10 Multiple Regression: Random x's

Throughout Chapters 7–9 we assumed that the x variables were fixed; that is, that they remain constant in repeated sampling. However, in many regression applications, they are random variables. In this chapter we obtain estimators and test statistics for a regression model with random x variables. Many of these estimators and test statistics are the same as those for fixed x's, but their properties are somewhat different.

In the random-x case, $k+1$ variables $y, x_1, x_2, \ldots, x_k$ are measured on each of the n subjects or experimental units in the sample. These n observation vectors yield the data

$$\begin{array}{ccccc} y_1 & x_{11} & x_{12} & \cdots & x_{1k} \\ y_2 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ y_n & x_{n1} & x_{n2} & \cdots & x_{nk}. \end{array} \tag{10.1}$$

The rows of this array are random vectors of the second type described in Section 3.1. The variables $y, x_1, x_2, \ldots, x_k$ in a row are typically correlated and have different variances; that is, for the random vector $(y, x_1, \ldots, x_k) = (y, \mathbf{x}')$, we have

$$\operatorname{cov}\begin{pmatrix} y \\ x_1 \\ \vdots \\ x_k \end{pmatrix} = \operatorname{cov}\begin{pmatrix} y \\ \mathbf{x} \end{pmatrix} = \mathbf{\Sigma},$$

where $\mathbf{\Sigma}$ is not a diagonal matrix. The vectors themselves [rows of the array in (10.1)] are ordinarily mutually independent (uncorrelated) if they arise from a random sample.

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

In Sections 10.1–10.5 we assume that y and the x variables have a multivariate normal distribution. Many of the results in Sections 10.6–10.8 do not require a normality assumption.

10.1 MULTIVARIATE NORMAL REGRESSION MODEL

The estimation and testing results in Sections 10.1–10.5 are based on the assumption that $(y, x_1, \ldots, x_k) = (y, \mathbf{x}')$ is distributed as $N_{k+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_y \\ \hline \mu_1 \\ \vdots \\ \mu_k \end{pmatrix} = \begin{pmatrix} \mu_y \\ \boldsymbol{\mu}_x \end{pmatrix}$$

$$\boldsymbol{\Sigma} = \left(\begin{array}{c|ccc} \sigma_{yy} & \sigma_{y1} & \cdots & \sigma_{yk} \\ \hline \sigma_{1y} & \sigma_{11} & \cdots & \sigma_{1k} \\ \vdots & \vdots & & \vdots \\ \sigma_{ky} & \sigma_{k1} & \cdots & \sigma_{kk} \end{array}\right) = \begin{pmatrix} \sigma_{yy} & \boldsymbol{\sigma}'_{yx} \\ \boldsymbol{\sigma}_{yx} & \boldsymbol{\Sigma}_{xx} \end{pmatrix}, \tag{10.3}$$

where $\boldsymbol{\mu}_x$ is the mean vector for the x's, $\boldsymbol{\sigma}_{yx}$ is the vector of covariances between y and the x's, and $\boldsymbol{\Sigma}_{xx}$ is the covariance matrix for the x's.

From Corollary 1 to Theorem 4.4d, we have

$$E(y|\mathbf{x}) = \mu_y + \boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x) \tag{10.4}$$

$$= \beta_0 + \boldsymbol{\beta}'_1\mathbf{x}, \tag{10.5}$$

where

$$\beta_0 = \mu_y - \boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x, \tag{10.6}$$

$$\boldsymbol{\beta}_1 = \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{yx}. \tag{10.7}$$

From Corollary 1 to Theorem 4.4d, we also obtain

$$\mathrm{var}(y|\mathbf{x}) = \sigma_{yy} - \boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{yx} = \sigma^2. \tag{10.8}$$

The mean, $E(y|\mathbf{x}) = \mu_y + \boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)$, is a linear function of $\mathbf{x}$, but the variance, $\sigma^2 = \sigma_{yy} - \boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{yx}$, is not a function $\mathbf{x}$. Thus under the multivariate normal assumption, (10.4) and (10.8) provide a linear model with constant variance, which is analogous to the fixed-x case. Note, however, that $E(y|\mathbf{x}) = \beta_0 + \boldsymbol{\beta}_1'\mathbf{x}$ in (10.5) does not allow for curvature such as $E(y) = \beta_0 + \beta_1 x + \beta_2 x^2$. Thus $E(y|\mathbf{x}) = \beta_0 + \boldsymbol{\beta}_1'\mathbf{x}$ represents a model that is linear in the x's as well as the β's. This differs from the linear model in the fixed-x case, which requires only linearity in the β's.

10.2 ESTIMATION AND TESTING IN MULTIVARIATE NORMAL REGRESSION

Before obtaining estimators of β_0, $\boldsymbol{\beta}_1$, and σ^2 in (10.6)–(10.8), we must first estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Maximum likelihood estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are given in the following theorem.

Theorme 10.2a. If $(y_1, \mathbf{x}_1'), (y_2, \mathbf{x}_2'), \ldots, (y_n, \mathbf{x}_n')$ [rows of the array in (10.1)] is a random sample from $N_{k+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as given in (10.2) and (10.3), the maximum likelihood estimators are

$$\hat{\boldsymbol{\mu}} = \begin{pmatrix} \hat{\mu}_y \\ \hat{\boldsymbol{\mu}}_x \end{pmatrix} = \begin{pmatrix} \bar{y} \\ \bar{\mathbf{x}} \end{pmatrix}, \tag{10.9}$$

$$\hat{\boldsymbol{\Sigma}} = \frac{n-1}{n}\mathbf{S} = \frac{n-1}{n}\begin{pmatrix} s_{yy} & \mathbf{s}_{yx}' \\ \mathbf{s}_{yx} & \mathbf{S}_{xx} \end{pmatrix}, \tag{10.10}$$

where the partitioning of $\hat{\boldsymbol{\mu}}$ and $\mathbf{S}$ is analogous to the partitioning of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in (10.2) and (10.3). The elements of the sample covariance matrix $\mathbf{S}$ are defined in (7.40) and in (10.14).

PROOF. Denote $(y_i, \mathbf{x}_i')$ by $\mathbf{v}_i', i = 1, 2, \ldots, n$. As noted below (10.1), $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are independent because they arise from a random sample. The likelihood function (joint density) is therefore given by the product

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \prod_{i=1}^{n} f(\mathbf{v}_i;\, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &= \prod_{i=1}^{n} \frac{1}{(\sqrt{2\pi})^{k+1}|\boldsymbol{\Sigma}|^{1/2}} e^{-(\mathbf{v}_i-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{v}_i-\boldsymbol{\mu})/2} \\ &= \frac{1}{(\sqrt{2\pi})^{n(k+1)}|\boldsymbol{\Sigma}|^{n/2}} e^{-\sum_{i=1}^{n}(\mathbf{v}_i-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{v}_i-\boldsymbol{\mu})/2}. \end{aligned} \tag{10.11}$$

Note that $L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{i=1}^{n} f(\mathbf{v}_i;\, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a product of n multivariate normal densities, each involving $k+1$ random variables. Thus there are $n(k+1)$ random variables as compared to the likelihood $L(\boldsymbol{\beta}, \sigma^2)$ in (7.50) that involves n random variables $y_1, y_2, \ldots, y_n$ [the x's are fixed in (7.50)].

To find the maximum likelihood estimator for $\boldsymbol{\mu}$, we expand and sum the exponent in (10.11) and then take the logarithm to obtain

$$\ln L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = -n(k+1)\ln\sqrt{2\pi} - \frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_i \mathbf{v}_i'\boldsymbol{\Sigma}^{-1}\mathbf{v}_i + \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\sum_i \mathbf{v}_i - \frac{n}{2}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}. \tag{10.12}$$

Differentiating (10.12) with respect to $\boldsymbol{\mu}$ using (2.112) and (2.113) and setting the result equal to $\mathbf{0}$, we obtain

$$\frac{\partial \ln L(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu}} = -\mathbf{0} - \mathbf{0} - \mathbf{0} + \boldsymbol{\Sigma}^{-1}\sum_i \mathbf{v}_i - \frac{2n}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = \mathbf{0},$$

which gives

$$\hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{v}_i = \bar{\mathbf{v}} = \begin{pmatrix} \bar{y} \\ \bar{\mathbf{x}} \end{pmatrix},$$

where $\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k)'$ is the vector of sample means of the x's. To find the maximum likelihood estimator of $\boldsymbol{\Sigma}$, we rewrite the exponent of (10.11) and then take the logarithm to obtain

$$\begin{aligned}
\ln L(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1}) &= -n(k+1)\ln\sqrt{2\pi} + \frac{n}{2}\ln|\boldsymbol{\Sigma}^{-1}| - \frac{1}{2}\sum_i (\mathbf{v}_i - \bar{\mathbf{v}})'\boldsymbol{\Sigma}^{-1}(\mathbf{v}_i - \bar{\mathbf{v}}) \\
&\quad - \frac{n}{2}(\bar{\mathbf{v}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{v}} - \boldsymbol{\mu}) \\
&= -n(k+1)\ln\sqrt{2\pi} + \frac{n}{2}\ln|\boldsymbol{\Sigma}^{-1}| - \frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\sum_i (\mathbf{v}_i - \bar{\mathbf{v}})(\mathbf{v}_i - \bar{\mathbf{v}})'\right] \\
&\quad - \frac{n}{2}\operatorname{tr}[\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{v}} - \boldsymbol{\mu})(\bar{\mathbf{v}} - \boldsymbol{\mu})'].
\end{aligned}$$

Differentiating this with respect to $\boldsymbol{\Sigma}^{-1}$ using (2.115) and (2.116), and setting the result equal to $\mathbf{0}$, we obtain

$$\begin{aligned}
\frac{\partial \ln L(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{-1})}{\partial \boldsymbol{\Sigma}^{-1}} &= n\boldsymbol{\Sigma} - \frac{n}{2}\operatorname{diag}(\boldsymbol{\Sigma}) - \sum_i (\mathbf{v}_i - \bar{\mathbf{v}})(\mathbf{v}_i - \bar{\mathbf{v}})' + \frac{1}{2}\operatorname{diag}\left[\sum_i (\mathbf{v}_i - \bar{\mathbf{v}})(\mathbf{v}_i - \bar{\mathbf{v}})\right] \\
&\quad - n(\bar{\mathbf{v}} - \boldsymbol{\mu})(\bar{\mathbf{v}} - \boldsymbol{\mu})' + \frac{n}{2}\operatorname{diag}[(\bar{\mathbf{v}} - \boldsymbol{\mu})(\bar{\mathbf{v}} - \boldsymbol{\mu})'] = \mathbf{0}.
\end{aligned}$$

Since $\hat{\boldsymbol{\mu}} = \bar{\mathbf{v}}$, the last two terms disappear and we obtain

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{v}_i - \bar{\mathbf{v}})(\mathbf{v}_i - \bar{\mathbf{v}})' = \frac{n-1}{n}\mathbf{S}. \tag{10.13}$$

See Problem 10.1 for verification that $\sum_i (\mathbf{v}_i - \bar{\mathbf{v}})(\mathbf{v}_i - \bar{\mathbf{v}})' = (n-1)\mathbf{S}$. □

In partitioned form, the sample covariance matrix $\mathbf{S}$ can be written as in (10.10)

$$\mathbf{S} = \begin{pmatrix} s_{yy} & \mathbf{s}'_{yx} \\ \mathbf{s}_{yx} & \mathbf{S}_{xx} \end{pmatrix} = \left(\begin{array}{c|ccc} s_{yy} & s_{y1} & \cdots & s_{yk} \\ \hline s_{1y} & s_{11} & \cdots & s_{1k} \\ \vdots & \vdots & & \vdots \\ s_{ky} & s_{k1} & \cdots & s_{kk} \end{array}\right), \tag{10.14}$$

where $\mathbf{s}_{yx}$ is the vector of sample covariances between y and the x's and $\mathbf{S}_{xx}$ is the sample covariance matrix for the x's. For example

$$s_{y1} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_{i1} - \bar{x}_1)}{n-1},$$

$$s_{11} = \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2}{n-1},$$

$$s_{12} = \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)}{(n-1)}$$

[see (7.41)–(7.43)]. By (5.7), $E(s_{yy}) = \sigma_{yy}$ and $E(s_{jj}) = \sigma_{jj}$. By (5.17), $E(s_{yj}) = \sigma_{yj}$ and $E(s_{ij}) = \sigma_{ij}$. Thus $E(\mathbf{S}) = \boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is given in (10.3). The maximum likelihood estimator $\hat{\boldsymbol{\Sigma}} = (n-1)\mathbf{S}/n$ is therefore biased.

In order to find maximum likelihood estimators of $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$, and σ^2 we first note the *invariance property* of maximum likelihood estimators.

Theorem 10.2b. The maximum likelihood estimator of a function of one or more parameters is the same function of the corresponding estimators; that is, if $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator of the vector or matrix of parameters $\boldsymbol{\theta}$, then $g(\hat{\boldsymbol{\theta}})$ is the maximum likelihood estimator of $g(\boldsymbol{\theta})$.

PROOF. See Hogg and Craig (1995, p. 265). □

Example 10.2. We illustrate the use of the invariance property in Theorem 10.2b by showing that the sample correlation matrix $\mathbf{R}$ is the maximum likelihood estimator of the population correlation matrix $\mathbf{P}_\rho$ when sampling from the multivariate normal distribution. By (3.30), the relationship between $\mathbf{P}_\rho$ and $\boldsymbol{\Sigma}$ is given by $\mathbf{P}_\rho = \mathbf{D}_{\boldsymbol{\sigma}}^{-1}\boldsymbol{\Sigma}\mathbf{D}_\sigma^{-1}$, where $\mathbf{D}_\sigma = [\text{diag}(\boldsymbol{\Sigma})]^{1/2}$, so that

$$\mathbf{D}_\sigma^{-1} = \text{diag}\left(\frac{1}{\sqrt{\sigma_{11}}}, \frac{1}{\sqrt{\sigma_{22}}}, \ldots, \frac{1}{\sqrt{\sigma_{pp}}}\right).$$

The maximum likelihood estimator of $1/\sqrt{\sigma_{jj}}$ is $1/\sqrt{\hat{\sigma}_{jj}}$, where $\hat{\sigma}_{jj} = (1/n)\Sigma_{i=1}^n (y_{ij} - \bar{y}_j)^2$. Thus $\hat{\mathbf{D}}_\sigma^{-1} = \text{diag}(1/\sqrt{\hat{\sigma}_{11}}, 1/\sqrt{\hat{\sigma}_{22}}, \ldots, 1/\sqrt{\hat{\sigma}_{pp}}$, and we obtain

$$\begin{aligned}
\hat{\mathbf{P}}_\rho = \hat{\mathbf{D}}_\sigma^{-1}\hat{\boldsymbol{\Sigma}}\hat{\mathbf{D}}_\sigma^{-1} &= \left(\frac{\hat{\sigma}_{jk}}{\sqrt{\hat{\sigma}_{jj}}\sqrt{\hat{\sigma}_{kk}}}\right) \\
&= \left(\frac{\sum_i (y_{ij} - \bar{y}_j)(y_{ik} - \bar{y}_k)/n}{\sqrt{\sum_i (y_{ij} - \bar{y}_j)^2/n}\sqrt{\sum_i (y_{ik} - \bar{y}_k)^2/n}}\right) \\
&= \left(\frac{\sum_i (y_{ij} - \bar{y}_j)(y_{ik} - \bar{y}_k)}{\sqrt{\sum_i (y_{ij} - \bar{y}_j)^2}\sqrt{\sum_i (y_{ik} - \bar{y}_k)^2}}\right) \\
&= (r_{jk}) = \mathbf{R}.
\end{aligned}$$

□

Maximum likelihood estimators of $\beta_0, \boldsymbol{\beta}_1$, and σ^2 are now given in the following theorem.

Theorem 10.2c. If $(y_1, \mathbf{x}_1'), (y_2, \mathbf{x}_2'), \ldots, (y_n, \mathbf{x}_n')$, is a random sample from $N_{k+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are given by (10.2) and (10.3), the maximum likelihood estimators for β_0, $\boldsymbol{\beta}_1$, and σ^2 in (10.6)–(10.8) are as follows:

$$\hat{\beta}_0 = \bar{y} - \mathbf{s}_{yx}'\mathbf{S}_{xx}^{-1}\bar{\mathbf{x}}, \tag{10.15}$$

$$\hat{\boldsymbol{\beta}}_1 = \mathbf{S}_{xx}^{-1}\mathbf{s}_{yx}, \tag{10.16}$$

$$\hat{\sigma}^2 = \frac{n-1}{n}s^2 \quad \text{where} \quad s^2 = s_{yy} - \mathbf{s}_{yx}'\mathbf{S}_{xx}^{-1}\mathbf{s}_{yx}. \tag{10.17}$$

The estimator s^2 is a bias-corrected estimator of σ^2.

Proof. By the invariance property of maximum likelihood estimators (Theorem 10.2b), we insert (10.9) and (10.10) into (10.6), (10.7), and (10.8) to obtain the desired results (using the unbiased estimator $\mathbf{S}$ in place of $\hat{\boldsymbol{\Sigma}}$). □

The estimators $\hat{\beta}_0$, $\hat{\boldsymbol{\beta}}_1$, and s^2 have a minimum variance property analogous to that of the corresponding estimators for the case of normal y's and fixed x's in Theorem 7.6d. It can be shown that $\hat{\boldsymbol{\mu}}$ and $\mathbf{S}$ in (10.9) and (10.10) are jointly sufficient for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ (see Problem 10.2). Then, with some additional properties that can be demonstrated, it follows that $\hat{\beta}_0$, $\hat{\boldsymbol{\beta}}_1$, and s^2 are minimum variance unbiased estimators for β_0, $\boldsymbol{\beta}_1$, and σ^2 (Graybill 1976, p. 380).

The maximum likelihood estimators $\hat{\beta}_0$ and $\hat{\boldsymbol{\beta}}_1$ in (10.15) and (10.16) are the same algebraic functions of the observations as the least-squares estimators given in (7.47) and (7.46) for the fixed-x case. The estimators in (10.15) and (10.16) are also identical to the maximum likelihood estimators for normal y's and fixed x's in Section 7.6.2 (see Problem 7.17). However, even though the estimators in the random-x case and fixed-x case are the same, their distributions differ. When y and the x's are multivariate normal, $\hat{\boldsymbol{\beta}}_1$ does not have a multivariate normal distribution as it does in the fixed-x case with normal y's [Theorem 7.6b(i)]. For large n, the distribution is similar to the multivariate normal, but for small n, the distribution has heavier tails than the multivariate normal.

In spite of the nonnormality of $\hat{\boldsymbol{\beta}}_1$ in the random-x model, the F tests and t tests and associated confidence regions and intervals of Chapter 8 (fixed-x model) are still appropriate. To see this, note that since the conditional distribution of y for a given value of $\mathbf{x}$ is normal (Corollary 1 to Theorem 4.4d), the conditional distribution of the vector of observations $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$ for a given value of the $\mathbf{X}$ matrix is multivariate normal. Therefore, a test statistic such as (8.35) is distributed conditionally as an F for the given value of $\mathbf{X}$ when H_0 is true. However, the central F distribution depends only on degrees of freedom; it does not depend on $\mathbf{X}$. Thus under H_0, the statistic has (unconditionally) an F distribution for all values of $\mathbf{X}$, and so tests can be carried out exactly as in the fixed-x case.

The main difference is that when H_0 is false, the noncentrality parameter is a function of $\mathbf{X}$, which is random. Hence the noncentral F distribution does not apply to the random-x case. This only affects such things as power calculations.

Confidence intervals for the β_j's in Section 8.6.2 and for linear functions of the β_j's in Section 8.6.3 are based on the central t distribution [e.g., see (8.48)]. Thus they also remain valid for the random-x case. However, the expected width of the interval differs in the two cases (random x's and fixed x's) because of randomness in $\mathbf{X}$.

In Section 10.5, we obtain the F test for $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ using the likelihood ratio approach.

10.3 STANDARDIZED REGRESSION COEFFICENTS

We now show that the regression coefficient vector $\hat{\boldsymbol{\beta}}_1$ in (10.16) can be expressed in terms of sample correlations. By analogy to (10.14), the sample correlation matrix can be written in partitioned form as

$$\mathbf{R} = \begin{pmatrix} 1 & \mathbf{r}'_{yx} \\ \mathbf{r}_{yx} & \mathbf{R}_{xx} \end{pmatrix} = \left(\begin{array}{c|cccc} 1 & r_{y1} & r_{y2} & \dots & r_{yk} \\ \hline r_{1y} & 1 & r_{12} & \dots & r_{1k} \\ r_{2y} & r_{21} & 1 & \dots & r_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{ky} & r_{k1} & r_{k2} & \dots & 1 \end{array} \right), \tag{10.18}$$

where $\mathbf{r}_{yx}$ is the vector of correlations between y and the x's and $\mathbf{R}_{xx}$ is the correlation matrix for the x's. For example

$$r_{y2} = \frac{s_{y2}}{\sqrt{s_y^2 s_2^2}} = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_{i2} - \bar{x}_2)}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (x_{i2} - \bar{x}_2)^2}},$$

$$r_{12} = \frac{s_{12}}{\sqrt{s_1^2 s_2^2}} = \frac{\sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)}{\sqrt{\sum_{i=1}^n (x_{i1} - \bar{x}_1)^2 \sum_{i=1}^n (x_{i2} - \bar{x}_2)^2}}.$$

By analogy to (3.31), $\mathbf{R}$ can be converted to $\mathbf{S}$ by

$$\mathbf{S} = \mathbf{DRD},$$

where $\mathbf{D} = [\mathrm{diag}(\mathbf{S})]^{1/2}$, which can be written in partitioned form as

$$\mathbf{D} = \left(\begin{array}{c|cccc} s_y & 0 & 0 & \dots & 0 \\ \hline 0 & \sqrt{s_{11}} & 0 & \dots & 0 \\ 0 & 0 & \sqrt{s_{22}} & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \sqrt{s_{kk}} \end{array} \right) = \begin{pmatrix} s_y & \mathbf{0}' \\ \mathbf{0} & \mathbf{D}_x \end{pmatrix}.$$

Using the partitioned form of $\mathbf{S}$ in (10.14), $\mathbf{S} = \mathbf{DRD}$ can be written as

$$\mathbf{S} = \begin{pmatrix} s_{yy} & \mathbf{s}'_{yx} \\ \mathbf{s}_{yx} & \mathbf{S}_{xx} \end{pmatrix} = \begin{pmatrix} s_y^2 & s_y \mathbf{r}'_{yx} \mathbf{D}_x \\ s_y \mathbf{D}_x \mathbf{r}_{yx} & \mathbf{D}_x \mathbf{R}_{xx} \mathbf{D}_x \end{pmatrix}, \tag{10.19}$$

so that

$$\mathbf{S}_{xx} = \mathbf{D}_x \mathbf{R}_{xx} \mathbf{D}_x, \tag{10.20}$$

$$\mathbf{s}_{yx} = s_y \mathbf{D}_x \mathbf{r}_{yx}, \tag{10.21}$$

where $\mathbf{D}_x = \mathrm{diag}(s_1, s_2, \ldots, s_k)$ and $s_y = \sqrt{s_y^2} = \sqrt{s_{yy}}$ is the sample standard deviation of y. When (10.20) and (10.21) are substituted into (10.16), we obtain an expression for $\hat{\boldsymbol{\beta}}_1$ in terms of correlations:

$$\hat{\boldsymbol{\beta}}_1 = s_y \mathbf{D}_x^{-1} \mathbf{R}_{xx}^{-1} \mathbf{r}_{yx}. \tag{10.22}$$

The regression coefficients $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$ in $\hat{\boldsymbol{\beta}}_1$ can be standardized so as to show the effect of standardized x values (sometimes called *z scores*). We illustrate this for $k = 2$. The model in centered form [see (7.30) and an expression following (7.38)] is

$$\hat{y}_i = \bar{y} + \hat{\beta}_1(x_{i1} - \bar{x}_1) + \hat{\beta}_2(x_{i2} - \bar{x}_2).$$

This can be expressed in terms of standardized variables as

$$\frac{\hat{y}_i - \bar{y}}{s_y} = \frac{s_1}{s_y}\hat{\beta}_1\left(\frac{x_{i1} - \bar{x}_1}{s_1}\right) + \frac{s_2}{s_y}\hat{\beta}_2\left(\frac{x_{i2} - \bar{x}_2}{s_2}\right), \tag{10.23}$$

where $s_j = \sqrt{s_{jj}}$ is the standard deviation of x_j. We thus define the standardized coefficients as

$$\hat{\beta}_j^* = \frac{s_j}{s_y}\hat{\beta}_j.$$

These coefficients are often referred to as *beta weights* or *beta coefficients*. Since they are used with standardized variables $(x_{ij} - \bar{x}_j)/s_j$ in (10.23), the $\hat{\beta}_j^*$'s can be readily compared to each other, whereas the $\hat{\beta}_j$'s cannot be so compared. [Division by s_y in (10.23) is customary but not necessary; the relative values of $s_1\hat{\beta}_1$ and $s_2\hat{\beta}_2$ are the same as those of $s_1\hat{\beta}_1/s_y$ and $s_2\hat{\beta}_2/s_y$.]

The beta weights can be expressed in vector form as

$$\hat{\boldsymbol{\beta}}_1^* = \frac{1}{s_y}\mathbf{D}_x\hat{\boldsymbol{\beta}}_1.$$

Using (10.22), this can be written as

$$\hat{\boldsymbol{\beta}}_1^* = \mathbf{R}_{xx}^{-1}\mathbf{r}_{yx}. \tag{10.24}$$

Note that $\hat{\boldsymbol{\beta}}_1^*$ in (10.24) is not the same as $\hat{\boldsymbol{\beta}}_1^*$ from the reduced model in (8.8). Note also the analogy of $\hat{\boldsymbol{\beta}}_1^* = \mathbf{R}_{xx}^{-1}\mathbf{r}_{yx}$ in (10.24) to $\hat{\boldsymbol{\beta}}_1 = \mathbf{S}_{xx}^{-1}\mathbf{s}_{yx}$ in (10.16). In effect, $\mathbf{R}_{xx}$ and $\mathbf{r}_{xy}$ are the covariance matrix and covariance vector for standardized variables.

Replacing $\mathbf{S}_{xx}^{-1}$ and $\mathbf{s}_{yx}$ by $\mathbf{R}_{xx}^{-1}$ and $\mathbf{r}_{yx}$ leads to regression coefficients for standardized variables.

Example 10.3. The following six hematology variables were measured on 51 workers (Royston 1983):

$y =$ lymphocyte count	$x_3 =$ white blood cell count ($\times .01$)
$x_1 =$ hemoglobin concentration	$x_4 =$ neutrophil count
$x_2 =$ packed-cell volume	$x_5 =$ serum lead concentration

The data are given in Table 10.1.

For $\bar{y}, \bar{\mathbf{x}}, \mathbf{S}_{xx}$ and $\mathbf{s}_{yx}$, we have

$$\bar{y} = 22.902, \quad \bar{\mathbf{x}}' = (15.108, 45.196, 53.824, 25.529, 21.039),$$

$$\mathbf{S}_{xx} = \begin{pmatrix} 0.691 & 1.494 & 3.255 & 0.422 & -0.268 \\ 1.494 & 5.401 & 10.155 & 1.374 & 1.292 \\ 3.255 & 10.155 & 200.668 & 64.655 & 4.067 \\ 0.422 & 1.374 & 64.655 & 56.374 & 0.579 \\ -0.268 & 1.292 & 4.067 & 0.579 & 18.078 \end{pmatrix},$$

$$\mathbf{s}_{yx} = \begin{pmatrix} 1.535 \\ 4.880 \\ 106.202 \\ 3.753 \\ 3.064 \end{pmatrix}.$$

By (10.15) to (10.17), we obtain

$$\hat{\boldsymbol{\beta}}_1 = \mathbf{S}_{xx}^{-1}\mathbf{s}_{yx} = \begin{pmatrix} -0.491 \\ -0.316 \\ 0.837 \\ -0.882 \\ 0.025 \end{pmatrix},$$

$$\hat{\beta}_0 = \bar{y} - \mathbf{s}'_{yx}\mathbf{S}_{xx}^{-1}\bar{\mathbf{x}} = 22.902 - 1.355 = 21.547,$$

$$s^2 = s_{yy} - \mathbf{s}'_{yx}\mathbf{S}_{xx}^{-1}\mathbf{s}_{yx} = 90.2902 - 83.3542 = 6.9360.$$

TABLE 10.1 Hematology Data

Observation Number	y	x_1	x_2	x_3	x_4	x_5
1	14	13.4	39	41	25	17
2	15	14.6	46	50	30	20
3	19	13.5	42	45	21	18
4	23	15.0	46	46	16	18
5	17	14.6	44	51	31	19
6	20	14.0	44	49	24	19
7	21	16.4	49	43	17	18
8	16	14.8	44	44	26	29
9	27	15.2	46	41	13	27
10	34	15.5	48	84	42	36
11	26	15.2	47	56	27	22
12	28	16.9	50	51	17	23
13	24	14.8	44	47	20	23
14	26	16.2	45	56	25	19
15	23	14.7	43	40	13	17
16	9	14.7	42	34	22	13
17	18	16.5	45	54	32	17
18	28	15.4	45	69	36	24
19	17	15.1	45	46	29	17
20	14	14.2	46	42	25	28
21	8	15.9	46	52	34	16
22	25	16.0	47	47	14	18
23	37	17.4	50	86	39	17
24	20	14.3	43	55	31	19
25	15	14.8	44	42	24	29
26	9	14.9	43	43	32	17
27	16	15.5	45	52	30	20
28	18	14.5	43	39	18	25
29	17	14.4	45	60	37	23
30	23	14.6	44	47	21	27
31	43	15.3	45	79	23	23
32	17	14.9	45	34	15	24
33	23	15.8	47	60	32	21
34	31	14.4	44	77	39	23
35	11	14.7	46	37	23	23
36	25	14.8	43	52	19	22
37	30	15.4	45	60	25	18
38	32	16.2	50	81	38	18
39	17	15.0	45	49	26	24
40	22	15.1	47	60	33	16
41	20	16.0	46	46	22	22
42	20	15.3	48	55	23	23

(*Continued*)

TABLE 10.1 *Continued*

Observation Number	y	x_1	x_2	x_3	x_4	x_5
43	20	14.5	41	62	36	21
44	26	14.2	41	49	20	20
45	40	15.0	45	72	25	25
46	22	14.2	46	58	31	22
47	61	14.9	45	84	17	17
48	12	16.2	48	31	15	18
49	20	14.5	45	40	18	20
50	35	16.4	49	69	22	24
51	38	14.7	44	78	34	16

The correlations are given by

$$\mathbf{R}_{xx} = \begin{pmatrix} 1.000 & 0.774 & 0.277 & 0.068 & -0.076 \\ 0.774 & 1.000 & 0.308 & 0.079 & 0.131 \\ 0.277 & 0.308 & 1.000 & 0.608 & 0.068 \\ 0.068 & 0.079 & 0.608 & 1.000 & 0.018 \\ -0.076 & 0.131 & 0.068 & 0.018 & 1.000 \end{pmatrix}, \quad \mathbf{r}_{yx} = \begin{pmatrix} 0.194 \\ 0.221 \\ 0.789 \\ 0.053 \\ 0.076 \end{pmatrix}.$$

By (10.24), the standardized coefficient vector is given by

$$\hat{\boldsymbol{\beta}}_1^* = \mathbf{R}_{xx}^{-1}\mathbf{r}_{yx} = \begin{pmatrix} -0.043 \\ -0.077 \\ 1.248 \\ -0.697 \\ 0.011 \end{pmatrix}.$$

□

10.4 R^2 IN MULTIVARIATE NORMAL REGRESSION

In the case of fixed x's, we defined R^2 as the proportion of variation in y due to regression [see (7.55)]. In the case of random x's, we obtain R as an estimate of a population multiple correlation between y and the x's. Then R^2 is the square of this sample multiple correlation.

The *population multiple correlation coefficient* $\rho_{y|\mathbf{x}}$ is defined as the correlation between y and the linear function $w = \mu_y + \boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)$:

$$\rho_{y|\mathbf{x}} = \text{corr}(y, w) = \frac{\sigma_{yw}}{\sigma_y \sigma_w}. \tag{10.25}$$

(We use the subscript $y|\mathbf{x}$ to distinguish $\rho_{y|\mathbf{x}}$ from ρ, the correlation between y and x in the bivariate normal case; see Sections 3.2, 6.4, and 10.5). By (10.4), w is equal to $E(y|\mathbf{x})$, which is the population analogue of $\hat{y} = \hat{\beta}_0 + \hat{\boldsymbol{\beta}}_1'\mathbf{x}_1$, the sample predicted value of y. As $\mathbf{x}$ varies randomly, the *population predicted value* $w = \mu_y + \boldsymbol{\sigma}_{yx}'\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)$ becomes a random variable.

It is easily established that $\text{cov}(y, w)$ and $\text{var}(w)$ have the same value:

$$\text{cov}(y, w) = \text{var}(w) = \boldsymbol{\sigma}_{yx}'\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{yx}. \tag{10.26}$$

Then the population multiple correlation $\rho_{y|\mathbf{x}}$ in (10.25) becomes

$$\rho_{y|\mathbf{x}} = \frac{\text{cov}(y, w)}{\sqrt{\text{var}(y)\text{var}(w)}} = \sqrt{\frac{\boldsymbol{\sigma}_{yx}'\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{yx}}{\boldsymbol{\sigma}_{yy}}},$$

and the *population coefficient of determination* or *population squared multiple correlation* $\rho_{y|\mathbf{x}}^2$ is given by

$$\rho_{y|\mathbf{x}}^2 = \frac{\boldsymbol{\sigma}_{yx}'\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{yx}}{\boldsymbol{\sigma}_{yy}}. \tag{10.27}$$

We now list some properties of $\rho_{y|\mathbf{x}}$ and $\rho_{y|\mathbf{x}}^2$.

1. $\rho_{y|\mathbf{x}}$ is the maximum correlation between y and any linear function of $\mathbf{x}$:

$$\rho_{y|\mathbf{x}} = \max_{\alpha} \rho_{y}, \boldsymbol{\alpha}'\mathbf{x}. \tag{10.28}$$

   This is an alternative definition of $\rho_{y|\mathbf{x}}$ that is not based on the multivariate normal distribution as is the definition in (10.25).

2. $\rho_{y|\mathbf{x}}^2$ can be expressed in terms of determinants:

$$\rho_{y|\mathbf{x}}^2 = 1 - \frac{|\boldsymbol{\Sigma}|}{\sigma_{yy}|\boldsymbol{\Sigma}_{xx}|}, \tag{10.29}$$

   where $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_{xx}$ are defined in (10.3).

3. $\rho_{y|\mathbf{x}}^2$ is invariant to linear transformations on y or on the x's; that is, if $u = ay$ and $\mathbf{v} = \mathbf{B}\mathbf{x}$, where $\mathbf{B}$ is nonsingular, then

$$\rho_{u|\mathbf{v}}^2 = \rho_{y|\mathbf{x}}^2. \tag{10.30}$$

   (Note that $\mathbf{v}$ here is not the same as $\mathbf{v}_i$ used in the proof of Theorem 10.2a.)

4. Using $\text{var}(w) = \boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{yx}$ in (10.26), $\rho^2_{y|\mathbf{x}}$ in (10.27) can be written in the form
$$\rho^2_{y|\mathbf{x}} = \frac{\text{var}(w)}{\text{var}(y)}. \tag{10.31}$$
Since $w = \mu_y + \boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)$ is the population regression equation, $\rho^2_{y|\mathbf{x}}$ in (10.31) represents the proportion of the variance of y that can be attributed to the regression relationship with the variables in $\mathbf{x}$. In this sense, $\rho^2_{y|\mathbf{x}}$ is analogous to R^2 in the fixed-x case in (7.55).
5. By (10.8) and (10.27), $\text{var}(y|\mathbf{x})$ can be expressed in terms of $\rho^2_{y|\mathbf{x}}$:
$$\begin{aligned}\text{var}(y|\mathbf{x}) &= \sigma_{yy} - \boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{yx} = \sigma_{yy} - \sigma_{yy}\rho^2_{y|\mathbf{x}} \\ &= \sigma_{yy}(1 - \rho^2_{y|\mathbf{x}}).\end{aligned} \tag{10.32}$$
6. If we consider $y - w$ as a residual or error term, then $y - w$ is uncorrelated with the x's
$$\text{cov}(y - w, \mathbf{x}) = \mathbf{0}' \tag{10.33}$$
(see Problem 10.8).

We can obtain a maximum likelihood estimator for $\rho^2_{y|\mathbf{x}}$ by substituting estimators from (10.14) for the parameters in (10.27):
$$R^2 = \frac{\mathbf{s}'_{yx}\mathbf{S}_{xx}^{-1}\mathbf{s}_{yx}}{s_{yy}} \tag{10.34}$$

We use the notation R^2 rather than $\hat{\rho}^2_{y|\mathbf{x}}$ because (10.34) is recognized as having the same form as R^2 for the fixed-x case in (7.59). We refer to R^2 as the *sample coefficient of determination* or as the *sample squared multiple correlation*. The square root of R^2
$$R = \sqrt{\frac{\mathbf{s}'_{yx}\mathbf{S}_{xx}^{-1}\mathbf{s}_{yx}}{s_{yy}}} \tag{10.35}$$
is the *sample multiple correlation coefficient*.

We now list several properties of R and R^2, some of which are analogous to properties of $\rho^2_{y|\mathbf{x}}$ above.

1. R is equal to the correlation between y and $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k = \hat{\beta}_0 + \hat{\boldsymbol{\beta}}'_1\boldsymbol{x}$:
$$R = r_{y\hat{y}}. \tag{10.36}$$
2. R is equal to the maximum correlation between y and any linear combination of the x's, $\mathbf{a}'\mathbf{x}$:
$$R = \max_{\mathbf{a}} \; r_{y,\mathbf{a}'\mathbf{x}}. \tag{10.37}$$

3. R^2 can be expressed in terms of correlations:

$$R^2 = \mathbf{r}'_{yx}\mathbf{R}_{xx}^{-1}\mathbf{r}_{yx}, \tag{10.38}$$

where $\mathbf{r}_{yx}$ and $\mathbf{R}_{xx}$ are from the sample correlation matrix $\mathbf{R}$ partitioned as in (10.18).

4. R^2 can be obtained from $\mathbf{R}^{-1}$:

$$R^2 = 1 - \frac{1}{r^{yy}}, \tag{10.39}$$

where r^{yy} is the first diagonal element of $\mathbf{R}^{-1}$. Using the other diagonal elements of $\mathbf{R}^{-1}$, this relationship can be extended to give the multiple correlation of any x_j with the other x's and y. Thus from $\mathbf{R}^{-1}$ we obtain multiple correlations, as opposed to the simple correlations in $\mathbf{R}$.

5. R^2 can be expressed in terms of determinants:

$$R^2 = 1 - \frac{|\mathbf{S}|}{s_{yy}|\mathbf{S}_{xx}|} \tag{10.40}$$

$$= 1 - \frac{|\mathbf{R}|}{|\mathbf{R}_{xx}|}, \tag{10.41}$$

where $\mathbf{S}_{xx}$ and $\mathbf{R}_{xx}$ are defined in (10.14) and (10.18).

6. From (10.24) and (10.38), we can express R^2 in terms of beta weights:

$$R^2 = \mathbf{r}'_{yx}\hat{\boldsymbol{\beta}}_1^*, \tag{10.42}$$

where $\hat{\boldsymbol{\beta}}_1^* = \mathbf{R}_{xx}^{-1}\mathbf{r}_{yx}$. This equation does *not* imply that R^2 is the sum of squared partial correlations (Section 10.8).

7. If $\rho^2_{y|\mathbf{x}} = 0$, the expected value of R^2 is given by

$$E(R^2) = \frac{k}{n-1}. \tag{10.43}$$

Thus R^2 is biased when $\rho^2_{y|\mathbf{x}}$ is 0 [this is analogous to (7.57)].

8. $R^2 \geq \max_j r^2_{yj}$, where r_{yj} is an element of $\mathbf{r}'_{yx} = (r_{y1}, r_{y2}, \ldots, r_{yk})$.
9. R^2 is invariant to full rank linear transformations on y or on the x's.

Example 10.4. For the hematology data in Table 10.1, $\mathbf{S}_{xx}, \mathbf{s}_{yx}, \mathbf{R}_{xx}$, and $\mathbf{r}_{yx}$ were obtained in Example 10.3. Using either (10.34) or (10.38), we obtain

$$R^2 = .9232.$$

□

10.5 TESTS AND CONFIDENCE INTERVALS FOR R^2

Note that by (10.27), $\rho^2_{y|\mathbf{x}} = 0$ becomes

$$\rho^2_{y|\mathbf{x}} = \frac{\boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}^{-1}_{xx}\boldsymbol{\sigma}_{yx}}{\sigma_{yy}} = 0,$$

which leads to $\boldsymbol{\sigma}_{yx} = \mathbf{0}$ since $\boldsymbol{\Sigma}_{xx}$ is positive definite. Then by (10.7), $\boldsymbol{\beta}_1 = \boldsymbol{\Sigma}^{-1}_{xx}\boldsymbol{\sigma}_{yx} = \mathbf{0}$, and $H_0 : \rho^2_{y|\mathbf{x}} = 0$ is equivalent to $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$.

The F statistic for fixed x's is given in (8.5), (8.22), and (8.23) as

$$F = \frac{(\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{y}^2)/k}{(\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y})/(n-k-1)}$$

$$= \frac{R^2/k}{(1-R^2)/(n-k-1)}. \tag{10.44}$$

The test statistic in (10.44) can be obtained by the likelihood ratio approach in the case of random x's (Anderson 1984, pp. 140–142):

Theorem 10.5. If $(y_1, \mathbf{x}'_1), (y_2, \mathbf{x}'_2), \ldots, (y_n, \mathbf{x}'_n)$ is a random sample from $N_{k+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are given by (10.2) and (10.3), the likelihood ratio test for $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$ or equivalently $H_0 : \rho^2_{y|\mathbf{x}} = 0$ can be based on F in (10.44). We reject H_0 if $F \geq F_{\alpha, k, n-k-1}$.

Proof. Using the notation $\mathbf{v}'_i = (y_i, \mathbf{x}'_i)$, as in the proof of Theorem 10.2a, the likelihood function $L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{i=1}^n f(\mathbf{v}_i;\ \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is given by (10.11), and the likelihood ratio is

$$\mathrm{LR} = \frac{\max_{H_0} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\max_{H_1} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})}.$$

Under H_1, the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are essentially unrestricted, and we have

$$\max_{H_1} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \max L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = L(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}),$$

where $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are the maximum likelihood estimators in (10.9) and (10.10). Since $(\mathbf{v}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{v}_i - \boldsymbol{\mu})$ is a scalar, the exponent of $L(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in (10.11) can be written as

$$\frac{\sum_{i=1}^{n} \text{tr}\left[(\mathbf{v}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{v}_i - \boldsymbol{\mu})\right]}{2} = \frac{\sum_{i=1}^{n} \text{tr}\left[\boldsymbol{\Sigma}^{-1}(\mathbf{v}_i - \boldsymbol{\mu})(\mathbf{v}_i - \boldsymbol{\mu})'\right]}{2}$$
$$= \frac{\text{tr}\left[\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}(\mathbf{v}_i - \boldsymbol{\mu})(\mathbf{v}_i - \boldsymbol{\mu})'\right]}{2}.$$

Then substitution of $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in $L(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ gives

$$\max_{H_1} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = L(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = \frac{1}{(\sqrt{2\pi})^{n(k+1)}|\hat{\boldsymbol{\Sigma}}|^{n/2}} e^{-\text{tr}(\hat{\boldsymbol{\Sigma}}^{-1} n\hat{\boldsymbol{\Sigma}}/2)}$$
$$= \frac{e^{-n(k+1)/2}}{(\sqrt{2\pi})^{n(k+1)}|\hat{\boldsymbol{\Sigma}}|^{n/2}}.$$

Under $H_0 : \rho^2_{y|\mathbf{x}} = 0$, we have $\boldsymbol{\sigma}_{yx} = \mathbf{0}$, and $\boldsymbol{\Sigma}$ in (10.3) becomes

$$\boldsymbol{\Sigma}_0 = \begin{pmatrix} \sigma_{yy} & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Sigma}_{xx} \end{pmatrix}, \tag{10.45}$$

whose maximum likelihood estimator is

$$\hat{\boldsymbol{\Sigma}}_0 = \begin{pmatrix} \hat{\sigma}_{yy} & \mathbf{0}' \\ \mathbf{0} & \hat{\boldsymbol{\Sigma}}_{xx} \end{pmatrix}. \tag{10.46}$$

Using $\hat{\boldsymbol{\Sigma}}_0$ in (10.46) and $\hat{\boldsymbol{\mu}} = \bar{\mathbf{v}}$ in (10.9), we have

$$\max_{H_0} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = L(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}_0) = \frac{1}{(\sqrt{2\pi})^{n(k+1)}|\hat{\boldsymbol{\Sigma}}_0|^{n/2}} e^{-\text{tr}(\hat{\boldsymbol{\Sigma}}_0^{-1} n\hat{\boldsymbol{\Sigma}}_0/2)}.$$

By (2.74), this becomes

$$L(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}_0) = \frac{e^{-n(k+1)/2}}{(\sqrt{2\pi})^{n(k+1)}\hat{\sigma}_{yy}^{n/2}|\hat{\boldsymbol{\Sigma}}_{xx}|^{n/2}}. \tag{10.47}$$

Thus

$$\text{LR} = \frac{|\hat{\boldsymbol{\Sigma}}|^{n/2}}{\hat{\sigma}_{yy}^{n/2}|\hat{\boldsymbol{\Sigma}}_{xx}|^{n/2}}. \tag{10.48}$$

Substituting $\hat{\boldsymbol{\Sigma}} = (n-1)\mathbf{S}/n$ and using (10.40), we obtain

$$\text{LR} = (1 - R^2)^{n/2}. \tag{10.49}$$

We reject H_0 for $(1 - R^2)^{n/2} \leq c$, which is equivalent to

$$F = \frac{R^2/k}{(1 - R^2)/(n - k - 1)} \geq F_{\alpha, k, n-k-1},$$

since $R^2/(1 - R^2)$ is a monotone increasing function of R^2 and F is distributed as $F(k, n - k - 1)$ when H_0 is true (Anderson 1984, pp. 138–139). $\square$

When $k = 1$, F in (10.44) reduces to $F = (n-2)r^2/(1 - r^2)$. Then, by Problem 5.16

$$t = \frac{\sqrt{n-2}\,r}{\sqrt{1 - r^2}}$$

[see (6.20)] has a t distribution with $n - 2$ degrees of freedom (df) when (y, x) has a bivariate normal distribution with $\rho = 0$.

If (y, x) is bivariate normal and $\rho \neq 0$, then $\text{var}(r) = (1 - \rho^2)^2/n$ and the function

$$u = \frac{\sqrt{n}(r - \rho)}{1 - \rho^2} \tag{10.50}$$

is approximately standard normal for large n. However, the distribution of u approaches normality very slowly as n increases (Kendall and Stuart 1969, p. 236). Its use is questionable for $n < 500$.

Fisher (1921) found a function of r that approaches normality much faster than does (10.50) and can thereby be used with much smaller n than that required for (10.50). In addition, the variance is almost independent of ρ. Fisher's function is

$$z = \frac{1}{2} \ln \frac{1 + r}{1 - r} = \tanh^{-1} r, \tag{10.51}$$

where $\tanh^{-1} r$ is the inverse hyperbolic tangent of r. The approximate mean and variance of z are

$$E(z) \cong \frac{1}{2} \ln \frac{1 + \rho}{1 - \rho} = \tanh^{-1} \rho, \tag{10.52}$$

$$\text{var}(z) \cong \frac{1}{n - 3}. \tag{10.53}$$

We can use Fisher's z transformation in (10.51) to test hypotheses such as $H_0: \rho = \rho_0$ or $H_0: \rho_1 = \rho_2$. To test $H_0: \rho = \rho_0$ vs. $H_1: \rho \neq \rho_0$, we calculate

$$v = \frac{z - \tanh^{-1}\rho_0}{\sqrt{1/(n-3)}}, \tag{10.54}$$

which is approximately distributed as the standard normal $N(0, 1)$. We reject H_0 if $|v| \geq z_{\alpha/2}$, where $z = \tanh^{-1} r$ and $z_{\alpha/2}$ is the upper $\alpha/2$ percentage point of the standard normal distribution. To test $H_0: \rho_1 = \rho_2$ vs. $H_1: \rho_1 \neq \rho_2$ for two independent samples of sizes n_1 and n_2 yielding sample correlations r_1 and r_2, we calculate

$$v = \frac{z_1 - z_2}{\sqrt{1/(n_1 - 3) + 1/(n_2 - 3)}} \tag{10.55}$$

and reject H_0 if $|v| \geq z_{\alpha/2}$, where $z_1 = \tanh^{-1} r_1$ and $z_2 = \tanh^{-1} r_2$. To test $H_0: \rho_1 = \cdots = \rho_q$ for $q > 2$, see Problem 10.18.

To obtain a confidence interval for ρ, we note that since z in (10.51) is approximately normal, we can write

$$P\left(-z_{\alpha/2} \leq \frac{z - \tanh^{-1}\rho}{1/\sqrt{n-3}} \leq z_{\alpha/2}\right) \cong 1 - \alpha. \tag{10.56}$$

Solving the inequality for ρ, we obtain the approximate $100(1-\alpha)\%$ confidence interval

$$\tanh\left(z - \frac{z_{\alpha/2}}{\sqrt{n-3}}\right) \leq \rho \leq \tanh\left(z + \frac{z_{\alpha/2}}{\sqrt{n-3}}\right). \tag{10.57}$$

A confidence interval for $\rho^2_{y|x}$ was given by Helland (1987).

Example 10.5a. For the hematology data in Table 10.1, we obtained R^2 in Example 10.4. The overall F test of $H_0: \boldsymbol{\beta}_1 = \mathbf{0}$ or $H_0: \rho^2_{y|\mathbf{x}} = 0$ is carried out using F in (10.44):

$$\begin{aligned} F &= \frac{R^2/k}{(1 - R^2)/(n - k - 1)} \\ &= \frac{.9232/5}{(1 - .9232)/45} = 108.158. \end{aligned}$$

The p value is less than 10^{-16}. $\square$

Example 10.5b. To illustrate Fisher's z transformation in (10.51) and its use to compare two independent correlations in (10.55), we divide the hematology data in Table 10.1 into two subsamples of sizes $n_1 = 26$ and $n_2 = 25$ (the first 26 observations and the last 25 observations). For the correlation between y and x_1 in each of the two subsamples, we obtain $r_1 = .4994$ and $r_2 = .0424$. The z transformation in (10.51) for each of these two values is given by

$$z_1 = \tanh^{-1} r_1 = .5485,$$
$$z_2 = \tanh^{-1} r_2 = .0425.$$

To test $H_0 : \rho_1 = \rho_2$, we use the approximate test statistic (10.55) to obtain

$$v = \frac{.5485 - .0425}{\sqrt{1/(26-3) + 1/(25-3)}} = 1.6969.$$

Since $1.6969 < z_{.025} = 1.96$, we do not reject H_0.

To obtain approximate 95% confidence limits for ρ_1, we use (10.57):

$$\text{Lower limit for } \rho_1 : \quad \tanh\left(.5485 - \frac{1.96}{\sqrt{23}}\right) = .1389,$$
$$\text{Upper limit for } \rho_1 : \quad \tanh\left(.5485 + \frac{1.96}{\sqrt{23}}\right) = .7430.$$

For ρ_2, the limits are given by

$$\text{Lower limit for } \rho_2 : \quad \tanh\left(.0425 - \frac{1.96}{\sqrt{22}}\right) = -.3587,$$
$$\text{Upper limit for } \rho_2 : \quad \tanh\left(.0425 + \frac{1.96}{\sqrt{22}}\right) = .4303.$$

□

10.6 EFFECT OF EACH VARIABLE ON R^2

The contribution of a variable x_j to the multiple correlation R will, in general, be different from its bivariate correlation with y; that is, the increase in R^2 when x_j is added is not equal to $r^2_{yx_j}$. This increase in R_2 can be either more or less than $r^2_{yx_j}$. It seems clear that relationships with other variables can render a variable partially redundant and thereby reduce the contribution of x_j to R^2, but it is not intuitively apparent how the contribution of x_j to R^2 can exceed $r^2_{yx_j}$. The latter phenomenon has been illustrated numerically by Flury (1989) and Hamilton (1987).

In this section, we provide a breakdown of the factors that determine how much each variable adds to R^2 and show how the increase in R^2 can exceed $r_{yx_j}^2$ (Rencher 1993). We first introduce some notation. The variable of interest is denoted by z, which can be one of the x's or a new variable added to the x's. We make the following additional notational definitions:

R_{yw}^2 = squared multiple correlation between y and $\mathbf{w} = (x_1, x_2, \ldots, x_k, z)'$.

R_{yx}^2 = squared multiple correlation between y and $\mathbf{x} = (x_1, x_2, \ldots, x_k)'$.

$R_{zx}^2 = \mathbf{s}_{zx}'\mathbf{S}_{xx}^{-1}\mathbf{s}_{zx}/s_z^2$ = squared multiple correlation between z and $\mathbf{x}$.

r_{yz} = simple correlation between y and z .

$\mathbf{r}_{yx} = (r_{yx_1}, r_{yx_2}, \ldots, r_{yx_k})'$ = vector of correlations between y and $\mathbf{x}$.

$\mathbf{r}_{zx} = (r_{zx_1}, r_{zx_2}, \ldots, r_{zx_k})'$ = vector of correlations between z and $\mathbf{x}$.

$\hat{\boldsymbol{\beta}}_{zx}^* = \mathbf{R}_{xx}^{-1}\mathbf{r}_{zx}$ is the vector of standardized regression coefficients (beta weights) of z regressed on $\mathbf{x}$ [see (10.24)].

The effect of z on R^2 is formulated in the following theorem.

Theorem 10.6. The increase in R^2 due to z can be expressed as

$$R_{yw}^2 - R_{yx}^2 = \frac{(\hat{r}_{yz} - r_{yz})^2}{1 - R_{zx}^2}, \tag{10.58}$$

where $\hat{r}_{yz} = \hat{\boldsymbol{\beta}}_{zx}^{*\prime}\mathbf{r}_{yx}$ is a "predicted" value of r_{yz} based on the relationship of z to the x's.

PROOF. See Problem 10.19. □

Since the right side of (10.58) is positive, R^2 cannot decrease with an additional variable, which is a verification of property 3 in Section 7.7. If z is orthogonal to $\mathbf{x}$ (i.e., if $\boldsymbol{r}_{zx} = \mathbf{0}$), then $\hat{\boldsymbol{\beta}}_{zx}^* = \mathbf{0}$, which implies that $\hat{r}_{yz} = 0$ and $R_{zx}^2 = 0$. In this case, (10.58) can be written as $R_{yw}^2 = R_{yx}^2 + r_{yz}^2$, which verifies property 5 of Section 7.7.

It is clear in Theorem 10.6 that the contribution of z to R^2 can either be less than or greater than r_{yz}^2. If $\hat{r}_{yz}$ is close to r_{yz}, the contribution of z is less than r_{yz}^2. There are three ways in which the contribution of z can exceed r_{yz}^2: (1) $\hat{r}_{yz}$ is substantially larger in absolute value than r_{yz}, (2) $\hat{r}_{yz}$ and r_{yz} are of opposite signs, and (3) R_{zx}^2 is large.

In many cases, the researcher may find it helpful to know why a variable contributed more than expected or less than expected. For example, admission to a university or professional school may be based on previous grades and the score on a standardized national test. An applicant for admission to a university with limited enrollment would submit high school grades and a national test score. These might be entered into a regression equation to obtain a predicted value of first-year grade-point average at the university. It is typically found that the standardized test increases R^2 only slightly above that based on high school grades alone. This small increase in R^2 would be disappointing to admissions officials who had hoped that the national test score might be a more useful predictor than high school grades. The designers of such standardized tests may find it beneficial to know precisely why the test makes such an unexpectedly small contribution relative to high school grades.

In Theorem 10.6, we have available the specific information needed by the designer of the standardized test. To illustrate the use of (10.58), let y be the grade-point average for the first year at the university, let z be the score on the standardized test, and let $x_1, x_2, \ldots, x_k$ be high school grades in key subject areas. By (10.58), the increase in R^2 due to z is $(\hat{r}_{yz} - r_{yz})^2/(1 - R^2_{zx})$, in which we see that z adds little to R^2 if $\hat{r}_{yz}$ is close to r_{yz}. We could examine the coefficients in $\hat{r}_{yz} = \hat{\boldsymbol{\beta}}^{*\prime}_{zx}\mathbf{r}_{yx}$ to determine which of the r_{yx_j}'s in $\mathbf{r}_{yx}$ have the most effect. This information could be used in redesigning the questions so as to reduce these particular r_{yx_j}'s. It may also be possible to increase the contribution of z to R^2_{yw} by increasing R^2_{zx} (thereby reducing $1 - R^2_{zx}$). This might be done by designing the questions in the standardized test so that the test score z is more correlated with high school grades, $x_1, x_2, \ldots, x_q$.

Theil and Chung (1988) proposed a measure of the relative importance of a variable in multiple regression based on information theory.

Example 10.6. For the hematology data in Table 10.1, the overall R^2_{yw} was found in Example 10.4 to be .92318. From Theorem 10.6, the increase in R^2 due to a variable z has the breakdown $R^2_{yw} - R^2_{yx} = (\hat{r}_{yz} - r_{yz})^2/(1 - R^2_{zx})$, where z represents any one of $x_1, x_2, \ldots, x_5$, and x represents the other four variables. The values of $\hat{r}_{yz}$, r_{yz}, R^2_{zx}, $R^2_{yw} - R^2_{yx}$, and F are given below for each variable in turn as z:

z	$\hat{r}_{yz}$	r_{yz}	R^2_{zx}	$R^2_{yw} - R^2_{yx}$	F	p value
x_1	.2101	.1943	.6332	.00068	0.4	.53
x_2	.2486	.2210	.6426	.00213	1.25	.26
x_3	.0932	.7890	.4423	.86820	508.6	0
x_4	.4822	.0526	.3837	.29945	175.4	0
x_5	.0659	.0758	.0979	.00011	0.064	.81

The F value is from the partial F test in (8.25), (8.37), or (8.39) for the significance of the increase in R^2 due to each variable.

An interesting variable here is x_4, whose value of r_{yz} is .0526, the smallest among the five variables. Despite this small individual correlation with y, x_4 contributes much more to R^2_{yw} than do all other variables except x_3 because $\hat{r}_{yz}$ is much greater for x_4 than for the other variables. This illustrates how the contribution of a variable can be augmented in the presence of other variables as reflected in $\hat{r}_{yz}$.

The difference between the two major contributors x_3 and x_4 may be very revealing to the researcher. The contribution of x_3 to R^2_{yw} is due mostly to its own correlation with y, whereas virtually all the effect of x_4 comes from its association with the other variables as reflected in $\hat{r}_{yz}$. □

10.7 PREDICTION FOR MULTIVARIATE NORMAL OR NONNORMAL DATA

In this section, we consider an approach to modeling and estimation in the random-x case that is somewhat reminiscent of least squares in the fixed-x case. Suppose that $(y, \mathbf{x}') = (y, x_1, x_2, \ldots, x_k)$ is not necessarily assumed to be multivariate normal and we wish to find a function $t(\mathbf{x})$ for predicting y. In order to find a predicted value $t(\mathbf{x})$ that is expected to be "close" to y, we will choose the function $t(\mathbf{x})$ that minimizes the mean squared error $E[y - t(\mathbf{x})]^2$, where the expectation is in the joint distribution of $y, x_1, \ldots, x_k$. This function is given in the following theorem.

Theorem 10.7. For the random vector $(y, \mathbf{x}')$, the function $t(\mathbf{x})$ that minimizes the mean squared error $E[y - t(\mathbf{x})]^2$ is given by $E(y|\mathbf{x})$.

PROOF. For notational simplicity, we use $k = 1$. By (4.28), the joint density $g(y, x)$ can be written as $g(y, x) = f(y|x)h(x)$. Then

$$\begin{aligned} E[y - t(x)]^2 &= \int\int [y - t(x)]^2 g(y, x)\, dy\, dx \\ &= \int\int [y - t(x)]^2 f(y|x)h(x)\, dy\, dx \\ &= \int h(x) \left\{ \int [y - t(x)]^2 f(y|x)\, dy \right\} dx. \end{aligned}$$

To find the function $t(x)$ that minimizes $E(y - t)^2$, we differentiate with respect to t and set the result equal to 0 [for a more general proof not involving differentiation, see Graybill (1976, pp. 432–434) or Christensen (1996, p. 119)]. Assuming that we can interchange integration and differentiation, we obtain

$$\frac{\partial E[y - t(x)]^2}{\partial t} = \int h(x) \left\{ \int 2(-1)[y - t(x)]\, f(y|x) dy \right\} dx = 0,$$

which gives

$$\begin{aligned} &2 \int h(x) \left[\int y f(y|x) dy - \int t(x) f(y|x) dy \right] dx = 0, \\ &2 \int h(x)[E(y|x) - t(x)] dx = 0. \end{aligned}$$

The left side is 0 if

$$t(x) = E(y|x).$$

□

In the case of the multivariate normal, the prediction function $E(y|\mathbf{x})$ is a linear function of $\mathbf{x}$ [see (10.4) and (10.5)]. However, in general, $E(y|\mathbf{x})$ is not linear. For an illustration of a nonlinear $E(y|x)$, see Example 3.2, in which we have $E(y|x) = \frac{1}{2}(1 + 4x - 2x^2)$.

If we restrict $t(\mathbf{x})$ to *linear* functions of $\mathbf{x}$, then the optimal result is the same linear function as in the multivariate normal case [see (10.6) and (10.7)].

Theorem 10.7b. The linear function $t(\mathbf{x})$ that minimizes $E[y - t(\mathbf{x})]^2$ is given by $t(\mathbf{x}) = \beta_0 + \boldsymbol{\beta}_1'\mathbf{x}$, where

$$\beta_0 = \mu_y - \boldsymbol{\sigma}_{yx}'\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\mu}_x, \tag{10.59}$$

$$\boldsymbol{\beta}_1 = \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{yx}. \tag{10.60}$$

PROOF. See Problem 10.21. □

We can find estimators $\hat{\beta}_0$ and $\hat{\boldsymbol{\beta}}_1$ for β_0 and $\boldsymbol{\beta}_1$ in (10.59) and (10.60) by minimizing the sample mean squared error, $\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\boldsymbol{\beta}}_1'\mathbf{x}_i)^2/n$. The results are given in the following theorem.

Theorem 10.7c. If $(y_1, \mathbf{x}_1'), (y_2, \mathbf{x}_2'), \ldots, (y_n, \mathbf{x}_n')$ is a random sample with mean vector and covariance matrix

$$\hat{\boldsymbol{\mu}} = \begin{pmatrix} \bar{y} \\ \bar{\mathbf{x}} \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} s_{yy} & \mathbf{s}_{yx}' \\ \mathbf{s}_{yx} & \mathbf{S}_{xx} \end{pmatrix},$$

then the estimators $\hat{\beta}_0$ and $\hat{\boldsymbol{\beta}}_1$ that minimize $\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\boldsymbol{\beta}}_1'\mathbf{x}_i)^2/n$ are given by

$$\hat{\beta}_0 = \bar{y} - \mathbf{s}_{yx}'\mathbf{S}_{xx}^{-1}\bar{\mathbf{x}}, \tag{10.61}$$

$$\hat{\boldsymbol{\beta}}_1 = \mathbf{S}_{xx}^{-1}\mathbf{s}_{yx}. \tag{10.62}$$

PROOF. See Problem 10.22. □

The estimators $\hat{\beta}_0$ and $\hat{\boldsymbol{\beta}}_1$ in (10.61) and (10.62) are the same as the maximum likelihood estimators in the normal case [see (10.15) and (10.16)].

10.8 SAMPLE PARTIAL CORRELATIONS

Partial correlations were introduced in Sections 4.5 and 7.10. Assuming multivariate normality, the population partial correlation $\rho_{ij\cdot rs\cdots q}$ is the correlation between y_i and y_j in the conditional distribution of $\mathbf{y}$ given $\mathbf{x}$, where y_i and y_j are in $\mathbf{y}$ and the subscripts $r, s, \ldots, q$ represent all the variables in $\mathbf{x}$. By (4.36), we obtain

$$\rho_{ij\cdot rs\ldots q} = \frac{\sigma_{ij\cdot rs\ldots q}}{\sqrt{\sigma_{ii\cdot rs\ldots q}\sigma_{jj\cdot rs\ldots q}}}, \tag{10.63}$$

where $\sigma_{ij\cdot rs\cdots q,}$ is the (ij) element of $\boldsymbol{\Sigma}_{y\cdot x} = \mathrm{cov}(\mathbf{y}|\mathbf{x})$. For normal populations, $\boldsymbol{\Sigma}_{y\cdot x}$ is given by (4.27) as $\boldsymbol{\Sigma}_{y\cdot x} = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}$, where $\boldsymbol{\Sigma}_{yy}, \boldsymbol{\Sigma}_{yx}, \boldsymbol{\Sigma}_{xx}$, and $\boldsymbol{\Sigma}_{yx}$ are from the partitioned covariance matrix

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

[see (3.33)]. The matrix of (population) partial correlations $\rho_{ij\cdot rs\ldots q}$ can be found by (4.37):

$$\mathbf{P}_{y\cdot x} = \mathbf{D}_{y\cdot x}^{-1}\boldsymbol{\Sigma}_{y\cdot x}\mathbf{D}_{y\cdot x}^{-1} = \mathbf{D}_{y\cdot x}^{-1}(\boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy})\mathbf{D}_{y\cdot x}^{-1}, \tag{10.64}$$

where $\mathbf{D}_{y\cdot x} = [\mathrm{diag}(\boldsymbol{\Sigma}_{y\cdot x})]^{1/2}$.

To obtain a maximum likelihood estimator $\mathbf{R}_{y\cdot x} = (r_{ij\cdot rs\ldots q})$ of $\mathbf{P}_{y\cdot x} = (\rho_{ij\cdot rs\ldots q})$ in (10.64), we use the invariance property of maximum likelihood estimators (Theorem 10.2b) to obtain

$$\mathbf{R}_{y\cdot x} = \mathbf{D}_s^{-1}(\mathbf{S}_{yy} - \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy})\mathbf{D}_s^{-1}, \tag{10.65}$$

where

$$\mathbf{D}_s = [\mathrm{diag}(\mathbf{S}_{yy} - \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy})]^{1/2}.$$

The matrices $\mathbf{S}_{yy}, \mathbf{S}_{yx}, \mathbf{S}_{xx}$, and $\mathbf{S}_{xy}$ are from the sample covariance matrix partitioned by analogy to $\boldsymbol{\Sigma}$ above

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{yy} & \mathbf{S}_{yx} \\ \mathbf{S}_{xy} & \mathbf{S}_{xx} \end{pmatrix},$$

where

$$\mathbf{S}_{yy} = \begin{pmatrix} s_{y_1}^2 & s_{y_1y_2} & \cdots & s_{y_1y_p} \\ s_{y_2y_1} & s_{y_2}^2 & \cdots & s_{y_2y_p} \\ \vdots & \vdots & & \vdots \\ s_{y_py_1} & s_{y_py_2} & \cdots & s_{y_p}^2 \end{pmatrix} \quad \text{and}$$

$$\mathbf{S}_{yx} = \begin{pmatrix} s_{y_1x_1} & s_{y_1x_2} & \cdots & s_{y_1x_q} \\ s_{y_2x_1} & s_{y_2x_2} & \cdots & s_{y_2x_q} \\ \vdots & \vdots & & \vdots \\ s_{y_px_1} & s_{y_px_2} & \cdots & s_{y_px_q} \end{pmatrix}$$

are estimators of $\boldsymbol{\Sigma}_{yy}$ and $\boldsymbol{\Sigma}_{yx}$. Thus the maximum likelihood estimator of $\rho_{ij\cdot rs\ldots q}$ in (10.63) is $r_{ij\cdot rs\ldots q}$, the (ij) th element of $\mathbf{R}_{y\cdot x}$ in (10.65).

We now consider two other expressions for partial correlation and show that they are equivalent to $r_{ij\cdot rs\ldots q}$ in (10.65). To simplify exposition, we illustrate with $r_{12\cdot 3}$. The sample partial correlation of y_1 and y_2 with y_3 held fixed is usually given as

$$r_{12\cdot 3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}, \tag{10.66}$$

where r_{12}, r_{13}, and r_{23} are the ordinary correlations between y_1 and y_2, y_1 and y_3, and y_2 and y_3, respectively. In the following theorem, we relate $r_{12\cdot 3}$ to two previous definitions of partial correlation.

Theorem 10.8a. The expression for $r_{12\cdot 3}$ in (10.66) is equivalent to an element of $\mathbf{R}_{y\cdot x}$ in (10.65) and is also equal to $r_{y_1-\hat{y}_1, y_2-\hat{y}_2}$ from (7.94), where $y_1 - \hat{y}_1$ and $y_2 - \hat{y}_2$ are residuals from regression of y_1 on y_3 and y_2 on y_3.

Proof. We first consider $r_{y_1-\hat{y}_1, y_2-\hat{y}_2}$, which is not a maximum likelihood estimator and can therefore be used when the data are not normal. We obtain $\hat{y}_1$ and $\hat{y}_1$ by regressing y_1 on y_3 and y_2 on y_3. Using the notation in Section 7.10, we indicate the predicted value of y_1 based on regression of y_1 on y_3 as $\hat{y}_1(y_3)$. With a similar definition of $\hat{y}_2(y_3)$, the residuals can be expressed as

$$u_1 = y_1 - \hat{y}_1(y_3) = y_1 - (\hat{\beta}_{01} + \hat{\beta}_{11}y_3),$$
$$u_2 = y_2 - \hat{y}_2(y_3) = y_2 - (\hat{\beta}_{02} + \hat{\beta}_{12}y_3),$$

where, by (6.5), $\hat{\beta}_{11}$ and $\hat{\beta}_{12}$ are the usual least-squares estimators

$$\hat{\beta}_{11} = \frac{\sum_{i=1}^n (y_{1i} - \bar{y}_1)(y_{3i} - \bar{y}_3)}{\sum_{i=1}^n (y_{3i} - \bar{y}_3)^2}, \tag{10.67}$$

$$\hat{\beta}_{12} = \frac{\sum_{i=1}^n (y_{2i} - \bar{y}_2)(y_{3i} - \bar{y}_3)}{\sum_{i=1}^n (y_{3i} - \bar{y}_3)^2}. \tag{10.68}$$

Then the sample correlation between $u_1 = y_1 - \hat{y}_1(y_3)$ and $u_2 = y_2 - \hat{y}_2(y_3)$ [see (7.94)] is

$$\begin{aligned} r_{u_1u_2} &= r_{y_1-\hat{y}_1, y_2-\hat{y}_2} \\ &= \frac{\widehat{\text{cov}}(u_1, u_2)}{\sqrt{\widehat{\text{var}}(u_1)\widehat{\text{var}}(u_2)}}. \end{aligned} \tag{10.69}$$

Since the sample mean of the residuals u_1 and u_2 is 0 [see (9.11)], $r_{u_1u_2}$ can be written as

$$r_{u_1u_2} = \frac{\sum_{i=1}^n u_{1i}u_{2i}}{\sqrt{\sum_{i=1}^n u_{1i}^2 \sum_{i=1}^n u_{2i}^2}} = \frac{\sum_{i=1}^n (y_{1i} - \hat{y}_{1i})(y_{2i} - \hat{y}_{2i})}{\sqrt{\sum_{i=1}^n (y_{1i} - \hat{y}_{1i})^2 \sum_{i=1}^n (y_{2i} - \hat{y}_{2i})^2}}. \tag{10.70}$$

We now show that $r_{u_1u_2}$ in (10.70) can be expressed as an element of $\mathbf{R}_{y\cdot x}$ in (10.65). Note that in this illustration, $\mathbf{R}_{y\cdot x}$ is 2×2. The numerator of (10.70) can be written as

$$\begin{aligned}\sum_{i=1}^n u_{1i}u_{2i} &= \sum_{i=1}^n (y_{1i} - \hat{y}_{1i})(y_{2i} - \hat{y}_{2i}) \\ &= \sum_{i=1}^n (y_{1i} - \hat{\beta}_{01} - \hat{\beta}_{11}y_{3i})(y_{2i} - \hat{\beta}_{02} - \hat{\beta}_{12}y_{3i}).\end{aligned}$$

Using $\hat{\beta}_{01} = \bar{y}_1 - \hat{\beta}_{11}\bar{y}_3$ and $\hat{\beta}_{02} = \bar{y}_2 - \hat{\beta}_{12}\bar{y}_3$, we obtain

$$\begin{aligned}\sum_{i=1}^n u_{1i}u_{2i} &= \sum_{i=1}^n [y_{1i} - \bar{y}_1 - \hat{\beta}_{11}(y_{3i} - \bar{y}_3)][y_{2i} - \bar{y}_2 - \hat{\beta}_{12}(y_{3i} - \bar{y}_3)] \\ &= \sum_i (y_{1i} - \bar{y}_1)(y_{2i} - \bar{y}_2) - \hat{\beta}_{11}\hat{\beta}_{12}\sum_i (y_{3i} - \bar{y}_3)^2.\end{aligned} \tag{10.71}$$

The other two terms in (10.71) sum to zero. Using (10.67) and (10.68), the second term on the right side of (10.71) can be written as

$$\hat{\beta}_{11}\hat{\beta}_{12}\sum_i (y_{3i} - \bar{y}_3)^2 = \frac{\left[\sum_{i=1}^n (y_{1i} - \bar{y}_1)(y_{3i} - \bar{y}_3)\right]\left[\sum_{i=1}^n (y_{2i} - \bar{y}_2)(y_{3i} - \bar{y}_3)\right]}{\sum_{i=1}^n (y_{3i} - \bar{y}_3)^2}. \tag{10.72}$$

If we divide (10.71) by $n - 1$, divide numerator and denominator of (10.72) by $n - 1$, and substitute (10.72) into (10.71), we obtain

$$\widehat{\text{cov}}(u_1, u_2) = \widehat{\text{cov}}(y_1 - \hat{y}_1, y_2 - \hat{y}_2) = s_{12} - \frac{s_{13}s_{23}}{s_{33}}.$$

This is the element in the first row and second column of $\mathbf{S}_{yy} - \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}$ in (10.65), where $\mathbf{S}_{yy} = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix}$, $\mathbf{S}_{yx} = \mathbf{s}_{yx} = \begin{pmatrix} s_{13} \\ s_{23} \end{pmatrix}$, $\mathbf{S}_{xx} = s_{33}$, and $\mathbf{S}_{xy} = \mathbf{s}'_{yx}$. In this case, the 2×2 matrix $\mathbf{S}_{yy} - \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}$ is given by

$$\mathbf{S}_{yy} - \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy} = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} - \frac{1}{s_{33}}\begin{pmatrix} s_{13} \\ s_{23} \end{pmatrix}(s_{13}, s_{23})$$
$$= \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} - \frac{1}{s_{33}}\begin{pmatrix} s_{13}^2 & s_{13}s_{23} \\ s_{23}s_{13} & s_{23}^2 \end{pmatrix}.$$

Thus $r_{u_1u_2}$, as based on residuals in (10.69), is equivalent to the maximum likelihood estimator in (10.65).

We now use (10.71) to convert $r_{u_1u_2}$ in (10.69) into the familiar formula for $r_{12\cdot3}$ given in (10.66). By (10.70), we obtain

$$r_{u_1u_2} = \frac{\sum_i u_{1i}u_{2i}}{\sqrt{\sum_i u_{1i}^2 \sum_i u_{2i}^2}}. \tag{10.73}$$

By an extension of (10.71), we further obtain

$$\sum_{i=1}^n u_{1i}^2 = \sum_i (y_{1i} - \bar{y}_1)^2 - \hat{\beta}_{11}^2 \sum_i (y_{3i} - \bar{y}_3)^2, \tag{10.74}$$

$$\sum_{i=1}^n u_{2i}^2 = \sum_i (y_{2i} - \bar{y}_2)^2 - \hat{\beta}_{12}^2 \sum_i (y_{3i} - \bar{y}_3)^2. \tag{10.75}$$

Then (10.73) becomes

$$r_{u_1u_2} = \frac{\sum_i (y_{1i} - \bar{y}_1)(y_{2i} - \bar{y}_2) - \hat{\beta}_{11}\hat{\beta}_{12}\sum_i (y_{3i} - \bar{y}_3)^2}{\sqrt{\left[\sum_i (y_{1i} - \bar{y}_1)^2 - \hat{\beta}_{11}^2 \sum_i (y_{3i} - \bar{y}_3)^2\right]\left[\sum_i (y_{2i} - \bar{y}_2)^2 - \hat{\beta}_{12}^2 \sum_i (y_{3i} - \bar{y}_3)^2\right]}}. \tag{10.76}$$

We now substitute for $\hat{\beta}_{11}$ and $\hat{\beta}_{12}$ as defined in (10.67) and (10.68) and divide numerator and denominator by $\sqrt{\sum_i (y_{1i} - \bar{y}_1)^2 \sum_i (y_{2i} - \bar{y}_2)^2}$ to obtain

$$r_{u_1u_2} = r_{12\cdot3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}. \tag{10.77}$$

Thus $r_{u_1u_2}$ based on residuals as in (10.69) is equivalent to the usual formulation $r_{12\cdot3}$ in (10.66). $\square$

For the general case $r_{ij\cdot rs\ldots q}$, where i and j are subscripts pertaining to $\mathbf{y}$ and $r, s, \ldots, q$ are all the subscripts associated with $\mathbf{x}$, we define a residual vector $\mathbf{y}_i - \hat{\mathbf{y}}_i(\mathbf{x})$, where $\hat{\mathbf{y}}_i(\mathbf{x})$ is the vector of predicted values from the regression of $\mathbf{y}$ on $\mathbf{x}$. [Note that i is used differently in $r_{ij\cdot rs\ldots q}$ and $\mathbf{y}_i - \hat{\mathbf{y}}_i(\mathbf{x})$.] In Theorem 10.8a, $r_{12\cdot 3}$ was found to be equal to $r_{y_1-\hat{y}_1, y_2-\hat{y}_2}$, the ordinary correlation of the two residuals, and to be equivalent to the partial correlation defined as an element of $\mathbf{R}_{y\cdot x}$ in (10.65). In the following theorem, this is extended to the vectors $\mathbf{y}$ and $\mathbf{x}$.

Theorem 10.8b. The sample covariance matrix of the residual vector $\mathbf{y}_i - \hat{\mathbf{y}}_i(\mathbf{x})$ is equivalent to $\mathbf{S}_{yy} - \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}$ in (10.65), that is, $\mathbf{S}_{\mathbf{y}-\hat{\mathbf{y}}} = \mathbf{S}_{yy} - \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}$.

PROOF. The sample predicted value $\hat{\mathbf{y}}_i(\mathbf{x})$ is an estimator of $E(\mathbf{y}|\mathbf{x}_i) = \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_x)$ given in (4.26). For $\hat{\mathbf{y}}_i(\mathbf{x})$, we use the maximum likelihood estimator of $E(\mathbf{y}|\mathbf{x}_i)$:

$$\hat{\mathbf{y}}_i(\mathbf{x}) = \bar{\mathbf{y}} + \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}). \tag{10.78}$$

[The same result can be obtained without reference to normality; see Rencher (1998, p. 304).]

Since the sample mean of $\mathbf{y}_i - \hat{\mathbf{y}}_i(\mathbf{x})$ is $\mathbf{0}$ (see Problem 10.26), the sample covariance matrix of $\mathbf{y}_i - \hat{\mathbf{y}}_i(\mathbf{x})$ is defined as

$$\mathbf{S}_{\mathbf{y}-\hat{\mathbf{y}}} = \frac{1}{n-1}\sum_{i=1}^{n}[\mathbf{y}_i - \hat{\mathbf{y}}_i(\mathbf{x})][\mathbf{y}_i - \hat{\mathbf{y}}_i(\mathbf{x})]' \tag{10.79}$$

(see Problem 10.1).We first note that by extension of (10.13), we have $\mathbf{S}_{yy} = \sum_i (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})'/(n-1)$, $\mathbf{S}_{yx} = \sum_i (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{x}_i - \bar{\mathbf{x}})'/(n-1)$, and $\mathbf{S}_{xx} = \sum_i (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'/(n-1)$ (see Problem 10.27). Using these expressions, after substituting (10.78) in (10.79), we obtain

$$\begin{aligned}
\mathbf{S}_{\mathbf{y}-\hat{\mathbf{y}}} &= \frac{1}{n-1}\sum_{i=1}^{n}[\mathbf{y}_i - \bar{\mathbf{y}} - \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})][\mathbf{y}_i - \bar{\mathbf{y}} - \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})]' \\
&= \frac{1}{n-1}\left[\sum_{i=1}^{n}(\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})' - \sum_{i=1}^{n}(\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\right. \\
&\quad \left. -\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{y}_i - \bar{\mathbf{y}})' + \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\right] \\
&= \mathbf{S}_{yy} - \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy} - \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy} + \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy} \\
&= \mathbf{S}_{yy} - \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}.
\end{aligned}$$

Thus the covariance matrix of residuals gives the same result as the maximum likelihood estimator of conditional covariances and correlations in (10.65). □

Example 10.8. We illustrate some partial correlations for the hematology data in Table 10.1. To find $r_{y1\cdot 2345}$, for example, we use (10.65), $\mathbf{R}_{y\cdot x} = \mathbf{D}_s^{-1}(\mathbf{S}_{yy} - \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy})\mathbf{D}_s^{-1}$. In this case, $\mathbf{y} = (y, x_1)'$ and $\mathbf{x} = (x_2, x_3, x_4, x_5)'$. The matrix $\mathbf{S}$ is therefore partitioned as

$$\mathbf{S} = \left(\begin{array}{cc|cccc} 90.290 & 1.535 & 4.880 & 106.202 & 3.753 & 3.064 \\ 1.535 & 0.691 & 1.494 & 3.255 & 0.422 & -0.268 \\ \hline 4.880 & 1.494 & 5.401 & 10.155 & 1.374 & 1.292 \\ 106.202 & 3.255 & 10.155 & 200.668 & 64.655 & 4.067 \\ 3.753 & 0.422 & 1.374 & 64.655 & 56.374 & 0.579 \\ 3.064 & -0.268 & 1.292 & 4.067 & 0.579 & 18.078 \end{array}\right)$$

$$= \begin{pmatrix} \mathbf{S}_{yy} & \mathbf{S}_{yx} \\ \mathbf{S}_{xy} & \mathbf{S}_{xx} \end{pmatrix}.$$

The matrix $\mathbf{D}_s = [\text{diag}(\mathbf{S}_{yy} - \mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy})]^{1/2}$ is given by

$$\mathbf{D}_s = \begin{pmatrix} 2.645 & 0 \\ 0 & .503 \end{pmatrix},$$

and we have

$$\mathbf{R}_{y\cdot x} = \begin{pmatrix} 1.0000 & -0.0934 \\ -0.0934 & 1.000 \end{pmatrix}.$$

Thus, $r_{y1\cdot 2345} = -.0934$. On the other hand, $r_{y1} = .1934$.

To find $r_{y2\cdot 1345}$, we have $\mathbf{y} = (y, x_2)'$ and $\mathbf{x} = (x_1, x_3, x_4, x_5)'$. Thus

$$\mathbf{S}_{yy} = \begin{pmatrix} 90.290 & 4.880 \\ 4.880 & 5.401 \end{pmatrix},$$

and there are corresponding matrices for $\mathbf{S}_{yx}$, $\mathbf{S}_{xy}$, and $\mathbf{S}_{xx}$. The diagonal matrix $\mathbf{D}_s$ is given by $\mathbf{D}_s = \text{diag}(2.670, 1.389)$, and we have

$$\mathbf{R}_{y\cdot x} = \begin{pmatrix} 1.000 & -0.164 \\ -0.164 & 1.000 \end{pmatrix}.$$

Thus, $r_{y2\cdot 1345} = -.164$, which can be compared with $r_{y2} = .221$.

To find $r_{y3\cdot 45}$, we have $\mathbf{y} = (y, x_1, x_2, x_3)'$ and $\mathbf{x} = (x_4, x_5)'$. Then, for example, we obtain

$$\mathbf{S}_{yy} = \begin{pmatrix} 90.290 & 1.535 & 4.880 & 106.202 \\ 1.535 & 0.691 & 1.494 & 3.255 \\ 4.880 & 1.494 & 5.401 & 10.155 \\ 106.202 & 3.255 & 10.155 & 200.668 \end{pmatrix}.$$

The diagonal matrix $\mathbf{D}_s$ is given by

$$\mathbf{D}_s = \mathrm{diag}(9.462, .827, 2.297, 11.219),$$

and we have

$$\mathbf{R}_{y\cdot x} = \begin{pmatrix} 1.000 & 0.198 & 0.210 & 0.954 \\ 0.198 & 1.000 & 0.792 & 0.304 \\ 0.210 & 0.792 & 1.000 & 0.324 \\ 0.954 & 0.304 & 0.324 & 1.000 \end{pmatrix}.$$

Thus, for example, $r_{y1\cdot 45} = .198, r_{y3\cdot 45} = .954, r_{12\cdot 45} = .792$, and $r_{23\cdot 45} = .324$. In this case, $\mathbf{R}_{y\cdot x}$ is little changed from $\mathbf{R}_{yy}$:

$$\mathbf{R}_{yy} = \begin{pmatrix} 1.000 & 0.194 & 0.221 & 0.789 \\ 0.194 & 1.000 & 0.774 & 0.277 \\ 0.221 & 0.774 & 1.000 & 0.308 \\ 0.789 & 0.277 & 0.308 & 1.000 \end{pmatrix}.$$

□

PROBLEMS

10.1 Show that $\mathbf{S}$ in (10.14) can be found as $\mathbf{S} = \sum_{i=1}^{n} (\mathbf{v}_i - \bar{\mathbf{v}})(\mathbf{v}_i - \bar{\mathbf{v}})'/(n-1)$ as in (10.13).

10.2 Show that $\hat{\boldsymbol{\mu}}$ and $\mathbf{S}$ in (10.9) and (10.10) are jointly sufficient for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, as noted following Theorem 10.2c.

10.3 Show that $\mathbf{S} = \mathbf{DRD}$ gives the partitioned result in (10.19).

10.4 Show that $\mathrm{cov}(y, w) = \boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}^{-1}_{xx}\boldsymbol{\sigma}_{yx}$ and $\mathrm{var}(w) = \boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}^{-1}_{xx}\boldsymbol{\sigma}_{yx}$ as in (10.26), where $w = \mu_y + \boldsymbol{\sigma}'_{yx}\boldsymbol{\Sigma}^{-1}_{xx}(\mathbf{x} - \boldsymbol{\mu}_x)$.

10.5 Show that $\rho^2_{y|\mathbf{x}}$ in (10.27) is the maximum squared correlation between y and any linear function of $\mathbf{x}$, as in (10.28).

10.6 Show that $\rho^2_{y|\mathbf{x}}$ can be expressed as $\rho^2_{y|\mathbf{x}} = 1 - |\boldsymbol{\Sigma}|/(\sigma_{yy}|\boldsymbol{\Sigma}_{xx}|)$ as in (10.29).

10.7 Show that $\rho^2_{y|\mathbf{x}}$ is invariant to linear transformations $u = ay$ and $\mathbf{v} = \mathbf{Bx}$, where $\mathbf{B}$ is nonsingular, as in (10.30).

10.8 Show that $\text{cov}(y - w, \mathbf{x}) = \mathbf{0}'$ as in (10.33).

10.9 Verify that $R^2 = r^2_{y\hat{y}}$, as in (10.36), using the following two definitions of $r^2_{y\hat{y}}$:

(**a**) $r^2_{y\hat{y}} = \left[\sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{\hat{y}})\right]^2 / \left[\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2\right]$
(**b**) $r_{y\hat{y}} = s_{y\hat{y}}/(s_y s_{\hat{y}})$

10.10 Show that $R^2 = \max_{\mathbf{a}} r^2_{y,\mathbf{a}'\mathbf{x}}$ as in (10.37).

10.11 Show that $R^2 = \mathbf{r}'_{yx}\mathbf{R}^{-1}_{xx}\mathbf{r}_{yx}$ as in (10.38) .

10.12 Show that $R^2 = 1 - 1/r^{yy}$ as in (10.39), where r^{yy} is the upper left-hand diagonal element of $\mathbf{R}^{-1}$, with $\mathbf{R}$ partitioned as in (10.18).

10.13 Verify that R^2 can be expressed in terms of determinants as in (10.40) and (10.41).

10.14 Show that R^2 is invariant to full-rank linear transformations on y or the x's, as in property 9 in Section 10.4.

10.15 Show that $\hat{\boldsymbol{\Sigma}}_0$ in (10.46) is the maximum likelihood estimator of $\boldsymbol{\Sigma}_0$ in (10.45) and that $\max_{H_0} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is given by (10.47).

10.16 Show that LR in (10.48) is equal to $\text{LR} = (1 - R^2)^{n/2}$ in (10.49).

10.17 Obtain the confidence interval in (10.57) from the inequality in (10.56).

10.18 Suppose that we have three independent samples of bivariate normal data. The three sample correlations are r_1, r_2, and r_3 based, respectively, on sample sizes n_1, n_2, and n_3.

(**a**) Find the covariance matrix $\mathbf{V}$ of $\mathbf{z} = (z_1 \quad z_2 \quad z_3)'$ where $z_i = \frac{1}{2}\ln[(1 + r_i)/(1 - r_i)]$.
(**b**) Let $\boldsymbol{\mu}'_z = (\tanh^{-1}\rho_1, \tanh^{-1}\rho_2, \tanh^{-1}\rho_3)$, and let

$$\mathbf{C} = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}.$$

Find the distribution of $[\mathbf{C}(\mathbf{z} - \boldsymbol{\mu}_z)]' \, [\mathbf{CVC}']^{-1}[\mathbf{C}(\mathbf{z} - \boldsymbol{\mu}_z)]$.
(**c**) Using (b), propose a test for $H_0 : \rho_1 = \rho_2 = \rho_3$ or equivalently $H_0 : \mathbf{C}\boldsymbol{\mu}_z = \mathbf{0}$.

10.19 Prove Theorem 10.6.

10.20 Show that if z were orthogonal to the x's, (10.58) could be written in the form $R^2_{yw} = R^2_{yx} + r^2_{yz}$, as noted following Theorem 10.6.

10.21 Prove Theorem 10.7b.

10.22 Prove Theorem 10.7c.

10.23 Show that $\sum_{i=1}^{n} u_{1i}u_{2i} = \sum_{i=1}^{n} (y_{1i} - \bar{y}_1)(y_{2i} - \bar{y}_2) - \hat{\beta}_{11}\hat{\beta}_{12} \sum_{i=1}^{n} (y_{3i} - \bar{y}_3)^2$ as in (10.71).

10.24 Show that $\sum_{i=1}^{n} u_{1i}^2 = \sum_{i=1}^{n} (y_{1i} - \bar{y}_1)^2 - \hat{\beta}_{11}^2 \sum_{i=1}^{n} (y_{3i} - \bar{y}_3)^2$ as in (10.74).

10.25 Obtain $r_{12\cdot 3}$ in (10.77) from $r_{u_1u_2}$ in (10.76).

10.26 Show that $\Sigma_{i=1}^{n}[\mathbf{y}_i - \hat{\mathbf{y}}_i(\mathbf{x})] = \mathbf{0}$, as noted following (10.78).

10.27 Show that $\mathbf{S}_{yy} = \sum_i (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})'/(n-1), \mathbf{S}_{yx} = \sum_i (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{x}_i - \bar{\mathbf{x}})'/(n-1)$, and $\mathbf{S}_{xx} = \sum_i (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'/(n-1)$, as noted following (10.79).

10.28 In an experiment with rats, the concentration of a particular drug in the liver was of interest. For 19 rats the following variables were observed:

$$\begin{aligned} y &= \text{percentage of the dose in the liver} \\ x_1 &= \text{body weight} \\ x_2 &= \text{liver weight} \\ x_3 &= \text{relative dose} \end{aligned}$$

The data are given in Table 10.2 (Weisberg 1985, p. 122).

(a) Find $\mathbf{S}_{xx}, \mathbf{s}_{yx}, \hat{\boldsymbol{\beta}}_1, \hat{\beta}_0$, and s^2.
(b) Find $\mathbf{R}_{xx}, \mathbf{r}_{yx}$, and $\hat{\boldsymbol{\beta}}_1^*$.
(c) Find R^2.
(d) Test $H_0 : \boldsymbol{\beta}_1 = \mathbf{0}$.

10.29 Use the hematology data in Table 10.1 as divided into two subsamples of sizes 26 and 25 in Example 10.5b (the first 26 observations and the last 25 observations). For each pair of variables below, find r_1 and r_2 for the two subsamples, find z_1 and z_2 as in (10.51), test $H_0 : \rho_1 = \rho_2$ as in (10.55), and find confidence limits for ρ_1 and ρ_2 as in (10.57).

(a) y and x_2
(b) y and x_3

TABLE 10.2 Rat Data

y	x_1	x_2	x_3	y	x_1	x_2	x_3
.42	176	6.5	0.88	.27	158	6.9	.80
.25	176	9.5	0.88	.36	148	7.3	.74
.56	190	9.0	1.00	.21	149	5.2	.75
.23	176	8.9	0.88	.28	163	8.4	.81
.23	200	7.2	1.00	.34	170	7.2	.85
.32	167	8.9	0.83	.28	186	6.8	.94
.37	188	8.0	0.94	.30	164	7.3	.73
.41	195	10.0	0.98	.37	181	9.0	.90
.33	176	8.0	0.88	.46	149	6.4	.75
.38	165	7.9	0.84				

(**c**) y and x_4

(**d**) y and x_5

10.30 For the rat data in Table 10.2, check the effect of each variable on R^2 as in Section 10.6.

10.31 Using the rat data in Table 10.2.

(**a**) Find $r_{y1\cdot 23}$ and compare to r_{y1}.

(**b**) Find $r_{y2\cdot 13}$

(**c**) Find $\mathbf{R}_{y\cdot x}$, where $\mathbf{y} = (y, x_1, x_2)'$ and $\mathbf{x} = x_3$, in order to obtain $r_{y1\cdot 3}, r_{y2\cdot 3}$, and $r_{12\cdot 3}$.