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,

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

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

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

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

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

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

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

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

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

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,

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

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,

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

and

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

$$= E\left(\begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_p - \mu_p \end{bmatrix} [X_1 - \mu_1, X_2 - \mu_2, \ldots, X_p - \mu_p]\right)$$

$$= E\begin{bmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) & \cdots & (X_1 - \mu_1)(X_p - \mu_p) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 & \cdots & (X_2 - \mu_2)(X_p - \mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ (X_p - \mu_p)(X_1 - \mu_1) & (X_p - \mu_p)(X_2 - \mu_2) & \cdots & (X_p - \mu_p)^2 \end{bmatrix}$$

$$= \begin{bmatrix} E(X_1 - \mu_1)^2 & E(X_1 - \mu_1)(X_2 - \mu_2) & \cdots & E(X_1 - \mu_1)(X_p - \mu_p) \\ E(X_2 - \mu_2)(X_1 - \mu_1) & E(X_2 - \mu_2)^2 & \cdots & E(X_2 - \mu_2)(X_p - \mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_p - \mu_p)(X_1 - \mu_1) & E(X_p - \mu_p)(X_2 - \mu_2) & \cdots & E(X_p - \mu_p)^2 \end{bmatrix}$$

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

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

$$= E\begin{bmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 \end{bmatrix}$$

$$= \begin{bmatrix} E(X_1 - \mu_1)^2 & E(X_1 - \mu_1)(X_2 - \mu_2) \\ E(X_2 - \mu_2)(X_1 - \mu_1) & E(X_2 - \mu_2)^2 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} .69 & -.08 \\ -.08 & .16 \end{bmatrix}$$

■

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

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

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

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

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

From the definitions of the transpose and matrix multiplication,

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

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

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

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}, $i = 1, 2, \ldots, q$, $j = q + 1, q + 2, \ldots, p$, 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,

$$\underset{(p\times p)}{\boldsymbol{\Sigma}} = E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' = \begin{matrix} & q & p-q \\ q & \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ p-q & \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{matrix} \quad (p\times p)$$

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

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

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

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

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