4 Multivariate Normal Distribution

In order to make inferences, we often assume that the random vector of interest has a multivariate normal distribution. Before developing the multivariate normal density function and its properties, we first review the univariate normal distribution.

4.1 UNIVARIATE NORMAL DENSITY FUNCTION

We begin with a standard normal random variable z with mean 0 and variance 1. We then transform z to a random variable y with arbitrary mean μ and variance σ^2, and we find the density of y from that of z. In Section 4.2, we will follow an analogous procedure to obtain the density of a multivariate normal random vector.

The standard normal density is given by

$$g(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \qquad -\infty < z < \infty, \tag{4.1}$$

for which $E(z) = 0$ and $\text{var}(z) = 1$. When z has the density (4.1), we say that z is distributed as $N(0, 1)$, or simply that z is $N(0,1)$.

To obtain a normal random variable y with arbitrary mean μ and variance σ^2, we use the transformation $z = (y - \mu)/\sigma$ or $y = \sigma z + \mu$, so that $E(y) = \mu$ and $\text{var}(y) = \sigma^2$. We now find the density $f(y)$ from $g(z)$ in (4.1). For a continuous increasing function (such as $y = \sigma z + \mu$) or for a continuous decreasing function, the change-of-variable technique for a definite integral gives

$$f(y) = g(z)\left|\frac{dz}{dy}\right|, \tag{4.2}$$

where $|dz/dy|$ is the absolute value of dz/dy (Hogg and Craig 1995, p. 169). To use (4.2) to find the density of y, it is clear that both z and dz/dy on the right side must be expressed in terms of y.

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

Let us apply (4.2) to $y = \sigma z + \mu$. The density $g(z)$ is given in (4.1), and for $z = (y - \mu)/\sigma$, we have $|dz/dy| = 1/\sigma$. Thus

$$\begin{aligned} f(y) &= g(z)\left|\frac{dz}{dy}\right| = g\left(\frac{y-\mu}{\sigma}\right)\frac{1}{\sigma} \\ &= \frac{1}{\sqrt{2\pi}\sigma}\, e^{-(y-\mu)^2/2\sigma^2}, \end{aligned} \tag{4.3}$$

which is the normal density with $E(y) = \mu$ and $\text{var}(y) = \sigma^2$. When y has the density (4.3), we say that y is distributed as $N(\mu, \sigma^2)$ or simply that y is $N(\mu, \sigma^2)$.

In Section 4.2, we use a multivariate extension of this technique to find the multivariate normal density function.

4.2 MULTIVARIATE NORMAL DENSITY FUNCTION

We begin with independent standard normal random variables $z_1, z_2, \ldots, z_p$, with $\mu_i = 0$ and $\sigma_i^2 = 1$ for all i and $\sigma_{ij} = 0$ for $i \neq j$, and we then transform the z_i's to multivariate normal variables $y_1, y_2, \ldots, y_p$, with arbitrary means, variances, and covariances. We thus start with a random vector $\mathbf{z} = (z_1, z_2, \ldots, z_p)'$, where $E(\mathbf{z}) = \mathbf{0}$, $\text{cov}(\mathbf{z}) = \mathbf{I}$, and each z_i has the $N(0,1)$ density in (4.1). We wish to transform $\mathbf{z}$ to a multivariate normal random vector $\mathbf{y} = (y_1, y_2 \ldots, y_p)'$ with $E(\mathbf{y}) = \boldsymbol{\mu}$ and $\text{cov}(\mathbf{y}) = \boldsymbol{\Sigma}$, where $\boldsymbol{\mu}$ is any $p \times 1$ vector and $\boldsymbol{\Sigma}$ is any $p \times p$ positive definite matrix.

By (4.1) and an extension of (3.12), we have

$$\begin{aligned} g(z_1, z_2, \ldots, z_p) = g(\mathbf{z}) &= g_1(z_1)g_2(z_2)\cdots g_p(z_p) \\ &= \frac{1}{\sqrt{2\pi}}\, e^{-z_1^2/2}\frac{1}{\sqrt{2\pi}}\, e^{-z_2^2/2}\cdots\frac{1}{\sqrt{2\pi}}\, e^{-z_p^2/2} \\ &= \frac{1}{(\sqrt{2\pi})^p}\, e^{-\sum_i z_i^2/2} \\ &= \frac{1}{(\sqrt{2\pi})^p}\, e^{-\mathbf{z}'\mathbf{z}/2} \quad \text{[by (2.20)]}. \end{aligned} \tag{4.4}$$

If $\mathbf{z}$ has the density (4.4), we say that $\mathbf{z}$ has a multivariate normal density with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{I}$ or simply that $\mathbf{z}$ is distributed as $N_p(\mathbf{0}, \mathbf{I})$, where p is the dimension of the distribution and corresponds to the number of variables in $\mathbf{y}$.

To transform $\mathbf{z}$ to $\mathbf{y}$ with arbitrary mean vector $E(\mathbf{y}) = \boldsymbol{\mu}$ and arbitrary (positive definite) covariance matrix $\text{cov}(\mathbf{y}) = \boldsymbol{\Sigma}$, we define the transformation

$$\mathbf{y} = \boldsymbol{\Sigma}^{1/2}\mathbf{z} + \boldsymbol{\mu}, \tag{4.5}$$

where $\boldsymbol{\Sigma}^{1/2}$ is the (symmetric) square root matrix defined in (2.109). By (3.41) and (3.46), we obtain

$$E(\mathbf{y}) = E(\boldsymbol{\Sigma}^{1/2}\mathbf{z} + \boldsymbol{\mu}) = \boldsymbol{\Sigma}^{1/2}E(\mathbf{z}) + \boldsymbol{\mu} = \boldsymbol{\Sigma}^{1/2}\mathbf{0} + \boldsymbol{\mu} = \boldsymbol{\mu},$$
$$\mathrm{cov}(\mathbf{y}) = \mathrm{cov}(\boldsymbol{\Sigma}^{1/2}\mathbf{z} + \boldsymbol{\mu}) = \boldsymbol{\Sigma}^{1/2}\mathrm{cov}(\mathbf{z})(\boldsymbol{\Sigma}^{1/2})' = \boldsymbol{\Sigma}^{1/2}\mathbf{I}\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Sigma}.$$

Note the analogy of (4.5) to $y = \sigma z + \mu$ in Section 4.1.

Let us now find the density of $\mathbf{y} = \boldsymbol{\Sigma}^{1/2}\mathbf{z} + \boldsymbol{\mu}$ from the density of $\mathbf{z}$ in (4.4). By (4.2), the density of $y = \sigma z + \mu$ is $f(y) = g(z)|dz/dy = g(z)|1/\sigma|$. The analogous expression for the multivariate linear transformation $\mathbf{y} = \boldsymbol{\Sigma}^{1/2}\mathbf{z} + \boldsymbol{\mu}$ is

$$f(\mathbf{y}) = g(\mathbf{z})\mathrm{abs}(|\boldsymbol{\Sigma}^{-1/2}|), \tag{4.6}$$

where $\boldsymbol{\Sigma}^{-1/2}$ is defined as $(\boldsymbol{\Sigma}^{1/2})^{-1}$ and $\mathrm{abs}(|\boldsymbol{\Sigma}^{-1/2}|)$ represents the absolute value of the determinant of $\boldsymbol{\Sigma}^{-1/2}$, which parallels the absolute value expression $|dz/dy| = |1/\sigma|$ in the univariate case. (The determinant $|\boldsymbol{\Sigma}^{-1/2}|$ is the *Jacobian* of the transformation; see any advanced calculus text.) Since $\boldsymbol{\Sigma}^{-1/2}$ is positive definite, we can dispense with the absolute value and write (4.6) as

$$f(\mathbf{y}) = g(\mathbf{z})|\boldsymbol{\Sigma}^{-1/2}| \tag{4.7}$$

$$= g(\mathbf{z})|\boldsymbol{\Sigma}|^{-1/2} \qquad \text{[by (2.67)]}. \tag{4.8}$$

In order to express $\mathbf{z}$ in terms of $\mathbf{y}$, we use (4.5) to obtain $\mathbf{z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$. Then using (4.4) and (4.8), we can write the density of $\mathbf{y}$ as

$$\begin{aligned} f(\mathbf{y}) = g(\mathbf{z})|\boldsymbol{\Sigma}|^{-1/2} &= \frac{1}{(\sqrt{2\pi})^p|\boldsymbol{\Sigma}|^{1/2}}\, e^{-\mathbf{z}'\mathbf{z}/2} \\ &= \frac{1}{(\sqrt{2\pi})^p|\boldsymbol{\Sigma}|^{1/2}}\, e^{-[\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})]'[\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})]/2} \\ &= \frac{1}{(\sqrt{2\pi})^p|\boldsymbol{\Sigma}|^{1/2}}\, e^{-(\mathbf{y}-\boldsymbol{\mu})'(\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2})^{-1}(\mathbf{y}-\boldsymbol{\mu})/2} \\ &= \frac{1}{(\sqrt{2\pi})^p|\boldsymbol{\Sigma}|^{1/2}}\, e^{-(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})/2}, \end{aligned} \tag{4.9}$$

which is the multivariate normal density function with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. When $\mathbf{y}$ has the density (4.9), we say that $\mathbf{y}$ is distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ or simply that $\mathbf{y}$ is $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The subscript p is the dimension of the p-variate normal distribution and indicates the number of variables, that is, $\mathbf{y}$ is $p \times 1$, $\boldsymbol{\mu}$ is $p \times 1$, and $\boldsymbol{\Sigma}$ is $p \times p$.

A comparison of (4.9) and (4.3) shows the standardized distance $(\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})$ in place of $(y - \mu)^2/\sigma^2$ in the exponent and the square root of the generalized variance $|\boldsymbol{\Sigma}|$ in place of the square root of σ^2 in the denominator. [For standardized distance, see (3.27), and for generalized variance, see (3.26).] These distance and variance functions serve analogous purposes in the densities (4.9) and (4.3). In (4.9), $f(\mathbf{y})$ decreases as the distance from $\mathbf{y}$ to $\boldsymbol{\mu}$ increases, and a small value of $|\boldsymbol{\Sigma}|$ indicates that the $\mathbf{y}$'s are concentrated closer to $\boldsymbol{\mu}$ than is the case when $|\boldsymbol{\Sigma}|$ is large. A small value of $|\boldsymbol{\Sigma}|$ may also indicate a high degree of multicollinearity among the variables. High *multicollinearity* indicates that the variables are highly intercorrelated, in which case the $\mathbf{y}$'s tend to occupy a subspace of the p dimensions.

4.3 MOMENT GENERATING FUNCTIONS

We now review moment generating functions, which can be used to obtain some of the properties of multivariate normal random variables. We begin with the univariate case.

The *moment generating function* for a univariate random variable y is defined as

$$M_y(t) = E(e^{ty}), \tag{4.10}$$

provided $E(e^{ty})$ exists for every real number t in the neighborhood $-h < t < h$ for some positive number h. For the univariate normal $N(\mu, \sigma^2)$, the moment generating function of y is given by

$$M_y(t) = e^{t\mu + t^2\sigma^2/2}. \tag{4.11}$$

Moment generating functions characterize a distribution in some important ways that prove very useful (see the two properties at the end of this section). As their name implies, moment generating functions can also be used to generate moments. We now demonstrate this. For a continuous random variable y, the moment generating function can be written as $M_y(t) = E(e^{ty}) = \int_{-\infty}^{\infty} e^{ty} f(y)\, dy$. Then, provided we can interchange the order of integration and differentiation,we have

$$\frac{dM_y(t)}{dt} = M_y'(t) = \int_{-\infty}^{\infty} y e^{ty} f(y)\, dy. \tag{4.12}$$

Setting $t = 0$ gives the first moment or mean:

$$M_y'(0) = \int_{-\infty}^{\infty} y f(y)\, dy = E(y). \tag{4.13}$$

Similarly, the kth moment can be obtained using the kth derivative evaluated at 0:

$$M_y^{(k)}(0) = E(y^k). \tag{4.14}$$

The second moment, $E(y^2)$, can be used to find the variance [see (3.8)].

For a random vector $\mathbf{y}$, the moment generating function is defined as

$$M_{\mathbf{y}}(\mathbf{t}) = E(e^{t_1 y_1 + t_2 y_2 + \cdots + t_p y_p}) = E(e^{\mathbf{t}'\mathbf{y}}). \tag{4.15}$$

By analogy with (4.13), we have

$$\frac{\partial M_{\mathbf{y}}(\mathbf{0})}{\partial \mathbf{t}} = E(\mathbf{y}), \tag{4.16}$$

where the notation $\partial M_{\mathbf{y}}(\mathbf{0})/\partial \mathbf{t}$ indicates that $\partial M_{\mathbf{y}}(\mathbf{t})/\partial \mathbf{t}$ is evaluated at $\mathbf{t} = \mathbf{0}$. Similarly, $\partial^2 M_{\mathbf{y}}(\mathbf{t})/\partial t_r \partial t_s$ evaluated at $t_r = t_s = 0$ gives $E(y_r y_s)$:

$$\frac{\partial^2 M_{\mathbf{y}}(\mathbf{0})}{\partial t_r \partial t_s} = E(y_r y_s). \tag{4.17}$$

For a multivariate normal random vector $\mathbf{y}$, the moment generating function is given in the following theorem.

Theorem 4.3. If $\mathbf{y}$ is distributed as $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, its moment generating function is given by

$$M_{\mathbf{y}}(\mathbf{t}) = e^{\mathbf{t}'\boldsymbol{\mu} + \mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}/2}. \tag{4.18}$$

Proof. By (4.15) and (4.9), the moment generating function is

$$M_{\mathbf{y}}(\mathbf{t}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} k e^{\mathbf{t}'\mathbf{y} - (\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})/2} \, d\mathbf{y}, \tag{4.19}$$

where $k = 1/(\sqrt{2\pi})^p |\boldsymbol{\Sigma}|^{1/2}$ and $d\mathbf{y} = dy_1 \, dy_2 \cdots dy_p$. By rewriting the exponent, we obtain

$$M_{\mathbf{y}}(\mathbf{t}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} k \; e^{\mathbf{t}'\boldsymbol{\mu} + \mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}/2 - (\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\Sigma}\mathbf{t})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu}-\boldsymbol{\Sigma}\mathbf{t})/2} \, d\mathbf{y} \tag{4.20}$$

$$= e^{\mathbf{t}'\boldsymbol{\mu} + \mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}/2} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} k \; e^{-[\mathbf{y}-(\boldsymbol{\mu}+\boldsymbol{\Sigma}\mathbf{t})]'\boldsymbol{\Sigma}^{-1}[\mathbf{y}-(\boldsymbol{\mu}+\boldsymbol{\Sigma}\mathbf{t})]/2} \, d\mathbf{y} \tag{4.21}$$

$$= e^{\mathbf{t}'\boldsymbol{\mu} + \mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}/2}.$$

The multiple integral in (4.21) is equal to 1 because the multivariate normal density in (4.9) integrates to 1 for any mean vector, including $\boldsymbol{\mu} + \boldsymbol{\Sigma}\mathbf{t}$. □

Corollary 1. The moment generating function for $\mathbf{y} - \boldsymbol{\mu}$ is

$$M_{\mathbf{y}-\boldsymbol{\mu}}(\mathbf{t}) = e^{\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}/2}. \tag{4.22}$$

□

We now list two important properties of moment generating functions.

1. If two random vectors have the same moment generating function, they have the same density.
2. Two random vectors are independent if and only if their joint moment generating function factors into the product of their two separate moment generating functions; that is, if $\mathbf{y}' = (\mathbf{y}_1', \mathbf{y}_2')$ and $\mathbf{t}' = (\mathbf{t}_1', \mathbf{t}_2')$, then $\mathbf{y}_1$ and $\mathbf{y}_2$ are independent if and only if

$$M_{\mathbf{y}}(\mathbf{t}) = M_{\mathbf{y}_1}(\mathbf{t}_1)M_{\mathbf{y}_2}(\mathbf{t}_2). \tag{4.23}$$

4.4 PROPERTIES OF THE MULTIVARIATE NORMAL DISTRIBUTION

We first consider the distribution of linear functions of multivariate normal random variables.

Theorem 4.4a. Let the $p \times 1$ random vector $\mathbf{y}$ be $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, let $\mathbf{a}$ be any $p \times 1$ vector of constants, and let $\mathbf{A}$ be any $k \times p$ matrix of constants with rank $k \leq p$. Then

(i) $z = \mathbf{a}'\mathbf{y}$ is $N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$

(ii) $\mathbf{z} = \mathbf{A}\mathbf{y}$ is $N_k(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$.

PROOF

(i) The moment generating function for $z = \mathbf{a}'\mathbf{y}$ is given by

$$\begin{aligned} M_z(t) &= E(e^{tz}) = E(e^{t\mathbf{a}'\mathbf{y}}) = E(e^{(t\mathbf{a})'\mathbf{y}}) \\ &= e^{(t\mathbf{a})'\boldsymbol{\mu} + (t\mathbf{a})'\boldsymbol{\Sigma}(t\mathbf{a})/2} \qquad \text{[by (4.18)]} \\ &= e^{(\mathbf{a}'\boldsymbol{\mu})t + (\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})t^2/2}. \end{aligned} \tag{4.24}$$

On comparing (4.24) with (4.11), it is clear that $z = \mathbf{a}'\mathbf{y}$ is univariate normal with mean $\mathbf{a}'\boldsymbol{\mu}$ and variance $\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}$.

(ii) The moment generating function for $\mathbf{z} = \mathbf{A}\mathbf{y}$ is given by

$$M_{\mathbf{z}}(\mathbf{t}) = E(e^{\mathbf{t}'\mathbf{z}}) = E(e^{\mathbf{t}'\mathbf{A}\mathbf{y}}),$$

which becomes

$$M_{\mathbf{z}}(\mathbf{t}) = e^{\mathbf{t}'(\mathbf{A}\boldsymbol{\mu}) + \mathbf{t}'(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')\mathbf{t}/2} \tag{4.25}$$

(see Problem 4.7). By Corollary 1 to Theorem 2.6b, the covariance matrix $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'$ is positive definite. Thus, by (4.18) and (4.25), the $k \times 1$ random vector $\mathbf{z} = \mathbf{A}\mathbf{y}$ is distributed as the k-variate normal $N_k(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$. □

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

$$\mathbf{z} = \mathbf{A}\mathbf{y} + \mathbf{b} \quad \text{is} \quad N_k(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}').$$ □

The marginal distributions of multivariate normal variables are also normal, as shown in the following theorem.

Theorem 4.4b. If $\mathbf{y}$ is $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then any $r \times 1$ subvector of $\mathbf{y}$ has an r-variate normal distribution with the same means, variances, and covariances as in the original p-variate normal distribution.

Proof. Without loss of generality, let $\mathbf{y}$ be partitioned as $\mathbf{y}' = (\mathbf{y}_1', \mathbf{y}_2')$, where $\mathbf{y}_1$ is the $r \times 1$ subvector of interest. Let $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ be partitioned accordingly:

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

Define $\mathbf{A} = (\mathbf{I}_r, \mathbf{O})$, where $\mathbf{I}_r$ is an $r \times r$ identity matrix and $\mathbf{O}$ is an $r \times (p - r)$ matrix of 0s. Then $\mathbf{A}\mathbf{y} = \mathbf{y}_1$, and by Theorem 4.4a (ii), $\mathbf{y}_1$ is distributed as $N_r(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$. □

Corollary 1. If $\mathbf{y}$ is $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then any individual variable y_i in $\mathbf{y}$ is distributed as $N(\mu_i, \sigma_{ii})$. □

For the next two theorems, we use the notation of Section 3.5, in which the random vector $\mathbf{v}$ is partitioned into two subvectors denoted by $\mathbf{y}$ and $\mathbf{x}$, where $\mathbf{y}$ is $p \times 1$ and $\mathbf{x}$

is $q \times 1$, with a corresponding partitioning of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ [see (3.32) and (3.33)]:

$$\mathbf{v} = \begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix}, \quad \boldsymbol{\mu} = E\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{pmatrix}, \quad \boldsymbol{\Sigma} = \mathrm{cov}\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{pmatrix}.$$

By (3.15), if two random variables y_i and y_j are independent, then $\sigma_{ij} = 0$. The converse of this is not true, as illustrated in Example 3.2. By extension, if two random vectors $\mathbf{y}$ and $\mathbf{x}$ are independent (i.e., each y_i is independent of each x_j), then $\boldsymbol{\Sigma}_{yx} = \mathbf{O}$ (the covariance of each y_i with each x_j is 0). The converse is not true in general, but it is true for multivariate normal random vectors.

Theorem 4.4c. If $\mathbf{v} = \begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix}$ is $N_{p+q}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\mathbf{y}$ and $\mathbf{x}$ are independent if $\boldsymbol{\Sigma}_{yx} = \mathbf{O}$.

Proof. Suppose $\boldsymbol{\Sigma}_{yx} = \mathbf{O}$. Then

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{yy} & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\Sigma}_{xx} \end{pmatrix},$$

and the exponent of the moment generating function in (4.18) becomes

$$\begin{aligned} \mathbf{t}'\boldsymbol{\mu} + \tfrac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t} &= (\mathbf{t}_y', \mathbf{t}_x')\begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{pmatrix} + \tfrac{1}{2}(\mathbf{t}_y', \mathbf{t}_x')\begin{pmatrix} \boldsymbol{\Sigma}_{yy} & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\Sigma}_{xx} \end{pmatrix}\begin{pmatrix} \mathbf{t}_y \\ \mathbf{t}_x \end{pmatrix} \\ &= \mathbf{t}_y'\boldsymbol{\mu}_y + \mathbf{t}_x'\boldsymbol{\mu}_x + \tfrac{1}{2}\mathbf{t}_y'\boldsymbol{\Sigma}_{yy}\mathbf{t}_y + \tfrac{1}{2}\mathbf{t}_x'\boldsymbol{\Sigma}_{xx}\mathbf{t}_x. \end{aligned}$$

The moment generating function can then be written as

$$M_{\mathbf{v}}(\mathbf{t}) = e^{\mathbf{t}_y'\boldsymbol{\mu}_y + \mathbf{t}_y'\boldsymbol{\Sigma}_{yy}\mathbf{t}_y/2} e^{\mathbf{t}_x'\boldsymbol{\mu}_x + \mathbf{t}_x'\boldsymbol{\Sigma}_{xx}\mathbf{t}_x/2},$$

which is the product of the moment generating functions of $\mathbf{y}$ and $\mathbf{x}$. Hence, by (4.23), $\mathbf{y}$ and $\mathbf{x}$ are independent. □

Corollary 1. If $\mathbf{y}$ is $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then any two individual variables y_i and y_j are independent if $\sigma_{ij} = 0$.

Corollary 2. If $\mathbf{y}$ is $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and if $\mathrm{cov}(\mathbf{Ay}, \mathbf{By}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{B}' = \mathbf{O}$, then $\mathbf{Ay}$ and $\mathbf{By}$ are independent. □

The relationship between subvectors $\mathbf{y}$ and $\mathbf{x}$ when they are not independent ($\boldsymbol{\Sigma}_{yx} \neq \mathbf{O}$) is given in the following theorem.

Theorem 4.4d. If $\mathbf{y}$ and $\mathbf{x}$ are jointly multivariate normal with $\boldsymbol{\Sigma}_{yx} \neq \mathbf{O}$, then the conditional distribution of $\mathbf{y}$ given $\mathbf{x}$, $f(\mathbf{y}|x)$, is multivariate normal with mean vector and covariance matrix

$$E(\mathbf{y}|\mathbf{x}) = \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x), \tag{4.26}$$

$$\mathrm{cov}(\mathbf{y}|\mathbf{x}) = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}. \tag{4.27}$$

PROOF. By an extension of (3.18), the conditional density of $\mathbf{y}$ given $\mathbf{x}$ is

$$f(\mathbf{y}|\mathbf{x}) = \frac{g(\mathbf{y}, \mathbf{x})}{h(\mathbf{x})}, \tag{4.28}$$

where $g(\mathbf{y}, \mathbf{x})$ is the joint density of $\mathbf{y}$ and $\mathbf{x}$, and $h(\mathbf{x})$ is the marginal density of $\mathbf{x}$. The proof can be carried out by directly evaluating the ratio on the right hand side of (4.28), using results (2.50) and (2.71) (see Problem 4.13). For variety, we use an alternative approach that avoids working explicitly with $g(\mathbf{y}, \mathbf{x})$ and $h(\mathbf{x})$ and the resulting partitioned matrix formulas.

Consider the function

$$\begin{pmatrix} \mathbf{w} \\ \mathbf{u} \end{pmatrix} = \mathbf{A}\left[\begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{pmatrix}\right], \tag{4.29}$$

where

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{I} & -\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1} \\ \mathbf{O} & \mathbf{I} \end{pmatrix}.$$

To be conformal, the identity matrix in $\mathbf{A}_1$ is $p \times p$ while the identity in $\mathbf{A}_2$ is $q \times q$. Simplifying and rearranging (4.29), we obtain $\mathbf{w} = \mathbf{y} - [\boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)]$ and $\mathbf{u} = \mathbf{x} - \boldsymbol{\mu}_x$. Using the multivariate change-of-variable technique [referred to in (4.6], the joint density of $(\mathbf{w}, \mathbf{u})$ is

$$p(\mathbf{w}, \mathbf{u}) = g(\mathbf{y}, \mathbf{x})|\mathbf{A}^{-1}| = g(\mathbf{y}, \mathbf{x})$$

[employing Theorem 2.9a (ii) and (vi)]. Similarly, the marginal density of $\mathbf{u}$ is

$$q(\mathbf{u}) = h(\mathbf{x})|\mathbf{I}^{-1}| = h(\mathbf{x}).$$

Using (3.45), it also turns out that

$$\mathrm{cov}(\mathbf{w}, \mathbf{u}) = \mathbf{A}_1\boldsymbol{\Sigma}\mathbf{A}_2 = \boldsymbol{\Sigma}_{yx} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xx} = \mathbf{O} \tag{4.30}$$

(see Problem 4.14). Thus, by Theorem 4.4c, $\mathbf{w}$ is independent of $\mathbf{u}$. Hence

$$p(\mathbf{w}, \mathbf{u}) = r(\mathbf{w})q(\mathbf{u}),$$

where $r(\mathbf{w})$ is the density of $\mathbf{w}$. Since $p(\mathbf{w}, \mathbf{u}) = g(\mathbf{y}, \mathbf{x})$ and $q(\mathbf{u}) = h(\mathbf{x})$, we also have

$$g(\mathbf{y}, \mathbf{x}) = r(\mathbf{w})h(\mathbf{x}),$$

and by (4.28),

$$r(\mathbf{w}) = \frac{g(\mathbf{y}, \mathbf{x})}{h(\mathbf{x})} = f(\mathbf{y}|\mathbf{x}).$$

Hence we obtain $f(\mathbf{y}|\mathbf{x})$ simply by finding $r(\mathbf{w})$. By Corollary 1 to Theorem 4.4a, $r(\mathbf{w})$ is the multivariate normal density with

$$\boldsymbol{\mu}_w = \mathbf{A}_1\left[\begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{pmatrix} - \begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{pmatrix}\right] = \mathbf{0}, \tag{4.31}$$

$$\begin{aligned} \boldsymbol{\Sigma}_{ww} &= \mathbf{A}_1\boldsymbol{\Sigma}\mathbf{A}_1' \\ &= (\mathbf{I}, \ -\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1})\begin{pmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{pmatrix}\begin{pmatrix} \mathbf{I} \\ -\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy} \end{pmatrix} \\ &= \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}. \end{aligned} \tag{4.32}$$

Thus $r(\mathbf{w}) = r(\mathbf{y} - [\boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)])$ is of the form $N_p(\mathbf{0}, \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy})$. Equivalently, $\mathbf{y}|\mathbf{x}$ is $N_p[\boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x), \ \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}]$. □

Since $E(\mathbf{y}|\mathbf{x}) = \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)$ in (4.26) is a linear function of $\mathbf{x}$, any pair of variables y_i and y_j in a multivariate normal vector exhibits a linear trend $E(y_i|y_j) = \mu_i + (\sigma_{ij}/\sigma_{jj})(y_j - \mu_j)$. Thus the covariance σ_{ij} is related to the slope of the line representing the trend, and σ_{ij} is a useful measure of relationship between two normal variables. In the case of nonnormal variables that exhibit a curved trend, σ_{ij} may give a very misleading indication of the relationship, as illustrated in Example 3.2.

The conditional covariance matrix $\text{cov}(\mathbf{y}|\mathbf{x}) = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}$ in (4.27) does not involve $\mathbf{x}$. For some nonnormal distributions, on the other hand, $\text{cov}(\mathbf{y}|\mathbf{x})$ is a function of $\mathbf{x}$.

If there is only one y, so that $\mathbf{v}$ is partitioned in the form $\mathbf{v} = (y, x_1, x_2, \ldots, x_q) = (y, \mathbf{x}')$, then $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ have the form

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_y \\ \boldsymbol{\mu}_x \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_y^2 & \boldsymbol{\sigma}_{yx}' \\ \boldsymbol{\sigma}_{yx} & \boldsymbol{\Sigma}_{xx} \end{pmatrix},$$

where μ_y and σ_y^2 are the mean and variance of y, $\boldsymbol{\sigma}_{yx}' = (\sigma_{y1}, \sigma_{y2}, \ldots, \sigma_{yq})$ contains the covariances $\sigma_{yi} = \text{cov}(y, x_i)$, and $\boldsymbol{\Sigma}_{xx}$ contains the variances and covariances of the x variables. The conditional distribution is given in the following corollary to Theorem 4.4d.

Corollary 1. If $\mathbf{v} = (y, x_1, x_2, \ldots, x_q) = (y, \mathbf{x}')$, with

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_y \\ \boldsymbol{\mu}_x \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_y^2 & \boldsymbol{\sigma}'_{yx} \\ \boldsymbol{\sigma}_{yx} & \boldsymbol{\Sigma}_{xx} \end{pmatrix},$$

then $y|\mathbf{x}$ is normal with

$$E(y|\mathbf{x}) = \mu_y + \boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x), \tag{4.33}$$

$$\text{var}(y|\mathbf{x}) = \sigma_y^2 - \boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{yx}. \tag{4.34}$$

□

In (4.34), $\boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{yx} \geq 0$ because $\boldsymbol{\Sigma}_{xx}^{-1}$ is positive definite. Therefore

$$\text{var}(y|\mathbf{x}) \leq \text{var}(y). \tag{4.35}$$

Example 4.4a. To illustrate Theorems 4.4a–c, suppose that $\mathbf{y}$ is $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} 4 & 0 & 2 \\ 0 & 1 & -1 \\ 2 & -1 & 3 \end{pmatrix}.$$

For $z = y_1 - 2y_2 + y_3 = (1, -2, 1)\mathbf{y} = \mathbf{a}'\mathbf{y}$, we have $\mathbf{a}'\boldsymbol{\mu} = 3$ and $\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} = 19$. Hence by Theorem 4.4a(i), z is $N(3, 19)$.

The linear functions

$$z_1 = y_1 - y_2 + y_3, \quad z_2 = 3y_1 + y_2 - 2y_3$$

can be written as

$$\mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 1 \\ 3 & 1 & -2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \mathbf{A}\mathbf{y}.$$

Then by Theorem 3.6b(i) and Theorem 3.6d(i), we obtain

$$\mathbf{A}\boldsymbol{\mu} = \begin{pmatrix} 4 \\ 6 \end{pmatrix}, \quad \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}' = \begin{pmatrix} 14 & 4 \\ 4 & 29 \end{pmatrix},$$

and by Theorem 4.4a(ii), we have

$$\mathbf{z} \text{ is } N_2\left[\begin{pmatrix} 4 \\ 6 \end{pmatrix}, \begin{pmatrix} 14 & 4 \\ 4 & 29 \end{pmatrix}\right].$$

To illustrate the marginal distributions in Theorem 4.4b, note that y_1 is $N(3, 4)$, y_3 is

$$N(2, 3), \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \text{ is } N_2\left[\begin{pmatrix} 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}\right], \text{ and } \begin{pmatrix} y_1 \\ y_3 \end{pmatrix} \text{ is } N_2\left[\begin{pmatrix} 3 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 & 2 \\ 2 & 3 \end{pmatrix}\right].$$

To illustrate Theorem 4.4c, we note that $\sigma_{12} = 0$, and therefore y_1 and y_2 are independent. □

Example 4.4b. To illustrate Theorem 4.4d, let the random vector $\mathbf{v}$ be $N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = \begin{pmatrix} 2 \\ 5 \\ -2 \\ 1 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} 9 & 0 & 3 & 3 \\ 0 & 1 & -1 & 2 \\ 3 & -1 & 6 & -3 \\ 3 & 2 & -3 & 7 \end{pmatrix}.$$

If $\mathbf{v}$ is partitioned as $\mathbf{v} = (y_1, y_2, x_1, x_2)'$, then $\boldsymbol{\mu}_y = \begin{pmatrix} 2 \\ 5 \end{pmatrix}$, $\boldsymbol{\mu}_x = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$, $\boldsymbol{\Sigma}_{yy} = \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix}$, $\boldsymbol{\Sigma}_{yx} = \begin{pmatrix} 3 & 3 \\ -1 & 2 \end{pmatrix}$, and $\boldsymbol{\Sigma}_{xx} = \begin{pmatrix} 6 & -3 \\ -3 & 7 \end{pmatrix}$. By (4.26), we obtain

$$\begin{aligned} E(\mathbf{y}|\mathbf{x}) &= \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x) \\ &= \begin{pmatrix} 2 \\ 5 \end{pmatrix} + \begin{pmatrix} 3 & 3 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 6 & -3 \\ -3 & 7 \end{pmatrix}^{-1}\begin{pmatrix} x_1 + 2 \\ x_2 - 1 \end{pmatrix} \\ &= \begin{pmatrix} 2 \\ 5 \end{pmatrix} + \frac{1}{33}\begin{pmatrix} 30 & 27 \\ -1 & 9 \end{pmatrix}\begin{pmatrix} x_1 + 2 \\ x_2 - 1 \end{pmatrix} \\ &= \begin{pmatrix} 3 + \frac{10}{11}x_1 + \frac{9}{11}x_2 \\ \frac{14}{3} - \frac{1}{33}x_1 + \frac{3}{11}x_2 \end{pmatrix}. \end{aligned}$$

By (4.27), we have

$$\begin{aligned} \mathrm{cov}(\mathbf{y}|\mathbf{x}) &= \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy} \\ &= \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 3 & 3 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 6 & -3 \\ -3 & 7 \end{pmatrix}^{-1}\begin{pmatrix} 3 & -1 \\ 3 & 2 \end{pmatrix} \\ &= \begin{pmatrix} 9 & 0 \\ 0 & 1 \end{pmatrix} - \tfrac{1}{33}\begin{pmatrix} 171 & 24 \\ 24 & 19 \end{pmatrix} \\ &= \tfrac{1}{33}\begin{pmatrix} 126 & -24 \\ -24 & 14 \end{pmatrix}. \end{aligned}$$

Thus

$$\mathbf{y}|\mathbf{x} \text{ is } N_2\left[\begin{pmatrix} 3 + \frac{10}{11}x_1 + \frac{9}{11}x_2 \\ \frac{14}{3} - \frac{1}{33}x_1 + \frac{3}{11}x_2 \end{pmatrix}, \frac{1}{33}\begin{pmatrix} 126 & -24 \\ -24 & 14 \end{pmatrix}\right].$$

□

Example 4.4c. To illustrate Corollary 1 to Theorem 4.4d, let $\mathbf{v}$ be $N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are as given in Example 4.4b. If $\mathbf{v}$ is partitioned as $\mathbf{v} = (y, x_1, x_2, x_3)'$, then $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are partitioned as follows:

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_y \\ \boldsymbol{\mu}_x \end{pmatrix} = \left(\begin{array}{c} 2 \\ 5 \\ \hline -2 \\ 1 \end{array}\right),$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_y^2 & \boldsymbol{\sigma}'_{yx} \\ \boldsymbol{\sigma}_{yx} & \boldsymbol{\Sigma}_{xx} \end{pmatrix} = \left(\begin{array}{c|ccc} 9 & 0 & 3 & 3 \\ 0 & 1 & -1 & 2 \\ \hline 3 & -1 & 6 & -3 \\ 3 & 2 & -3 & 7 \end{array}\right).$$

By (4.33), we have

$$\begin{aligned} E(y|x_1, x_2, x_3) &= \mu_y + \boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x) \\ &= 2 + (0, 3, 3)\begin{pmatrix} 1 & -1 & 2 \\ -1 & 6 & -3 \\ 2 & -3 & 7 \end{pmatrix}^{-1}\begin{pmatrix} x_1 - 5 \\ x_2 + 2 \\ x_3 + 1 \end{pmatrix} \\ &= \tfrac{95}{7} - \tfrac{12}{7}x_1 + \tfrac{6}{7}x_2 + \tfrac{9}{7}x_3. \end{aligned}$$

By (4.34), we obtain

$$\begin{aligned} \text{var}(y|x_1, x_2, x_3) &= \sigma_y^2 - \boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{yx} \\ &= 9 - (0, 3, 3)\begin{pmatrix} 1 & -1 & 2 \\ -1 & 6 & -3 \\ 2 & -3 & 7 \end{pmatrix}^{-1}\begin{pmatrix} 0 \\ 3 \\ 3 \end{pmatrix} \\ &= 9 - \tfrac{45}{7} = \tfrac{18}{7}. \end{aligned}$$

Hence $y|x_1, x_2, x_3$ is $N(\frac{95}{7} - \frac{12}{7}x_1 + \frac{6}{7}x_2 + \frac{9}{7}x_3, \frac{18}{7})$. Note that $\text{var}(y|x_1, x_2, x_3) = \frac{18}{7}$ is less than $\text{var}(y) = 9$, which illustrates (4.35). □

4.5 PARTIAL CORRELATION

We now define the partial correlation of y_i and y_j adjusted for a subset of other y variables. For convenience, we use the notation of Theorems 4.4c and 4.4d. The subset of y's containing y_i and y_j is denoted by $\mathbf{y}$, and the other subset of y's is denoted by $\mathbf{x}$.

Let $\mathbf{v}$ be $N_{p+q}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and let $\mathbf{v}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$ be partitioned as in Theorem 4.4c and 4.4d:

$$\mathbf{v} = \begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_x \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{pmatrix}.$$

The covariance of y_i and y_j in the conditional distribution of $\mathbf{y}$ given $\mathbf{x}$ will be denoted by $\sigma_{ij\cdot rs\ldots q}$, where y_i and y_j are two of the variables in $\mathbf{y}$ and $y_r, y_s, \ldots, y_q$ are all the variables in $\mathbf{x}$. Thus $\sigma_{ij\cdot rs\ldots q}$ is the (ij)th element of $\mathrm{cov}(\mathbf{y}|\mathbf{x}) = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}$. For example, $\sigma_{13\cdot 567}$ represents the covariance between y_1 and y_3 in the conditional distribution of y_1, y_2, y_3, y_4 given y_5, y_6, and y_7 [in this case $\mathbf{x} = (y_5, y_6, y_7)'$]. Similarly, $\sigma_{22\cdot 567}$ represents the variance of y_2 in the conditional distribution of y_1, y_2, y_3, y_4 given y_5, y_6, y_7.

We now define the *partial correlation coefficient* $\rho_{ij\cdot rs\ldots q}$ to be the correlation between y_i and y_j in the conditional distribution of $\mathbf{y}$ given $\mathbf{x}$, where $\mathbf{x} = (y_r, y_s, \ldots, y_q)'$. From the usual definition of a correlation [see (3.19)], we can obtain $\rho_{ij\cdot rs\ldots q}$ from $\sigma_{ij\cdot rs\ldots q}$:

$$\rho_{ij\cdot rs\ldots q} = \frac{\sigma_{ij\cdot rs\ldots q}}{\sqrt{\sigma_{ii\cdot rs\ldots q}\sigma_{jj\cdot rs\ldots q}}}. \tag{4.36}$$

This is the population partial correlation. The sample partial correlation $r_{ij\cdot rs} \cdots q$ is discussed in Section 10.7, including a formulation that does not require normality.

The matrix of partial correlations, $\mathbf{P}_{y\cdot x} = (\rho_{ij\cdot rs\ldots q})$ can be found by (3.30) and (4.27) as

$$\mathbf{P}_{y\cdot x} = \mathbf{D}_{y\cdot x}^{-1}\boldsymbol{\Sigma}_{y\cdot x}\mathbf{D}_{y\cdot x}^{-1}, \tag{4.37}$$

where $\boldsymbol{\Sigma}_{y\cdot x} = \mathrm{cov}(\mathbf{y}|\mathbf{x}) = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}$ and $\mathbf{D}_{y\cdot x} = [\mathrm{diag}(\boldsymbol{\Sigma}_{y\cdot x})]^{1/2}$.

Unless $\mathbf{y}$ and $\mathbf{x}$ are independent ($\boldsymbol{\Sigma}_{yx} = \mathbf{O}$), the partial correlation $\rho_{ij\cdot rs\ldots q}$ is different from the usual correlation $\rho_{ij} = \sigma_{ij}/\sqrt{\sigma_{ii}\sigma_{jj}}$. In fact, $\rho_{ij\cdot rs\ldots q}$ and ρ_{ij} can be of opposite signs (for an illustration, see Problem 4.16 g, h). To show this, we express $\sigma_{ij\cdot rs\ldots q}$ in terms of σ_{ij}. We first write $\boldsymbol{\Sigma}_{yx}$ in terms of its rows

$$\boldsymbol{\Sigma}_{yx} = \mathrm{cov}(\mathbf{y}, \mathbf{x}) = \begin{pmatrix} \sigma_{y_1x_1} & \sigma_{y_1x_2} & \cdots & \sigma_{y_1x_q} \\ \sigma_{y_2x_1} & \sigma_{y_2x_2} & \cdots & \sigma_{y_2x_q} \\ \vdots & \vdots & & \vdots \\ \sigma_{y_px_1} & \sigma_{y_px_2} & \cdots & \sigma_{y_px_q} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\sigma}'_{1x} \\ \boldsymbol{\sigma}'_{2x} \\ \vdots \\ \boldsymbol{\sigma}'_{px} \end{pmatrix}, \tag{4.38}$$

where $\boldsymbol{\sigma}'_{ix} = (\sigma_{y_i x_1}, \sigma_{y_i x_2}, \ldots, \sigma_{y_i x_q})$. Then $\sigma_{ij \cdot rs \ldots q}$, the (ij)th element of $\boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}$, can be written as

$$\sigma_{ij \cdot rs \ldots q} = \sigma_{ij} - \boldsymbol{\sigma}'_{ix}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{jx}. \tag{4.39}$$

Suppose that σ_{ij} is positive. Then $\sigma_{ij \cdot rs \ldots q}$ is negative if $\boldsymbol{\sigma}'_{ix}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{jx} > \sigma_{ij}$. Note also that since $\boldsymbol{\Sigma}_{xx}^{-1}$ is positive definite, (4.39) shows that

$$\sigma_{ii \cdot rs \ldots q} = \sigma_{ii} - \boldsymbol{\sigma}'_{ix}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{ix} \leq \sigma_{ii}.$$

Example 4.5. We compare ρ_{12} and $\rho_{12 \cdot 34}$ using $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in Example 4.4b. From $\boldsymbol{\Sigma}$, we obtain

$$\rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}} = \frac{0}{\sqrt{(9)(1)}} = 0.$$

From $\text{cov}(\mathbf{y}|\mathbf{x}) = \frac{1}{33}\begin{pmatrix} 126 & -24 \\ -24 & 14 \end{pmatrix}$ in Example 4.4b, we obtain

$$\rho_{12 \cdot 34} = \frac{\sigma_{12 \cdot 34}}{\sqrt{\sigma_{11 \cdot 34}\sigma_{22 \cdot 34}}} = \frac{-24/33}{\sqrt{(126/33)(14/33)}} = \frac{-24}{\sqrt{(36)(49)}}$$

$$= \frac{-4}{7} = -.571.$$

□

PROBLEMS

4.1 Show that $E(z) = 0$ and $\text{var}(z) = 1$ when z has the standard normal density (4.1).

4.2 Obtain (4.8) from (4.7); that is, show that $|\boldsymbol{\Sigma}^{-1/2}| = |\boldsymbol{\Sigma}|^{-1/2}$.

4.3 Show that $\partial M_{\mathbf{y}}(\mathbf{0})/\partial \mathbf{t} = E(\mathbf{y})$ as in (4.16).

4.4 Show that $\partial^2 M_{\mathbf{y}}(\mathbf{0})/\partial t_r \partial t_s = E(y_r y_s)$ as in (4.17).

4.5 Show that the exponent in (4.19) can be expressed as in (4.20); that is, show that $\mathbf{t}'\mathbf{y} - (\mathbf{y} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})/2 = \mathbf{t}'\boldsymbol{\mu} + \mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}/2 - (\mathbf{y} - \boldsymbol{\mu} - \boldsymbol{\Sigma}\mathbf{t})'\boldsymbol{\Sigma}^{-1}$ $(\mathbf{y} - \boldsymbol{\mu} - \boldsymbol{\Sigma}\mathbf{t})/2$.

4.6 Prove Corollary 1 to Theorem 4.3.

4.7 Show that $E(e^{\mathbf{t}'\mathbf{A}\mathbf{y}}) = e^{\mathbf{t}'(\mathbf{A}\boldsymbol{\mu}) + \mathbf{t}'(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')\mathbf{t}/2}$ as in (4.25).

4.8 Consider a random variable with moment generating function $M(t)$. Show that the second derivative of $\ln[M(t)]$ evaluated at $t = 0$ is the variance of the random variable.

4.9 Assuming that $\mathbf{y}$ is $N_p(\boldsymbol{\mu}, \sigma^2\mathbf{I})$ and $\mathbf{C}$ is an orthogonal matrix, show that $\mathbf{Cy}$ is $N_p(\mathbf{C}\boldsymbol{\mu}, \sigma^2\mathbf{I})$.

4.10 Prove Corollary 1 to Theorem 4.4a.

4.11 Let $\mathbf{A} = (\mathbf{I}_r, \mathbf{O})$, as defined in the proof of Theorem 4.4b. Show that $\mathbf{Ay} = \mathbf{y}_1$, $\mathbf{A}\boldsymbol{\mu} = \boldsymbol{\mu}_1$, and $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}' = \boldsymbol{\Sigma}_{11}$.

4.12 Prove Corollary 2 to Theorem 4.4c.

4.13 Prove Theorem 4.4d by direct evaluation of (4.28).

4.14 Given $\mathbf{w} = \mathbf{y} - \mathbf{Bx}$, show that $\text{cov}(\mathbf{w}, \mathbf{x}) = \boldsymbol{\Sigma}_{yx} - \mathbf{B}\boldsymbol{\Sigma}_{xx}$, as in (4.30).

4.15 Show that $E(\mathbf{y} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\mathbf{x}) = \boldsymbol{\mu}_y - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x$ as in (4.31) and that $\text{cov}(\mathbf{y} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\mathbf{x}) = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}$ as in (4.32).

4.16 Suppose that $\mathbf{y}$ is $N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ -2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} 4 & 2 & -1 & 2 \\ 2 & 6 & 3 & -2 \\ -1 & 3 & 5 & -4 \\ 2 & -2 & -4 & 4 \end{pmatrix}.$$

Find the following.

(a) The joint marginal distribution of y_1 and y_3
(b) The marginal distribution of y_2
(c) The distribution of $z = y_1 + 2y_2 - y_3 + 3y_4$
(d) The joint distribution of $z_1 = y_1 + y_2 - y_3 - y_4$ and $z_2 = -3y_1 + y_2 + 2y_3 - 2y_4$
(e) $f(y_1, y_2 | y_3, y_4)$
(f) $f(y_1, y_3 | y_2, y_4)$
(g) ρ_{13}
(h) $\rho_{13\cdot 24}$
(i) $f(y_1 | y_2, y_3, y_4)$

4.17 Let $\mathbf{y}$ be distributed as $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 3 \end{pmatrix}.$$

Find the following.

(a) The distribution of $z = 4y_1 - 6y_2 + y_3$
(b) The distribution of $\mathbf{z} = \begin{pmatrix} y_1 - y_2 + y_3 \\ 2y_1 + y_2 - y_3 \end{pmatrix}$
(c) $f(y_2 | y_1, y_3)$
(d) $f(y_1, y_2 | y_3)$
(e) ρ_{12} and $\rho_{12\cdot 3}$

4.18 If $\mathbf{y}$ is $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\Sigma} = \begin{pmatrix} 2 & 0 & -1 \\ 0 & 4 & 0 \\ -1 & 0 & 3 \end{pmatrix},$$

which variables are independent? (See Corollary 1 to Theorem 4.4a)

4.19 If $\mathbf{y}$ is $N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & -4 \\ 0 & 0 & -4 & 6 \end{pmatrix},$$

which variables are independent?

4.20 Show that $\sigma_{ij \cdot rs} \cdots q = \sigma_{ij} - \boldsymbol{\sigma}'_{ix} \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\sigma}_{jx}$ as in (4.39).