times3)}{\mathbf{A}} = \begin{bmatrix} 0 & 3 & 1 \\ 1 & -1 & 1 \end{bmatrix} \quad \text{and} \quad \underset{(2\times3)}{\mathbf{B}} = \begin{bmatrix} 1 & -2 & -3 \\ 2 & 5 & 1 \end{bmatrix}$$

then

$$\underset{(2\times3)}{4\mathbf{A}} = \begin{bmatrix} 0 & 12 & 4 \\ 4 & -4 & 4 \end{bmatrix} \quad \text{and}$$

$$\underset{(2\times3)}{\mathbf{A}} + \underset{(2\times3)}{\mathbf{B}} = \begin{bmatrix} 0+1 & 3-2 & 1-3 \\ 1+2 & -1+5 & 1+1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & -2 \\ 3 & 4 & 2 \end{bmatrix}$$

■

It is also possible to define the multiplication of two matrices if the dimensions of the matrices conform in the following manner: When $\mathbf{A}$ is $(n \times k)$ and $\mathbf{B}$ is $(k \times p)$, so that the number of elements in a row of $\mathbf{A}$ is the same as the number of elements in a column of $\mathbf{B}$, we can form the matrix product $\mathbf{AB}$. An element of the new matrix $\mathbf{AB}$ is formed by taking the inner product of each row of $\mathbf{A}$ with each column of $\mathbf{B}$.

The *matrix product* **AB** is

$$\underset{(n\times k)(k\times p)}{\mathbf{A}\ \ \mathbf{B}} = \text{the } (n \times p) \text{ matrix whose entry in the } i\text{th row and } j\text{th column is the inner product of the } i\text{th row of } \mathbf{A} \text{ and the } j\text{th column of } \mathbf{B}$$

or

$$(i, j) \text{ entry of } \mathbf{AB} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ik}b_{kj} = \sum_{\ell=1}^{k} a_{i\ell}b_{\ell j} \qquad \text{(2-10)}$$

When $k = 4$, we have four products to add for each entry in the matrix **AB**. Thus,

$$\underset{(n\times 4)(4\times p)}{\mathbf{A}\ \ \mathbf{B}} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & a_{i3} & a_{i4} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & a_{n4} \end{bmatrix} \begin{bmatrix} b_{11} & \cdots & b_{1j} & \cdots & b_{1p} \\ b_{21} & \cdots & b_{2j} & \cdots & b_{2p} \\ b_{31} & \cdots & b_{3j} & \cdots & b_{3p} \\ b_{41} & \cdots & b_{4j} & \cdots & b_{4p} \end{bmatrix}$$

$$\begin{array}{c} \text{Column} \\ j \end{array}$$

$$= \text{Row } i \begin{bmatrix} & & \vdots & \\ \cdots (a_{i1}b_{1j} + a_{i2}b_{2j} & + & a_{i3}b_{3j} + a_{i4}b_{4j}) \cdots \\ & & \vdots & \end{bmatrix}$$

Example 2.5 (Matrix multiplication) If

$$\mathbf{A} = \begin{bmatrix} 3 & -1 & 2 \\ 1 & 5 & 4 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} -2 \\ 7 \\ 9 \end{bmatrix}, \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} 2 & 0 \\ 1 & -1 \end{bmatrix}$$

then

$$\underset{(2\times 3)(3\times 1)}{\mathbf{A}\ \ \mathbf{B}} = \begin{bmatrix} 3 & -1 & 2 \\ 1 & 5 & 4 \end{bmatrix} \begin{bmatrix} -2 \\ 7 \\ 9 \end{bmatrix} = \begin{bmatrix} 3(-2) + (-1)(7) + 2(9) \\ 1(-2) + 5(7) \qquad + 4(9) \end{bmatrix}$$

$$= \underset{(2\times 1)}{\begin{bmatrix} 5 \\ 69 \end{bmatrix}}$$

and

$$\underset{(2\times 2)(2\times 3)}{\mathbf{C}\ \ \mathbf{A}} = \begin{bmatrix} 2 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 3 & -1 & 2 \\ 1 & 5 & 4 \end{bmatrix}$$

$$= \begin{bmatrix} 2(3) + 0(1) & 2(-1) + 0(5) & 2(2) + 0(4) \\ 1(3) - 1(1) & 1(-1) - 1(5) & 1(2) - 1(4) \end{bmatrix}$$

$$= \underset{(2\times 3)}{\begin{bmatrix} 6 & -2 & 4 \\ 2 & -6 & -2 \end{bmatrix}}$$

■

When a matrix **B** consists of a single column, it is customary to use the lowercase **b** vector notation.

Example 2.6 (Some typical products and their dimensions) Let

$$\mathbf{A} = \begin{bmatrix} 1 & -2 & 3 \\ 2 & 4 & -1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 7 \\ -3 \\ 6 \end{bmatrix} \quad \mathbf{c} = \begin{bmatrix} 5 \\ 8 \\ -4 \end{bmatrix} \quad \mathbf{d} = \begin{bmatrix} 2 \\ 9 \end{bmatrix}$$

Then $\mathbf{A}\,\mathbf{b}$, $\mathbf{b}\mathbf{c}'$, $\mathbf{b}'\mathbf{c}$, and $\mathbf{d}'\mathbf{A}\,\mathbf{b}$ are typical products.

$$\mathbf{A}\,\mathbf{b} = \begin{bmatrix} 1 & -2 & 3 \\ 2 & 4 & -1 \end{bmatrix} \begin{bmatrix} 7 \\ -3 \\ 6 \end{bmatrix} = \begin{bmatrix} 31 \\ -4 \end{bmatrix}$$

The product $\mathbf{A}\,\mathbf{b}$ is a vector with dimension equal to the number of rows of $\mathbf{A}$.

$$\mathbf{b}'\mathbf{c} = [7 \quad -3 \quad 6] \begin{bmatrix} 5 \\ 8 \\ -4 \end{bmatrix} = [-13]$$

The product $\mathbf{b}'\mathbf{c}$ is a 1×1 vector or a single number, here -13.

$$\mathbf{b}\,\mathbf{c}' = \begin{bmatrix} 7 \\ -3 \\ 6 \end{bmatrix} [5 \quad 8 \quad -4] = \begin{bmatrix} 35 & 56 & -28 \\ -15 & -24 & 12 \\ 30 & 48 & -24 \end{bmatrix}$$

The product $\mathbf{b}\,\mathbf{c}'$ is a matrix whose row dimension equals the dimension of $\mathbf{b}$ and whose column dimension equals that of $\mathbf{c}$. This product is unlike $\mathbf{b}'\mathbf{c}$, which is a single number.

$$\mathbf{d}'\mathbf{A}\,\mathbf{b} = [2 \quad 9] \begin{bmatrix} 1 & -2 & 3 \\ 2 & 4 & -1 \end{bmatrix} \begin{bmatrix} 7 \\ -3 \\ 6 \end{bmatrix} = [26]$$

The product $\mathbf{d}'\mathbf{A}\,\mathbf{b}$ is a 1×1 vector or a single number, here 26. ■

Square matrices will be of special importance in our development of statistical methods. A square matrix is said to be *symmetric* if $\mathbf{A} = \mathbf{A}'$ or $a_{ij} = a_{ji}$ for all i and j.

Example 2.7 (A symmetric matrix) The matrix

$$\begin{bmatrix} 3 & 5 \\ 5 & -2 \end{bmatrix}$$

is symmetric; the matrix

$$\begin{bmatrix} 3 & 6 \\ 4 & -2 \end{bmatrix}$$

is not symmetric. ■

When two square matrices $\mathbf{A}$ and $\mathbf{B}$ are of the same dimension, both products $\mathbf{AB}$ and $\mathbf{BA}$ are defined, although they need not be equal. (See Supplement 2A.) If we let $\mathbf{I}$ denote the square matrix with ones on the diagonal and zeros elsewhere, it follows from the definition of matrix multiplication that the (i, j)th entry of $\mathbf{AI}$ is $a_{i1} \times 0 + \cdots + a_{i,j-1} \times 0 + a_{ij} \times 1 + a_{i,j+1} \times 0 + \cdots + a_{ik} \times 0 = a_{ij}$, so $\mathbf{AI} = \mathbf{A}$. Similarly, $\mathbf{IA} = \mathbf{A}$, so

$$\underset{(k\times k)(k\times k)}{\mathbf{I}\;\;\mathbf{A}} = \underset{(k\times k)(k\times k)}{\mathbf{A}\;\;\mathbf{I}} = \underset{(k\times k)}{\mathbf{A}} \quad \text{for any } \underset{(k\times k)}{\mathbf{A}} \tag{2-11}$$

The matrix $\mathbf{I}$ acts like 1 in ordinary multiplication $(1 \cdot a = a \cdot 1 = a)$, so it is called the *identity* matrix.

The fundamental scalar relation about the existence of an inverse number a^{-1} such that $a^{-1}a = aa^{-1} = 1$ if $a \neq 0$ has the following matrix algebra extension: If there exists a matrix $\mathbf{B}$ such that

$$\underset{(k\times k)(k\times k)}{\mathbf{B}\;\;\mathbf{A}} = \underset{(k\times k)(k\times k)}{\mathbf{A}\;\;\mathbf{B}} = \underset{(k\times k)}{\mathbf{I}}$$

then $\mathbf{B}$ is called the *inverse* of $\mathbf{A}$ and is denoted by $\mathbf{A}^{-1}$.

The technical condition that an inverse exists is that the k columns $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$ of $\mathbf{A}$ are linearly independent. That is, the existence of $\mathbf{A}^{-1}$ is equivalent to

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_k\mathbf{a}_k = \mathbf{0} \quad \text{only if} \quad c_1 = \cdots = c_k = 0 \tag{2-12}$$

(See Result 2A.9 in Supplement 2A.)

Example 2.8 (The existence of a matrix inverse) For

$$\mathbf{A} = \begin{bmatrix} 3 & 2 \\ 4 & 1 \end{bmatrix}$$

you may verify that

$$\begin{bmatrix} -.2 & .4 \\ .8 & -.6 \end{bmatrix}\begin{bmatrix} 3 & 2 \\ 4 & 1 \end{bmatrix} = \begin{bmatrix} (-.2)3 + (.4)4 & (-.2)2 + (.4)1 \\ (.8)3 + (-.6)4 & (.8)2 + (-.6)1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

so

$$\begin{bmatrix} -.2 & .4 \\ .8 & -.6 \end{bmatrix}$$

is $\mathbf{A}^{-1}$. We note that

$$c_1 \begin{bmatrix} 3 \\ 4 \end{bmatrix} + c_2 \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

implies that $c_1 = c_2 = 0$, so the columns of $\mathbf{A}$ are linearly independent. This confirms the condition stated in (2-12). ■

A method for computing an inverse, when one exists, is given in Supplement 2A. The routine, but lengthy, calculations are usually relegated to a computer, especially when the dimension is greater than three. Even so, you must be forewarned that if the column sum in (2-12) is *nearly* $\mathbf{0}$ for some constants $c_1, \ldots, c_k$, then the computer may produce incorrect inverses due to extreme errors in rounding. It is always good to check the products $\mathbf{A}\mathbf{A}^{-1}$ and $\mathbf{A}^{-1}\mathbf{A}$ for equality with $\mathbf{I}$ when $\mathbf{A}^{-1}$ is produced by a computer package. (See Exercise 2.10.)

Diagonal matrices have inverses that are easy to compute. For example,

$$\begin{bmatrix} a_{11} & 0 & 0 & 0 & 0 \\ 0 & a_{22} & 0 & 0 & 0 \\ 0 & 0 & a_{33} & 0 & 0 \\ 0 & 0 & 0 & a_{44} & 0 \\ 0 & 0 & 0 & 0 & a_{55} \end{bmatrix} \quad \text{has inverse} \quad \begin{bmatrix} \frac{1}{a_{11}} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{a_{22}} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{a_{33}} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{a_{44}} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{a_{55}} \end{bmatrix}$$

if all the $a_{ii} \neq 0$.

Another special class of square matrices with which we shall become familiar are the *orthogonal* matrices, characterized by

$$\mathbf{QQ}' = \mathbf{Q}'\mathbf{Q} = \mathbf{I} \quad \text{or} \quad \mathbf{Q}' = \mathbf{Q}^{-1} \tag{2-13}$$

The name derives from the property that if $\mathbf{Q}$ has ith row $\mathbf{q}_i'$, then $\mathbf{QQ}' = \mathbf{I}$ implies that $\mathbf{q}_i'\mathbf{q}_i = 1$ and $\mathbf{q}_i'\mathbf{q}_j = 0$ for $i \neq j$, so the rows have unit length and are mutually perpendicular (orthogonal). According to the condition $\mathbf{Q}'\mathbf{Q} = \mathbf{I}$, the columns have the same property.

We conclude our brief introduction to the elements of matrix algebra by introducing a concept fundamental to multivariate statistical analysis. A square matrix $\mathbf{A}$ is said to have an *eigenvalue* λ, with corresponding *eigenvector* $\mathbf{x} \neq \mathbf{0}$, if

$$\mathbf{Ax} = \lambda\mathbf{x} \tag{2-14}$$

Ordinarily, we normalize $\mathbf{x}$ so that it has length unity; that is, $1 = \mathbf{x}'\mathbf{x}$. It is convenient to denote normalized eigenvectors by $\mathbf{e}$, and we do so in what follows. Sparing you the details of the derivation (see [1]), we state the following basic result:

Let $\mathbf{A}$ be a $k \times k$ square symmetric matrix. Then $\mathbf{A}$ has k pairs of eigenvalues and eigenvectors namely,

$$\lambda_1, \mathbf{e}_1 \qquad \lambda_2, \mathbf{e}_2 \quad \ldots \quad \lambda_k, \mathbf{e}_k \tag{2-15}$$

The eigenvectors can be chosen to satisfy $1 = \mathbf{e}_1'\mathbf{e}_1 = \cdots = \mathbf{e}_k'\mathbf{e}_k$ and be mutually perpendicular. The eigenvectors are unique unless two or more eigenvalues are equal.

Example 2.9 (Verifying eigenvalues and eigenvectors) Let

$$\mathbf{A} = \begin{bmatrix} 1 & -5 \\ -5 & 1 \end{bmatrix}$$

Then, since

$$\begin{bmatrix} 1 & -5 \\ -5 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} = 6 \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

$\lambda_1 = 6$ is an eigenvalue, and

$$\mathbf{e}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

is its corresponding normalized eigenvector. You may wish to show that a second eigenvalue–eigenvector pair is $\lambda_2 = -4$, $\mathbf{e}_2' = [1/\sqrt{2}, 1/\sqrt{2}]$. ■

A method for calculating the λ's and $\mathbf{e}$'s is described in Supplement 2A. It is instructive to do a few sample calculations to understand the technique. We usually rely on a computer when the dimension of the square matrix is greater than two or three.

2.3 Positive Definite Matrices

The study of the variation and interrelationships in multivariate data is often based upon distances and the assumption that the data are multivariate normally distributed. Squared distances (see Chapter 1) and the multivariate normal density can be expressed in terms of matrix products called *quadratic forms* (see Chapter 4). Consequently, it should not be surprising that quadratic forms play a central role in multivariate analysis. In this section, we consider quadratic forms that are always nonnegative and the associated *positive definite* matrices.

Results involving quadratic forms and symmetric matrices are, in many cases, a direct consequence of an expansion for symmetric matrices known as the *spectral decomposition*. The spectral decomposition of a $k \times k$ symmetric matrix $\mathbf{A}$ is given by[1]

$$\underset{(k\times k)}{\mathbf{A}} = \lambda_1 \underset{(k\times 1)}{\mathbf{e}_1} \underset{(1\times k)}{\mathbf{e}_1'} + \lambda_2 \underset{(k\times 1)}{\mathbf{e}_2} \underset{(1\times k)}{\mathbf{e}_2'} + \cdots + \lambda_k \underset{(k\times 1)}{\mathbf{e}_k} \underset{(1\times k)}{\mathbf{e}_k'} \tag{2-16}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the eigenvalues of $\mathbf{A}$ and $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k$ are the associated normalized eigenvectors. (See also Result 2A.14 in Supplement 2A). Thus, $\mathbf{e}_i'\mathbf{e}_i = 1$ for $i = 1, 2, \ldots, k$, and $\mathbf{e}_i'\mathbf{e}_j = 0$ for $i \neq j$.

Example 2.10 (The spectral decomposition of a matrix) Consider the symmetric matrix

$$\mathbf{A} = \begin{bmatrix} 13 & -4 & 2 \\ -4 & 13 & -2 \\ 2 & -2 & 10 \end{bmatrix}$$

The eigenvalues obtained from the characteristic equation $|\mathbf{A} - \lambda\mathbf{I}| = 0$ are $\lambda_1 = 9$, $\lambda_2 = 9$, and $\lambda_3 = 18$ (Definition 2A.30). The corresponding eigenvectors $\mathbf{e}_1, \mathbf{e}_2$, and $\mathbf{e}_3$ are the (normalized) solutions of the equations $\mathbf{A}\mathbf{e}_i = \lambda_i\mathbf{e}_i$ for $i = 1, 2, 3$. Thus, $\mathbf{A}\mathbf{e}_1 = \lambda\mathbf{e}_1$ gives

$$\begin{bmatrix} 13 & -4 & 2 \\ -4 & 13 & -2 \\ 2 & -2 & 10 \end{bmatrix} \begin{bmatrix} e_{11} \\ e_{21} \\ e_{31} \end{bmatrix} = 9 \begin{bmatrix} e_{11} \\ e_{21} \\ e_{31} \end{bmatrix}$$

or

$$\begin{aligned} 13e_{11} - 4e_{21} + 2e_{31} &= 9e_{11} \\ -4e_{11} + 13e_{21} - 2e_{31} &= 9e_{21} \\ 2e_{11} - 2e_{21} + 10e_{31} &= 9e_{31} \end{aligned}$$

Moving the terms on the right of the equals sign to the left yields three homogeneous equations in three unknowns, but two of the equations are redundant. Selecting one of the equations and arbitrarily setting $e_{11} = 1$ and $e_{21} = 1$, we find that $e_{31} = 0$. Consequently, the normalized eigenvector is $\mathbf{e}_1' = [1/\sqrt{1^2 + 1^2 + 0^2}, 1/\sqrt{1^2 + 1^2 + 0^2}, 0/\sqrt{1^2 + 1^2 + 0^2}] = [1/\sqrt{2}, 1/\sqrt{2}, 0]$, since the sum of the squares of its elements is unity. You may verify that $\mathbf{e}_2' = [1/\sqrt{18}, -1/\sqrt{18}, -4/\sqrt{18}]$ is also an eigenvector for $9 = \lambda_2$, and $\mathbf{e}_3' = [2/3, -2/3, 1/3]$ is the normalized eigenvector corresponding to the eigenvalue $\lambda_3 = 18$. Moreover, $\mathbf{e}_i'\mathbf{e}_j = 0$ for $i \neq j$.

[1]A proof of Equation (2-16) is beyond the scope of this book. The interested reader will find a proof in [6], Chapter 8.

The spectral decomposition of $\mathbf{A}$ is then

$$\mathbf{A} = \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2' + \lambda_3 \mathbf{e}_3 \mathbf{e}_3'$$

or

$$\begin{bmatrix} 13 & -4 & 2 \\ -4 & 13 & -2 \\ 2 & -2 & 10 \end{bmatrix} = 9 \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix}$$

$$+ 9 \begin{bmatrix} \frac{1}{\sqrt{18}} \\ \frac{-1}{\sqrt{18}} \\ \frac{-4}{\sqrt{18}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{18}} & \frac{-1}{\sqrt{18}} & \frac{-4}{\sqrt{18}} \end{bmatrix} + 18 \begin{bmatrix} \frac{2}{3} \\ -\frac{2}{3} \\ \frac{1}{3} \end{bmatrix} \begin{bmatrix} \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

$$= 9 \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix} + 9 \begin{bmatrix} \frac{1}{18} & -\frac{1}{18} & -\frac{4}{18} \\ -\frac{1}{18} & \frac{1}{18} & \frac{4}{18} \\ -\frac{4}{18} & \frac{4}{18} & \frac{16}{18} \end{bmatrix}$$

$$+ 18 \begin{bmatrix} \frac{4}{9} & -\frac{4}{9} & \frac{2}{9} \\ -\frac{4}{9} & \frac{4}{9} & -\frac{2}{9} \\ \frac{2}{9} & -\frac{2}{9} & \frac{1}{9} \end{bmatrix}$$

as you may readily verify. ■

The spectral decomposition is an important analytical tool. With it, we are very easily able to demonstrate certain statistical results. The first of these is a matrix explanation of distance, which we now develop.

Because $\mathbf{x}'\mathbf{A}\mathbf{x}$ has only squared terms x_i^2 and product terms $x_i x_k$, it is called a *quadratic form*. When a $k \times k$ symmetric matrix $\mathbf{A}$ is such that

$$0 \leq \mathbf{x}'\mathbf{A}\mathbf{x} \tag{2-17}$$

for all $\mathbf{x}' = [x_1, x_2, \dots, x_k]$, both the matrix $\mathbf{A}$ and the quadratic form are said to be *nonnegative definite*. If equality holds in (2-17) only for the vector $\mathbf{x}' = [0, 0, \dots, 0]$, then $\mathbf{A}$ or the quadratic form is said to be *positive definite*. In other words, $\mathbf{A}$ is positive definite if

$$0 < \mathbf{x}'\mathbf{A}\mathbf{x} \tag{2-18}$$

for all vectors $\mathbf{x} \neq \mathbf{0}$.

Example 2.11 (A positive definite matrix and quadratic form) Show that the matrix for the following quadratic form is positive definite:

$$3x_1^2 + 2x_2^2 - 2\sqrt{2}\, x_1 x_2$$

To illustrate the general approach, we first write the quadratic form in matrix notation as

$$[x_1 \quad x_2]\begin{bmatrix} 3 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{x}'\mathbf{A}\mathbf{x}$$

By Definition 2A.30, the eigenvalues of $\mathbf{A}$ are the solutions of the equation $|\mathbf{A} - \lambda\mathbf{I}| = 0$, or $(3 - \lambda)(2 - \lambda) - 2 = 0$. The solutions are $\lambda_1 = 4$ and $\lambda_2 = 1$. Using the spectral decomposition in (2-16), we can write

$$\underset{(2\times 2)}{\mathbf{A}} = \lambda_1 \underset{(2\times 1)}{\mathbf{e}_1} \underset{(1\times 2)}{\mathbf{e}_1'} + \lambda_2 \underset{(2\times 1)}{\mathbf{e}_2} \underset{(1\times 2)}{\mathbf{e}_2'}$$

$$= 4\underset{(2\times 1)}{\mathbf{e}_1} \underset{(1\times 2)}{\mathbf{e}_1'} + \underset{(2\times 1)}{\mathbf{e}_2} \underset{(1\times 2)}{\mathbf{e}_2'}$$

where $\mathbf{e}_1$ and $\mathbf{e}_2$ are the normalized and orthogonal eigenvectors associated with the eigenvalues $\lambda_1 = 4$ and $\lambda_2 = 1$, respectively. Because 4 and 1 are scalars, premultiplication and postmultiplication of $\mathbf{A}$ by $\mathbf{x}'$ and $\mathbf{x}$, respectively, where $\mathbf{x}' = [x_1, x_2]$ is any *nonzero* vector, give

$$\underset{(1\times 2)}{\mathbf{x}'} \underset{(2\times 2)}{\mathbf{A}} \underset{(2\times 1)}{\mathbf{x}} = 4\underset{(1\times 2)}{\mathbf{x}'} \underset{(2\times 1)}{\mathbf{e}_1} \underset{(1\times 2)}{\mathbf{e}_1'} \underset{(2\times 1)}{\mathbf{x}} + \underset{(1\times 2)}{\mathbf{x}'} \underset{(2\times 1)}{\mathbf{e}_2} \underset{(1\times 2)}{\mathbf{e}_2'} \underset{(2\times 1)}{\mathbf{x}}$$

$$= 4y_1^2 + y_2^2 \geq 0$$

with

$$y_1 = \mathbf{x}'\mathbf{e}_1 = \mathbf{e}_1'\mathbf{x} \quad \text{and} \quad y_2 = \mathbf{x}'\mathbf{e}_2 = \mathbf{e}_2'\mathbf{x}$$

We now show that y_1 and y_2 are not both zero and, consequently, that $\mathbf{x}'\mathbf{A}\mathbf{x} = 4y_1^2 + y_2^2 > 0$, or $\mathbf{A}$ is *positive definite.*

From the definitions of y_1 and y_2, we have

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1' \\ \mathbf{e}_2' \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

or

$$\underset{(2\times 1)}{\mathbf{y}} = \underset{(2\times 2)}{\mathbf{E}} \underset{(2\times 1)}{\mathbf{x}}$$

Now $\mathbf{E}$ is an orthogonal matrix and hence has inverse $\mathbf{E}'$. Thus, $\mathbf{x} = \mathbf{E}'\mathbf{y}$. But $\mathbf{x}$ is a nonzero vector, and $\mathbf{0} \neq \mathbf{x} = \mathbf{E}'\mathbf{y}$ implies that $\mathbf{y} \neq \mathbf{0}$. ■

Using the spectral decomposition, we can easily show that a $k \times k$ symmetric matrix $\mathbf{A}$ is a positive definite matrix if and only if every eigenvalue of $\mathbf{A}$ is positive. (See Exercise 2.17.) $\mathbf{A}$ is a nonnegative definite matrix if and only if all of its eigenvalues are greater than or equal to zero.

Assume for the moment that the p elements $x_1, x_2, \ldots, x_p$ of a vector $\mathbf{x}$ are realizations of p random variables $X_1, X_2, \ldots, X_p$. As we pointed out in Chapter 1, we can regard these elements as the coordinates of a point in p-dimensional space, and the "distance" of the point $[x_1, x_2, \ldots, x_p]'$ to the origin can, and in this case should, be interpreted in terms of standard deviation units. In this way, we can account for the inherent uncertainty (variability) in the observations. Points with the same associated "uncertainty" are regarded as being at the same distance from the origin.

If we use the distance formula introduced in Chapter 1 [see Equation (1-22)], the distance from the origin satisfies the general formula

$$(\text{distance})^2 = a_{11}x_1^2 + a_{22}x_2^2 + \cdots + a_{pp}x_p^2 + 2(a_{12}x_1x_2 + a_{13}x_1x_3 + \cdots + a_{p-1,p}x_{p-1}x_p)$$

provided that $(\text{distance})^2 > 0$ for all $[x_1, x_2, \ldots, x_p] \neq [0, 0, \ldots, 0]$. Setting $a_{ij} = a_{ji}$, $i \neq j$, $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, p$, we have

$$0 < (\text{distance})^2 = [x_1, x_2, \ldots, x_p] \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$$

or

$$0 < (\text{distance})^2 = \mathbf{x}'\mathbf{A}\mathbf{x} \qquad \text{for } \mathbf{x} \neq \mathbf{0} \tag{2-19}$$

From (2-19), we see that the $p \times p$ symmetric matrix $\mathbf{A}$ is positive definite. In sum, distance is determined from a positive definite quadratic form $\mathbf{x}'\mathbf{A}\mathbf{x}$. Conversely, a positive definite quadratic form can be interpreted as a squared distance.

Comment. Let the square of the distance from the point $\mathbf{x}' = [x_1, x_2, \ldots, x_p]$ to the origin be given by $\mathbf{x}'\mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is a $p \times p$ symmetric positive definite matrix. Then the square of the distance from $\mathbf{x}$ to an arbitrary fixed point $\boldsymbol{\mu}' = [\mu_1, \mu_2, \ldots, \mu_p]$ is given by the general expression $(\mathbf{x} - \boldsymbol{\mu})'\mathbf{A}(\mathbf{x} - \boldsymbol{\mu})$.

Expressing distance as the square root of a positive definite quadratic form allows us to give a geometrical interpretation based on the eigenvalues and eigenvectors of the matrix $\mathbf{A}$. For example, suppose $p = 2$. Then the points $\mathbf{x}' = [x_1, x_2]$ of constant distance c from the origin satisfy

$$\mathbf{x}'\mathbf{A}\mathbf{x} = a_{11}x_1^2 + a_{22}x_2^2 + 2a_{12}x_1x_2 = c^2$$

By the spectral decomposition, as in Example 2.11,

$$\mathbf{A} = \lambda_1\mathbf{e}_1\mathbf{e}_1' + \lambda_2\mathbf{e}_2\mathbf{e}_2' \quad \text{so} \quad \mathbf{x}'\mathbf{A}\mathbf{x} = \lambda_1(\mathbf{x}'\mathbf{e}_1)^2 + \lambda_2(\mathbf{x}'\mathbf{e}_2)^2$$

Now, $c^2 = \lambda_1 y_1^2 + \lambda_2 y_2^2$ is an ellipse in $y_1 = \mathbf{x}'\mathbf{e}_1$ and $y_2 = \mathbf{x}'\mathbf{e}_2$ because $\lambda_1, \lambda_2 > 0$ when $\mathbf{A}$ is positive definite. (See Exercise 2.17.) We easily verify that $\mathbf{x} = c\lambda_1^{-1/2}\mathbf{e}_1$ satisfies $\mathbf{x}'\mathbf{A}\mathbf{x} = \lambda_1(c\lambda_1^{-1/2}\mathbf{e}_1'\mathbf{e}_1)^2 = c^2$. Similarly, $\mathbf{x} = c\lambda_2^{-1/2}\mathbf{e}_2$ gives the appropriate distance in the $\mathbf{e}_2$ direction. Thus, the points at distance c lie on an ellipse whose axes are given by the eigenvectors of $\mathbf{A}$ with lengths proportional to the reciprocals of the square roots of the eigenvalues. The constant of proportionality is c. The situation is illustrated in Figure 2.6.

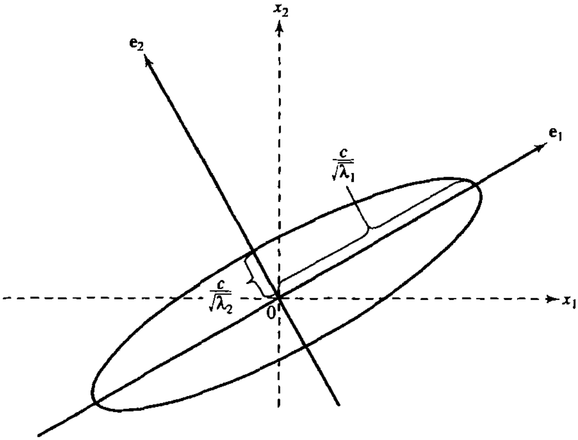

Figure 2.6 Points a constant distance c from the origin ($p = 2, 1 \leq \lambda_1 < \lambda_2$).

If $p > 2$, the points $\mathbf{x}' = [x_1, x_2, \ldots, x_p]$ a constant distance $c = \sqrt{\mathbf{x}'\mathbf{A}\mathbf{x}}$ from the origin lie on hyperellipsoids $c^2 = \lambda_1(\mathbf{x}'\mathbf{e}_1)^2 + \cdots + \lambda_p(\mathbf{x}'\mathbf{e}_p)^2$, whose axes are given by the eigenvectors of $\mathbf{A}$. The half-length in the direction $\mathbf{e}_i$ is equal to $c/\sqrt{\lambda_i}$, $i = 1, 2, \ldots, p$, where $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the eigenvalues of $\mathbf{A}$.

2.4 A Square-Root Matrix

The spectral decomposition allows us to express the inverse of a square matrix in terms of its eigenvalues and eigenvectors, and this leads to a useful *square-root matrix*.

Let $\mathbf{A}$ be a $k \times k$ positive definite matrix with the spectral decomposition $\mathbf{A} = \sum_{i=1}^{k} \lambda_i \mathbf{e}_i \mathbf{e}_i'$. Let the normalized eigenvectors be the columns of another matrix $\mathbf{P} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k]$. Then

$$\underset{(k\times k)}{\mathbf{A}} = \sum_{i=1}^{k} \lambda_i \underset{(k\times 1)}{\mathbf{e}_i} \underset{(1\times k)}{\mathbf{e}_i'} = \underset{(k\times k)}{\mathbf{P}} \underset{(k\times k)}{\mathbf{\Lambda}} \underset{(k\times k)}{\mathbf{P}'} \tag{2-20}$$

where $\mathbf{PP}' = \mathbf{P}'\mathbf{P} = \mathbf{I}$ and $\mathbf{\Lambda}$ is the diagonal matrix

$$\underset{(k\times k)}{\mathbf{\Lambda}} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix} \quad \text{with } \lambda_i > 0$$

Thus,

$$\mathbf{A}^{-1} = \mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}' = \sum_{i=1}^{k} \frac{1}{\lambda_i} \mathbf{e}_i\mathbf{e}_i' \tag{2-21}$$

since $(\mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}')\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}' = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'(\mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}') = \mathbf{P}\mathbf{P}' = \mathbf{I}$.

Next, let $\boldsymbol{\Lambda}^{1/2}$ denote the diagonal matrix with $\sqrt{\lambda_i}$ as the ith diagonal element. The matrix $\sum_{i=1}^{k} \sqrt{\lambda_i}\, \mathbf{e}_i\mathbf{e}_i' = \mathbf{P}\boldsymbol{\Lambda}^{1/2}\mathbf{P}'$ is called the *square root* of $\mathbf{A}$ and is denoted by $\mathbf{A}^{1/2}$.

The square-root matrix, of a positive definite matrix $\mathbf{A}$,

$$\mathbf{A}^{1/2} = \sum_{i=1}^{k} \sqrt{\lambda_i}\, \mathbf{e}_i\mathbf{e}_i' = \mathbf{P}\boldsymbol{\Lambda}^{1/2}\mathbf{P}' \tag{2-22}$$

has the following properties:

1. $(\mathbf{A}^{1/2})' = \mathbf{A}^{1/2}$ (that is, $\mathbf{A}^{1/2}$ is symmetric).
2. $\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}$.
3. $(\mathbf{A}^{1/2})^{-1} = \sum_{i=1}^{k} \frac{1}{\sqrt{\lambda_i}} \mathbf{e}_i\mathbf{e}_i' = \mathbf{P}\boldsymbol{\Lambda}^{-1/2}\mathbf{P}'$, where $\boldsymbol{\Lambda}^{-1/2}$ is a diagonal matrix with $1/\sqrt{\lambda_i}$ as the ith diagonal element.
4. $\mathbf{A}^{1/2}\mathbf{A}^{-1/2} = \mathbf{A}^{-1/2}\mathbf{A}^{1/2} = \mathbf{I}$, and $\mathbf{A}^{-1/2}\mathbf{A}^{-1/2} = \mathbf{A}^{-1}$, where $\mathbf{A}^{-1/2} = (\mathbf{A}^{1/2})^{-1}$.

2.5 Random Vectors and Matrices

A *random vector* is a vector whose elements are random variables. Similarly, a *random matrix* is a matrix whose elements are random variables. The expected value of a random matrix (or vector) is the matrix (vector) consisting of the expected values of each of its elements. Specifically, let $\mathbf{X} = \{X_{ij}\}$ be an $n \times p$ random matrix. Then the expected value of $\mathbf{X}$, denoted by $E(\mathbf{X})$, is the $n \times p$ matrix of numbers (if they exist)

$$E(\mathbf{X}) = \begin{bmatrix} E(X_{11}) & E(X_{12}) & \cdots & E(X_{1p}) \\ E(X_{21}) & E(X_{22}) & \cdots & E(X_{2p}) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_{n1}) & E(X_{n2}) & \cdots & E(X_{np}) \end{bmatrix} \tag{2-23}$$

where, for each element of the matrix,[2]

$$E(X_{ij}) = \begin{cases} \displaystyle\int_{-\infty}^{\infty} x_{ij} f_{ij}(x_{ij})\,dx_{ij} & \text{if } X_{ij} \text{ is a continuous random variable with probability density function } f_{ij}(x_{ij}) \\ \displaystyle\sum_{\text{all } x_{ij}} x_{ij} p_{ij}(x_{ij}) & \text{if } X_{ij} \text{ is a discrete random variable with probability function } p_{ij}(x_{ij}) \end{cases}$$

Example 2.12 (Computing expected values for discrete random variables) Suppose $p = 2$ and $n = 1$, and consider the random vector $\mathbf{X}' = [X_1, X_2]$. Let the discrete random variable X_1 have the following probability function:

x_1	-1	0	1
$p_1(x_1)$	.3	.3	.4

Then $E(X_1) = \sum_{\text{all } x_1} x_1 p_1(x_1) = (-1)(.3) + (0)(.3) + (1)(.4) = .1.$

Similarly, let the discrete random variable X_2 have the probability function

x_2	0	1
$p_2(x_2)$	.8	.2

Then $E(X_2) = \sum_{\text{all } x_2} x_2 p_2(x_2) = (0)(.8) + (1)(.2) = .2.$

Thus,

$$E(\mathbf{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \end{bmatrix} = \begin{bmatrix} .1 \\ .2 \end{bmatrix}$$

■

Two results involving the expectation of sums and products of matrices follow directly from the definition of the expected value of a random matrix and the univariate properties of expectation, $E(X_1 + Y_1) = E(X_1) + E(Y_1)$ and $E(cX_1) = cE(X_1)$. Let $\mathbf{X}$ and $\mathbf{Y}$ be random matrices of the same dimension, and let $\mathbf{A}$ and $\mathbf{B}$ be conformable matrices of constants. Then (see Exercise 2.40)

$$E(\mathbf{X} + \mathbf{Y}) = E(\mathbf{X}) + E(\mathbf{Y}) \tag{2-24}$$
$$E(\mathbf{AXB}) = \mathbf{A}E(\mathbf{X})\mathbf{B}$$

[2]If you are unfamiliar with calculus, you should concentrate on the interpretation of the expected value and, eventually, variance. Our development is based primarily on the properties of expectation rather than its particular evaluation for continuous or discrete random variables.

2.6 Mean Vectors and Covariance Matrices

Suppose $\mathbf{X}' = [X_1, X_2, \ldots, X_p]$ is a $p \times 1$ random vector. Then each element of $\mathbf{X}$ is a random variable with its own marginal probability distribution. (See Example 2.12.) The marginal means μ_i and variances σ_i^2 are defined as $\mu_i = E(X_i)$ and $\sigma_i^2 = E(X_i - \mu_i)^2$, $i = 1, 2, \ldots, p$, respectively. Specifically,

$$\mu_i = \begin{cases} \displaystyle\int_{-\infty}^{\infty} x_i f_i(x_i)\, dx_i & \text{if } X_i \text{ is a continuous random variable with probability density function } f_i(x_i) \\ \displaystyle\sum_{\text{all } x_i} x_i p_i(x_i) & \text{if } X_i \text{ is a discrete random variable with probability function } p_i(x_i) \end{cases}$$

$$\sigma_i^2 = \begin{cases} \displaystyle\int_{-\infty}^{\infty} (x_i - \mu_i)^2 f_i(x_i)\, dx_i & \text{if } X_i \text{ is a continuous random variable with probability density function } f_i(x_i) \\ \displaystyle\sum_{\text{all } x_i} (x_i - \mu_i)^2 p_i(x_i) & \text{if } X_i \text{ is a discrete random variable with probability function } p_i(x_i) \end{cases} \tag{2-25}$$

It will be convenient in later sections to denote the marginal variances by σ_{ii} rather than the more traditional σ_i^2, and consequently, we shall adopt this notation.

The behavior of any pair of random variables, such as X_i and X_k, is described by their joint probability function, and a measure of the linear association between them is provided by the covariance

$$\sigma_{ik} = E(X_i - \mu_i)(X_k - \mu_k)$$

$$= \begin{cases} \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x_i - \mu_i)(x_k - \mu_k) f_{ik}(x_i, x_k)\,dx_i\,dx_k & \text{if } X_i, X_k \text{ are continuous random variables with the joint density function } f_{ik}(x_i, x_k) \\ \displaystyle\sum_{\text{all } x_i}\sum_{\text{all } x_k} (x_i - \mu_i)(x_k - \mu_k) p_{ik}(x_i, x_k) & \text{if } X_i, X_k \text{ are discrete random variables with joint probability function } p_{ik}(x_i, x_k) \end{cases} \tag{2-26}$$

and μ_i and μ_k, $i, k = 1, 2, \ldots, p$, are the marginal means. When $i = k$, the covariance becomes the marginal variance.

More generally, the collective behavior of the p random variables $X_1, X_2, \ldots, X_p$ or, equivalently, the random vector $\mathbf{X}' = [X_1, X_2, \ldots, X_p]$, is described by a joint probability density function $f(x_1, x_2, \ldots, x_p) = f(\mathbf{x})$. As we have already noted in this book, $f(\mathbf{x})$ will often be the multivariate normal density function. (See Chapter 4.)

If the joint probability $P[X_i \le x_i \text{ and } X_k \le x_k]$ can be written as the product of the corresponding marginal probabilities, so that

$$P[X_i \le x_i \text{ and } X_k \le x_k] = P[X_i \le x_i]P[X_k \le x_k] \tag{2-27}$$

for all pairs of values x_i, x_k, then X_i and X_k are said to be *statistically independent*. When X_i and X_k are continuous random variables with joint density $f_{ik}(x_i, x_k)$ and marginal densities $f_i(x_i)$ and $f_k(x_k)$, the independence condition becomes

$$f_{ik}(x_i, x_k) = f_i(x_i)f_k(x_k)$$

for all pairs (x_i, x_k).

The p continuous random variables $X_1, X_2, \ldots, X_p$ are *mutually statistically independent* if their joint density can be factored as

$$f_{12\cdots p}(x_1, x_2, \ldots, x_p) = f_1(x_1)f_2(x_2)\cdots f_p(x_p) \tag{2-28}$$

for all p-tuples $(x_1, x_2, \ldots, x_p)$.

Statistical independence has an important implication for covariance. The factorization in (2-28) implies that $\text{Cov}(X_i, X_k) = 0$. Thus,

$$\text{Cov}(X_i, X_k) = 0 \qquad \text{if } X_i \text{ and } X_k \text{ are independent} \tag{2-29}$$

The converse of (2-29) is not true in general; there are situations where $\text{Cov}(X_i, X_k) = 0$, but X_i and X_k are not independent. (See [5].)

The means and covariances of the $p \times 1$ random vector $\mathbf{X}$ can be set out as matrices. The expected value of each element is contained in the vector of means $\boldsymbol{\mu} = E(\mathbf{X})$, and the p variances σ_{ii} and the $p(p-1)/2$ distinct covariances $\sigma_{ik}(i < k)$ are contained in the symmetric variance-covariance matrix $\boldsymbol{\Sigma} = E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'$. Specifically,

$$E(\mathbf{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \boldsymbol{\mu} \tag{2-30}$$

and

$$\begin{aligned}
\boldsymbol{\Sigma} &= E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' \\
&= E\left(\begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_p - \mu_p \end{bmatrix} [X_1 - \mu_1, X_2 - \mu_2, \ldots, X_p - \mu_p] \right) \\
&= E\begin{bmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) & \cdots & (X_1 - \mu_1)(X_p - \mu_p) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 & \cdots & (X_2 - \mu_2)(X_p - \mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ (X_p - \mu_p)(X_1 - \mu_1) & (X_p - \mu_p)(X_2 - \mu_2) & \cdots & (X_p - \mu_p)^2 \end{bmatrix} \\
&= \begin{bmatrix} E(X_1 - \mu_1)^2 & E(X_1 - \mu_1)(X_2 - \mu_2) & \cdots & E(X_1 - \mu_1)(X_p - \mu_p) \\ E(X_2 - \mu_2)(X_1 - \mu_1) & E(X_2 - \mu_2)^2 & \cdots & E(X_2 - \mu_2)(X_p - \mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_p - \mu_p)(X_1 - \mu_1) & E(X_p - \mu_p)(X_2 - \mu_2) & \cdots & E(X_p - \mu_p)^2 \end{bmatrix}
\end{aligned}$$

or

$$\boldsymbol{\Sigma} = \text{Cov}(\mathbf{X}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix} \tag{2-31}$$

Example 2.13 (Computing the covariance matrix) Find the covariance matrix for the two random variables X_1 and X_2 introduced in Example 2.12 when their joint probability function $p_{12}(x_1, x_2)$ is represented by the entries in the body of the following table:

x_1 \ x_2	0	1	$p_1(x_1)$
-1	.24	.06	.3
0	.16	.14	.3
1	.40	.00	.4
$p_2(x_2)$	.8	.2	1

We have already shown that $\mu_1 = E(X_1) = .1$ and $\mu_2 = E(X_2) = .2$. (See Example 2.12.) In addition,

$$\sigma_{11} = E(X_1 - \mu_1)^2 = \sum_{\text{all } x_1} (x_1 - .1)^2 p_1(x_1)$$

$$= (-1 - .1)^2(.3) + (0 - .1)^2(.3) + (1 - .1)^2(.4) = .69$$

$$\sigma_{22} = E(X_2 - \mu_2)^2 = \sum_{\text{all } x_2} (x_2 - .2)^2 p_2(x_2)$$

$$= (0 - .2)^2(.8) + (1 - .2)^2(.2)$$

$$= .16$$

$$\sigma_{12} = E(X_1 - \mu_1)(X_2 - \mu_2) = \sum_{\text{all pairs } (x_1, x_2)} (x_1 - .1)(x_2 - .2)p_{12}(x_1, x_2)$$

$$= (-1 - .1)(0 - .2)(.24) + (-1 - .1)(1 - .2)(.06)$$

$$+ \cdots + (1 - .1)(1 - .2)(.00) = -.08$$

$$\sigma_{21} = E(X_2 - \mu_2)(X_1 - \mu_1) = E(X_1 - \mu_1)(X_2 - \mu_2) = \sigma_{12} = -.08$$

Consequently, with $\mathbf{X}' = [X_1, X_2]$,

$$\boldsymbol{\mu} = E(\mathbf{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} .1 \\ .2 \end{bmatrix}$$

and

$$\begin{aligned}
\boldsymbol{\Sigma} &= E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' \\
&= E\begin{bmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 \end{bmatrix} \\
&= \begin{bmatrix} E(X_1 - \mu_1)^2 & E(X_1 - \mu_1)(X_2 - \mu_2) \\ E(X_2 - \mu_2)(X_1 - \mu_1) & E(X_2 - \mu_2)^2 \end{bmatrix} \\
&= \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} .69 & -.08 \\ -.08 & .16 \end{bmatrix}
\end{aligned}$$

■

We note that the computation of means, variances, and covariances for *discrete* random variables involves summation (as in Examples 2.12 and 2.13), while analogous computations for *continuous* random variables involve integration.

Because $\sigma_{ik} = E(X_i - \mu_i)(X_k - \mu_k) = \sigma_{ki}$, it is convenient to write the matrix appearing in (2-31) as

$$\boldsymbol{\Sigma} = E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix} \tag{2-32}$$

We shall refer to $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as the *population mean* (vector) and *population variance–covariance* (matrix), respectively.

The multivariate normal distribution is completely specified once the mean vector $\boldsymbol{\mu}$ and variance–covariance matrix $\boldsymbol{\Sigma}$ are given (see Chapter 4), so it is not surprising that these quantities play an important role in many multivariate procedures.

It is frequently informative to separate the information contained in variances σ_{ii} from that contained in measures of association and, in particular, the measure of association known as the *population correlation coefficient* ρ_{ik}. The correlation coefficient ρ_{ik} is defined in terms of the covariance σ_{ik} and variances σ_{ii} and σ_{kk} as

$$\rho_{ik} = \frac{\sigma_{ik}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{kk}}} \tag{2-33}$$

The correlation coefficient measures the amount of *linear* association between the random variables X_i and X_k. (See, for example, [5].)

Let the population correlation matrix be the $p \times p$ symmetric matrix

$$\boldsymbol{\rho} = \begin{bmatrix} \dfrac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{11}}} & \dfrac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \cdots & \dfrac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} \\ \dfrac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \dfrac{\sigma_{22}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{22}}} & \cdots & \dfrac{\sigma_{2p}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{pp}}} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} & \dfrac{\sigma_{2p}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{pp}}} & \cdots & \dfrac{\sigma_{pp}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{pp}}} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix} \tag{2-34}$$

and let the $p \times p$ *standard deviation* matrix be

$$\mathbf{V}^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{\sigma_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\sigma_{pp}} \end{bmatrix} \tag{2-35}$$

Then it is easily verified (see Exercise 2.23) that

$$\mathbf{V}^{1/2}\boldsymbol{\rho}\mathbf{V}^{1/2} = \boldsymbol{\Sigma} \tag{2-36}$$

and

$$\boldsymbol{\rho} = (\mathbf{V}^{1/2})^{-1}\boldsymbol{\Sigma}(\mathbf{V}^{1/2})^{-1} \tag{2-37}$$

That is, $\boldsymbol{\Sigma}$ can be obtained from $\mathbf{V}^{1/2}$ and $\boldsymbol{\rho}$, whereas $\boldsymbol{\rho}$ can be obtained from $\boldsymbol{\Sigma}$. Moreover, the expression of these relationships in terms of matrix operations allows the calculations to be conveniently implemented on a computer.

Example 2.14 (Computing the correlation matrix from the covariance matrix) Suppose

$$\boldsymbol{\Sigma} = \begin{bmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 25 \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix}$$

Obtain $\mathbf{V}^{1/2}$ and $\boldsymbol{\rho}$.

Here

$$\mathbf{V}^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & 0 \\ 0 & \sqrt{\sigma_{22}} & 0 \\ 0 & 0 & \sqrt{\sigma_{33}} \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

and

$$(\mathbf{V}^{1/2})^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{5} \end{bmatrix}$$

Consequently, from (2-37), the correlation matrix $\boldsymbol{\rho}$ is given by

$$(\mathbf{V}^{1/2})^{-1}\boldsymbol{\Sigma}(\mathbf{V}^{1/2})^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{5} \end{bmatrix} \begin{bmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 25 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{5} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \frac{1}{6} & \frac{1}{5} \\ \frac{1}{6} & 1 & -\frac{1}{5} \\ \frac{1}{5} & -\frac{1}{5} & 1 \end{bmatrix}$$

■

Partitioning the Covariance Matrix

Often, the characteristics measured on individual trials will fall naturally into two or more groups. As examples, consider measurements of variables representing consumption and income or variables representing personality traits and physical characteristics. One approach to handling these situations is to let the characteristics defining the distinct groups be subsets of the *total* collection of characteristics. If the total collection is represented by a $(p \times 1)$-dimensional random vector $\mathbf{X}$, the subsets can be regarded as components of $\mathbf{X}$ and can be sorted by partitioning $\mathbf{X}$.

In general, we can partition the p characteristics contained in the $p \times 1$ random vector $\mathbf{X}$ into, for instance, two groups of size q and $p - q$, respectively. For example, we can write

$$\mathbf{X} = \begin{bmatrix} \left.\begin{matrix} X_1 \\ \vdots \\ X_q \end{matrix}\right\} q \\ \hdashline \left.\begin{matrix} X_{q+1} \\ \vdots \\ X_p \end{matrix}\right\} p - q \end{bmatrix} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \hdashline \mathbf{X}^{(2)} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\mu} = E(\mathbf{X}) = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_q \\ \hdashline \mu_{q+1} \\ \vdots \\ \mu_p \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}^{(1)} \\ \hdashline \boldsymbol{\mu}^{(2)} \end{bmatrix} \tag{2-38}$$

From the definitions of the transpose and matrix multiplication,

$$(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})'$$

$$= \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_q - \mu_q \end{bmatrix} [X_{q+1} - \mu_{q+1}, X_{q+2} - \mu_{q+2}, \dots, X_p - \mu_p]$$

$$= \begin{bmatrix} (X_1 - \mu_1)(X_{q+1} - \mu_{q+1}) & (X_1 - \mu_1)(X_{q+2} - \mu_{q+2}) & \cdots & (X_1 - \mu_1)(X_p - \mu_p) \\ (X_2 - \mu_2)(X_{q+1} - \mu_{q+1}) & (X_2 - \mu_2)(X_{q+2} - \mu_{q+2}) & \cdots & (X_2 - \mu_2)(X_p - \mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ (X_q - \mu_q)(X_{q+1} - \mu_{q+1}) & (X_q - \mu_q)(X_{q+2} - \mu_{q+2}) & \cdots & (X_q - \mu_q)(X_p - \mu_p) \end{bmatrix}$$

Upon taking the expectation of the matrix $(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})'$, we get

$$E(\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)})(\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)})' = \begin{bmatrix} \sigma_{1,q+1} & \sigma_{1,q+2} & \cdots & \sigma_{1p} \\ \sigma_{2,q+1} & \sigma_{2,q+2} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{q,q+1} & \sigma_{q,q+2} & \cdots & \sigma_{qp} \end{bmatrix} = \boldsymbol{\Sigma}_{12} \quad (2\text{-}39)$$

which gives all the covariances, σ_{ij}, $i = 1, 2, \dots, q$, $j = q + 1, q + 2, \dots, p$, between a component of $\mathbf{X}^{(1)}$ and a component of $\mathbf{X}^{(2)}$. Note that the matrix $\boldsymbol{\Sigma}_{12}$ is not necessarily symmetric or even square.

Making use of the partitioning in Equation (2-38), we can easily demonstrate that

$$(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'$$

$$= \begin{bmatrix} (\underset{(q\times 1)}{\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)}})(\underset{(1\times q)}{\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)}})' & (\underset{(q\times 1)}{\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)}})(\underset{(1\times(p-q))}{\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)}})' \\ (\underset{((p-q)\times 1)}{\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)}})(\underset{(1\times q)}{\mathbf{X}^{(1)} - \boldsymbol{\mu}^{(1)}})' & (\underset{((p-q)\times 1)}{\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)}})(\underset{(1\times(p-q))}{\mathbf{X}^{(2)} - \boldsymbol{\mu}^{(2)}})' \end{bmatrix}$$

and consequently,

$$\underset{(p\times p)}{\boldsymbol{\Sigma}} = E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' = \begin{array}{c} \\ q \\ p-q \end{array} \overset{\begin{array}{cc} q & p-q \end{array}}{\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}}_{(p\times p)}$$

$$= \left[\begin{array}{ccc|ccc} \sigma_{11} & \cdots & \sigma_{1q} & \sigma_{1,q+1} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{q1} & \cdots & \sigma_{qq} & \sigma_{q,q+1} & \cdots & \sigma_{qp} \\ \hline \sigma_{q+1,1} & \cdots & \sigma_{q+1,q} & \sigma_{q+1,q+1} & \cdots & \sigma_{q+1,p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pq} & \sigma_{p,q+1} & \cdots & \sigma_{pp} \end{array}\right] \quad (2\text{-}40)$$

Note that $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}'_{21}$. The covariance matrix of $\mathbf{X}^{(1)}$ is $\boldsymbol{\Sigma}_{11}$, that of $\mathbf{X}^{(2)}$ is $\boldsymbol{\Sigma}_{22}$, and that of elements from $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ is $\boldsymbol{\Sigma}_{12}$ (or $\boldsymbol{\Sigma}_{21}$).

It is sometimes convenient to use the Cov $(\mathbf{X}^{(1)}, \mathbf{X}^{(2)})$ notation where

$$\text{Cov}\,(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = \boldsymbol{\Sigma}_{12}$$

is a matrix containing all of the covariances between a component of $\mathbf{X}^{(1)}$ and a component of $\mathbf{X}^{(2)}$.

The Mean Vector and Covariance Matrix for Linear Combinations of Random Variables

Recall that if a single random variable, such as X_1, is multiplied by a constant c, then

$$E(cX_1) = cE(X_1) = c\mu_1$$

and

$$\text{Var}\,(cX_1) = E(cX_1 - c\mu_1)^2 = c^2\text{Var}\,(X_1) = c^2\sigma_{11}$$

If X_2 is a second random variable and a and b are constants, then, using additional properties of expectation, we get

$$\begin{aligned}
\text{Cov}(aX_1, bX_2) &= E(aX_1 - a\mu_1)(bX_2 - b\mu_2) \\
&= abE(X_1 - \mu_1)(X_2 - \mu_2) \\
&= ab\text{Cov}\,(X_1, X_2) = ab\sigma_{12}
\end{aligned}$$

Finally, for the linear combination $aX_1 + bX_2$, we have

$$\begin{aligned}
E(aX_1 + bX_2) &= aE(X_1) + bE(X_2) = a\mu_1 + b\mu_2 \\
\text{Var}\,(aX_1 + bX_2) &= E[(aX_1 + bX_2) - (a\mu_1 + b\mu_2)]^2 \\
&= E[a(X_1 - \mu_1) + b(X_2 - \mu_2)]^2 \\
&= E[a^2(X_1 - \mu_1)^2 + b^2(X_2 - \mu_2)^2 + 2ab(X_1 - \mu_1)(X_2 - \mu_2)] \\
&= a^2\text{Var}\,(X_1) + b^2\text{Var}\,(X_2) + 2ab\text{Cov}\,(X_1, X_2) \\
&= a^2\sigma_{11} + b^2\sigma_{22} + 2ab\sigma_{12} \qquad (2\text{-}41)
\end{aligned}$$

With $\mathbf{c}' = [a, b]$, $aX_1 + bX_2$ can be written as

$$[a \quad b]\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \mathbf{c}'\mathbf{X}$$

Similarly, $E(aX_1 + bX_2) = a\mu_1 + b\mu_2$ can be expressed as

$$[a \quad b]\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \mathbf{c}'\boldsymbol{\mu}$$

If we let

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

be the variance–covariance matrix of $\mathbf{X}$, Equation (2-41) becomes

$$\operatorname{Var}(aX_1 + bX_2) = \operatorname{Var}(\mathbf{c}'\mathbf{X}) = \mathbf{c}'\boldsymbol{\Sigma}\mathbf{c} \tag{2-42}$$

since

$$\mathbf{c}'\boldsymbol{\Sigma}\mathbf{c} = [a \quad b]\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = a^2\sigma_{11} + 2ab\,\sigma_{12} + b^2\sigma_{22}$$

The preceding results can be extended to a linear combination of p random variables:

The linear combination $\mathbf{c}'\mathbf{X} = c_1X_1 + \cdots + c_pX_p$ has

$$\begin{aligned} \text{mean} &= E(\mathbf{c}'\mathbf{X}) = \mathbf{c}'\boldsymbol{\mu} \\ \text{variance} &= \operatorname{Var}(\mathbf{c}'\mathbf{X}) = \mathbf{c}'\boldsymbol{\Sigma}\mathbf{c} \end{aligned} \tag{2-43}$$

where $\boldsymbol{\mu} = E(\mathbf{X})$ and $\boldsymbol{\Sigma} = \operatorname{Cov}(\mathbf{X})$.

In general, consider the q linear combinations of the p random variables $X_1, \ldots, X_p$:

$$\begin{aligned} Z_1 &= c_{11}X_1 + c_{12}X_2 + \cdots + c_{1p}X_p \\ Z_2 &= c_{21}X_1 + c_{22}X_2 + \cdots + c_{2p}X_p \\ &\vdots \\ Z_q &= c_{q1}X_1 + c_{q2}X_2 + \cdots + c_{qp}X_p \end{aligned}$$

or

$$\mathbf{Z} = \underset{(q\times 1)}{\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_q \end{bmatrix}} = \underset{(q\times p)}{\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{q1} & c_{q2} & \cdots & c_{qp} \end{bmatrix}} \underset{(p\times 1)}{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}} = \mathbf{CX} \tag{2-44}$$

The linear combinations $\mathbf{Z} = \mathbf{CX}$ have

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{Z}} &= E(\mathbf{Z}) = E(\mathbf{CX}) = \mathbf{C}\boldsymbol{\mu}_{\mathbf{X}} \\ \boldsymbol{\Sigma}_{\mathbf{Z}} &= \operatorname{Cov}(\mathbf{Z}) = \operatorname{Cov}(\mathbf{CX}) = \mathbf{C}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{C}' \end{aligned} \tag{2-45}$$

where $\boldsymbol{\mu}_{\mathbf{X}}$ and $\boldsymbol{\Sigma}_{\mathbf{X}}$ are the mean vector and variance-covariance matrix of $\mathbf{X}$, respectively. (See Exercise 2.28 for the computation of the off-diagonal terms in $\mathbf{C}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{C}'$.)

We shall rely heavily on the result in (2-45) in our discussions of principal components and factor analysis in Chapters 8 and 9.

Example 2.15 (Means and covariances of linear combinations) Let $\mathbf{X}' = [X_1, X_2]$ be a random vector with mean vector $\boldsymbol{\mu}'_{\mathbf{X}} = [\mu_1, \mu_2]$ and variance–covariance matrix

$$\boldsymbol{\Sigma}_{\mathbf{X}} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

Find the mean vector and covariance matrix for the linear combinations

$$Z_1 = X_1 - X_2$$
$$Z_2 = X_1 + X_2$$

or

$$\mathbf{Z} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \mathbf{CX}$$

in terms of $\boldsymbol{\mu}_{\mathbf{X}}$ and $\boldsymbol{\Sigma}_{\mathbf{X}}$.

Here

$$\boldsymbol{\mu}_{\mathbf{Z}} = E(\mathbf{Z}) = \mathbf{C}\boldsymbol{\mu}_{\mathbf{X}} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_1 + \mu_2 \end{bmatrix}$$

and

$$\boldsymbol{\Sigma}_{\mathbf{Z}} = \text{Cov}(\mathbf{Z}) = \mathbf{C}\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{C}' = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} - 2\sigma_{12} + \sigma_{22} & \sigma_{11} - \sigma_{22} \\ \sigma_{11} - \sigma_{22} & \sigma_{11} + 2\sigma_{12} + \sigma_{22} \end{bmatrix}$$

Note that if $\sigma_{11} = \sigma_{22}$—that is, if X_1 and X_2 have equal variances—the off-diagonal terms in $\boldsymbol{\Sigma}_{\mathbf{Z}}$ vanish. This demonstrates the well-known result that the sum and difference of two random variables with identical variances are uncorrelated. ■

Partitioning the Sample Mean Vector and Covariance Matrix

Many of the matrix results in this section have been expressed in terms of population means and variances (covariances). The results in (2-36), (2-37), (2-38), and (2-40) also hold if the population quantities are replaced by their appropriately defined sample counterparts.

Let $\bar{\mathbf{x}}' = [\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p]$ be the vector of sample averages constructed from n observations on p variables $X_1, X_2, \ldots, X_p$, and let

$$\mathbf{S}_n = \begin{bmatrix} s_{11} & \cdots & s_{1p} \\ \vdots & \ddots & \vdots \\ s_{1p} & \cdots & s_{pp} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{n}\sum_{j=1}^{n}(x_{j1} - \bar{x}_1)^2 & \cdots & \frac{1}{n}\sum_{j=1}^{n}(x_{j1} - \bar{x}_1)(x_{jp} - \bar{x}_p) \\ \vdots & \ddots & \vdots \\ \frac{1}{n}\sum_{j=1}^{n}(x_{j1} - \bar{x}_1)(x_{jp} - \bar{x}_p) & \cdots & \frac{1}{n}\sum_{j=1}^{n}(x_{jp} - \bar{x}_p)^2 \end{bmatrix}$$

be the corresponding sample variance–covariance matrix.

The sample mean vector and the covariance matrix can be partitioned in order to distinguish quantities corresponding to groups of variables. Thus,

$$\underset{(p\times 1)}{\bar{\mathbf{x}}} = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_q \\ \hdashline \bar{x}_{q+1} \\ \vdots \\ \bar{x}_p \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{x}}^{(1)} \\ \hdashline \bar{\mathbf{x}}^{(2)} \end{bmatrix} \tag{2-46}$$

and

$$\underset{(p\times p)}{\mathbf{S}_n} = \left[\begin{array}{ccc:ccc} s_{11} & \cdots & s_{1q} & s_{1,q+1} & \cdots & s_{1p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ s_{q1} & \cdots & s_{qq} & s_{q,q+1} & \cdots & s_{qp} \\ \hdashline s_{q+1,1} & \cdots & s_{q+1,q} & s_{q+1,q+1} & \cdots & s_{q+1,p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ s_{p1} & \cdots & s_{pq} & s_{p,q+1} & \cdots & s_{pp} \end{array}\right]$$

$$= \begin{array}{c} \\ q \\ p-q \end{array} \begin{array}{c} \begin{array}{cc} q & p-q \end{array} \\ \left[\begin{array}{c:c} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \hdashline \mathbf{S}_{21} & \mathbf{S}_{22} \end{array}\right] \end{array} \tag{2-47}$$

where $\bar{\mathbf{x}}^{(1)}$ and $\bar{\mathbf{x}}^{(2)}$ are the sample mean vectors constructed from observations $\mathbf{x}^{(1)} = [x_1, \ldots, x_q]'$ and $\mathbf{x}^{(2)} = [x_{q+1}, \ldots, x_p]'$, respectively; $\mathbf{S}_{11}$ is the sample covariance matrix computed from observations $\mathbf{x}^{(1)}$; $\mathbf{S}_{22}$ is the sample covariance matrix computed from observations $\mathbf{x}^{(2)}$; and $\mathbf{S}_{12} = \mathbf{S}_{21}'$ is the sample covariance matrix for elements of $\mathbf{x}^{(1)}$ and elements of $\mathbf{x}^{(2)}$.

2.7 Matrix Inequalities and Maximization

Maximization principles play an important role in several multivariate techniques. Linear discriminant analysis, for example, is concerned with allocating observations to predetermined groups. The allocation rule is often a linear function of measurements that *maximizes* the separation between groups relative to their within-group variability. As another example, principal components are linear combinations of measurements with *maximum* variability.

The matrix inequalities presented in this section will easily allow us to derive certain maximization results, which will be referenced in later chapters.

Cauchy-Schwarz Inequality. Let $\mathbf{b}$ and $\mathbf{d}$ be *any* two $p \times 1$ vectors. Then

$$(\mathbf{b}'\mathbf{d})^2 \le (\mathbf{b}'\mathbf{b})(\mathbf{d}'\mathbf{d}) \tag{2-48}$$

with equality if and only if $\mathbf{b} = c\mathbf{d}$ (or $\mathbf{d} = c\mathbf{b}$) for some constant c.

Proof. The inequality is obvious if either $\mathbf{b} = \mathbf{0}$ or $\mathbf{d} = \mathbf{0}$. Excluding this possibility, consider the vector $\mathbf{b} - x\mathbf{d}$, where x is an arbitrary scalar. Since the length of $\mathbf{b} - x\mathbf{d}$ is positive for $\mathbf{b} - x\mathbf{d} \neq \mathbf{0}$, in this case

$$0 < (\mathbf{b} - x\mathbf{d})'(\mathbf{b} - x\mathbf{d}) = \mathbf{b}'\mathbf{b} - x\mathbf{d}'\mathbf{b} - \mathbf{b}'(x\mathbf{d}) + x^2\mathbf{d}'\mathbf{d}$$
$$= \mathbf{b}'\mathbf{b} - 2x(\mathbf{b}'\mathbf{d}) + x^2(\mathbf{d}'\mathbf{d})$$

The last expression is quadratic in x. If we complete the square by adding and subtracting the scalar $(\mathbf{b}'\mathbf{d})^2/\mathbf{d}'\mathbf{d}$, we get

$$0 < \mathbf{b}'\mathbf{b} - \frac{(\mathbf{b}'\mathbf{d})^2}{\mathbf{d}'\mathbf{d}} + \frac{(\mathbf{b}'\mathbf{d})^2}{\mathbf{d}'\mathbf{d}} - 2x(\mathbf{b}'\mathbf{d}) + x^2(\mathbf{d}'\mathbf{d})$$
$$= \mathbf{b}'\mathbf{b} - \frac{(\mathbf{b}'\mathbf{d})^2}{\mathbf{d}'\mathbf{d}} + (\mathbf{d}'\mathbf{d})\left(x - \frac{\mathbf{b}'\mathbf{d}}{\mathbf{d}'\mathbf{d}}\right)^2$$

The term in brackets is zero if we choose $x = \mathbf{b}'\mathbf{d}/\mathbf{d}'\mathbf{d}$, so we conclude that

$$0 < \mathbf{b}'\mathbf{b} - \frac{(\mathbf{b}'\mathbf{d})^2}{\mathbf{d}'\mathbf{d}}$$

or $(\mathbf{b}'\mathbf{d})^2 < (\mathbf{b}'\mathbf{b})(\mathbf{d}'\mathbf{d})$ if $\mathbf{b} \neq x\mathbf{d}$ for some x.

Note that if $\mathbf{b} = c\mathbf{d}$, $0 = (\mathbf{b} - c\mathbf{d})'(\mathbf{b} - c\mathbf{d})$, and the same argument produces $(\mathbf{b}'\mathbf{d})^2 = (\mathbf{b}'\mathbf{b})(\mathbf{d}'\mathbf{d})$. ■

A simple, but important, extension of the Cauchy–Schwarz inequality follows directly.

Extended Cauchy–Schwarz Inequality. Let $\underset{(p\times 1)}{\mathbf{b}}$ and $\underset{(p\times 1)}{\mathbf{d}}$ be any two vectors, and let $\underset{(p\times p)}{\mathbf{B}}$ be a positive definite matrix. Then

$$(\mathbf{b}'\mathbf{d})^2 \leq (\mathbf{b}'\mathbf{B}\,\mathbf{b})(\mathbf{d}'\mathbf{B}^{-1}\mathbf{d}) \tag{2-49}$$

with equality if and only if $\mathbf{b} = c\mathbf{B}^{-1}\mathbf{d}$ (or $\mathbf{d} = c\mathbf{B}\mathbf{b}$) for some constant c.

Proof. The inequality is obvious when $\mathbf{b} = \mathbf{0}$ or $\mathbf{d} = \mathbf{0}$. For cases other than these, consider the square-root matrix $\mathbf{B}^{1/2}$ defined in terms of its eigenvalues λ_i and the normalized eigenvectors $\mathbf{e}_i$ as $\mathbf{B}^{1/2} = \sum_{i=1}^{p} \sqrt{\lambda_i}\,\mathbf{e}_i\mathbf{e}_i'$. If we set [see also (2-22)]

$$\mathbf{B}^{-1/2} = \sum_{i=1}^{p} \frac{1}{\sqrt{\lambda_i}}\,\mathbf{e}_i\mathbf{e}_i'$$

it follows that

$$\mathbf{b}'\mathbf{d} = \mathbf{b}'\mathbf{I}\mathbf{d} = \mathbf{b}'\mathbf{B}^{1/2}\mathbf{B}^{-1/2}\mathbf{d} = (\mathbf{B}^{1/2}\mathbf{b})'(\mathbf{B}^{-1/2}\mathbf{d})$$

and the proof is completed by applying the Cauchy–Schwarz inequality to the vectors $(\mathbf{B}^{1/2}\mathbf{b})$ and $(\mathbf{B}^{-1/2}\mathbf{d})$. ■

The extended Cauchy–Schwarz inequality gives rise to the following maximization result.

Maximization Lemma. Let $\underset{(p\times p)}{\mathbf{B}}$ be positive definite and $\underset{(p\times 1)}{\mathbf{d}}$ be a given vector. Then, for an arbitrary nonzero vector $\underset{(p\times 1)}{\mathbf{x}}$,

$$\max_{\mathbf{x}\neq \mathbf{0}} \frac{(\mathbf{x}'\mathbf{d})^2}{\mathbf{x}'\mathbf{B}\mathbf{x}} = \mathbf{d}'\mathbf{B}^{-1}\mathbf{d} \tag{2-50}$$

with the maximum attained when $\underset{(p\times 1)}{\mathbf{x}} = c\underset{(p\times p)(p\times 1)}{\mathbf{B}^{-1}\ \mathbf{d}}$ for any constant $c \neq 0$.

Proof. By the extended Cauchy-Schwarz inequality, $(\mathbf{x}'\mathbf{d})^2 \leq (\mathbf{x}'\mathbf{B}\mathbf{x})(\mathbf{d}'\mathbf{B}^{-1}\mathbf{d})$. Because $\mathbf{x} \neq \mathbf{0}$ and $\mathbf{B}$ is positive definite, $\mathbf{x}'\mathbf{B}\mathbf{x} > 0$. Dividing both sides of the inequality by the positive scalar $\mathbf{x}'\mathbf{B}\mathbf{x}$ yields the upper bound

$$\frac{(\mathbf{x}'\mathbf{d})^2}{\mathbf{x}'\mathbf{B}\mathbf{x}} \leq \mathbf{d}'\mathbf{B}^{-1}\mathbf{d}$$

Taking the maximum over $\mathbf{x}$ gives Equation (2-50) because the bound is attained for $\mathbf{x} = c\mathbf{B}^{-1}\mathbf{d}$. ■

A final maximization result will provide us with an interpretation of eigenvalues.

Maximization of Quadratic Forms for Points on the Unit Sphere. Let $\underset{(p\times p)}{\mathbf{B}}$ be a positive definite matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ and associated normalized eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p$. Then

$$\begin{aligned} \max_{\mathbf{x}\neq \mathbf{0}} \frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} &= \lambda_1 \qquad (\text{attained when } \mathbf{x} = \mathbf{e}_1) \\ \min_{\mathbf{x}\neq \mathbf{0}} \frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} &= \lambda_p \qquad (\text{attained when } \mathbf{x} = \mathbf{e}_p) \end{aligned} \tag{2-51}$$

Moreover,

$$\max_{\mathbf{x}\perp \mathbf{e}_1,\ldots,\mathbf{e}_k} \frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \lambda_{k+1} \qquad (\text{attained when } \mathbf{x} = \mathbf{e}_{k+1}, k = 1, 2, \ldots, p-1) \tag{2-52}$$

where the symbol $\perp$ is read "is perpendicular to."

Proof. Let $\underset{(p\times p)}{\mathbf{P}}$ be the orthogonal matrix whose columns are the eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p$ and $\mathbf{\Lambda}$ be the diagonal matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$ along the main diagonal. Let $\mathbf{B}^{1/2} = \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}'$ [see (2-22)] and $\underset{(p\times 1)}{\mathbf{y}} = \underset{(p\times p)(p\times 1)}{\mathbf{P}'\ \mathbf{x}}$.

Consequently, $\mathbf{x} \neq \mathbf{0}$ implies $\mathbf{y} \neq \mathbf{0}$. Thus,

$$\frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \frac{\mathbf{x}'\mathbf{B}^{1/2}\mathbf{B}^{1/2}\mathbf{x}}{\mathbf{x}'\underbrace{\mathbf{P}\mathbf{P}'}_{\underset{(p\times p)}{\mathbf{I}}}\mathbf{x}} = \frac{\mathbf{x}'\mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}'\mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}'\mathbf{x}}{\mathbf{y}'\mathbf{y}} = \frac{\mathbf{y}'\mathbf{\Lambda}\mathbf{y}}{\mathbf{y}'\mathbf{y}}$$

$$= \frac{\sum_{i=1}^{p} \lambda_i y_i^2}{\sum_{i=1}^{p} y_i^2} \leq \lambda_1 \frac{\sum_{i=1}^{p} y_i^2}{\sum_{i=1}^{p} y_i^2} = \lambda_1 \tag{2-53}$$