3 Random Vectors and Matrices

3.1 INTODUCTION

As we work with linear models, it is often convenient to express the observed data (or data that will be observed) in the form of a vector or matrix. A *random vector* or *random matrix* is a vector or matrix whose elements are random variables. Informally, a *random variable* is defined as a variable whose value depends on the outcome of a chance experiment. (Formally, a random variable is a function defined for each element of a sample space.)

In terms of experimental structure, we can distinguish two kinds of random vectors:

1. A vector containing a measurement on each of n different individuals or experimental units. In this case, where the same variable is observed on each of n units selected at random, the n random variables $y_1, y_2, \ldots, y_n$ in the vector are typically uncorrelated and have the same variance.
2. A vector consisting of p different measurements on one individual or experimental unit. The p random variables thus obtained are typically correlated and have different variances.

To illustrate the first type of random vector, consider the multiple regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1, 2, \ldots, n,$$

as given in (1.2). In Chapters 7–9, we treat the x variables as constants, in which case we have two random vectors:

$$\mathrm{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \text{and} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}. \tag{3.1}$$

Linear Models in Statistics, *Second Edition*, by Alvin C. Rencher and G. Bruce Schaalje

The y_i values are observable, but the ε_i's are not observable unless the β's are known.

To illustrate the second type of random vector, consider regression of y on several random x variables (this regression case is discussed in Chapter 10). For the ith individual in the sample, we observe the $k+1$ random variables $y_i, x_{i1}, x_{i2}, \ldots, x_{ik}$, which constitute the random vector $(y_i, x_{i1}, \ldots, x_{ik})'$. In some cases, the $k+1$ variables $y_i, x_{i1}, \ldots, x_{ik}$ are all measured using the same units or scale of measurement, but typically the scales differ.

3.2 MEANS, VARIANCES, COVARIANCES, AND CORRELATIONS

In this section, we review some properties of univariate and bivariate random variables. We begin with a univariate random variable y. We do not distinguish notationally between the random variable y and an observed value of y. In many texts, an uppercase letter is used for the random variable and the corresponding lowercase letter represents a realization of the random variable, as in the expression $P(Y \leq y)$. This practice is convenient in a univariate context but would be confusing in the present text where we use uppercase letters for matrices and lowercase letters for vectors.

If $f(y)$ is the *density* of the random variable y, the *mean* or *expected value* of y is defined as

$$\mu = E(y) = \int_{-\infty}^{\infty} y f(y)\, dy. \tag{3.2}$$

This is the population mean. Later (beginning in Chapter 5), we also use the sample mean of y, obtained from a random sample of n observed values of y.

The expected value of a function of y such as y^2 can be found directly without first finding the density of y^2. In general, for a function $u(y)$, we have

$$E[u(y)] = \int_{-\infty}^{\infty} u(y) f(y)\, dy. \tag{3.3}$$

For a constant a and functions $u(y)$ and $v(y)$, it follows from (3.3) that

$$E(ay) = aE(y), \tag{3.4}$$

$$E[u(y) + v(y)] = E[u(y)] + E[v(y)]. \tag{3.5}$$

The *variance* of a random variable y is defined as

$$\sigma^2 = \mathrm{var}(y) = E(y - \mu)^2, \tag{3.6}$$

This is the population variance. Later (beginning in Chapter 5), we also use the sample variance of y, obtained from a random sample of n observed values of y. The square root of the variance is known as the *standard deviation*:

$$\sigma = \sqrt{\mathrm{var}(y)} = \sqrt{E(y-\mu)^2}. \tag{3.7}$$

Using (3.4) and (3.5), we can express the variance of y in the form

$$\sigma^2 = \mathrm{var}(y) = E(y^2) - \mu^2. \tag{3.8}$$

If a is a constant, we can use (3.4) and (3.6) to show that

$$\mathrm{var}(ay) = a^2\mathrm{var}(y) = a^2\sigma^2. \tag{3.9}$$

For any two variables y_i and y_j in the random vector $\mathbf{y}$ in (3.1), we define the *covariance* as

$$\sigma_{ij} = \mathrm{cov}(y_i, y_j) = E[(y_i - \mu_i)(y_j - \mu_j)], \tag{3.10}$$

where $\mu_i = E(y_i)$ and $\mu_j = E(y_j)$. Using (3.4) and (3.5), we can express σ_{ij} in the form

$$\sigma_{ij} = \mathrm{cov}(y_i, y_j) = E(y_i y_j) - \mu_i\mu_j. \tag{3.11}$$

Two random variables y_i and y_j are said to be *independent* if their joint density factors into the product of their marginal densities

$$f(y_i, y_j) = f_i(y_i)\,f_j(y_j), \tag{3.12}$$

where the marginal density $f_i(y_i)$ is defined as

$$f_i(y_i) = \int_{-\infty}^{\infty} f(y_i, y_j)\,dy_j. \tag{3.13}$$

From the definition of independence in (3.12), we obtain the following properties:

1. $E(y_i, y_j) = E(y_i)E(y_j)$ if y_i and y_j are independent. (3.14)

2. $\sigma_{ij} = \mathrm{cov}(y_i, y_j) = 0$ if y_i and y_j are independent. (3.15)

The second property follows from the first.

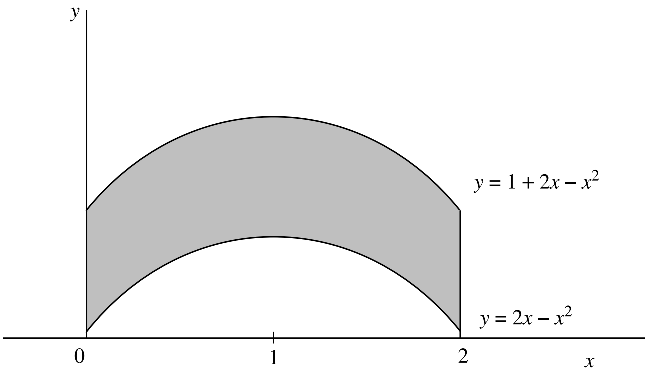

Figure 3.1 Region for $f(x, y)$ in Example 3.2.

In the first type of random vector defined in Section 3.1, the variables $y_1, y_2, \ldots, y_n$ would typically be independent if obtained from a random sample, and we would thus have $\sigma_{ij} = 0$ for all $i \neq j$. However, for the variables in the second type of random vector, we would typically have $\sigma_{ij} \neq 0$ for at least some values of i and j.

The converse of the property in (3.15) is not true; that is, $\sigma_{ij} = 0$ does not imply independence. This is illustrated in the following example.

Example 3.2. Suppose that the bivariate random variable (x, y) is distributed uniformly over the region $0 \leq x \leq 2$, $2x - x^2 \leq y \leq 1 + 2x - x^2$; see Figure 3.1.

The area of the region is given by

$$\text{Area} = \int_0^2 \int_{2x-x^2}^{1+2x-x^2} dy\, dx = 2.$$

Hence, for a uniform distribution over the region, we set

$$f(x, y) = \tfrac{1}{2}, \quad 0 \leq x \leq 2, \quad 2x - x^2 \leq y \leq 1 + 2x - x^2,$$

so that $\int\int f(x, y) dx\, dy = 1$.

To find σ_{xy} using (3.11), we need $E(xy)$, $E(x)$, and $E(y)$. The first of these is given by

$$\begin{aligned} E(xy) &= \int_0^2 \int_{2x-x^2}^{1+2x-x^2} xy\left(\tfrac{1}{2}\right) dy\, dx \\ &= \int_0^2 \frac{x}{4}(1 + 4x - 2x^2) dx = \frac{7}{6}. \end{aligned}$$

To find $E(x)$ and $E(y)$, we first find the marginal distributions of x and y. For $f_1(x)$, we have, by (3.13),

$$f_1(x) = \int_{2x-x^2}^{1+2x-x^2} \tfrac{1}{2}\,dy = \tfrac{1}{2}, \quad 0 \le x \le 2.$$

For $f_2(y)$, we obtain different results for $0 \le y \le 1$ and $1 \le y \le 2$:

$$f_2(y) = \int_0^{1-\sqrt{1-y}} \tfrac{1}{2}\,dx + \int_{1+\sqrt{1-y}}^{2} \tfrac{1}{2}\,dx = 1 - \sqrt{1-y}, \quad 0 \le y \le 1, \tag{3.16}$$

$$f_2(y) = \int_{1-\sqrt{2-y}}^{1+\sqrt{2-y}} \tfrac{1}{2}\,dx = \sqrt{2-y}, \quad 1 \le y \le 2. \tag{3.17}$$

Then

$$\begin{aligned} E(x) &= \int_0^2 x\left(\tfrac{1}{2}\right)dx = 1, \\ E(y) &= \int_0^1 y(1-\sqrt{1-y})dy + \int_1^2 y\sqrt{2-y} \quad dy = \tfrac{7}{6}. \end{aligned}$$

Now by (3.11), we obtain

$$\begin{aligned} \sigma_{xy} &= E(xy) - E(x)E(y) \\ &= \tfrac{7}{6} - (1)\left(\tfrac{7}{6}\right) = 0. \end{aligned}$$

However, x and y are clearly dependent since the range of y for each x depends on the value of x.

As a further indication of the dependence of y on x, we examine $E(y|x)$, the expected value of y for a given value of x, which is found as

$$E(y|x) = \int y f(y|x)dy.$$

The conditional density $f(y|x)$ is defined as

$$f(y|x) = \frac{f(x,y)}{f_1(x)}, \tag{3.18}$$

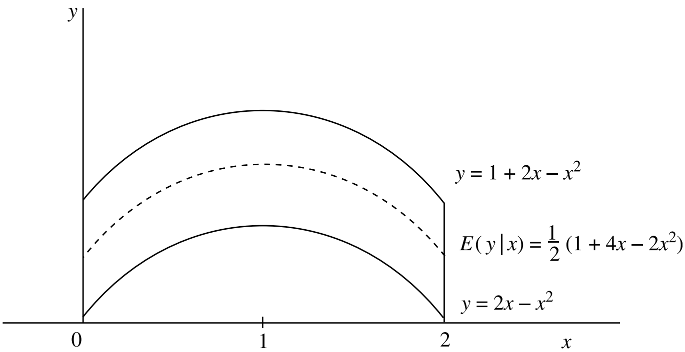

Figure 3.2 $E(y|x)$ in Example 3.2.

which becomes

$$f(y|x) = \frac{\frac{1}{2}}{\frac{1}{2}} = 1, \quad 2x - x^2 \leq y \leq 1 + 2x - x^2.$$

Thus

$$E(y|x) = \int_{2x-x^2}^{1+2x-x^2} y(1)dy$$
$$= \tfrac{1}{2}(1 + 4x - 2x^2).$$

Since $E(y|x)$ depends on x, the two variables are dependent. Note that $E(y|x) = \frac{1}{2}(1 + 4x - 2x^2)$ is the average of the two curves $y = 2x - x^2$ and $y = 1 + 2x - x^2$. This is illustrated in Figure 3.2. □

In Example 3.2 we have two dependent random variables x and y for which $\sigma_{xy} = 0$. In cases such as this, σ_{xy} is not a good measure of relationship. However, if x and y have a bivariate normal distribution (see Section 4.2), then $\sigma_{xy} = 0$ implies independence of x and y (see Corollary 1 to Theorem 4.4c). In the bivariate normal case, $E(y|x)$ is a linear function of x (see Theorem 4.4d), and curves such as $E(y|x) = \frac{1}{2}(1 + 4x - 2x^2)$ do not occur.

The covariance σ_{ij} as defined in (3.10) depends on the scale of measurement of both y_i and y_j. To standardize σ_{ij}, we divide it by (the product of) the standard deviations of y_i and y_j to obtain the *correlation*:

$$\rho_{ij} = \text{corr}(y_i, y_j) = \frac{\sigma_{ij}}{\sigma_i \sigma_j}. \tag{3.19}$$

3.3 MEAN VECTORS AND COVARIANCE MATRICES FOR RANDOM VECTORS

3.3.1 Mean Vectors

The expected value of a $p \times 1$ random vector $\mathbf{y}$ is defined as the vector of expected values of the p random variables $y_1, y_2, \ldots, y_p$ in $\mathbf{y}$:

$$E(\mathbf{y}) = E\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix} = \begin{pmatrix} E(y_1) \\ E(y_2) \\ \vdots \\ E(y_p) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} = \boldsymbol{\mu}, \tag{3.20}$$

where $E(y_i) = \mu_i$ is obtained as $E(y_i) = \int y_i f_i(y_i)\, dy_i$, using $f_i(y_i)$, the marginal density of y_i.

If $\mathbf{x}$ and $\mathbf{y}$ are $p \times 1$ random vectors, it follows from (3.20) and (3.5) that the expected value of their sum is the sum of their expected values:

$$E(\mathbf{x} + \mathbf{y}) = E(\mathbf{x}) + E(\mathbf{y}). \tag{3.21}$$

3.3.2 Covariance Matrix

The variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2$ of $y_1, y_2, \ldots, y_p$ and the covariances σ_{ij} for all $i \neq j$ can be conveniently displayed in the *covariance matrix*, which is denoted by $\boldsymbol{\Sigma}$, the uppercase version of σ_{ij}:

$$\boldsymbol{\Sigma} = \text{cov}(\mathbf{y}) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix}. \tag{3.22}$$

The ith row of $\boldsymbol{\Sigma}$ contains the variance of y_i and the covariance of y_i with each of the other y variables. To be consistent with the notation σ_{ij}, we have used $\sigma_{ii} = \sigma_i^2$, $i = 1, 2, \ldots, p$, for the variances. The variances are on the diagonal of $\boldsymbol{\Sigma}$, and the covariances occupy off-diagonal positions. There is a distinction in the font used for $\boldsymbol{\Sigma}$ as the covariance matrix and $\sum$ as the summation symbol. Note also the distinction in meaning between the notation $\text{cov}(\mathbf{y}) = \boldsymbol{\Sigma}$ and $\text{cov}(y_i, y_j) = \sigma_{ij}$.

The covariance matrix $\boldsymbol{\Sigma}$ is symmetric because $\sigma_{ij} = \sigma_{ji}$ [see (3.10)]. In many applications, $\boldsymbol{\Sigma}$ is assumed to be positive definite. This will ordinarily hold if the y variables are continuous random variables and if there are no linear relationships among them. (If there are linear relationships among the y variables, $\boldsymbol{\Sigma}$ will be positive semidefinite.)

By analogy with (3.20), we define the expected value of a random matrix $\mathbf{Z}$ as the matrix of expected values:

$$E(\mathbf{Z}) = E\begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & z_{2p} \\ \vdots & \vdots & & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{np} \end{pmatrix} = \begin{pmatrix} E(z_{11}) & E(z_{12}) & \cdots & E(z_{1p}) \\ E(z_{21}) & E(z_{22}) & \cdots & E(z_{2p}) \\ \vdots & \vdots & & \vdots \\ E(z_{n1}) & E(z_{n2}) & \cdots & E(z_{np}) \end{pmatrix}. \tag{3.23}$$

We can express $\boldsymbol{\Sigma}$ as the expected value of a random matrix. By (2.21), the (ij)th element of the matrix $(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})'$ is $(y_i-\mu_i)(y_j-\mu_j)$. Thus, by (3.10) and (3.23), the (ij)th element of $E\,[(\mathbf{y}-\boldsymbol{\mu})\,(\mathbf{y}-\boldsymbol{\mu})']$ is $E[\,(y_i-\mu_i)\,(y_j-\mu_j)] = \sigma_{ij}$. Hence

$$E[(\boldsymbol{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu}')] = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix} = \boldsymbol{\Sigma}. \tag{3.24}$$

We illustrate (3.24) for $p = 3$:

$$\begin{aligned}
\boldsymbol{\Sigma} &= E[(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'] \\
&= E\left[\begin{pmatrix} y_1 - \mu_1 \\ y_2 - \mu_2 \\ y_3 - \mu_3 \end{pmatrix}(y_1 - \mu_1, y_2 - \mu_2, y_3 - y_3)\right] \\
&= E\begin{bmatrix} (y_1 - \mu_1)^2 & (y_1 - \mu_1)(y_2 - \mu_2) & (y_1 - \mu_1)(y_3 - \mu_3) \\ (y_2 - \mu_2)(y_1 - \mu_1) & (y_2 - \mu_2)^2 & (y_2 - \mu_2)(y_3 - \mu_3) \\ (y_3 - \mu_3)(y_1 - \mu_1) & (y_3 - \mu_3)(y_2 - \mu_2) & (y_3 - \mu_3)^2 \end{bmatrix} \\
&= \begin{bmatrix} E(y_1 - \mu_1)^2 & E[(y_1 - \mu_1)(y_2 - \mu_2)] & E[(y_1 - \mu_1)(y_3 - \mu_3)] \\ E[(y_2 - \mu_2)(y_1 - \mu_1)] & E(y_2 - \mu_2)^2 & E[(y_2 - \mu_2)(y_3 - \mu_3)] \\ E[(y_3 - \mu_3)(y_1 - \mu_1)] & E[(y_3 - \mu_3)(y_2 - \mu_2)] & E(y_3 - \mu_3)^2 \end{bmatrix} \\
&= \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{pmatrix}.
\end{aligned}$$

We can write (3.24) in the form

$$\boldsymbol{\Sigma} = E[(- \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'] = E(\mathbf{y}\mathbf{y}') - \boldsymbol{\mu}\boldsymbol{\mu}', \tag{3.25}$$

which is analogous to (3.8) and (3.11).

3.3.3 Generalized Variance

A measure of overall variability in the population of $\mathbf{y}$ variables can be defined as the determinant of $\boldsymbol{\Sigma}$:

$$\text{Generalized variance} = |\boldsymbol{\Sigma}|. \tag{3.26}$$

If $|\boldsymbol{\Sigma}|$ is small, the $\mathbf{y}$ variables are concentrated closer to $\boldsymbol{\mu}$ than if $|\boldsymbol{\Sigma}|$ is large. A small value of $|\boldsymbol{\Sigma}|$ may also indicate that the variables $y_1, y_2, \ldots, y_p$ in $\mathbf{y}$ are highly intercorrelated, in which case the $\mathbf{y}$ variables tend to occupy a subspace of the p dimensions [this corresponds to one or more small eigenvalues; see Rencher (1998, Section 2.1.3)].

3.3.4 Standardized Distance

To obtain a useful measure of distance between $\mathbf{y}$ and $\boldsymbol{\mu}$, we need to account for the variances and covariances of the y_i variables in $\mathbf{y}$. By analogy with the univariate standardized variable $(y - \mu)/\sigma$, which has mean 0 and variance 1, the *standardized distance* is defined as

$$\text{Standardized distance} = (\mathbf{y} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu}). \tag{3.27}$$

The use of $\boldsymbol{\Sigma}^{-1}$ standardizes the (transformed) y_i variables so that they have means equal to 0 and variances equal to 1 and are also uncorrelated (see Problem 3.11). A distance such as (3.27) is often called a *Mahalanobis distance* (Mahalanobis 1936).

3.4 CORRELATION MATRICES

By analogy with $\boldsymbol{\Sigma}$ in (3.22), the *correlation matrix* is defined as

$$\mathbf{P}_\rho = (\rho_{ij}) = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{pmatrix}, \tag{3.28}$$

where $\rho_{ij} = \sigma_{ij}/\sigma_i\sigma_j$ is the correlation of y_i and y_j defined in (3.19). The second row of $\mathbf{P}_\rho$, for example, contains the correlation of y_2 with each of the other y variables. We use the subscript ρ in $\mathbf{P}_\rho$ to emphasize that $\mathbf{P}$ is the uppercase version of ρ.

If we define

$$\mathbf{D}_\sigma = [\text{diag}(\boldsymbol{\Sigma})]^{1/2} = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_p), \tag{3.29}$$

then by (2.31), we can obtain $\mathbf{P}_\rho$ from $\boldsymbol{\Sigma}$ and vice versa:

$$\mathbf{P}_\rho = \mathbf{D}_\sigma^{-1}\boldsymbol{\Sigma}\mathbf{D}_\sigma^{-1}, \tag{3.30}$$

$$\boldsymbol{\Sigma} = \mathbf{D}_\sigma\mathbf{P}_\rho\mathbf{D}_\sigma. \tag{3.31}$$

3.5 MEAN VECTORS AND COVARIANCE MATRICES FOR PARTITIONED RANDOM VECTORS

Suppose that the random vector $\mathbf{v}$ is partitioned into two subsets of variables, which we denote by $\mathbf{y}$ and $\mathbf{x}$:

$$\mathbf{v} = \begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_p \\ x_1 \\ \vdots \\ x_q \end{pmatrix}.$$

Thus there are $p + q$ random variables in $\mathbf{v}$.

The mean vector and covariance matrix for $\mathbf{v}$ partitioned as above can be expressed in the following form

$$\boldsymbol{\mu} = E(\mathbf{v}) = E\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} = \begin{pmatrix} E(\mathbf{y}) \\ E(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{pmatrix}, \tag{3.32}$$

$$\boldsymbol{\Sigma} = \operatorname{cov}(\mathbf{v}) = \operatorname{cov}\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{pmatrix}, \tag{3.33}$$

where $\boldsymbol{\Sigma}_{xy} = \boldsymbol{\Sigma}_{yx}'$. In (3.32), the submatrix $\boldsymbol{\mu}_y = [E(y_1), E(y_2), \ldots, E(y_p)]'$ contains the means of $y_1, y_2, \ldots, y_p$. Similarly $\boldsymbol{\mu}_x$ contains the means of the x variables. In (3.33), the submatrix $\boldsymbol{\Sigma}_{yy} = \operatorname{cov}(\mathbf{y})$ is a $p \times p$ covariance matrix for $\mathbf{y}$ containing the variances of y_1, $y_2, \ldots,$ y_p on the diagonal and the covariance of each y_i with each y_j $(i \neq j)$ off the diagonal:

$$\boldsymbol{\Sigma}_{yy} = \begin{pmatrix} \sigma_{y_1}^2 & \sigma_{y_1y_2} & \cdots & \sigma_{y_1y_p} \\ \sigma_{y_2y_1} & \sigma_{y_2}^2 & \cdots & \sigma_{y_2y_p} \\ \vdots & \vdots & & \vdots \\ \sigma_{y_py_1} & \sigma_{y_py_2} & \cdots & \sigma_{y_p}^2 \end{pmatrix}.$$

Similarly, $\boldsymbol{\Sigma}_{xx} = \text{cov}(\mathbf{x})$ is the $q \times q$ covariance matrix of $x_1, x_2, \ldots, x_q$. The matrix $\boldsymbol{\Sigma}_{yx}$ in (3.33) is $p \times q$ and contains the covariance of each y_i with each x_j:

$$\boldsymbol{\Sigma}_{yx} = \begin{pmatrix} \sigma_{y_1x_1} & \sigma_{y_1x_2} & \cdots & \sigma_{y_1x_q} \\ \sigma_{y_2x_1} & \sigma_{y_2x_2} & \cdots & \sigma_{y_2x_q} \\ \vdots & \vdots & & \vdots \\ \sigma_{y_px_1} & \sigma_{y_px_2} & \cdots & \sigma_{y_px_q} \end{pmatrix}.$$

Thus $\boldsymbol{\Sigma}_{yx}$ is rectangular unless $p = q$. The covariance matrix $\boldsymbol{\Sigma}_{yx}$ is also denoted by cov($\mathbf{y}$, $\mathbf{x}$) and can be defined as

$$\boldsymbol{\Sigma}_{yx} = \text{cov}(\mathbf{y}, \mathbf{x}) = E[(\mathbf{y} - \boldsymbol{\mu}_y)(\mathbf{x} - \boldsymbol{\mu}_x)']. \tag{3.34}$$

Note the difference in meaning between $\text{cov}\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix}$ in (3.33) and $\text{cov}(\mathbf{y}, \mathbf{x}) = \boldsymbol{\Sigma}_{yx}$ in (3.34). We have now used the notation cov in three ways: (1) cov(y_i, y_j), (2) cov($\mathbf{y}$), and (3) cov($\mathbf{y}$, $\mathbf{x}$). The first of these is a scalar, the second is a symmetric (usually positive definite) matrix, and the third is a rectangular matrix.

3.6 LINEAR FUNCTIONS OF RANDOM VECTORS

We often use linear combinations of the variables $y_1, y_2, \ldots, y_p$ from a random vector $\mathbf{y}$. Let $\mathbf{a} = (a_1, a_2, \ldots, a_p)'$ be a vector of constants. Then, by an expression preceding (2.18), the linear combination using the a terms as coefficients can be written as

$$z = a_1y_1 + a_2y_2 + \cdots + a_py_p = \mathbf{a}'\mathbf{y}. \tag{3.35}$$

We consider the means, variances, and covariances of such linear combinations in Sections 3.6.1 and 3.6.2.

3.6.1 Means

Since $\mathbf{y}$ is a random vector, the linear combination $z = \mathbf{a}'\mathbf{y}$ is a (univariate) random variable. The mean of $\mathbf{a}'\mathbf{y}$ is given the following theorem.

Theorem 3.6a. If $\mathbf{a}$ is a $p \times 1$ vector of constants and $\mathbf{y}$ is a $p \times 1$ random vector with mean vector $\boldsymbol{\mu}$, then

$$\mu_z = E(\mathbf{a}'\mathbf{y}) = \mathbf{a}'E(\mathbf{y}) = \mathbf{a}'\boldsymbol{\mu}. \tag{3.36}$$

PROOF. Using (3.4), (3.5), and (3.35), we obtain

$$\begin{aligned} E(\mathbf{a}'\mathbf{y}) &= E(a_1 y_1 + a_2 y_2 + \cdots + a_p y_p) \\ &= E(a_1 y_1) + E(a_2 y_2) + \cdots + E(a_p y_p) \\ &= a_1 E(y_1) + a_2 E(y_2) + \cdots + a_p E(y_p) \\ &= (a_1, a_2, \ldots, a_p) \begin{pmatrix} E(y_1) \\ E(y_2) \\ \vdots \\ E(y_p) \end{pmatrix} \\ &= \mathbf{a}'E(\mathbf{y}) = \mathbf{a}'\boldsymbol{\mu}. \end{aligned}$$

□

Suppose that we have several linear combinations of $\mathbf{y}$ with constant coefficients:

$$\begin{aligned} z_1 &= a_{11}y_1 + a_{12}y_2 + \ldots + a_{1p}y_p = \mathbf{a}_1'\mathbf{y} \\ z_2 &= a_{21}y_1 + a_{22}y_2 + \ldots + a_{2p}y_p = \mathbf{a}_2'\mathbf{y} \\ &\vdots \qquad\qquad\qquad\qquad\qquad\qquad \vdots \\ z_k &= a_{k1}y_1 + a_{k2}y_2 + \ldots + a_{kp}y_p = \mathbf{a}_k'\mathbf{y}, \end{aligned}$$

where $\mathbf{a}_i' = (a_{i1}, a_{i2}, \ldots, a_{ip})$ and $\mathbf{y} = (y_1, y_2, \ldots, y_p)'$. These k linear functions can be written in the form

$$\mathbf{z} = \mathbf{A}\mathbf{y}, \tag{3.37}$$

where

$$\mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_k \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \\ \vdots \\ \mathbf{a}_k' \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \dots & a_{kp} \end{pmatrix}.$$

It is possible to have $k > p$, but we typically have $k \leq p$ with the rows of $\mathbf{A}$ linearly independent, so that $\mathbf{A}$ is full-rank. Since $\mathbf{y}$ is a random vector, each $z_i = \mathbf{a}_i'\mathbf{y}$ is a random variable and $\mathbf{z} = (z_1, z_2, \dots, z_k)'$ is a random vector. The expected value of $\mathbf{z} = \mathbf{Ay}$ is given in the following theorem, as well as some extensions.

Theorem 3.6b. Suppose that $\mathbf{y}$ is a random vector, $\mathbf{X}$ is a random matrix, $\mathbf{a}$ and $\mathbf{b}$ are vectors of constants, and $\mathbf{A}$ and $\mathbf{B}$ are matrices of constants. Then, assuming the matrices and vectors in each product are conformal, we have the following expected values:

(i) $E(\mathbf{Ay}) = \mathbf{A}E(\mathbf{y})$. (3.38)

(ii) $E(\mathbf{a}'\mathbf{Xb}) = \mathbf{a}'E(\mathbf{X})\mathbf{b}$. (3.39)

(iii) $E(\mathbf{AXB}) = \mathbf{A}E(\mathbf{X})\mathbf{B}$. (3.40)

PROOF. These results follow from Theorem 3.6A (see Problem 3.14). □

Corollary 1. If $\mathbf{A}$ is a $k \times p$ matrix of constants, $\mathbf{b}$ is a $k \times 1$ vector of constants, and $\mathbf{y}$ is a $p \times 1$ random vector, then

$$E(\mathbf{Ay} + \mathbf{b}) = \mathbf{A}E(\mathbf{y}) + \mathbf{b}. \quad (3.41)$$

□

3.6.2 Variances and Covariances

The variance of the random variable $z = \mathbf{a}'\mathbf{y}$ is given in the following theorem.

Theorem 3.6c. If $\mathbf{a}$ is a $p \times 1$ vector of constants and $\mathbf{y}$ is a $p \times 1$ random vector with covariance matrix $\boldsymbol{\Sigma}$, then the variance of $z = \mathbf{a}'\mathbf{y}$ is given by

$$\sigma_z^2 = \text{var}(\mathbf{a}'\mathbf{y}) = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}. \quad (3.42)$$

PROOF. By (3.6) and Theorem 3.6a, we obtain

$$\begin{aligned}
\mathrm{var}(\mathbf{a}'\mathbf{y}) &= E(\mathbf{a}'\mathbf{y} - \mathbf{a}'\boldsymbol{\mu})^2 = E[\mathbf{a}'(\mathbf{y} - \boldsymbol{\mu})]^2 \\
&= E[\mathbf{a}'(\mathbf{y} - \boldsymbol{\mu})\mathbf{a}'(\mathbf{y} - \boldsymbol{\mu})] \\
&= E[\mathbf{a}'(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'\mathbf{a}] && \text{[by (2.18)]} \\
&= \mathbf{a}'E[(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})']\mathbf{a} && \text{[by Theorem 3.6b(ii)]} \\
&= \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} && \text{[by(3.24)].}
\end{aligned}$$

□

We illustrate 3.42 for $p = 3$:

$$\begin{aligned}
\mathrm{var}(\mathbf{a}'\mathbf{y}) &= \mathrm{var}(a_1y_1 + a_2y_2 + a_3y_3) = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} \\
&= a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + a_3^2\sigma_3^2 + 2a_1a_2\sigma_{12} + 2a_1a_3\sigma_{13} + 2a_2a_3\sigma_{23}.
\end{aligned}$$

Thus, $\mathrm{var}(\mathbf{a}'\mathbf{y}) = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}$ involves all the variances and covariances of y_1, y_2, and y_3.

The covariance of two linear combinations is given in the following corollary to Theorem 3.6c.

Corollary 1. If $\mathbf{a}$ and $\mathbf{b}$ are $p \times 1$ vectors of constants, then

$$\mathrm{cov}(\mathbf{a}'\mathbf{y}, \mathbf{b}'\mathbf{y}) = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{b}. \tag{3.43}$$

□

Each variable z_i in the random vector $\mathbf{z} = (z_1, z_2, \ldots, z_k)' = \mathbf{Ay}$ in (3.37) has a variance, and each pair z_i and z_j $(i \neq j)$ has a covariance. These variances and covariances are found in the covariance matrix for $\mathbf{z}$, which is given in the following theorem, along with $\mathrm{cov}(\mathbf{z}, \mathbf{w})$, where $\mathbf{w} = \mathbf{By}$ is another set of linear functions.

Theorem 3.6d. Let $\mathbf{z} = \mathbf{Ay}$ and $\mathbf{w} = \mathbf{By}$, where $\mathbf{A}$ is a $k \times p$ matrix of constants, $\mathbf{B}$ is an $m \times p$ matrix of constants, and $\mathbf{y}$ is a $p \times 1$ random vector with covariance matrix $\boldsymbol{\Sigma}$. Then

$$\text{(i)}\ \mathrm{cov}(\mathbf{z}) = \mathrm{cov}(\mathbf{Ay}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}', \tag{3.44}$$

$$\text{(ii)}\ \mathrm{cov}(\mathbf{z}, \mathbf{w}) = \mathrm{cov}(\mathbf{Ay}, \mathbf{By}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{B}'. \tag{3.45}$$

□

Typically, $k \leq p$, and the $k \times p$ matrix $\mathbf{A}$ is full rank, in which case, by Corollary 1 to 2.6b, $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'$ is positive definite (assuming $\boldsymbol{\Sigma}$ to be positive definite). If $k > p$, then by Corollary 2 to Theorem 2.6b, $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'$ is positive semidefinite. In this case, $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'$ is still a covariance matrix, but it cannot be used in either the numerator or denominator of the multivariate normal density given in (4.9) in Chapter 4.

Note that $\text{cov}(\mathbf{z}, \mathbf{w}) = \mathbf{A\Sigma B}'$ is a $k \times m$ rectangular matrix containing the covariance of each z_i with each w_j, that is, $\text{cov}(\mathbf{z}, \mathbf{w})$ contains $\text{cov}(z_i, w_j)$, $i = 1, 2, \ldots, k$, $j = 1, 2, \ldots, m$. These km covariances can also be found individually by (3.43).

Corollary 1. If $\mathbf{b}$ is a $k \times 1$ vector of constants, then

$$\text{cov}(\mathbf{Ay} + \mathbf{b}) = \mathbf{A\Sigma A}'. \tag{3.46}$$

□

The covariance matrix of linear functions of two different random vectors is given in the following theorem.

Theorem 3.6e. Let $\mathbf{y}$ be a $p \times 1$ random vector and $\mathbf{x}$ be a $q \times 1$ random vector such that $\text{cov}(\mathbf{y}, \mathbf{x}) = \mathbf{\Sigma}_{yx}$. Let $\mathbf{A}$ be a $k \times p$ matrix of constants and $\mathbf{B}$ be an $h \times q$ matrix of constants. Then

$$\text{cov}(\mathbf{Ay}, \mathbf{Bx}) = \mathbf{A\Sigma}_{yx}\mathbf{B}'. \tag{3.47}$$

PROOF. Let

$$\mathbf{v} = \begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} \quad \text{and} \quad \mathbf{C} = \begin{pmatrix} \mathbf{A} & \mathbf{O} \\ \mathbf{O} & \mathbf{B} \end{pmatrix}.$$

Use Theorem 3.6d(i) to obtain $\text{cov}(\mathbf{Cv})$. The result follows. □

PROBLEMS

3.1 Show that $E(ay) = aE(y)$ as in (3.4).

3.2 Show that $E(y - \mu)^2 = E(y^2) - \mu^2$ as in (3.8).

3.3 Show that $\text{var}(ay) = a^2\sigma^2$ as in (3.9).

3.4 Show that $\text{cov}(y_i, y_j) = E(y_i y_j) - \mu_i \mu_j$ as in (3.11).

3.5 Show that if y_i and y_j are independent, then $E(y_i y_j) = E(y_i)E(y_j)$ as in (3.14).

3.6 Show that if y_i and y_j are independent, then $\sigma_{ij} = 0$ as in (3.15).

3.7 Establish the following results in Example 3.2:

(a) Show that $f_2(y) = 1 - \sqrt{1 - y}$ for $0 \leq y \leq 1$ and $f_2(y) = \sqrt{2 - y}$ for $1 \leq y \leq 2$.
(b) Show that $E(y) = \frac{7}{6}$ and $E(xy) = \frac{7}{6}$.
(c) Show that $E(y|x) = \frac{1}{2}(1 + 4x - 2x^2)$.

3.8 Suppose the bivariate random variable (x, y) is uniformly distributed over the region bounded below by $y = x - 1$ for $1 \leq x \leq 2$ and by $y = 3 - x$ for $2 \leq x \leq 3$ and bounded above by $y = x$ for $1 \leq x \leq 2$ and by $y = 4 - x$ for $2 \leq x \leq 3$.

(a) Show that the area of this region is 2, so that $f(x, y) = \frac{1}{2}$.

(b) Find $f_1(x), f_2(y)$, $E(x)$, $E(y)$, $E(xy)$, and σ_{xy}, as was done in Example 3.2. Are x and y independent?

(c) Find $f(y|x)$ and $E(y|x)$.

3.9 Show that $E(\mathbf{x} + \mathbf{y}) = E(\mathbf{x}) + E(\mathbf{y})$ as in (3.21).

3.10 Show that $E[(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})'] = E(\mathbf{y}\mathbf{y}') - \boldsymbol{\mu}\boldsymbol{\mu}'$ as in (3.25).

3.11 Show that the standardized distance transforms the variables so that they are uncorrelated and have means equal to 0 and variances equal to 1 as noted following (3.27).

3.12 Illustrate $\mathbf{P}_\rho = \mathbf{D}_\sigma^{-1}\boldsymbol{\Sigma}\mathbf{D}_\sigma^{-1}$ in (3.30) for $p = 3$.

3.13 Using (3.24), show that

$$\text{cov}(\mathbf{v}) = \text{cov}\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{pmatrix}$$

as in (3.33).

3.14 Prove Theorem 3.6b.

3.15 Prove Corollary 1 to Theorem 3.6b.

3.16 Prove Corollary 1 to Theorem 3.6c.

3.17 Prove Theorem 3.6d.

3.18 Prove Corollary 1 to Theorem 3.6d.

3.19 Consider four $k \times 1$ random vectors $\mathbf{y}$, $\mathbf{x}$, $\mathbf{v}$, and $\mathbf{w}$, and four $h \times k$ constant matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$. Find $\text{cov}(\mathbf{Ay} + \mathbf{Bx}, \mathbf{Cv} + \mathbf{Dw})$.

3.20 Let $\mathbf{y} = (y_1, y_2, y_3)'$ be a random vector with mean vector and covariance matrix

$$\boldsymbol{\mu} = \begin{pmatrix} 1 \\ -1 \\ 3 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 3 \\ 0 & 3 & 10 \end{pmatrix}.$$

(a) Let $z = 2y_1 - 3y_2 + y_3$. Find $E(z)$ and $\text{var}(z)$.

(b) Let $z_1 = y_1 + y_2 + y_3$ and $z_2 = 3y_1 + y_2 - 2y_3$. Find $E(\mathbf{z})$ and $\text{cov}(\mathbf{z})$, where $\mathbf{z} = (z_1, z_2)'$.

3.21 Let $\mathbf{y}$ be a random vector with mean vector and covariance matrix $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as given in Problem 3.20, and define $\mathbf{w} = (w_1, w_2, w_3)'$ as follows:

$$\begin{aligned} w_1 &= 2y_1 - y_2 + y_3 \\ w_2 &= y_1 + 2y_2 - 3y_3 \\ w_3 &= y_1 + y_2 + 2y_3. \end{aligned}$$

(a) Find $E(\mathbf{w})$ and $\text{cov}(\mathbf{w})$.

(b) Using $\mathbf{z}$ as defined in Problem 3.20b, find $\text{cov}(\mathbf{z}, \mathbf{w})$.