Chapter 5

Matrix Approach to Simple Linear Regression Analysis

Matrix algebra is widely used for mathematical and statistical analysis. The matrix approach is practically a necessity in multiple regression analysis, since it permits extensive systems of equations and large arrays of data to be denoted compactly and operated upon efficiently.

In this chapter, we first take up a brief introduction to matrix algebra. (A more comprehensive treatment of matrix algebra may be found in specialized texts such as Reference 5.1.) Then we apply matrix methods to the simple linear regression model discussed in previous chapters. Although matrix algebra is not really required for simple linear regression, the application of matrix methods to this case will provide a useful transition to multiple regression, which will be taken up in Parts II and III.

Readers familiar with matrix algebra may wish to scan the introductory parts of this chapter and focus upon the later parts dealing with the use of matrix methods in regression analysis.

5.1 Matrices

Definition of Matrix

A matrix is a rectangular array of elements arranged in rows and columns. An example of a matrix is:

$$\begin{array}{l} \\ \\ \text{Row 1} \\ \text{Row 2} \\ \text{Row 3} \end{array} \begin{array}{cc} \text{Column} & \text{Column} \\ 1 & 2 \\ \left[\begin{array}{c} 16{,}000 \\ 33{,}000 \\ 21{,}000 \end{array}\right. & \left.\begin{array}{c} 23 \\ 47 \\ 35 \end{array}\right] \end{array}$$

The *elements* of this particular matrix are numbers representing income (column 1) and age (column 2) of three persons. The elements are arranged by row (person) and column (characteristic of person). Thus, the element in the first row and first column (16,000) represents the income of the first person. The element in the first row and second column (23) represents the age of the first person. The *dimension* of the matrix is 3×2, i.e., 3 rows by

2 columns. If we wanted to present income and age for 1,000 persons in a matrix with the same format as the one earlier, we would require a $1{,}000 \times 2$ matrix.

Other examples of matrices are:

$$\begin{bmatrix} 1 & 0 \\ 5 & 10 \end{bmatrix} \qquad \begin{bmatrix} 4 & 7 & 12 & 16 \\ 3 & 15 & 9 & 8 \end{bmatrix}$$

These two matrices have dimensions of 2×2 and 2×4, respectively. Note that in giving the dimension of a matrix, we always specify the number of rows first and then the number of columns. As in ordinary algebra, we may use symbols to identify the elements of a matrix:

$$\begin{matrix} & j=1 & j=2 & j=3 \\ i=1 & a_{11} & a_{12} & a_{13} \\ i=2 & a_{21} & a_{22} & a_{23} \end{matrix}$$

Note that the first subscript identifies the row number and the second the column number. We shall use the general notation a_{ij} for the element in the ith row and the jth column. In our above example, $i = 1, 2$ and $j = 1, 2, 3$.

A matrix may be denoted by a symbol such as **A**, **X**, or **Z**. The symbol is in **boldface** to identify that it refers to a matrix. Thus, we might define for the above matrix:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

Reference to the matrix **A** then implies reference to the 2×3 array just given.

Another notation for the matrix **A** just given is:

$$\mathbf{A} = [a_{ij}] \qquad i = 1, 2; j = 1, 2, 3$$

This notation avoids the need for writing out all elements of the matrix by stating only the general element. It can only be used, of course, when the elements of a matrix are symbols.

To summarize, a matrix with r rows and c columns will be represented either in full:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1c} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2c} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{ic} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rj} & \cdots & a_{rc} \end{bmatrix} \tag{5.1}$$

or in abbreviated form:

$$\mathbf{A} = [a_{ij}] \qquad i = 1, \ldots, r; j = 1, \ldots, c$$

or simply by a boldface symbol, such as **A**.

Comments

1. Do not think of a matrix as a number. It is a set of elements arranged in an array. Only when the matrix has dimension 1×1 is there a single number in a matrix, in which case one *can* think of it interchangeably as either a matrix or a number.

2. The following is *not* a matrix:

$$\begin{bmatrix} & 14 & \\ & 8 & \\ 10 & & 15 \\ 9 & & 16 \end{bmatrix}$$

since the numbers are not arranged in columns and rows. ■

Square Matrix

A matrix is said to be square if the number of rows equals the number of columns. Two examples are:

$$\begin{bmatrix} 4 & 7 \\ 3 & 9 \end{bmatrix} \qquad \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Vector

A matrix containing only one column is called a *column vector* or simply a *vector*. Two examples are:

$$\mathbf{A} = \begin{bmatrix} 4 \\ 7 \\ 10 \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{bmatrix}$$

The vector **A** is a 3×1 matrix, and the vector **C** is a 5×1 matrix.

A matrix containing only one row is called a *row vector*. Two examples are:

$$\mathbf{B}' = [15 \quad 25 \quad 50] \qquad \mathbf{F}' = [f_1 \quad f_2]$$

We use the prime symbol for row vectors for reasons to be seen shortly. Note that the row vector $\mathbf{B}'$ is a 1×3 matrix and the row vector $\mathbf{F}'$ is a 1×2 matrix.

A single subscript suffices to identify the elements of a vector.

Transpose

The transpose of a matrix **A** is another matrix, denoted by $\mathbf{A}'$, that is obtained by interchanging corresponding columns and rows of the matrix **A**.

For example, if:

$$\underset{3\times 2}{\mathbf{A}} = \begin{bmatrix} 2 & 5 \\ 7 & 10 \\ 3 & 4 \end{bmatrix}$$

then the transpose $\mathbf{A}'$ is:

$$\underset{2\times 3}{\mathbf{A}'} = \begin{bmatrix} 2 & 7 & 3 \\ 5 & 10 & 4 \end{bmatrix}$$

Note that the first column of **A** is the first row of $\mathbf{A}'$, and similarly the second column of **A** is the second row of $\mathbf{A}'$. Correspondingly, the first row of **A** has become the first column of $\mathbf{A}'$, and so on. Note that the dimension of $\mathbf{A}$, indicated under the symbol $\mathbf{A}$, becomes reversed for the dimension of $\mathbf{A}'$.

As another example, consider:

$$\underset{3\times 1}{\mathbf{C}} = \begin{bmatrix} 4 \\ 7 \\ 10 \end{bmatrix} \qquad \underset{1\times 3}{\mathbf{C}'} = [4 \quad 7 \quad 10]$$

Thus, the transpose of a column vector is a row vector, and vice versa. This is the reason why we used the symbol $\mathbf{B}'$ earlier to identify a row vector, since it may be thought of as the transpose of a column vector $\mathbf{B}$.

In general, we have:

$$\underset{r\times c}{\mathbf{A}} = \begin{bmatrix} a_{11} & \cdots & a_{1c} \\ \vdots & & \vdots \\ a_{r1} & \cdots & a_{rc} \end{bmatrix} = [a_{ij}] \qquad i = 1, \dots, r; j = 1, \dots, c \tag{5.2}$$

Row index ↗ ↖ Column index

$$\underset{c\times r}{\mathbf{A}'} = \begin{bmatrix} a_{11} & \cdots & a_{r1} \\ \vdots & & \vdots \\ a_{1c} & \cdots & a_{rc} \end{bmatrix} = [a_{ji}] \qquad j = 1, \dots, c; i = 1, \dots, r \tag{5.3}$$

Row index ↗ ↖ Column index

Thus, the element in the ith row and the jth column in $\mathbf{A}$ is found in the jth row and ith column in $\mathbf{A}'$.

Equality of Matrices

Two matrices $\mathbf{A}$ and $\mathbf{B}$ are said to be equal if they have the same dimension and if all corresponding elements are equal. Conversely, if two matrices are equal, their corresponding elements are equal. For example, if:

$$\underset{3\times 1}{\mathbf{A}} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \qquad \underset{3\times 1}{\mathbf{B}} = \begin{bmatrix} 4 \\ 7 \\ 3 \end{bmatrix}$$

then $\mathbf{A} = \mathbf{B}$ implies:

$$a_1 = 4 \qquad a_2 = 7 \qquad a_3 = 3$$

Similarly, if:

$$\underset{3\times 2}{\mathbf{A}} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \qquad \underset{3\times 2}{\mathbf{B}} = \begin{bmatrix} 17 & 2 \\ 14 & 5 \\ 13 & 9 \end{bmatrix}$$

then $\mathbf{A} = \mathbf{B}$ implies:

$$\begin{array}{ll} a_{11} = 17 & a_{12} = 2 \\ a_{21} = 14 & a_{22} = 5 \\ a_{31} = 13 & a_{32} = 9 \end{array}$$

Regression Examples

In regression analysis, one basic matrix is the vector $\mathbf{Y}$, consisting of the n observations on the response variable:

$$\underset{n\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \tag{5.4}$$

Note that the transpose $\mathbf{Y}'$ is the row vector:

$$\underset{1\times n}{\mathbf{Y}'} = [Y_1 \quad Y_2 \quad \cdots \quad Y_n] \tag{5.5}$$

Another basic matrix in regression analysis is the $\mathbf{X}$ matrix, which is defined as follows for simple linear regression analysis:

$$\underset{n\times 2}{\mathbf{X}} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \tag{5.6}$$

The matrix $\mathbf{X}$ consists of a column of 1s and a column containing the n observations on the predictor variable X. Note that the transpose of $\mathbf{X}$ is:

$$\underset{2\times n}{\mathbf{X}'} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \tag{5.7}$$

The $\mathbf{X}$ matrix is often referred to as the *design matrix.*

5.2 Matrix Addition and Subtraction

Adding or subtracting two matrices requires that they have the same dimension. The sum, or difference, of two matrices is another matrix whose elements each consist of the sum, or difference, of the corresponding elements of the two matrices. Suppose:

$$\underset{3\times 2}{\mathbf{A}} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \qquad \underset{3\times 2}{\mathbf{B}} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{bmatrix}$$

then:

$$\underset{3\times 2}{\mathbf{A}+\mathbf{B}} = \begin{bmatrix} 1+1 & 4+2 \\ 2+2 & 5+3 \\ 3+3 & 6+4 \end{bmatrix} = \begin{bmatrix} 2 & 6 \\ 4 & 8 \\ 6 & 10 \end{bmatrix}$$

Similarly:

$$\underset{3\times2}{\mathbf{A}-\mathbf{B}} = \begin{bmatrix} 1-1 & 4-2 \\ 2-2 & 5-3 \\ 3-3 & 6-4 \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ 0 & 2 \\ 0 & 2 \end{bmatrix}$$

In general, if:

$$\underset{r\times c}{\mathbf{A}} = [a_{ij}] \qquad \underset{r\times c}{\mathbf{B}} = [b_{ij}] \qquad i = 1, \ldots, r; j = 1, \ldots, c$$

then:

$$\underset{r\times c}{\mathbf{A}+\mathbf{B}} = [a_{ij} + b_{ij}] \qquad \text{and} \qquad \underset{r\times c}{\mathbf{A}-\mathbf{B}} = [a_{ij} - b_{ij}] \tag{5.8}$$

Formula (5.8) generalizes in an obvious way to addition and subtraction of more than two matrices. Note also that $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$, as in ordinary algebra.

Regression Example

The regression model:

$$Y_i = E\{Y_i\} + \varepsilon_i \qquad i = 1, \ldots, n$$

can be written compactly in matrix notation. First, let us define the vector of the mean responses:

$$\underset{n\times1}{\mathbf{E}\{\mathbf{Y}\}} = \begin{bmatrix} E\{Y_1\} \\ E\{Y_2\} \\ \vdots \\ E\{Y_n\} \end{bmatrix} \tag{5.9}$$

and the vector of the error terms:

$$\underset{n\times1}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{5.10}$$

Recalling the definition of the observations vector $\mathbf{Y}$ in (5.4), we can write the regression model as follows:

$$\underset{n\times1}{\mathbf{Y}} = \underset{n\times1}{\mathbf{E}\{\mathbf{Y}\}} + \underset{n\times1}{\boldsymbol{\varepsilon}}$$

because:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} E\{Y_1\} \\ E\{Y_2\} \\ \vdots \\ E\{Y_n\} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} E\{Y_1\} + \varepsilon_1 \\ E\{Y_2\} + \varepsilon_2 \\ \vdots \\ E\{Y_n\} + \varepsilon_n \end{bmatrix}$$

Thus, the observations vector $\mathbf{Y}$ equals the sum of two vectors, a vector containing the expected values and another containing the error terms.

5.3 Matrix Multiplication

Multiplication of a Matrix by a Scalar

A *scalar* is an ordinary number or a symbol representing a number. In multiplication of a matrix by a scalar, every element of the matrix is multiplied by the scalar. For example, suppose the matrix **A** is given by:

$$\mathbf{A} = \begin{bmatrix} 2 & 7 \\ 9 & 3 \end{bmatrix}$$

Then 4**A**, where 4 is the scalar, equals:

$$4\mathbf{A} = 4\begin{bmatrix} 2 & 7 \\ 9 & 3 \end{bmatrix} = \begin{bmatrix} 8 & 28 \\ 36 & 12 \end{bmatrix}$$

Similarly, $k\mathbf{A}$ equals:

$$k\mathbf{A} = k\begin{bmatrix} 2 & 7 \\ 9 & 3 \end{bmatrix} = \begin{bmatrix} 2k & 7k \\ 9k & 3k \end{bmatrix}$$

where k denotes a scalar.

If every element of a matrix has a common factor, this factor can be taken outside the matrix and treated as a scalar. For example:

$$\begin{bmatrix} 9 & 27 \\ 15 & 18 \end{bmatrix} = 3\begin{bmatrix} 3 & 9 \\ 5 & 6 \end{bmatrix}$$

Similarly:

$$\begin{bmatrix} \frac{5}{k} & \frac{2}{k} \\ \frac{3}{k} & \frac{8}{k} \end{bmatrix} = \frac{1}{k}\begin{bmatrix} 5 & 2 \\ 3 & 8 \end{bmatrix}$$

In general, if $\mathbf{A} = [a_{ij}]$ and k is a scalar, we have:

$$k\mathbf{A} = \mathbf{A}k = [ka_{ij}] \tag{5.11}$$

Multiplication of a Matrix by a Matrix

Multiplication of a matrix by a matrix may appear somewhat complicated at first, but a little practice will make it a routine operation.

Consider the two matrices:

$$\underset{2\times2}{\mathbf{A}} = \begin{bmatrix} 2 & 5 \\ 4 & 1 \end{bmatrix} \qquad \underset{2\times2}{\mathbf{B}} = \begin{bmatrix} 4 & 6 \\ 5 & 8 \end{bmatrix}$$

The product **AB** will be a 2×2 matrix whose elements are obtained by finding the cross products of rows of **A** with columns of **B** and summing the cross products. For instance, to find the element in the first row and the first column of the product **AB**, we work with the first row of **A** and the first column of **B**, as follows:

$$\begin{array}{c} \\ \text{Row 1} \\ \text{Row 2} \end{array} \overset{\mathbf{A}}{\begin{bmatrix} \boxed{2 \quad 5} \\ 4 \quad 1 \end{bmatrix}} \qquad \underset{\text{Col. 1 Col. 2}}{\overset{\mathbf{B}}{\begin{bmatrix} \boxed{\begin{matrix} 4 \\ 5 \end{matrix}} & \begin{matrix} 6 \\ 8 \end{matrix} \end{bmatrix}}} \qquad \text{Row 1} \underset{\text{Col. 1}}{\overset{\mathbf{AB}}{\begin{bmatrix} 33 & \\ & \end{bmatrix}}}$$

We take the cross products and sum:

$$2(4) + 5(5) = 33$$

The number 33 is the element in the first row and first column of the matrix **AB**.

To find the element in the first row and second column of **AB**, we work with the first row of **A** and the second column of **B**:

$$\begin{array}{c} \\ \text{Row 1} \\ \text{Row 2} \end{array} \overset{\mathbf{A}}{\begin{bmatrix} \boxed{2 \quad 5} \\ 4 \quad 1 \end{bmatrix}} \qquad \underset{\text{Col. 1 Col. 2}}{\overset{\mathbf{B}}{\begin{bmatrix} \begin{matrix} 4 \\ 5 \end{matrix} & \boxed{\begin{matrix} 6 \\ 8 \end{matrix}} \end{bmatrix}}} \qquad \text{Row 1} \underset{\text{Col. 1 Col. 2}}{\overset{\mathbf{AB}}{\begin{bmatrix} 33 & 52 \\ & \end{bmatrix}}}$$

The sum of the cross products is:

$$2(6) + 5(8) = 52$$

Continuing this process, we find the product **AB** to be:

$$\underset{2\times 2}{\mathbf{AB}} = \begin{bmatrix} 2 & 5 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} 4 & 6 \\ 5 & 8 \end{bmatrix} = \begin{bmatrix} 33 & 52 \\ 21 & 32 \end{bmatrix}$$

Let us consider another example:

$$\underset{2\times 3}{\mathbf{A}} = \begin{bmatrix} 1 & 3 & 4 \\ 0 & 5 & 8 \end{bmatrix} \qquad \underset{3\times 1}{\mathbf{B}} = \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix}$$

$$\underset{2\times 1}{\mathbf{AB}} = \begin{bmatrix} 1 & 3 & 4 \\ 0 & 5 & 8 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix} = \begin{bmatrix} 26 \\ 41 \end{bmatrix}$$

When obtaining the product **AB**, we say that **A** is *postmultiplied* by **B** or **B** is *premultiplied* by **A**. The reason for this precise terminology is that multiplication rules for ordinary algebra do not apply to matrix algebra. In ordinary algebra, $xy = yx$. In matrix algebra, $\mathbf{AB} \neq \mathbf{BA}$ usually. In fact, even though the product **AB** may be defined, the product **BA** may not be defined at all.

In general, the product **AB** is defined only when the number of columns in **A** equals the number of rows in **B** so that there will be corresponding terms in the cross products. Thus, in our previous two examples, we had:

$$\begin{array}{ccccc} & \text{Equal} & & & \\ \mathbf{A} & \swarrow \searrow & \mathbf{B} & = & \mathbf{AB} \\ 2\times 2 & & 2\times 2 & & 2\times 2 \\ \nwarrow & & \nearrow & & \\ & \text{Dimension of product} & & & \end{array} \qquad \begin{array}{ccccc} & \text{Equal} & & & \\ \mathbf{A} & \swarrow \searrow & \mathbf{B} & = & \mathbf{AB} \\ 2\times 3 & & 3\times 1 & & 2\times 1 \\ \nwarrow & & \nearrow & & \\ & \text{Dimension of product} & & & \end{array}$$

Note that the dimension of the product **AB** is given by the number of rows in **A** and the number of columns in **B**. Note also that in the second case the product **BA** would not be defined since the number of columns in **B** is not equal to the number of rows in **A**:

$$\underset{3\times 1}{\mathbf{B}} \ \overset{\text{Unequal}}{\swarrow\searrow} \ \underset{2\times 3}{\mathbf{A}}$$

Here is another example of matrix multiplication:

$$\mathbf{AB} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{bmatrix}$$

In general, if **A** has dimension $r \times c$ and **B** has dimension $c \times s$, the product **AB** is a matrix of dimension $r \times s$ whose element in the ith row and jth column is:

$$\sum_{k=1}^{c} a_{ik}b_{kj}$$

so that:

$$\underset{r\times s}{\mathbf{AB}} = \left[\sum_{k=1}^{c} a_{ik}b_{kj}\right] \qquad i = 1, \ldots, r; j = 1, \ldots, s \tag{5.12}$$

Thus, in the foregoing example, the element in the first row and second column of the product **AB** is:

$$\sum_{k=1}^{3} a_{1k}b_{k2} = a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32}$$

as indeed we found by taking the cross products of the elements in the first row of **A** and second column of **B** and summing.

Additional Examples

1.
$$\begin{bmatrix} 4 & 2 \\ 5 & 8 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 4a_1 + 2a_2 \\ 5a_1 + 8a_2 \end{bmatrix}$$

2.
$$\begin{bmatrix} 2 & 3 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} = [2^2 + 3^2 + 5^2] = [38]$$

Here, the product is a 1×1 matrix, which is equivalent to a scalar. Thus, the matrix product here equals the number 38.

3.
$$\begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ \beta_0 + \beta_1 X_3 \end{bmatrix}$$

Regression Examples

A product frequently needed is $\mathbf{Y}'\mathbf{Y}$, where $\mathbf{Y}$ is the vector of observations on the response variable as defined in (5.4):

$$\underset{1\times 1}{\mathbf{Y}'\mathbf{Y}} = [Y_1 \quad Y_2 \quad \cdots \quad Y_n] \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = [Y_1^2 + Y_2^2 + \cdots + Y_n^2] = [\textstyle\sum Y_i^2] \quad \textbf{(5.13)}$$

Note that $\mathbf{Y}'\mathbf{Y}$ is a 1×1 matrix, or a scalar. We thus have a compact way of writing a sum of squared terms: $\mathbf{Y}'\mathbf{Y} = \sum Y_i^2$.

We also will need $\mathbf{X}'\mathbf{X}$, which is a 2×2 matrix, where $\mathbf{X}$ is defined in (5.6):

$$\underset{2\times 2}{\mathbf{X}'\mathbf{X}} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} \quad \textbf{(5.14)}$$

and $\mathbf{X}'\mathbf{Y}$, which is a 2×1 matrix:

$$\underset{2\times 1}{\mathbf{X}'\mathbf{Y}} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix} \quad \textbf{(5.15)}$$

5.4 Special Types of Matrices

Certain special types of matrices arise regularly in regression analysis. We consider the most important of these.

Symmetric Matrix

If $\mathbf{A} = \mathbf{A}'$, $\mathbf{A}$ is said to be symmetric. Thus, $\mathbf{A}$ below is symmetric:

$$\underset{3\times 3}{\mathbf{A}} = \begin{bmatrix} 1 & 4 & 6 \\ 4 & 2 & 5 \\ 6 & 5 & 3 \end{bmatrix} \qquad \underset{3\times 3}{\mathbf{A}'} = \begin{bmatrix} 1 & 4 & 6 \\ 4 & 2 & 5 \\ 6 & 5 & 3 \end{bmatrix}$$

A symmetric matrix necessarily is square. Symmetric matrices arise typically in regression analysis when we premultiply a matrix, say, $\mathbf{X}$, by its transpose, $\mathbf{X}'$. The resulting matrix, $\mathbf{X}'\mathbf{X}$, is symmetric, as can readily be seen from (5.14).

Diagonal Matrix

A diagonal matrix is a square matrix whose off-diagonal elements are all zeros, such as:

$$\underset{3\times 3}{\mathbf{A}} = \begin{bmatrix} a_1 & 0 & 0 \\ 0 & a_2 & 0 \\ 0 & 0 & a_3 \end{bmatrix} \qquad \underset{4\times 4}{\mathbf{B}} = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix}$$

We will often not show all zeros for a diagonal matrix, presenting it in the form:

$$\underset{3\times 3}{\mathbf{A}} = \begin{bmatrix} a_1 & & 0 \\ & a_2 & \\ 0 & & a_3 \end{bmatrix} \qquad \underset{4\times 4}{\mathbf{B}} = \begin{bmatrix} 4 & & & 0 \\ & 1 & & \\ & & 10 & \\ 0 & & & 5 \end{bmatrix}$$

Two important types of diagonal matrices are the identity matrix and the scalar matrix.

Identity Matrix. The identity matrix or unit matrix is denoted by **I**. It is a diagonal matrix whose elements on the main diagonal are all 1s. Premultiplying or postmultiplying any $r \times r$ matrix **A** by the $r \times r$ identity matrix **I** leaves **A** unchanged. For example:

$$\mathbf{IA} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Similarly, we have:

$$\mathbf{AI} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Note that the identity matrix **I** therefore corresponds to the number 1 in ordinary algebra, since we have there that $1 \cdot x = x \cdot 1 = x$.

In general, we have for any $r \times r$ matrix **A**:

$$\mathbf{AI} = \mathbf{IA} = \mathbf{A} \tag{5.16}$$

Thus, the identity matrix can be inserted or dropped from a matrix expression whenever it is convenient to do so.

Scalar Matrix. A scalar matrix is a diagonal matrix whose main-diagonal elements are the same. Two examples of scalar matrices are:

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \qquad \begin{bmatrix} k & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & k \end{bmatrix}$$

A scalar matrix can be expressed as $k\mathbf{I}$, where k is the scalar. For instance:

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = 2\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 2\mathbf{I}$$

$$\begin{bmatrix} k & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & k \end{bmatrix} = k\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = k\mathbf{I}$$

Multiplying an $r \times r$ matrix **A** by the $r \times r$ scalar matrix $k\mathbf{I}$ is equivalent to multiplying **A** by the scalar k.

Vector and Matrix with All Elements Unity

A column vector with all elements 1 will be denoted by **1**:

$$\underset{r\times 1}{\mathbf{1}} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \tag{5.17}$$

and a square matrix with all elements 1 will be denoted by **J**:

$$\underset{r\times r}{\mathbf{J}} = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{bmatrix} \tag{5.18}$$

For instance, we have:

$$\underset{3\times 1}{\mathbf{1}} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \qquad \underset{3\times 3}{\mathbf{J}} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Note that for an $n \times 1$ vector **1** we obtain:

$$\underset{1\times 1}{\mathbf{1}'\mathbf{1}} = [1 \quad \cdots \quad 1] \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = [n] = n$$

and:

$$\underset{n\times n}{\mathbf{1}\mathbf{1}'} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} [1 \quad \cdots \quad 1] = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{bmatrix} = \underset{n\times n}{\mathbf{J}}$$

Zero Vector

A zero vector is a vector containing only zeros. The zero column vector will be denoted by **0**:

$$\underset{r\times 1}{\mathbf{0}} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{5.19}$$

For example, we have:

$$\underset{3\times 1}{\mathbf{0}} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

5.5 Linear Dependence and Rank of Matrix

Linear Dependence

Consider the following matrix:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 5 & 1 \\ 2 & 2 & 10 & 6 \\ 3 & 4 & 15 & 1 \end{bmatrix}$$

Let us think now of the columns of this matrix as vectors. Thus, we view **A** as being made up of four column vectors. It happens here that the columns are interrelated in a special manner. Note that the third column vector is a multiple of the first column vector:

$$\begin{bmatrix} 5 \\ 10 \\ 15 \end{bmatrix} = 5 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

We say that the columns of **A** are linearly dependent. They contain redundant information, so to speak, since one column can be obtained as a linear combination of the others.

We define the set of c column vectors $\mathbf{C}_1, \ldots, \mathbf{C}_c$ in an $r \times c$ matrix to be linearly dependent if one vector can be expressed as a linear combination of the others. If no vector in the set can be so expressed, we define the set of vectors to be linearly independent. A more general, though equivalent, definition is:

When c scalars $k_1, \ldots, k_c$, not all zero, can be found such that:

$$k_1\mathbf{C}_1 + k_2\mathbf{C}_2 + \cdots + k_c\mathbf{C}_c = \mathbf{0}$$

where $\mathbf{0}$ denotes the zero column vector, the c column vectors are *linearly dependent*. If the only set of scalars for which the equality holds is $k_1 = 0, \ldots, k_c = 0$, the set of c column vectors is *linearly independent*. **(5.20)**

To illustrate for our example, $k_1 = 5$, $k_2 = 0$, $k_3 = -1$, $k_4 = 0$ leads to:

$$5\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + 0\begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix} - 1\begin{bmatrix} 5 \\ 10 \\ 15 \end{bmatrix} + 0\begin{bmatrix} 1 \\ 6 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Hence, the column vectors are linearly dependent. Note that some of the k_j equal zero here. For linear dependence, it is only required that not all k_j be zero.

Rank of Matrix

The rank of a matrix is defined to be the maximum number of linearly independent columns in the matrix. We know that the rank of **A** in our earlier example cannot be 4, since the four columns are linearly dependent. We can, however, find three columns (1, 2, and 4) which are linearly independent. There are no scalars k_1, k_2, k_4 such that $k_1\mathbf{C}_1 + k_2\mathbf{C}_2 + k_4\mathbf{C}_4 = \mathbf{0}$ other than $k_1 = k_2 = k_4 = 0$. Thus, the rank of **A** in our example is 3.

The rank of a matrix is unique and can equivalently be defined as the maximum number of linearly independent rows. It follows that the rank of an $r \times c$ matrix cannot exceed $\min(r, c)$, the minimum of the two values r and c.

When a matrix is the product of two matrices, its rank cannot exceed the smaller of the two ranks for the matrices being multiplied. Thus, if $\mathbf{C} = \mathbf{AB}$, the rank of $\mathbf{C}$ cannot exceed min(rank $\mathbf{A}$, rank $\mathbf{B}$).

5.6 Inverse of a Matrix

In ordinary algebra, the inverse of a number is its reciprocal. Thus, the inverse of 6 is $\frac{1}{6}$. A number multiplied by its inverse always equals 1:

$$6 \cdot \frac{1}{6} = \frac{1}{6} \cdot 6 = 1$$

$$x \cdot \frac{1}{x} = x \cdot x^{-1} = x^{-1} \cdot x = 1$$

In matrix algebra, the inverse of a matrix $\mathbf{A}$ is another matrix, denoted by $\mathbf{A}^{-1}$, such that:

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I} \tag{5.21}$$

where $\mathbf{I}$ is the identity matrix. Thus, again, the identity matrix $\mathbf{I}$ plays the same role as the number 1 in ordinary algebra. An inverse of a matrix is defined only for square matrices. Even so, many square matrices do not have inverses. If a square matrix does have an inverse, the inverse is unique.

Examples

1. The inverse of the matrix:

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

is:

$$\underset{2\times 2}{\mathbf{A}^{-1}} = \begin{bmatrix} -.1 & .4 \\ .3 & -.2 \end{bmatrix}$$

since:

$$\mathbf{A}^{-1}\mathbf{A} = \begin{bmatrix} -.1 & .4 \\ .3 & -.2 \end{bmatrix} \begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}$$

or:

$$\mathbf{A}\mathbf{A}^{-1} = \begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} -.1 & .4 \\ .3 & -.2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}$$

2. The inverse of the matrix:

$$\underset{3\times 3}{\mathbf{A}} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

is:

$$\underset{3\times 3}{\mathbf{A}^{-1}} = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix}$$

since:

$$\mathbf{A}^{-1}\mathbf{A} = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \mathbf{I}$$

Note that the inverse of a diagonal matrix is a diagonal matrix consisting simply of the reciprocals of the elements on the diagonal.

Finding the Inverse

Up to this point, the inverse of a matrix **A** has been given, and we have only checked to make sure it is the inverse by seeing whether or not $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. But how does one find the inverse, and when does it exist?

An inverse of a square $r \times r$ matrix exists if the rank of the matrix is r. Such a matrix is said to be *nonsingular* or of *full rank*. An $r \times r$ matrix with rank less than r is said to be *singular* or *not of full rank*, and does not have an inverse. The inverse of an $r \times r$ matrix of full rank also has rank r.

Finding the inverse of a matrix can often require a large amount of computing. We shall take the approach in this book that the inverse of a 2×2 matrix and a 3×3 matrix can be calculated by hand. For any larger matrix, one ordinarily uses a computer to find the inverse, unless the matrix is of a special form such as a diagonal matrix. It can be shown that the inverses for 2×2 and 3×3 matrices are as follows:

1. If:

$$\underset{2\times 2}{\mathbf{A}} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

then:

$$\underset{2\times 2}{\mathbf{A}^{-1}} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \begin{bmatrix} \frac{d}{D} & \frac{-b}{D} \\ \frac{-c}{D} & \frac{a}{D} \end{bmatrix} \tag{5.22}$$

where:

$$D = ad - bc \tag{5.22a}$$

D is called the *determinant* of the matrix **A**. If **A** were singular, its determinant would equal zero and no inverse of **A** would exist.

2. If:

$$\underset{3\times 3}{\mathbf{B}} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix}$$

then:

$$\underset{3\times 3}{\mathbf{B}^{-1}} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix}^{-1} = \begin{bmatrix} A & B & C \\ D & E & F \\ G & H & K \end{bmatrix} \tag{5.23}$$

where:

$$A = (ek - fh)/Z \quad B = -(bk - ch)/Z \quad C = (bf - ce)/Z$$
$$D = -(dk - fg)/Z \quad E = (ak - cg)/Z \quad F = -(af - cd)/Z \tag{5.23a}$$
$$G = (dh - eg)/Z \quad H = -(ah - bg)/Z \quad K = (ae - bd)/Z$$

and:

$$Z = a(ek - fh) - b(dk - fg) + c(dh - eg) \tag{5.23b}$$

Z is called the determinant of the matrix **B**.

Let us use (5.22) to find the inverse of:

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

We have:

$$a = 2 \quad b = 4$$
$$c = 3 \quad d = 1$$

$$D = ad - bc = 2(1) - 4(3) = -10$$

Hence:

$$\mathbf{A}^{-1} = \begin{bmatrix} \frac{1}{-10} & \frac{-4}{-10} \\ \frac{-3}{-10} & \frac{2}{-10} \end{bmatrix} = \begin{bmatrix} -.1 & .4 \\ .3 & -.2 \end{bmatrix}$$

as was given in an earlier example.

When an inverse $\mathbf{A}^{-1}$ has been obtained by hand calculations or from a computer program for which the accuracy of inverting a matrix is not known, it may be wise to compute $\mathbf{A}^{-1}\mathbf{A}$ to check whether the product equals the identity matrix, allowing for minor rounding departures from 0 and 1.

Regression Example

The principal inverse matrix encountered in regression analysis is the inverse of the matrix **X′X** in (5.14):

$$\underset{2\times 2}{\mathbf{X'X}} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}$$

Using rule (5.22), we have:

$$a = n \qquad b = \sum X_i$$
$$c = \sum X_i \quad d = \sum X_i^2$$

so that:

$$D = n\sum X_i^2 - \left(\sum X_i\right)\left(\sum X_i\right) = n\left[\sum X_i^2 - \frac{(\sum X_i)^2}{n}\right] = n\sum (X_i - \bar{X})^2$$

Hence:

$$\underset{2\times 2}{(\mathbf{X}'\mathbf{X})^{-1}} = \begin{bmatrix} \dfrac{\sum X_i^2}{n\sum(X_i-\bar{X})^2} & \dfrac{-\sum X_i}{n\sum(X_i-\bar{X})^2} \\ \dfrac{-\sum X_i}{n\sum(X_i-\bar{X})^2} & \dfrac{n}{n\sum(X_i-\bar{X})^2} \end{bmatrix}$$

Since $\sum X_i = n\bar{X}$ and $\sum(X_i-\bar{X})^2 = \sum X_i^2 - n\bar{X}^2$, we can simplify (5.24):

$$\underset{2\times 2}{(\mathbf{X}'\mathbf{X})^{-1}} = \begin{bmatrix} \dfrac{1}{n} + \dfrac{\bar{X}^2}{\sum(X_i-\bar{X})^2} & \dfrac{-\bar{X}}{\sum(X_i-\bar{X})^2} \\ \dfrac{-\bar{X}}{\sum(X_i-\bar{X})^2} & \dfrac{1}{\sum(X_i-\bar{X})^2} \end{bmatrix}$$

Uses of Inverse Matrix

In ordinary algebra, we solve an equation of the type:

$$5y = 20$$

by multiplying both sides of the equation by the inverse of 5, namely:

$$\frac{1}{5}(5y) = \frac{1}{5}(20)$$

and we obtain the solution:

$$y = \frac{1}{5}(20) = 4$$

In matrix algebra, if we have an equation:

$$\mathbf{AY} = \mathbf{C}$$

we correspondingly premultiply both sides by $\mathbf{A}^{-1}$, assuming $\mathbf{A}$ has an invers

$$\mathbf{A}^{-1}\mathbf{AY} = \mathbf{A}^{-1}\mathbf{C}$$

Since $\mathbf{A}^{-1}\mathbf{AY} = \mathbf{IY} = \mathbf{Y}$, we obtain the solution:

$$\mathbf{Y} = \mathbf{A}^{-1}\mathbf{C}$$

To illustrate this use, suppose we have two simultaneous equations:

$$2y_1 + 4y_2 = 20$$
$$3y_1 + y_2 = 10$$

which can be written as follows in matrix notation:

$$\begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 20 \\ 10 \end{bmatrix}$$

The solution of these equations then is:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 20 \\ 10 \end{bmatrix}$$

Earlier we found the required inverse, so we obtain:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} -.1 & .4 \\ .3 & -.2 \end{bmatrix} \begin{bmatrix} 20 \\ 10 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

Hence, $y_1 = 2$ and $y_2 = 4$ satisfy these two equations.

5.7 Some Basic Results for Matrices

We list here, without proof, some basic results for matrices which we will utilize in later work.

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A} \tag{5.25}$$

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}) \tag{5.26}$$

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}) \tag{5.27}$$

$$\mathbf{C}(\mathbf{A} + \mathbf{B}) = \mathbf{CA} + \mathbf{CB} \tag{5.28}$$

$$k(\mathbf{A} + \mathbf{B}) = k\mathbf{A} + k\mathbf{B} \tag{5.29}$$

$$(\mathbf{A}')' = \mathbf{A} \tag{5.30}$$

$$(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}' \tag{5.31}$$

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}' \tag{5.32}$$

$$(\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}' \tag{5.33}$$

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \tag{5.34}$$

$$(\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1} \tag{5.35}$$

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A} \tag{5.36}$$

$$(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})' \tag{5.37}$$

5.8 Random Vectors and Matrices

A random vector or a random matrix contains elements that are random variables. Thus, the observations vector $\mathbf{Y}$ in (5.4) is a random vector since the Y_i elements are random variables.

Expectation of Random Vector or Matrix

Suppose we have $n = 3$ observations in the observations vector $\mathbf{Y}$:

$$\underset{3\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$$

The expected value of $\mathbf{Y}$ is a vector, denoted by $\mathbf{E}\{\mathbf{Y}\}$, that is defined as follows:

$$\underset{3\times 1}{\mathbf{E}\{\mathbf{Y}\}} = \begin{bmatrix} E\{Y_1\} \\ E\{Y_2\} \\ E\{Y_3\} \end{bmatrix}$$

Thus, the expected value of a random vector is a vector whose elements are the expected values of the random variables that are the elements of the random vector. Similarly, the expectation of a random matrix is a matrix whose elements are the expected values of the corresponding random variables in the original matrix. We encountered a vector of expected values earlier in (5.9).

In general, for a random vector $\mathbf{Y}$ the expectation is:

$$\underset{n\times 1}{\mathbf{E}\{\mathbf{Y}\}} = [E\{Y_i\}] \qquad i = 1, \dots, n \tag{5.38}$$

and for a random matrix $\mathbf{Y}$ with dimension $n \times p$, the expectation is:

$$\underset{n\times p}{\mathbf{E}\{\mathbf{Y}\}} = [E\{Y_{ij}\}] \qquad i = 1, \dots, n; j = 1, \dots, p \tag{5.39}$$

Regression Example

Suppose the number of cases in a regression application is $n = 3$. The three error terms ε_1, ε_2, ε_3 each have expectation zero. For the error terms vector:

$$\underset{3\times 1}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}$$

we have:

$$\underset{3\times 1}{\mathbf{E}\{\boldsymbol{\varepsilon}\}} = \underset{3\times 1}{\mathbf{0}}$$

since:

$$\begin{bmatrix} E\{\varepsilon_1\} \\ E\{\varepsilon_2\} \\ E\{\varepsilon_3\} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Variance-Covariance Matrix of Random Vector

Consider again the random vector $\mathbf{Y}$ consisting of three observations Y_1, Y_2, Y_3. The variances of the three random variables, $\sigma^2\{Y_i\}$, and the covariances between any two of the random variables, $\sigma\{Y_i, Y_j\}$, are assembled in the *variance-covariance matrix of* $\mathbf{Y}$, denoted by $\boldsymbol{\sigma}^2\{\mathbf{Y}\}$, in the following form:

$$\boldsymbol{\sigma}^2\{\mathbf{Y}\} = \begin{bmatrix} \sigma^2\{Y_1\} & \sigma\{Y_1, Y_2\} & \sigma\{Y_1, Y_3\} \\ \sigma\{Y_2, Y_1\} & \sigma^2\{Y_2\} & \sigma\{Y_2, Y_3\} \\ \sigma\{Y_3, Y_1\} & \sigma\{Y_3, Y_2\} & \sigma^2\{Y_3\} \end{bmatrix} \tag{5.40}$$

Note that the variances are on the main diagonal, and the covariance $\sigma\{Y_i, Y_j\}$ is found in the ith row and jth column of the matrix. Thus, $\sigma\{Y_2, Y_1\}$ is found in the second row, first column, and $\sigma\{Y_1, Y_2\}$ is found in the first row, second column. Remember, of course, that $\sigma\{Y_2, Y_1\} = \sigma\{Y_1, Y_2\}$. Since $\sigma\{Y_i, Y_j\} = \sigma\{Y_j, Y_i\}$ for all $i \neq j$, $\boldsymbol{\sigma}^2\{\mathbf{Y}\}$ is a symmetric matrix.

It follows readily that:

$$\boldsymbol{\sigma}^2\{\mathbf{Y}\} = \mathbf{E}\{[\mathbf{Y} - \mathbf{E}\{\mathbf{Y}\}][\mathbf{Y} - \mathbf{E}\{\mathbf{Y}\}]'\} \tag{5.41}$$

For our illustration, we have:

$$\boldsymbol{\sigma}^2\{\mathbf{Y}\} = \mathbf{E}\left\{\begin{bmatrix} Y_1 - E\{Y_1\} \\ Y_2 - E\{Y_2\} \\ Y_3 - E\{Y_3\} \end{bmatrix} [Y_1 - E\{Y_1\} \quad Y_2 - E\{Y_2\} \quad Y_3 - E\{Y_3\}]\right\}$$

Multiplying the two matrices and then taking expectations, we obtain:

Location in Product	Term	Expected Value
Row 1, column 1	$(Y_1 - E\{Y_1\})^2$	$\sigma^2\{Y_1\}$
Row 1, column 2	$(Y_1 - E\{Y_1\})(Y_2 - E\{Y_2\})$	$\sigma\{Y_1, Y_2\}$
Row 1, column 3	$(Y_1 - E\{Y_1\})(Y_3 - E\{Y_3\})$	$\sigma\{Y_1, Y_3\}$
Row 2, column 1	$(Y_2 - E\{Y_2\})(Y_1 - E\{Y_1\})$	$\sigma\{Y_2, Y_1\}$
etc.	etc.	etc.

This, of course, leads to the variance-covariance matrix in (5.40). Remember the definitions of variance and covariance in (A.15) and (A.21), respectively, when taking expectations.

To generalize, the variance-covariance matrix for an $n \times 1$ random vector $\mathbf{Y}$ is:

$$\underset{n\times n}{\boldsymbol{\sigma}^2\{\mathbf{Y}\}} = \begin{bmatrix} \sigma^2\{Y_1\} & \sigma\{Y_1, Y_2\} & \cdots & \sigma\{Y_1, Y_n\} \\ \sigma\{Y_2, Y_1\} & \sigma^2\{Y_2\} & \cdots & \sigma\{Y_2, Y_n\} \\ \vdots & \vdots & & \vdots \\ \sigma\{Y_n, Y_1\} & \sigma\{Y_n, Y_2\} & \cdots & \sigma^2\{Y_n\} \end{bmatrix} \tag{5.42}$$

Note again that $\boldsymbol{\sigma}^2\{\mathbf{Y}\}$ is a symmetric matrix.

Regression Example

Let us return to the example based on $n = 3$ cases. Suppose that the three error terms have constant variance, $\sigma^2\{\varepsilon_i\} = \sigma^2$, and are uncorrelated so that $\sigma\{\varepsilon_i, \varepsilon_j\} = 0$ for $i \neq j$. The variance-covariance matrix for the random vector $\boldsymbol{\varepsilon}$ of the previous example is therefore as follows:

$$\underset{3\times 3}{\boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\}} = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}$$

Note that all variances are σ^2 and all covariances are zero. Note also that this variance-covariance matrix is a scalar matrix, with the common variance σ^2 the scalar. Hence, we can express the variance-covariance matrix in the following simple fashion:

$$\underset{3\times 3}{\boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\}} = \sigma^2 \underset{3\times 3}{\mathbf{I}}$$

since:

$$\sigma^2\mathbf{I} = \sigma^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}$$

Some Basic Results

Frequently, we shall encounter a random vector **W** that is obtained by premultiplying the random vector **Y** by a constant matrix **A** (a matrix whose elements are fixed):

$$\mathbf{W} = \mathbf{AY} \tag{5.43}$$

Some basic results for this case are:

$$\mathbf{E}\{\mathbf{A}\} = \mathbf{A} \tag{5.44}$$

$$\mathbf{E}\{\mathbf{W}\} = \mathbf{E}\{\mathbf{AY}\} = \mathbf{A}\mathbf{E}\{\mathbf{Y}\} \tag{5.45}$$

$$\boldsymbol{\sigma}^2\{\mathbf{W}\} = \boldsymbol{\sigma}^2\{\mathbf{AY}\} = \mathbf{A}\boldsymbol{\sigma}^2\{\mathbf{Y}\}\mathbf{A}' \tag{5.46}$$

where $\boldsymbol{\sigma}^2\{\mathbf{Y}\}$ is the variance-covariance matrix of **Y**.

Example

As a simple illustration of the use of these results, consider:

$$\underset{2\times 1}{\mathbf{W}} \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} = \underset{2\times 2}{\mathbf{A}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \underset{2\times 1}{\mathbf{Y}} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} Y_1 - Y_2 \\ Y_1 + Y_2 \end{bmatrix}$$

We then have by (5.45):

$$\underset{2\times 1}{\mathbf{E}\{\mathbf{W}\}} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} E\{Y_1\} \\ E\{Y_2\} \end{bmatrix} = \begin{bmatrix} E\{Y_1\} - E\{Y_2\} \\ E\{Y_1\} + E\{Y_2\} \end{bmatrix}$$

and by (5.46):

$$\underset{2\times 2}{\boldsymbol{\sigma}^2\{\mathbf{W}\}} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma^2\{Y_1\} & \sigma\{Y_1, Y_2\} \\ \sigma\{Y_2, Y_1\} & \sigma^2\{Y_2\} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2\{Y_1\} + \sigma^2\{Y_2\} - 2\sigma\{Y_1, Y_2\} & \sigma^2\{Y_1\} - \sigma^2\{Y_2\} \\ \sigma^2\{Y_1\} - \sigma^2\{Y_2\} & \sigma^2\{Y_1\} + \sigma^2\{Y_2\} + 2\sigma\{Y_1, Y_2\} \end{bmatrix}$$

Thus:

$$\sigma^2\{W_1\} = \sigma^2\{Y_1 - Y_2\} = \sigma^2\{Y_1\} + \sigma^2\{Y_2\} - 2\sigma\{Y_1, Y_2\}$$

$$\sigma^2\{W_2\} = \sigma^2\{Y_1 + Y_2\} = \sigma^2\{Y_1\} + \sigma^2\{Y_2\} + 2\sigma\{Y_1, Y_2\}$$

$$\sigma\{W_1, W_2\} = \sigma\{Y_1 - Y_2, Y_1 + Y_2\} = \sigma^2\{Y_1\} - \sigma^2\{Y_2\}$$

Multivariate Normal Distribution

Density Function. The density function for the multivariate normal distribution is best given in matrix form. We first need to define some vectors and matrices. The observations vector $\mathbf{Y}$ containing an observation on each of the p Y variables is defined as usual:

$$\underset{p\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{bmatrix} \tag{5.47}$$

The mean vector $\mathbf{E}\{\mathbf{Y}\}$, denoted by $\boldsymbol{\mu}$, contains the expected values for each of the p Y variables:

$$\underset{p\times 1}{\boldsymbol{\mu}} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \tag{5.48}$$

Finally, the variance-covariance matrix $\boldsymbol{\sigma}^2\{\mathbf{Y}\}$ is denoted by $\boldsymbol{\Sigma}$ and contains as always the variances and covariances of the p Y variables:

$$\underset{p\times p}{\boldsymbol{\Sigma}} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{bmatrix} \tag{5.49}$$

Here, σ_1^2 denotes the variance of Y_1, σ_{12} denotes the covariance of Y_1 and Y_2, and the like.

The density function of the multivariate normal distribution can now be stated as follows:

$$f(\mathbf{Y}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{Y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})\right] \tag{5.50}$$

Here, $|\boldsymbol{\Sigma}|$ is the determinant of the variance-covariance matrix $\boldsymbol{\Sigma}$. When there are $p = 2$ variables, the multivariate normal density function (5.50) simplifies to the bivariate normal density function (2.74).

The multivariate normal density function has properties that correspond to the ones described for the bivariate normal distribution. For instance, if $Y_1, \ldots, Y_p$ are jointly normally distributed (i.e., they follow the multivariate normal distribution), the marginal probability distribution of each variable Y_k is normal, with mean μ_k and standard deviation σ_k.

Simple Linear Regression Model in Matrix Terms

We are now ready to develop simple linear regression in matrix terms. Remember again that we will not present any new results, but shall only state in matrix terms the results obtained earlier. We begin with the normal error regression model (2.1):

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad i = 1, \ldots, n \tag{5.51}$$

This implies:

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_1 + \varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_2 + \varepsilon_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 X_n + \varepsilon_n \end{aligned} \tag{5.51a}$$

We defined earlier the observations vector **Y** in (5.4), the **X** matrix in (5.6), and the **ε** vector in (5.10). Let us repeat these definitions and also define the **β** vector of the regression coefficients:

$$\underset{n\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \underset{n\times 2}{\mathbf{X}} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \quad \underset{2\times 1}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \underset{n\times 1}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{5.52}$$

Now we can write (5.51a) in matrix terms compactly as follows:

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times 2}{\mathbf{X}}\ \underset{2\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}} \tag{5.53}$$

since:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$= \begin{bmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ \vdots \\ \beta_0 + \beta_1 X_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_1 + \varepsilon_1 \\ \beta_0 + \beta_1 X_2 + \varepsilon_2 \\ \vdots \\ \beta_0 + \beta_1 X_n + \varepsilon_n \end{bmatrix}$$

Note that **Xβ** is the vector of the expected values of the Y_i observations since $E\{Y_i\} = \beta_0 + \beta_1 X_i$; hence:

$$\underset{n\times 1}{\mathbf{E}\{\mathbf{Y}\}} = \underset{n\times 1}{\mathbf{X}\boldsymbol{\beta}} \tag{5.54}$$

where **E{Y}** is defined in (5.9).

The column of 1s in the **X** matrix may be viewed as consisting of the constant $X_0 \equiv 1$ in the alternative regression model (1.5):

$$Y_i = \beta_0 X_0 + \beta_1 X_i + \varepsilon_i \qquad \text{where } X_0 \equiv 1$$

Thus, the **X** matrix may be considered to contain a column vector consisting of 1s and another column vector consisting of the predictor variable observations X_i.

With respect to the error terms, regression model (2.1) assumes that $E\{\varepsilon_i\} = 0, \sigma^2\{\varepsilon_i\} = \sigma^2$, and that the ε_i are independent normal random variables. The condition $E\{\varepsilon_i\} = 0$ in matrix terms is:

$$\underset{n\times 1}{\mathbf{E}\{\boldsymbol{\varepsilon}\}} = \underset{n\times 1}{\mathbf{0}} \tag{5.55}$$

since (5.55) states:

$$\begin{bmatrix} E\{\varepsilon_1\} \\ E\{\varepsilon_2\} \\ \vdots \\ E\{\varepsilon_n\} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

The condition that the error terms have constant variance σ^2 and that all covariances $\sigma\{\varepsilon_i, \varepsilon_j\}$ for $i \neq j$ are zero (since the ε_i are independent) is expressed in matrix terms through the variance-covariance matrix of the error terms:

$$\underset{n\times n}{\boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\}} = \begin{bmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{bmatrix} \tag{5.56}$$

Since this is a scalar matrix, we know from the earlier example that it can be expressed in the following simple fashion:

$$\underset{n\times n}{\boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\}} = \sigma^2 \underset{n\times n}{\mathbf{I}} \tag{5.56a}$$

Thus, the normal error regression model (2.1) in matrix terms is:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{5.57}$$

where:

$\boldsymbol{\varepsilon}$ is a vector of independent normal random variables with $\mathbf{E}\{\boldsymbol{\varepsilon}\} = \mathbf{0}$ and $\boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\} = \sigma^2\mathbf{I}$

5.10 Least Squares Estimation of Regression Parameters

Normal Equations

The normal equations (1.9):

$$\begin{aligned} nb_0 + b_1 \sum X_i &= \sum Y_i \\ b_0 \sum X_i + b_1 \sum X_i^2 &= \sum X_i Y_i \end{aligned} \tag{5.58}$$

in matrix terms are:

$$\underset{2\times 2}{\mathbf{X'X}}\ \underset{2\times 1}{\mathbf{b}} = \underset{2\times 1}{\mathbf{X'Y}} \tag{5.59}$$

where $\mathbf{b}$ is the vector of the least squares regression coefficients:

$$\underset{2\times 1}{\mathbf{b}} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \tag{5.59a}$$

To see this, recall that we obtained **X'X** in (5.14) and **X'Y** in (5.15). Equation (5.59) thus states:

$$\begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

or:

$$\begin{bmatrix} nb_0 + b_1 \sum X_i \\ b_0 \sum X_i + b_1 \sum X_i^2 \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

These are precisely the normal equations in (5.58).

Estimated Regression Coefficients

To obtain the estimated regression coefficients from the normal equations (5.59) by matrix methods, we premultiply both sides by the inverse of **X'X** (we assume this exists):

$$(\mathbf{X'X})^{-1}\mathbf{X'Xb} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$$

We then find, since $(\mathbf{X'X})^{-1}\mathbf{X'X} = \mathbf{I}$ and $\mathbf{Ib} = \mathbf{b}$:

$$\underset{2\times 1}{\mathbf{b}} = \underset{2\times 2}{(\mathbf{X'X})^{-1}}\underset{2\times 1}{\mathbf{X'Y}} \tag{5.60}$$

The estimators b_0 and b_1 in **b** are the same as those given earlier in (1.10a) and (1.10b). We shall demonstrate this by an example.

Example

We shall use matrix methods to obtain the estimated regression coefficients for the Toluca Company example. The data on the Y and X variables were given in Table 1.1. Using these data, we define the **Y** observations vector and the **X** matrix as follows:

$$\text{(5.61a)} \quad \mathbf{Y} = \begin{bmatrix} 399 \\ 121 \\ \vdots \\ 323 \end{bmatrix} \qquad \text{(5.61b)} \quad \mathbf{X} = \begin{bmatrix} 1 & 80 \\ 1 & 30 \\ \vdots & \vdots \\ 1 & 70 \end{bmatrix} \tag{5.61}$$

We now require the following matrix products:

$$\mathbf{X'X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 80 & 30 & \cdots & 70 \end{bmatrix} \begin{bmatrix} 1 & 80 \\ 1 & 30 \\ \vdots & \vdots \\ 1 & 70 \end{bmatrix} = \begin{bmatrix} 25 & 1{,}750 \\ 1{,}750 & 142{,}300 \end{bmatrix} \tag{5.62}$$

$$\mathbf{X'Y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 80 & 30 & \cdots & 70 \end{bmatrix} \begin{bmatrix} 399 \\ 121 \\ \vdots \\ 323 \end{bmatrix} = \begin{bmatrix} 7{,}807 \\ 617{,}180 \end{bmatrix} \tag{5.63}$$

Using (5.22), we find the inverse of $\mathbf{X'X}$:

$$(\mathbf{X'X})^{-1} = \begin{bmatrix} .287475 & -.003535 \\ -.003535 & .00005051 \end{bmatrix} \tag{5.64}$$

In subsequent matrix calculations utilizing this inverse matrix and other matrix results, we shall actually utilize more digits for the matrix elements than are shown.

Finally, we employ (5.60) to obtain:

$$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = (\mathbf{X'X})^{-1}\mathbf{X'Y} = \begin{bmatrix} .287475 & -.003535 \\ -.003535 & .00005051 \end{bmatrix} \begin{bmatrix} 7{,}807 \\ 617{,}180 \end{bmatrix}$$

$$= \begin{bmatrix} 62.37 \\ 3.5702 \end{bmatrix} \tag{5.65}$$

or $b_0 = 62.37$ and $b_1 = 3.5702$. These results agree with the ones in Chapter 1. Any differences would have been due to rounding effects.

Comments

1. To derive the normal equations by the method of least squares, we minimize the quantity:

$$Q = \sum [Y_i - (\beta_0 + \beta_1 X_i)]^2$$

In matrix notation:

$$Q = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \tag{5.66}$$

Expanding, we obtain:

$$Q = \mathbf{Y'Y} - \boldsymbol{\beta}'\mathbf{X'Y} - \mathbf{Y'X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X'X}\boldsymbol{\beta}$$

since $(\mathbf{X}\boldsymbol{\beta})' = \boldsymbol{\beta}'\mathbf{X}'$ by (5.32). Note now that $\mathbf{Y'X}\boldsymbol{\beta}$ is 1×1, hence is equal to its transpose, which according to (5.33) is $\boldsymbol{\beta}'\mathbf{X'Y}$. Thus, we find:

$$Q = \mathbf{Y'Y} - 2\boldsymbol{\beta}'\mathbf{X'Y} + \boldsymbol{\beta}'\mathbf{X'X}\boldsymbol{\beta} \tag{5.67}$$

To find the value of $\boldsymbol{\beta}$ that minimizes Q, we differentiate with respect to β_0 and β_1. Let:

$$\frac{\partial}{\partial \boldsymbol{\beta}}(Q) = \begin{bmatrix} \dfrac{\partial Q}{\partial \beta_0} \\ \dfrac{\partial Q}{\partial \beta_1} \end{bmatrix} \tag{5.68}$$

Then it follows that:

$$\frac{\partial}{\partial \boldsymbol{\beta}}(Q) = -2\mathbf{X'Y} + 2\mathbf{X'X}\boldsymbol{\beta} \tag{5.69}$$

Equating to the zero vector, dividing by 2, and substituting $\mathbf{b}$ for $\boldsymbol{\beta}$ gives the matrix form of the least squares normal equations in (5.59).

2. A comparison of the normal equations and $\mathbf{X'X}$ shows that whenever the columns of $\mathbf{X'X}$ are linearly dependent, the normal equations will be linearly dependent also. No unique solutions can then be obtained for b_0 and b_1. Fortunately, in most regression applications, the columns of $\mathbf{X'X}$ are linearly independent, leading to unique solutions for b_0 and b_1. ■

5.11 Fitted Values and Residuals

Fitted Values

Let the vector of the fitted values $\hat{Y}_i$ be denoted by $\hat{\mathbf{Y}}$:

$$\underset{n\times 1}{\hat{\mathbf{Y}}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} \tag{5.70}$$

In matrix notation, we then have:

$$\underset{n\times 1}{\hat{\mathbf{Y}}} = \underset{n\times 2}{\mathbf{X}}\ \underset{2\times 1}{\mathbf{b}} \tag{5.71}$$

because:

$$\begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} b_0 + b_1 X_1 \\ b_0 + b_1 X_2 \\ \vdots \\ b_0 + b_1 X_n \end{bmatrix}$$

Example

For the Toluca Company example, we obtain the vector of fitted values using the matrices in (5.61b) and (5.65):

$$\hat{\mathbf{Y}} = \mathbf{Xb} = \begin{bmatrix} 1 & 80 \\ 1 & 30 \\ \vdots & \vdots \\ 1 & 70 \end{bmatrix} \begin{bmatrix} 62.37 \\ 3.5702 \end{bmatrix} = \begin{bmatrix} 347.98 \\ 169.47 \\ \vdots \\ 312.28 \end{bmatrix} \tag{5.72}$$

The fitted values are the same, of course, as in Table 1.2.

Hat Matrix. We can express the matrix result for $\hat{\mathbf{Y}}$ in (5.71) as follows by using the expression for **b** in (5.60):

$$\hat{\mathbf{Y}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

or, equivalently:

$$\underset{n\times 1}{\hat{\mathbf{Y}}} = \underset{n\times n}{\mathbf{H}}\ \underset{n\times 1}{\mathbf{Y}} \tag{5.73}$$

where:

$$\underset{n\times n}{\mathbf{H}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \tag{5.73a}$$

We see from (5.73) that the fitted values $\hat{Y}_i$ can be expressed as linear combinations of the response variable observations Y_i, with the coefficients being elements of the matrix **H**. The **H** matrix involves only the observations on the predictor variable X, as is evident from (5.73a).

The square $n \times n$ matrix **H** is called the *hat matrix*. It plays an important role in diagnostics for regression analysis, as we shall see in Chapter 10 when we consider whether regression results are unduly influenced by one or a few observations. The matrix **H** is symmetric and has the special property (called idempotency):

$$\mathbf{HH} = \mathbf{H} \tag{5.74}$$

In general, a matrix **M** is said to be *idempotent* if **MM** = **M**.

Residuals

Let the vector of the residuals $e_i = Y_i - \hat{Y}_i$ be denoted by **e**:

$$\underset{n\times 1}{\mathbf{e}} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \tag{5.75}$$

In matrix notation, we then have:

$$\underset{n\times 1}{\mathbf{e}} = \underset{n\times 1}{\mathbf{Y}} - \underset{n\times 1}{\hat{\mathbf{Y}}} = \underset{n\times 1}{\mathbf{Y}} - \underset{n\times 1}{\mathbf{Xb}} \tag{5.76}$$

Example

For the Toluca Company example, we obtain the vector of the residuals by using the results in (5.61a) and (5.72):

$$\mathbf{e} = \begin{bmatrix} 399 \\ 121 \\ \vdots \\ 323 \end{bmatrix} - \begin{bmatrix} 347.98 \\ 169.47 \\ \vdots \\ 312.28 \end{bmatrix} = \begin{bmatrix} 51.02 \\ -48.47 \\ \vdots \\ 10.72 \end{bmatrix} \tag{5.77}$$

The residuals are the same as in Table 1.2.

Variance-Covariance Matrix of Residuals. The residuals e_i, like the fitted values $\hat{Y}_i$, can be expressed as linear combinations of the response variable observations Y_i, using the result in (5.73) for $\hat{\mathbf{Y}}$:

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{HY} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$

We thus have the important result:

$$\underset{n\times 1}{\mathbf{e}} = \underset{n\times n}{(\mathbf{I} - \mathbf{H})}\ \underset{n\times 1}{\mathbf{Y}} \tag{5.78}$$

where **H** is the hat matrix defined in (5.53a). The matrix **I** − **H**, like the matrix **H**, is symmetric and idempotent.

The variance-covariance matrix of the vector of residuals **e** involves the matrix **I** − **H**:

$$\underset{n\times n}{\boldsymbol{\sigma}^2\{\mathbf{e}\}} = \sigma^2(\mathbf{I} - \mathbf{H}) \tag{5.79}$$

and is estimated by:

$$\underset{n\times n}{\mathbf{s}^2\{\mathbf{e}\}} = MSE(\mathbf{I} - \mathbf{H}) \tag{5.80}$$

Comment

The variance-covariance matrix of **e** in (5.79) can be derived by means of (5.46). Since $\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$, we obtain:

$$\boldsymbol{\sigma}^2\{\mathbf{e}\} = (\mathbf{I} - \mathbf{H})\boldsymbol{\sigma}^2\{\mathbf{Y}\}(\mathbf{I} - \mathbf{H})'$$

Now $\boldsymbol{\sigma}^2\{\mathbf{Y}\} = \boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\} = \sigma^2\mathbf{I}$ for the normal error model according to (5.56a). Also, $(\mathbf{I} - \mathbf{H})' = \mathbf{I} - \mathbf{H}$ because of the symmetry of the matrix. Hence:

$$\begin{aligned}\boldsymbol{\sigma}^2\{\mathbf{e}\} &= \sigma^2(\mathbf{I} - \mathbf{H})\mathbf{I}(\mathbf{I} - \mathbf{H}) \\ &= \sigma^2(\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H})\end{aligned}$$

In view of the fact that the matrix $\mathbf{I} - \mathbf{H}$ is idempotent, we know that $(\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H}) = \mathbf{I} - \mathbf{H}$ and we obtain formula (5.79). ■

5.12 Analysis of Variance Results

Sums of Squares

To see how the sums of squares are expressed in matrix notation, we begin with the total sum of squares *SSTO*, defined in (2.43). It will be convenient to use an algebraically equivalent expression:

$$SSTO = \sum(Y_i - \bar{Y})^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{n} \tag{5.81}$$

We know from (5.13) that:

$$\mathbf{Y}'\mathbf{Y} = \sum Y_i^2$$

The subtraction term $(\sum Y_i)^2/n$ in matrix form uses **J**, the matrix of 1s defined in (5.18), as follows:

$$\frac{(\sum Y_i)^2}{n} = \left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y} \tag{5.82}$$

For instance, if $n = 2$, we have:

$$\left(\frac{1}{2}\right)[Y_1 \quad Y_2]\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}\begin{bmatrix}Y_1\\Y_2\end{bmatrix} = \frac{(Y_1 + Y_2)(Y_1 + Y_2)}{2}$$

Hence, it follows that:

$$SSTO = \mathbf{Y}'\mathbf{Y} - \left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y} \tag{5.83}$$

Just as $\sum Y_i^2$ is represented by $\mathbf{Y}'\mathbf{Y}$ in matrix terms, so $SSE = \sum e_i^2 = \sum(Y_i - \hat{Y}_i)^2$ can be represented as follows:

$$SSE = \mathbf{e}'\mathbf{e} = (\mathbf{Y} - \mathbf{X}\mathbf{b})'(\mathbf{Y} - \mathbf{X}\mathbf{b}) \tag{5.84}$$

which can be shown to equal:

$$SSE = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y} \tag{5.84a}$$

Finally, it can be shown that:

$$SSR = \mathbf{b}'\mathbf{X}'\mathbf{Y} - \left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y} \tag{5.85}$$

Example

Let us find *SSE* for the Toluca Company example by matrix methods, using (5.84a). Using (5.61a), we obtain:

$$\mathbf{Y}'\mathbf{Y} = [399 \quad 121 \quad \cdots \quad 323]\begin{bmatrix}399\\121\\\vdots\\323\end{bmatrix} = 2{,}745{,}173$$

and using (5.65) and (5.63), we find:

$$\mathbf{b}'\mathbf{X}'\mathbf{Y} = [62.37 \quad 3.5702]\begin{bmatrix}7{,}807\\617{,}180\end{bmatrix} = 2{,}690{,}348$$

Hence:

$$SSE = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y} = 2{,}745{,}173 - 2{,}690{,}348 = 54{,}825$$

which is the same result as that obtained in Chapter 1. Any difference would have been due to rounding effects.

Comment

To illustrate the derivation of the sums of squares expressions in matrix notation, consider *SSE*:

$$SSE = \mathbf{e}'\mathbf{e} = (\mathbf{Y} - \mathbf{X}\mathbf{b})'(\mathbf{Y} - \mathbf{X}\mathbf{b}) = \mathbf{Y}'\mathbf{Y} - 2\mathbf{b}'\mathbf{X}'\mathbf{Y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}$$

In substituting for the rightmost **b** we obtain by (5.60):

$$\begin{aligned} SSE &= \mathbf{Y}'\mathbf{Y} - 2\mathbf{b}'\mathbf{X}'\mathbf{Y} + \mathbf{b}'\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ &= \mathbf{Y}'\mathbf{Y} - 2\mathbf{b}'\mathbf{X}'\mathbf{Y} + \mathbf{b}'\mathbf{I}\mathbf{X}'\mathbf{Y} \end{aligned}$$

In dropping **I** and subtracting, we obtain the result in (5.84a). ■

Sums of Squares as Quadratic Forms

The ANOVA sums of squares can be shown to be *quadratic forms*. An example of a quadratic form of the observations Y_i when $n = 2$ is:

$$5Y_1^2 + 6Y_1Y_2 + 4Y_2^2 \tag{5.86}$$

Note that this expression is a second-degree polynomial containing terms involving the squares of the observations and the cross product. We can express (5.86) in matrix terms as follows:

$$[Y_1 \quad Y_2]'\begin{bmatrix}5 & 3\\3 & 4\end{bmatrix}\begin{bmatrix}Y_1\\Y_2\end{bmatrix} = \mathbf{Y}'\mathbf{A}\mathbf{Y} \tag{5.86a}$$

where **A** is a symmetric matrix.

In general, a quadratic form is defined as:

$$\underset{1\times 1}{\mathbf{Y}'\mathbf{A}\mathbf{Y}} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}Y_iY_j \qquad \text{where } a_{ij} = a_{ji} \tag{5.87}$$

A is a symmetric $n \times n$ matrix and is called the *matrix of the quadratic form.*

The ANOVA sums of squares *SSTO*, *SSE*, and *SSR* are all quadratic forms, as can be seen by reexpressing $\mathbf{b}'\mathbf{X}'$. From (5.71), we know, using (5.32), that:

$$\mathbf{b}'\mathbf{X}' = (\mathbf{X}\mathbf{b})' = \hat{\mathbf{Y}}'$$

We now use the result in (5.73) to obtain:

$$\mathbf{b}'\mathbf{X}' = (\mathbf{H}\mathbf{Y})'$$

Since **H** is a symmetric matrix so that $\mathbf{H}' = \mathbf{H}$, we finally obtain, using (5.32):

$$\mathbf{b}'\mathbf{X}' = \mathbf{Y}'\mathbf{H} \tag{5.88}$$

This result enables us to express the ANOVA sums of squares as follows:

$$SSTO = \mathbf{Y}'\left[\mathbf{I} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y} \tag{5.89a}$$

$$SSE = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y} \tag{5.89b}$$

$$SSR = \mathbf{Y}'\left[\mathbf{H} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y} \tag{5.89c}$$

Each of these sums of squares can now be seen to be of the form $\mathbf{Y}'\mathbf{A}\mathbf{Y}$, where the three **A** matrices are:

$$\mathbf{I} - \left(\frac{1}{n}\right)\mathbf{J} \tag{5.90a}$$

$$\mathbf{I} - \mathbf{H} \tag{5.90b}$$

$$\mathbf{H} - \left(\frac{1}{n}\right)\mathbf{J} \tag{5.90c}$$

Since each of these **A** matrices is symmetric, *SSTO*, *SSE*, and *SSR* are quadratic forms, with the matrices of the quadratic forms given in (5.90). Quadratic forms play an important role in statistics because all sums of squares in the analysis of variance for linear statistical models can be expressed as quadratic forms.

5.13 Inferences in Regression Analysis

As we saw in earlier chapters, all interval estimates are of the following form: point estimator plus and minus a certain number of estimated standard deviations of the point estimator. Similarly, all tests require the point estimator and the estimated standard deviation of the point estimator or, in the case of analysis of variance tests, various sums of squares. Matrix algebra is of principal help in inference making when obtaining the estimated standard deviations and sums of squares. We have already given the matrix equivalents of the sums of squares for the analysis of variance. We focus here chiefly on the matrix expressions for the estimated variances of point estimators of interest.

Regression Coefficients

The variance-covariance matrix of **b**:

$$\underset{2\times2}{\boldsymbol{\sigma}^2\{\mathbf{b}\}} = \begin{bmatrix} \sigma^2\{b_0\} & \sigma\{b_0, b_1\} \\ \sigma\{b_1, b_0\} & \sigma^2\{b_1\} \end{bmatrix} \tag{5.91}$$

is:

$$\underset{2\times2}{\boldsymbol{\sigma}^2\{\mathbf{b}\}} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \tag{5.92}$$

or, from (5.24a):

$$\underset{2\times2}{\boldsymbol{\sigma}^2\{\mathbf{b}\}} = \begin{bmatrix} \dfrac{\sigma^2}{n} + \dfrac{\sigma^2\bar{X}^2}{\sum(X_i - \bar{X})^2} & \dfrac{-\bar{X}\sigma^2}{\sum(X_i - \bar{X})^2} \\ \dfrac{-\bar{X}\sigma^2}{\sum(X_i - \bar{X})^2} & \dfrac{\sigma^2}{\sum(X_i - \bar{X})^2} \end{bmatrix} \tag{5.92a}$$

When MSE is substituted for σ^2 in (5.92a), we obtain the estimated variance-covariance matrix of **b**, denoted by $\mathbf{s}^2\{\mathbf{b}\}$:

$$\underset{2\times2}{\mathbf{s}^2\{\mathbf{b}\}} = MSE(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} \dfrac{MSE}{n} + \dfrac{\bar{X}^2 MSE}{\sum(X_i - \bar{X})^2} & \dfrac{-\bar{X}MSE}{\sum(X_i - \bar{X})^2} \\ \dfrac{-\bar{X}MSE}{\sum(X_i - \bar{X})^2} & \dfrac{MSE}{\sum(X_i - \bar{X})^2} \end{bmatrix} \tag{5.93}$$

In (5.92a), you will recognize the variances of b_0 in (2.22b) and of b_1 in (2.3b) and the covariance of b_0 and b_1 in (4.5). Likewise, the estimated variances in (5.93) are familiar from earlier chapters.

Example

We wish to find $s^2\{b_0\}$ and $s^2\{b_1\}$ for the Toluca Company example by matrix methods. Using the results in Figure 2.2 and in (5.64), we obtain:

$$\mathbf{s}^2\{\mathbf{b}\} = MSE(\mathbf{X}'\mathbf{X})^{-1} = 2{,}384 \begin{bmatrix} .287475 & -.003535 \\ -.003535 & .00005051 \end{bmatrix}$$

$$= \begin{bmatrix} 685.34 & -8.428 \\ -8.428 & .12040 \end{bmatrix} \tag{5.94}$$

Thus, $s^2\{b_0\} = 685.34$ and $s^2\{b_1\} = .12040$. These are the same as the results obtained in Chapter 2.

Comment

To derive the variance-covariance matrix of **b**, recall that:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{A}\mathbf{Y}$$

where **A** is a constant matrix:

$$\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

Hence, by (5.46) we have:

$$\boldsymbol{\sigma}^2\{\mathbf{b}\} = \mathbf{A}\boldsymbol{\sigma}^2\{\mathbf{Y}\}\mathbf{A}'$$

Now $\boldsymbol{\sigma}^2\{\mathbf{Y}\} = \sigma^2\mathbf{I}$. Further, it follows from (5.32) and the fact that $(\mathbf{X'X})^{-1}$ is symmetric that:

$$\mathbf{A}' = \mathbf{X}(\mathbf{X'X})^{-1}$$

We find therefore:

$$\begin{aligned}\boldsymbol{\sigma}^2\{\mathbf{b}\} &= (\mathbf{X'X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X'X})^{-1}\\ &= \sigma^2(\mathbf{X'X})^{-1}\mathbf{X'X}(\mathbf{X'X})^{-1}\\ &= \sigma^2(\mathbf{X'X})^{-1}\mathbf{I}\\ &= \sigma^2(\mathbf{X'X})^{-1}\end{aligned}$$

■

Mean Response

To estimate the mean response at X_h, let us define the vector:

$$\underset{2\times 1}{\mathbf{X}_h} = \begin{bmatrix} 1 \\ X_h \end{bmatrix} \qquad \text{or} \qquad \underset{1\times 2}{\mathbf{X}'_h} = [1 \quad X_h] \tag{5.95}$$

The fitted value in matrix notation then is:

$$\hat{Y}_h = \mathbf{X}'_h\mathbf{b} \tag{5.96}$$

since:

$$\mathbf{X}'_h\mathbf{b} = [1 \quad X_h]\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = [b_0 + b_1X_h] = [\hat{Y}_h] = \hat{Y}_h$$

Note that $\mathbf{X}'_h\mathbf{b}$ is a 1×1 matrix; hence, we can write the final result as a scalar.

The variance of $\hat{Y}_h$, given earlier in (2.29b), in matrix notation is:

$$\sigma^2\{\hat{Y}_h\} = \sigma^2\mathbf{X}'_h(\mathbf{X'X})^{-1}\mathbf{X}_h \tag{5.97}$$

The variance of $\hat{Y}_h$ in (5.93) can be expressed as a function of $\boldsymbol{\sigma}^2\{\mathbf{b}\}$, the variance-covariance matrix of the estimated regression coefficients, by making use of the result in (5.92):

$$\sigma^2\{\hat{Y}_h\} = \mathbf{X}'_h\boldsymbol{\sigma}^2\{\mathbf{b}\}\mathbf{X}_h \tag{5.97a}$$

The estimated variance of $\hat{Y}_h$, given earlier in (2.30), in matrix notation is:

$$s^2\{\hat{Y}_h\} = MSE(\mathbf{X}'_h(\mathbf{X'X})^{-1}\mathbf{X}_h) \tag{5.98}$$

Example

We wish to find $s^2\{\hat{Y}_h\}$ for the Toluca Company example when $X_h = 65$. We define:

$$\mathbf{X}'_h = [1 \quad 65]$$

and use the result in (5.94) to obtain:

$$\begin{aligned}s^2\{\hat{Y}_h\} &= \mathbf{X}'_h\mathbf{s}^2\{\mathbf{b}\}\mathbf{X}_h\\ &= [1 \quad 65]\begin{bmatrix} 685.34 & -8.428 \\ -8.428 & .12040 \end{bmatrix}\begin{bmatrix} 1 \\ 65 \end{bmatrix} = 98.37\end{aligned}$$

This is the same result as that obtained in Chapter 2.

Comment

The result in (5.97a) can be derived directly by using (5.46), since $\hat{Y}_h = \mathbf{X}_h'\mathbf{b}$:

$$\sigma^2\{\hat{Y}_h\} = \mathbf{X}_h'\boldsymbol{\sigma}^2\{\mathbf{b}\}\mathbf{X}_h$$

Hence:

$$\sigma^2\{\hat{Y}_h\} = [1 \quad X_h]\begin{bmatrix} \sigma^2\{b_0\} & \sigma\{b_0, b_1\} \\ \sigma\{b_1, b_0\} & \sigma^2\{b_1\} \end{bmatrix}\begin{bmatrix} 1 \\ X_h \end{bmatrix}$$

or:

$$\sigma^2\{\hat{Y}_h\} = \sigma^2\{b_0\} + 2X_h\sigma\{b_0, b_1\} + X_h^2\sigma^2\{b_1\} \tag{5.99}$$

Using the results from (5.92a), we obtain:

$$\sigma^2\{\hat{Y}_h\} = \frac{\sigma^2}{n} + \frac{\sigma^2\bar{X}^2}{\sum(X_i - \bar{X})^2} + \frac{2X_h(-\bar{X})\sigma^2}{\sum(X_i - \bar{X})^2} + \frac{X_h^2\sigma^2}{\sum(X_i - \bar{X})^2}$$

which reduces to the familiar expression:

$$\sigma^2\{\hat{Y}_h\} = \sigma^2\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum(X_i - \bar{X})^2}\right] \tag{5.99a}$$

Thus, we see explicitly that the variance expression in (5.99a) contains contributions from $\sigma^2\{b_0\}$, $\sigma^2\{b_1\}$, and $\sigma\{b_0, b_1\}$, which it must according to (A.30b) since $\hat{Y}_h = b_0 + b_1X_h$ is a linear combination of b_0 and b_1. ■

Prediction of New Observation

The estimated variance $s^2\{\text{pred}\}$, given earlier in (2.38), in matrix notation is:

$$s^2\{\text{pred}\} = MSE(1 + \mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h) \tag{5.100}$$

Cited Reference

5.1. Graybill, F. A. *Matrices with Applications in Statistics.* 2nd ed. Belmont, Calif.: Wadsworth, 2002.

Problems

5.1. For the matrices below, obtain (1) $\mathbf{A} + \mathbf{B}$, (2) $\mathbf{A} - \mathbf{B}$, (3) $\mathbf{AC}$, (4) $\mathbf{AB}'$, (5) $\mathbf{B}'\mathbf{A}$.

$$\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 2 & 6 \\ 3 & 8 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 & 3 \\ 1 & 4 \\ 2 & 5 \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} 3 & 8 & 1 \\ 5 & 4 & 0 \end{bmatrix}$$

State the dimension of each resulting matrix.

5.2. For the matrices below, obtain (1) $\mathbf{A} + \mathbf{C}$, (2) $\mathbf{A} - \mathbf{C}$, (3) $\mathbf{B}'\mathbf{A}$, (4) $\mathbf{AC}'$, (5) $\mathbf{C}'\mathbf{A}$.

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 3 & 5 \\ 5 & 7 \\ 4 & 8 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 6 \\ 9 \\ 3 \\ 1 \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} 3 & 8 \\ 8 & 6 \\ 5 & 1 \\ 2 & 4 \end{bmatrix}$$

State the dimension of each resulting matrix.

5.3. Show how the following expressions are written in terms of matrices: (1) $Y_i - \hat{Y}_i = e_i$, (2) $\sum X_ie_i = 0$. Assume $i = 1, \ldots, 4$.