CHAPTER 4

Regression in Matrix Terms: Straight Line Case

We shall now present the steam data example given earlier in terms of matrix algebra. The use of matrices has many advantages, not the least of these being that once the problem is written and solved in matrix terms the solution can be applied to any regression problem no matter how many terms there are in the regression equation. Although there is a matrix algebra section in Chapter 0, some explanations are deliberately duplicated in this section.

Matrices

A matrix (plural matrices) is a rectangular array of symbols or numbers and is usually denoted by a single letter in **boldface** type, for example, $\mathbf{Q}$ or $\mathbf{q}$. There are several rules for manipulation of such arrays. Quite complicated expressions or equations can often be represented very simply by just a few letters properly defined and grouped.

We shall not introduce matrices formally but will use them in the context of the example. The reader with sound knowledge of matrices might wish to omit this chapter.

4.1. FITTING A STRAIGHT LINE IN MATRIX TERMS

We define $\mathbf{Y}$ to be the *vector of observations* Y_i, $\mathbf{X}$ to be the *matrix of predictor variables*, $\boldsymbol{\beta}$ to be the *vector of parameters to be estimated*, $\boldsymbol{\epsilon}$ to be a *vector of errors*, and $\mathbf{1}$ to be a vector of ones. In terms of the steam data in Table 1.1, and Eq. (1.2.3), we thus define

$$\mathbf{Y} = \begin{bmatrix} 10.98 \\ 11.13 \\ 12.51 \\ 8.40 \\ \vdots \\ 10.36 \\ 11.08 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & 35.3 \\ 1 & 29.7 \\ 1 & 30.8 \\ 1 & 58.8 \\ \vdots & \vdots \\ 1 & 33.4 \\ 1 & 28.6 \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \vdots \\ \epsilon_{24} \\ \epsilon_{25} \end{bmatrix} \quad \mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix}. \tag{4.1.1}$$

Note that

$\mathbf{Y}$ is a 25×1 vector.
$\mathbf{X}$ is a 25×2 matrix.
$\boldsymbol{\beta}$ is a 2×1 vector.
$\boldsymbol{\epsilon}$ is a 25×1 vector.
$\mathbf{1}$ is a 25×1 vector.

(Any matrix with one column is called a column vector; any matrix with one row is called a row vector. A 1×1 "matrix" is just an ordinary number or scalar.)

The dots in the matrices and vectors represent data not reproduced to save space in making the definition—a conventional procedure in matrix work. The column $\mathbf{1}$ of ones is not strictly needed at this stage, but it is convenient to define it here; it is extremely useful in matrix manipulations. Note that $\mathbf{X}$ is formed of two column vectors. The first is simply $\mathbf{1}$, the second is an (unnamed) vector of X-values in the data usually called the "X-column." Many writers call the column of ones in $\mathbf{X}$ the "X_0-column," pretending that there is a predictor variable X_0 that is always set at the value one. A variable chosen in some such arbitrary way is usually called a *dummy variable* and useful extended applications of such devices will be given in Chapters 14 and 23.

Manipulating Matrices

The rules of multiplication for matrices and vectors insist that two matrices must be *conformable*. For example, if $\mathbf{A}$ is an $n \times p$ matrix we can:

1. *Postmultiply* it by a $p \times q$ matrix to give as a result an $n \times p \times p \times q = n \times q$ matrix.
2. *Premultiply* it by an $m \times n$ matrix to give as a result an $m \times n \times n \times p = m \times p$ matrix.

Thus, for example, the multiplication $\boldsymbol{\beta}\mathbf{X}$ is not possible since $\boldsymbol{\beta}$ is 2×1 and $\mathbf{X}$ is 25×2. But $\mathbf{X}\boldsymbol{\beta}$ is possible as follows:

$$\underset{}{\mathbf{X}\boldsymbol{\beta}} = \underset{25 \times 2}{\begin{bmatrix} 1 & 35.3 \\ 1 & 29.7 \\ \vdots & \vdots \\ 1 & 28.6 \end{bmatrix}} \underset{2 \times 1}{\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}} = \underset{25 \times 1}{\begin{bmatrix} \beta_0 + 35.3\beta_1 \\ \beta_0 + 29.7\beta_1 \\ \vdots \\ \beta_0 + 28.6\beta_1 \end{bmatrix}}. \tag{4.1.2}$$

As a more general example consider the product

$$\underset{3 \times 3}{\overset{\mathbf{A}}{\begin{bmatrix} 1 & 2 & 4 \\ -1 & 0 & 1 \\ 2 & 3 & 1 \end{bmatrix}}} \underset{3 \times 2}{\overset{\mathbf{B}}{\begin{bmatrix} 1 & -1 \\ 2 & 1 \\ 3 & 5 \end{bmatrix}}} = \underset{3 \times 2}{\overset{\mathbf{C}}{\begin{bmatrix} 17 & 21 \\ 2 & 6 \\ 11 & 6 \end{bmatrix}}}.$$

To find the element in row i and column j of $\mathbf{C}$, we take row i of $\mathbf{A}$ and column j of $\mathbf{B}$, find the cross-product of corresponding elements, and add. For example,

$$\begin{array}{ll} \text{Row 2 of } \mathbf{A} \text{ is} & -1 \quad 0 \quad 1 \\ \text{Column 1 of } \mathbf{B} \text{ is} & 1 \quad 2 \quad 3 \end{array}$$

Thus the element in row 2, column 1 of **C** is

$$-1(1) + 0(2) + 1(3) = 2.$$

Orthogonality

Definition. If the sum of the cross-products of corresponding elements of row i and column j is zero, then row i is said to be *orthogonal* to column j. (The same definition applies to row and row or column and column.)

The Model in Matrix Form

The sum of two matrices or vectors is just the matrix whose elements are the sums of corresponding elements in the separate matrices or vectors. For example,

$$\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = \begin{bmatrix} \beta_0 + 35.3\beta_1 \\ \beta_0 + 29.7\beta_1 \\ \vdots \\ \beta_0 + 28.6\beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_{25} \end{bmatrix} = \begin{bmatrix} \beta_0 + 35.3\beta_1 + \epsilon_1 \\ \beta_0 + 29.7\beta_1 + \epsilon_2 \\ \vdots \\ \beta_0 + 28.6\beta_1 + \epsilon_{25} \end{bmatrix}. \tag{4.1.3}$$

The two matrices or vectors must have the same dimensions for this to be possible. (The difference between two matrices is similarly defined with differences instead of sums.) If two matrices or vectors are equal, corresponding elements are equal. Thus writing the matrix equation

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{4.1.4}$$

implies that

$$\begin{aligned} 10.98 &= \beta_0 + 35.3\beta_1 + \epsilon_1 \\ &\vdots \\ 11.08 &= \beta_0 + 28.6\beta_1 + \epsilon_{25} \end{aligned} \tag{4.1.5}$$

or

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \qquad (i = 1, \ldots, 25) \tag{4.1.6}$$

for each of the 25 observations. Thus the matrix equation, Eq. (4.1.4), and Eq. (4.1.6) express the same model. Equation (4.1.6) is identical to Eq. (1.2.3).

Setup for a Quadratic Model

In setting up the model in matrix form, only the choice of **X** usually presents any difficulty to the beginner. The simplest way of obtaining **X** is first to write down all of the parameters shown in the model as a vector $\boldsymbol{\beta}$, and then to see what corresponding X-columns would be needed to reproduce the model in its given algebraic form from the $\mathbf{X}\boldsymbol{\beta}$ product. For example, if the model is $Y = \beta_0 + \beta_1 X + \beta_{11} X^2 + \epsilon$, the vector $\boldsymbol{\beta}$ will be a column of three elements β_0, β_1, and β_{11} and the corresponding X-columns must necessarily be 1 (or X_0 if we use that notation), X, and X^2. Thus the ith row of $\mathbf{X}$ will consist of $(1, X_i, X_i^2)$, where X_i is the ith of the n observations. Note that a reordering of the elements of $\boldsymbol{\beta}$ requires a reordering of the columns of $\mathbf{X}$ to correspond.

Transpose

We now define the *transpose* of a matrix. It is the matrix obtained by writing all rows as columns in the order in which they occur so that the columns all become rows. The transpose of a matrix $\mathbf{M}$ is written $\mathbf{M}'$, for example,

$$\underset{3\times 2}{\mathbf{M} = \begin{bmatrix} 3 & 2 \\ 1 & 4 \\ 7 & 0 \end{bmatrix}}, \qquad \underset{2\times 3}{\mathbf{M}' = \begin{bmatrix} 3 & 1 & 7 \\ 2 & 4 & 0 \end{bmatrix}}.$$

Thus, for example,

$$\boldsymbol{\epsilon}' = (\epsilon_1, \epsilon_2, \dots, \epsilon_n).$$

Note that we can then write

$$\begin{aligned} \epsilon_1^2 + \epsilon_2^2 + \cdots + \epsilon_n^2 &= \boldsymbol{\epsilon}'\boldsymbol{\epsilon}, \\ Y_1^2 + Y_2^2 + \cdots + Y_n^2 &= \mathbf{Y}'\mathbf{Y}, \\ n\overline{Y} = Y_1 + Y_2 + \cdots + Y_n &= \mathbf{1}'\mathbf{Y}, \\ n\overline{Y}^2 = (\Sigma\, Y_i)^2/n &= \mathbf{Y}'\mathbf{1}\mathbf{1}'\mathbf{Y}/n. \end{aligned}$$

Furthermore

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 35.3 & 29.7 & \cdots & 28.6 \end{bmatrix} \begin{bmatrix} 1 & 35.3 \\ 1 & 29.7 \\ \vdots & \vdots \\ 1 & 28.6 \end{bmatrix} = \begin{bmatrix} 25 & 1315 \\ 1315 & 76323.42 \end{bmatrix}.$$

In general, for a straight line model, we see that

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \Sigma\, X_i \\ \Sigma\, X_i & \Sigma\, X_i^2 \end{bmatrix}. \tag{4.1.7}$$

In addition,

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 35.3 & 29.7 & \cdots & 28.6 \end{bmatrix} \begin{bmatrix} 10.98 \\ 11.13 \\ \vdots \\ 11.08 \end{bmatrix} = \begin{bmatrix} 235.60 \\ 11821.4320 \end{bmatrix}$$

so that, generally, for a straight line fit,

$$\mathbf{X'Y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \Sigma Y_i \\ \Sigma X_i Y_i \end{bmatrix}. \tag{4.1.8}$$

This means that the normal equations (1.2.8) can be written

$$\mathbf{X'Xb} = \mathbf{X'Y}, \tag{4.1.9}$$

where $\mathbf{b}' = (b_0, b_1)$, and these equations, when solved, provide the least squares estimates (b_0, b_1) of (β_0, β_1). How do we solve these equations in matrix form? To do so we define the *inverse* of a matrix. This exists only when a matrix is square and when the determinant of the matrix (a quantity that we shall not define here but of which we shall provide some examples) is nonzero. This latter condition is usually stated as *when the matrix is nonsingular*. This will be true in our applications unless otherwise stated. In regression work we wish to invert the $\mathbf{X'X}$ matrix. If it is *singular*, and so does not have an inverse, this will be reflected in the fact that some of the normal equations will be linear combinations of others; see, for example, Eq. (4.2.3). In this case there will be fewer equations than there are unknowns for which to solve. In such a case unique estimates are not possible unless some additional conditions on the parameters apply. (See Chapter 23 for additional comments on this point.)

Inverse of a Matrix

Suppose now that $\mathbf{M}$ is a nonsingular $p \times p$ matrix. The inverse of $\mathbf{M}$ is written $\mathbf{M}^{-1}$, is $p \times p$, and is such that

$$\mathbf{M}^{-1}\mathbf{M} = \mathbf{M}\mathbf{M}^{-1} = \mathbf{I}_p,$$

where $\mathbf{I}_p$ is the *unit matrix of order p*, which consists of unities (i.e., ones) in every position of the main diagonal (i.e., the diagonal running from the upper left corner to the lower right corner) and zeros elsewhere; for example,

$$\mathbf{I}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

(When the size of the unit matrix is obvious, the subscript is often omitted.) The unit matrix plays the same role in matrix multiplication that 1 does in ordinary multiplication—it leaves the multiplicand unchanged. The inverse of a matrix is unique.

Inverses of Small Matrices

The formulas for inverting matrices of sizes two and three are as follows:

$$\mathbf{M}^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \begin{bmatrix} d/D & -b/D \\ -c/D & a/D \end{bmatrix}, \tag{4.1.10}$$

where $D = ad - bc$ is the *determinant* of the 2×2 matrix $\mathbf{M}$.

$$\mathbf{Q}^{-1} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix}^{-1} = \begin{bmatrix} A & B & C \\ D & E & F \\ G & H & K \end{bmatrix}, \tag{4.1.11}$$

where

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

and where

$$\begin{aligned} Z &= a(ek - fh) - b(dk - fg) + c(dh - eg) \\ &= aek + bfg + cdh - ahf - dbk - gec \end{aligned}$$

is the determinant of $\mathbf{Q}$.

Matrix Symmetry for Square Matrices

Matrices of the form $\mathbf{X}'\mathbf{X}$ met in regression work are always symmetric, that is, the element in the ith row and jth column is the same as the element in the jth row and ith column. Thus the transpose of a symmetric matrix is the matrix itself. This is easy to see if we apply the general rule $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$ for transposes of a product. Because $(\mathbf{A}')' = \mathbf{A}$ itself, we can write $(\mathbf{X}'\mathbf{X})' = \mathbf{X}'\mathbf{X}$. (Working with a few simple numerical cases will clarify this point.) If the matrix $\mathbf{M}$ of size two above is symmetric, $b = c$ and the inverse also becomes symmetric. If the matrix $\mathbf{Q}$ above is symmetric, $b = d$, $c = g$, $f = h$. Then, relabeling the matrix $\mathbf{S}$, we obtain the symmetric inverse

$$\mathbf{S}^{-1} = \begin{bmatrix} a & b & c \\ b & e & f \\ c & f & k \end{bmatrix}^{-1} = \begin{bmatrix} A & B & C \\ B & E & F \\ C & F & K \end{bmatrix}, \tag{4.1.12}$$

where

$$\begin{array}{lll} A = (ek - f^2)/Y & B = -(bk - cf)/Y & C = (bf - ce)/Y \\ & E = (ak - c^2)/Y & F = -(af - bc)/Y \\ & & K = (ae - b^2)/Y \end{array}$$

and where

$$\begin{aligned} Y &= a(ek - f^2) - b(bk - cf) + c(bf - ce) \\ &= aek + 2bcf - af^2 - b^2k - c^2e \end{aligned}$$

is the determinant of $\mathbf{S}$. The inverse of any symmetric matrix is, itself, a symmetric matrix.

Diagonal Matrices

Matrices of sizes greater than three are usually cumbersome to invert unless they have a special form. One matrix that is easy to invert, no matter what its size, is a *diagonal* *matrix*, which consists of nonzero elements in the main upper-left to lower-right diagonal, and zeros elsewhere. The inverse is obtained by inverting all nonzero elements where they stand. For example,

$$\begin{bmatrix} a_1 & & & \mathbf{0} \\ & a_2 & & \\ & & \ddots & \\ \mathbf{0} & & & a_n \end{bmatrix}^{-1} = \begin{bmatrix} 1/a_1 & & & \mathbf{0} \\ & 1/a_2 & & \\ & & \ddots & \\ \mathbf{0} & & & 1/a_n \end{bmatrix}. \tag{4.1.13}$$

(Note, in this special case, the use of $\mathbf{0}$ to denote a large triangular block of zeros. This is often seen.)

Inverting Partitioned Matrices with Blocks of Zeros

Another type of simplification sometimes occurs when some columns of the $\mathbf{X}$ matrix are orthogonal to *all* other columns. The $\mathbf{X}'\mathbf{X}$ matrix then takes the *partitioned* form

$$\begin{bmatrix} \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix},$$

where, for example, $\mathbf{P}$ might be $p \times p$, $\mathbf{R}$ might be $r \times r$, and the symbol $\mathbf{0}$ is used to denote two differently shaped blocks of zeros, a $p \times r$ one in the top right-hand corner and an $r \times p$ one in the lower left-hand corner. The inverse of this matrix is then

$$\begin{bmatrix} \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{P}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}^{-1} \end{bmatrix}. \tag{4.1.14}$$

For example, if

$$\mathbf{P} = \begin{bmatrix} 1 & 3 \\ 2 & 8 \end{bmatrix}, \qquad \mathbf{P}^{-1} = \begin{bmatrix} 4 & -\frac{3}{2} \\ -1 & \frac{1}{2} \end{bmatrix},$$

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 3 & 2 \\ 4 & 1 & 1 \end{bmatrix}, \qquad \mathbf{R}^{-1} = \begin{bmatrix} -\frac{1}{9} & -\frac{1}{9} & \frac{1}{3} \\ -\frac{2}{3} & \frac{1}{3} & 0 \\ \frac{10}{9} & \frac{1}{9} & -\frac{1}{3} \end{bmatrix},$$

then

$$\begin{bmatrix} 1 & 3 & 0 & 0 & 0 \\ 2 & 8 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 2 & 3 & 2 \\ 0 & 0 & 4 & 1 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 4 & -\frac{3}{2} & 0 & 0 & 0 \\ -1 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{9} & -\frac{1}{9} & \frac{1}{3} \\ 0 & 0 & -\frac{2}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{10}{9} & \frac{1}{9} & -\frac{1}{3} \end{bmatrix}.$$

When there are more than two nonzero blocks, the obvious extension holds. It is important to note that the blocks must be on the main diagonal, and the off-diagonal blocks must consist entirely of zeros for the extension to apply.

Less Obvious Partitioning

The inverse formula (4.1.14) also applies even when the rows and columns containing nonzero elements are intermingled, provided that the matrix can be divided in such a way that the portions, such as **P** and **R** above, are completely separated from each other by zeros. For example, using the same numbers as above, the matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 3 & 0 \\ 2 & 0 & 3 & 0 & 2 \\ 0 & 2 & 0 & 8 & 0 \\ 4 & 0 & 1 & 0 & 1 \end{bmatrix}$$

can be partitioned and the separate portions inverted separately. Note that the second and fourth rows *and* columns are completely isolated, or insulated, from the first, third, and fifth rows and columns by zeros. Thus the nonzero elements in the second and fourth rows and columns comprise a 2×2 matrix, which can be separately inverted, whereas the other nonzero elements form a completely separate 3×3 matrix, which also can be separately inverted. Thus the inverse has the form

$$\begin{bmatrix} -\frac{1}{9} & 0 & -\frac{1}{9} & 0 & \frac{1}{3} \\ 0 & 4 & 0 & -\frac{3}{2} & 0 \\ -\frac{2}{3} & 0 & \frac{1}{3} & 0 & 0 \\ 0 & -1 & 0 & \frac{1}{2} & 0 \\ \frac{10}{9} & 0 & \frac{1}{9} & 0 & -\frac{1}{3} \end{bmatrix}.$$

Situations like this often occur when carefully designed experiments are analyzed using regression analysis.

The correctness of all these inverses can be confirmed by actually multiplying the inverse by the original, both before and behind. The result is an **I** matrix of appropriate size in every case. In practical situations, when the size of a matrix exceeds 3×3, and no simplified form is possible, finding the inverse can be a lengthy procedure. The work would usually be performed within an electronic computer.

Back to the Straight Line Case

We wish now to invert the $\mathbf{X'X}$ matrix of our example. This is of size 2×2 and of the general form of Eq. (4.1.7). Using Eq. (4.1.10) with $b = c$, we obtain the inverse as

$$(\mathbf{X'X})^{-1} = \begin{bmatrix} \dfrac{\Sigma X_i^2}{n\,\Sigma(X_i - \overline{X})^2} & \dfrac{-\overline{X}}{\Sigma(X_i - \overline{X})^2} \\ \dfrac{-\overline{X}}{\Sigma(X_i - \overline{X})^2} & \dfrac{1}{\Sigma(X_i - \overline{X})^2} \end{bmatrix}. \tag{4.1.15}$$

If *every* element of a matrix has a common factor it can be taken outside the matrix. (Conversely, if a matrix is multiplied by a constant C, every element of the matrix must be multiplied by C.) Thus an alternative form is

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{n\,\Sigma(X_i - \overline{X})^2}\begin{bmatrix} \Sigma X_i^2 & -\Sigma X_i \\ -\Sigma X_i & n \end{bmatrix}. \tag{4.1.16}$$

Since $\mathbf{X}'\mathbf{X}$ is symmetric, so is its inverse $(\mathbf{X}'\mathbf{X})^{-1}$ as mentioned earlier. The quantity taken outside the matrix is the determinant of $\mathbf{X}'\mathbf{X}$, written det $(\mathbf{X}'\mathbf{X})$ or $|\mathbf{X}'\mathbf{X}|$. Using the form of Eq. (4.1.15) on the steam data example we find that

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.4267941 & -0.0073535 \\ -0.0073535 & 0.0001398 \end{bmatrix}.$$

Solving the Normal Equations

If we premultiply Eq. (4.1.9) by $(\mathbf{X}'\mathbf{X})^{-1}$, we obtain

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

that is,

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{4.1.17}$$

since $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{I}$. This is an important result to remember since the solution of linear regression normal equations can *always* be written in this form, provided $\mathbf{X}'\mathbf{X}$ is nonsingular and the regression problem is properly expressed.

Using the data of our example we find that

$$\mathbf{b} = \begin{bmatrix} 0.4267941 & -0.0073535 \\ -0.0073535 & 0.0001398 \end{bmatrix}\begin{bmatrix} 235.60 \\ 11{,}821.4320 \end{bmatrix}$$

$$= \begin{bmatrix} 13.623790 \\ -0.079848 \end{bmatrix}.$$

A Small Sermon on Rounding Errors

Note that the results are not identical, to six places of decimals, to the values obtained in Section 1.2. Such discrepancies frequently occur because of the rounding off of numbers used in the calculation, and carelessness in such matters can cause serious errors, depending on the numbers involved. Here the numerical discrepancies are slight from a practical point of view, but they emphasize the fact that in general as many figures as possible should be carried in regression calculations. Sometimes, due to the magnitudes of the numbers in the calculation, the entire significance of the results can be lost through careless rounding.

Certain ways of performing the calculations (especially when they are done "by hand," i.e., on a pocket calculator) are better than others since they are less affected by round-off error. In particular, it is wise to postpone divisions to as late a stage as possible. For example, if we had employed the form of Eq. (4.1.16) instead of Eq. (4.1.15) to obtain $(\mathbf{X}'\mathbf{X})^{-1}$ we would have obtained

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{178{,}860.5}\begin{bmatrix} 76{,}323.42 & -1315 \\ -1315 & 25 \end{bmatrix}.$$

Then we could have obtained **b** from

$$\mathbf{b} = \frac{1}{178{,}860.5}\begin{bmatrix} 76{,}323.42 & -1315 \\ -1315 & 25 \end{bmatrix}\begin{bmatrix} 235.60 \\ 11{,}821.432 \end{bmatrix}$$

$$= \frac{1}{178{,}860.5}\begin{bmatrix} 2{,}436{,}614.672 \\ -14{,}278.2 \end{bmatrix}$$

$$= \begin{bmatrix} 13.622989 \\ -0.079829 \end{bmatrix},$$

the division being performed last of all.

Doing the calculations the three separate ways gives the following answers:

	Formulas (Section 1.2)	Inverse Matrix	Inverse Matrix (Division Last)
b_0	13.623005	13.623790	13.622989
b_1	−0.079829	−0.079848	−0.079829

As we have said, these differences are of slight consequence in this example. The third method is actually the most accurate. To see what the consequences of rounding can be, we suggest the reader make use of the inverse matrix in the second method, and round the elements in several ways—for example, rounding to 6, 5, 4, or 3 places of decimals or the same numbers of significant figures. Rounding errors provide a major share of disagreements when several people work the same regression problem using pocket calculators.

Modern computer programs vary somewhat in accuracy, but the variation is typically well below the horizon needed by the average user.

Section Summary

If we express the straight line model to be fitted to the data of our example in the form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

as in Eq. (4.1.4), then the least squares estimates of (β_0, β_1), that is, of

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix},$$

are given by the normal equations

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y},$$

which have the solution

$$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

This result is of great importance and should be memorized. Note that the fitted values $\hat{\mathbf{Y}}$ are obtained by evaluating

$$\hat{\mathbf{Y}} = \mathbf{Xb}.$$

4.2. SINGULARITY: WHAT HAPPENS IN REGRESSION TO MAKE X′X SINGULAR? AN EXAMPLE

In presenting the inverse matrix in Section 4.1, we said it need not exist. We think about this a little more via a simple example. We recall that, because terms in an inverse $(\mathbf{X}'\mathbf{X})^{-1}$ are always divided by the determinant of $\mathbf{X}'\mathbf{X}$, written $|\mathbf{X}'\mathbf{X}|$, the inverse "blows up" if the determinant is zero. We then say that $\mathbf{X}'\mathbf{X}$ is *singular* and that $(\mathbf{X}'\mathbf{X})^{-1}$ does not exist. The next question is how, in a regression situation, the determinant of $\mathbf{X}'\mathbf{X}$ can be zero. We illustrate this for a straight line case. Figure 4.1 shows several data points, all at the same value X_* of X. (There could be n points at X_* without affecting the argument below, so we use n in what follows.)

Suppose we are asked to fit a straight line to these data. Common sense tells us that we need data at two or more X-sites to determine a straight line "properly," that is, uniquely. So our first reaction might be that the fit cannot be made; but, of course, it can. *Any* straight line $b_0 + b_1X$ through the point $(X_*, \overline{Y})$ will minimize the sum of squares of deviations

$$S = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$$

because, for our data, $X_i = X_* = \overline{X}$, so that

$$\begin{aligned} S &= \sum_{i=1}^{n} (Y_i - \overline{Y} + \overline{Y} - \beta_0 - \beta_1 \overline{X})^2 \\ &= \sum_{i=1}^{n} (Y_i - \overline{Y})^2 + \sum_{i=1}^{n} (\overline{Y} - \beta_0 - \beta_1 \overline{X})^2 + \text{(zero cross-product)} \end{aligned}$$

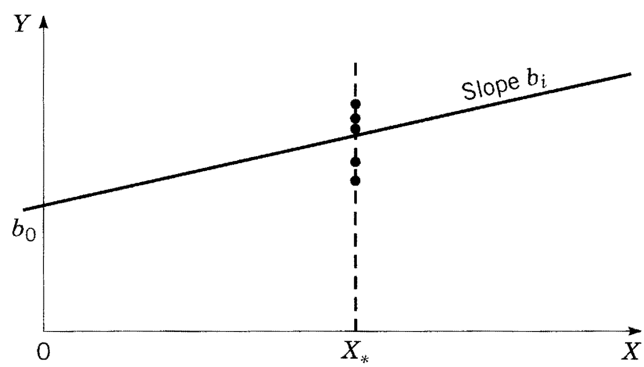

Figure 4.1.

and the second term vanishes when $\overline{Y} - \beta_0 - \beta_1\overline{X} = 0$. Thus any straight line $Y = b_0 + b_1 X$, which passes through or contains $(\overline{X}, \overline{Y})$, is a least squares line. The solution exists but is not unique. There is an infinity of solutions.

Let us follow this case through via matrix algebra. We see that

$$\mathbf{X}' = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_* & X_* & \cdots & X_* \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} 1 & X_* \\ 1 & X_* \\ \vdots & \vdots \\ 1 & X_* \end{bmatrix} \tag{4.2.1}$$

Note that the second column of $\mathbf{X}$ (i.e., the second row of $\mathbf{X}'$) is X_* times the first column, namely, a linear combination of it. Furthermore,

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} n & nX_* \\ nX_* & nX_*^2 \end{bmatrix} \tag{4.2.2}$$

with determinant $n(nX_*^2) - (nX_*)^2 = 0$. Thus we cannot evaluate $(\mathbf{X}'\mathbf{X})^{-1}$ nor carry out the formal computation $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{Y}$ because $\mathbf{X}'\mathbf{X}$ is singular. Note that the dependence in the columns of $\mathbf{X}$ has also been transmitted to the columns (and rows) of $\mathbf{X}'\mathbf{X}$; the second column of $\mathbf{X}'\mathbf{X}$ is X_* times the first column (and similarly with rows).

The implication of the singularity of $\mathbf{X}'\mathbf{X}$ is that we cannot estimate β_0 and β_1 *uniquely*. The normal equations can still be written down as $\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}\mathbf{Y}$, however. They just cannot be solved uniquely. For our simple example, the normal equations are

$$\begin{aligned} nb_0 + nX_*b_1 &= n\overline{Y}, \\ nX_*b_0 + nX_*^2b_1 &= nX_*\overline{Y}, \end{aligned} \tag{4.2.3}$$

and we see immediately that the second equation is X_* times the first, so that there is really only one distinct normal equation, not two. This single equation implies that $\overline{Y} = b_0 + b_1X_*$ and leads to an infinity of solutions expressible as

$$(b_0, b_1) = \{b_0, (\overline{Y} - b_0)/X_*\} \tag{4.2.4}$$

for any b_0. Essentially, we can pick any intercept value b_0 we please, and our nonunique least squares line so selected is created by joining the point $(0, b_0)$ to the point $(X_*, \overline{Y})$, giving a slope $b_1 = (\overline{Y} - b_0)/X_*$. (Alternatively, we could select a slope b_1 and determine b_0 via $b_0 = \overline{Y} - b_1X_*$.)

The basic computational difficulty inherent in our simple example occurs in more complex forms for multiple-X regression problems but, nevertheless, the overall principle is the same.

Singularity in the General Linear Regression Context

If, for a given set of data and a given model, the $\mathbf{X}$ matrix is such that any of its columns can be expressed as a linear combination of other columns, this dependency will be transferred to $\mathbf{X}'\mathbf{X}$ and so $\mathbf{X}'\mathbf{X}$ will have a zero determinant and be singular. This means that $(\mathbf{X}'\mathbf{X})^{-1}$ cannot be computed, and the least squares procedure will not give unique estimates but many alternative solutions. This arises because the data are inadequate for fitting the model or, what is the same thing, the model is too complex for the available data. One needs either more data, or a simpler model for the available data.

Computer programs are written so as to detect these problems. MINITAB, for example, leaves out of the regression fit any variable whose X-column is a linear combination of previous columns. Because all computations have rounding errors in them, columns that are not fully dependent on others, but almost so, are sometimes omitted. Recoding variables sometimes helps to avoid that possibility.

4.3. THE ANALYSIS OF VARIANCE IN MATRIX TERMS

We recall from Section 1.3 that, in a more general form of the analysis of variance table, we wrote

$$SS(b_1|b_0) = b_1\left[\Sigma X_iY_i - \frac{(\Sigma X_i)(\Sigma Y_i)}{n}\right] = b_1[\Sigma X_iY_i - n\overline{XY}],$$

$$SS(b_0) = \text{Correction for mean} = \frac{(\Sigma Y_i)^2}{n} = n\overline{Y}^2.$$

Each of these sums of squares has one degree of freedom. Now, adding these together,

$$\begin{aligned} SS(b_1|b_0) + SS(b_0) &= b_1\,\Sigma X_iY_i - b_1 n\,\overline{XY} + n\overline{Y}^2 \\ &= b_1\,\Sigma X_iY_i + n\overline{Y}(\overline{Y} - b_1\overline{X}) \\ &= b_1\,\Sigma X_iY_i + b_0\,\Sigma Y_i \qquad (4.3.1) \\ &= (b_0, b_1)\begin{bmatrix} \Sigma Y_i \\ \Sigma X_iY_i \end{bmatrix} \\ &= \mathbf{b'X'Y} \end{aligned}$$

in matrix terms, with two degrees of freedom. Thus we can write the analysis of variance table in matrix form as follows:

Source	df	SS	MS
$\mathbf{b}' = (b_0, b_1)$	2	$\mathbf{b'X'Y}$	
Residual	$n - 2$	$\mathbf{Y'Y} - \mathbf{b'X'Y}$	s^2
Total (uncorrected)	n	$\mathbf{Y'Y}$	

In this way we can split the total variation $\mathbf{Y'Y}$ into two portions, one due to the straight line we have estimated, namely, $\mathbf{b'X'Y}$, and a residual that shows the remaining variation of the points about the regression line. In order to find what portion of the total variation can be attributed to the addition of the term $\beta_1 X_i$ to the simpler model $Y_i = \beta_0 + \epsilon_i$, we would just subtract the correction factor $n\overline{Y}^2$ from the sum of squares $\mathbf{b'X'Y}$ in order to obtain $SS(b_1|b_0)$ as before. The quantity $n\overline{Y}^2$ would be $SS(b_0)$ if the model $Y_i = \beta_0 + \epsilon_i$ were fitted. The remainder of $\mathbf{b'X'Y}$ thus measures the *extra sum of squares removed by* b_1 when the model $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ is used. If an estimate of pure error from repeat points is available, it is subtracted from the residual sum of squares to provide the same breakup and the same tests as described in Section 2.1.

Example. For our steam data example we had

$$\mathbf{b} = \begin{bmatrix} 13.623790 \\ -0.079848 \end{bmatrix}, \qquad \mathbf{X'Y} = \begin{bmatrix} 235.60 \\ 11{,}821.4320 \end{bmatrix}.$$

Hence

$$\begin{aligned} SS(\mathbf{b}) &= \mathbf{b'X'Y} = 2265.8472, \\ SS(b_0) &= (\Sigma Y_i)^2/n = 2220.2944, \\ SS(b_1|b_0) &= SS(b_1, \text{after allowance for } b_0) \\ &= \mathbf{b'X'Y} - (\Sigma Y_i)^2/n = 45.5528. \end{aligned}$$

We see that this result differs from the value in Table 1.6 due to rounding error in the second decimal place.

Note that $SS(b_1|b_0)$ can be written in matrix form as

$$\begin{aligned} SS(b_1|b_0) &= \mathbf{b'X'Y} - \mathbf{Y'11'Y}/n \\ &= \mathbf{Y'X(X'X)^{-1}X'Y} - \mathbf{Y'11'Y}/n \\ &= \mathbf{Y'(H - 11'}/n)\mathbf{Y} \end{aligned}$$

where $\mathbf{H} = \mathbf{X(X'X)^{-1}X'}$ is a useful symmetric, idempotent ($\mathbf{H}^2 = \mathbf{H}$) matrix that occurs repeatedly in regression work. In replacing $\mathbf{b'}$ by $\mathbf{Y'X(X'X)^{-1}}$ above, we have made use of the important rule that the transpose of a product of matrices is the product of the transposes *in the reverse order*. In symbols, for example,

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

If we apply this rule to $\mathbf{b} = \mathbf{(X'X)^{-1}X'Y}$, identification of $\mathbf{A} = \mathbf{(X'X)^{-1}}$, $\mathbf{B} = \mathbf{X'}$, and $\mathbf{C} = \mathbf{Y}$, plus the realizations that $\mathbf{(X'X)^{-1}}$ is symmetric so that it is its own transpose, and

$$(\mathbf{X'})' = \mathbf{X}$$

(i.e., if we transpose a transpose, we are back to our starting point) bring the quoted result. (The transpose rule applies to a product of any size, by the way.)

Some other matrix regression results on which the reader might wish to test his or her matrix manipulative skills are the following:

$$\begin{aligned} \mathbf{e} &= (\mathbf{I} - \mathbf{H})\mathbf{Y}, \\ \mathbf{e'1} &= \mathbf{1'e} = 0, \\ \mathbf{e'\hat{Y}} &= \mathbf{\hat{Y}'e} = 0. \end{aligned}$$

4.4. THE VARIANCES AND COVARIANCE OF b_0 AND b_1 FROM THE MATRIX CALCULATION

We recall that $V(b_1) = \sigma^2/\Sigma(X_i - \overline{X})^2$. Also,

$$V(b_0) = V(\overline{Y} - b_1\overline{X}) = \sigma^2\left[\frac{1}{n} + \frac{\overline{X}^2}{\Sigma(X_i - X)^2}\right] = \frac{\sigma^2\Sigma X_i^2}{n\Sigma(X_i - \overline{X})^2} \qquad (4.4.1)$$

since, as we showed earlier, $\overline{Y}$ and b_1 have zero covariance, and the X's are regarded as constants. In addition,

$$\begin{aligned}\text{cov}(b_0, b_1) &= \text{cov}[(\bar{Y} - b_1\bar{X}), b_1] \\ &= -\bar{X}\,V(b_1) \\ &= -\bar{X}\sigma^2/\Sigma(X_i - \bar{X})^2.\end{aligned} \tag{4.4.2}$$

Thus we can write the *variance–covariance* matrix of the vector **b** as follows:

$$\begin{aligned}\mathbf{V}(\mathbf{b}) = \mathbf{V}\begin{pmatrix} b_0 \\ b_1 \end{pmatrix} &= \begin{bmatrix} V(b_0) & \text{cov}(b_0, b_1) \\ \text{cov}(b_0, b_1) & V(b_1) \end{bmatrix} \\ &= \begin{bmatrix} \dfrac{\sigma^2 \Sigma X_i^2}{n\,\Sigma(X_i - \bar{X})^2} & -\dfrac{\bar{X}\sigma^2}{\Sigma(X_i - \bar{X})^2} \\ -\dfrac{\bar{X}\sigma^2}{\Sigma(X_i - \bar{X})^2} & \dfrac{\sigma^2}{\Sigma(X_i - \bar{X})^2} \end{bmatrix}.\end{aligned} \tag{4.4.3}$$

Now if every element of a matrix has a common factor, we can remove it and set it outside the matrix, so that we can remove σ^2. The matrix that remains is seen to be $(\mathbf{X}'\mathbf{X})^{-1}$ from Eq. (4.1.15). Thus

$$\mathbf{V}(\mathbf{b}) = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2. \tag{4.4.4}$$

This is an important result and should be remembered. When σ^2 is unknown we use, instead, s^2, the estimate of σ^2 obtained from the analysis of variance table, if there is no lack of fit, or s_e^2, the pure error mean square if lack of fit is shown. This provides us with the *estimated variance–covariance matrix of* **b**. The standard errors of the regression coefficients are the square roots of the diagonal entries.

Correlation Between b_0 and b_1

The correlation between b_0 and b_1 can be obtained directly from the $(\mathbf{X}'\mathbf{X})^{-1}\sigma^2$ matrix, or the $(\mathbf{X}'\mathbf{X})^{-1}s^2$ matrix or even just the $(\mathbf{X}'\mathbf{X})^{-1}$ matrix because any common factor cancels anyway. We obtain

$$\text{Correlation}\,(b_0, b_1) = \frac{\text{cov}\,(b_0, b_1)}{\{V(b_0)V(b_1)\}^{1/2}} \tag{4.4.5}$$

$$= \frac{-\bar{X}\sigma^2/S_{XX}}{\left\{\dfrac{\sigma^2 \Sigma X_i^2}{nS_{XX}}\,\dfrac{\sigma^2}{S_{XX}}\right\}^{1/2}} \tag{4.4.6}$$

$$= -\bar{X}(n/\Sigma X_i^2)^{1/2}. \tag{4.4.7}$$

For the steam data, $\bar{X} = 52.6$, $n = 25$, $\Sigma X_i^2 = 76{,}323.42$, and the correlation is -0.952. Or we can use directly the numbers below (4.1.16) to give

$$\text{Correlation}\,(b_0, b_1) = -0.0073535/\{0.4267941)(0.0001398)\}^{1/2} = -0.952. \tag{4.4.8}$$

This is a relatively high value. This high negative correlation shows up in the relative position of the joint confidence region for (β_0, β_1) compared with the rectangular formed by the individual (marginal) confidence intervals. The negative sign implies the upper-left to lower-right slant of the larger axis of the ellipse while the high numerical value of 0.952 implies that it will run essentially from corner to corner of the rectangle. An accurate diagram is shown in Figure 5.3.

4.5. VARIANCE OF $\hat{Y}$ USING THE MATRIX DEVELOPMENT

Let X_0 be a selected value of X. The predicted mean value of Y for this value of X is

$$\hat{Y}_0 = b_0 + b_1 X_0.$$

Let us define the vector $\mathbf{X}_0$ as

$$\mathbf{X}_0' = (1, X_0).$$

We can then write

$$\hat{Y}_0 = (1, X_0)\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \mathbf{X}_0'\mathbf{b} = \mathbf{b}'\mathbf{X}_0.$$

Since $\hat{Y}_0$ is a linear combination of the random variables b_0 and b_1, it follows that

$$V(\hat{Y}_0) = V(b_0) + 2X_0 \operatorname{cov}(b_0, b_1) + X_0^2 V(b_1).$$

As can be verified by working out the indicated matrix and vector products, the above quantity can be expressed in the alternative form

$$\begin{aligned} V(\hat{Y}_0) &= [1, X_0]\begin{bmatrix} V(b_0) & \operatorname{cov}(b_0, b_1) \\ \operatorname{cov}(b_0, b_1) & V(b_1) \end{bmatrix}\begin{bmatrix} 1 \\ X_0 \end{bmatrix} \\ &= \mathbf{X}_0'(\mathbf{X}'\mathbf{X})^{-1}\sigma^2\mathbf{X}_0 \\ &= \mathbf{X}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0\sigma^2. \end{aligned}$$

Although now given in a different form, this is identical in value to Eq. (3.1.3). This important matrix result should be remembered. With suitable redefinition of $\mathbf{X}_0$ and $\mathbf{X}$, it is applicable to the general linear regression situation. A estimated variance is obtained when σ^2 is replaced by an estimate s^2.

4.6. SUMMARY OF MATRIX APPROACH TO FITTING A STRAIGHT LINE (NONSINGULAR CASE)

1. Set down the model in the form $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$.

2. Find $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ to obtain the least squares estimate $\mathbf{b}$ of $\boldsymbol{\beta}$ provided by the data. (This solves the normal equations $\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y}$.)

3. Construct $\mathbf{b}'\mathbf{X}'\mathbf{Y}$ the sum of squares due to coefficients and hence obtain the basic analysis of variance as follows:

Source	df	SS	MS
Regression	2	$\mathbf{b}'\mathbf{X}'\mathbf{Y}$	
Residual	$n - 2$	$\mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y}$	s^2 (estimates σ^2 if the model is correct)
Total	n	$\mathbf{Y}'\mathbf{Y}$	

Additional subdivision of the sum of squares is achieved by finding $SS(b_1|b_0)$, the extra sum of squares due to b_1, and introducing pure error. The more detailed analysis of variance table will take the following form:

Source		df	SS	MS	
$SS(\mathbf{b})$	$SS(b_0)$	1	$n\overline{Y}^2$		
	$SS(b_1\|b_0)$	1	$\mathbf{b}'\mathbf{X}'\mathbf{Y} - n\overline{Y}^2$		
Residual	Lack of fit	$n - 2 - n_e$	$\mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y} - SS(pe)$	MS_L	s^2
	Pure error	n_e	$SS(pe)$	s_e^2	
Total		n	$\mathbf{Y}'\mathbf{Y}$		

The second table is often rewritten with the *corrected* total sum of squares at the bottom, omitting the sum of squares due to the mean $n\overline{Y}^2$. [Incidentally, as we noted previously, we can write $n\overline{Y}^2$ in matrix form as $\mathbf{Y}'\mathbf{1}\mathbf{1}'\mathbf{Y}/n$ if we wish, although this is not usually done. This computation is, in fact, less subject to round-off error if performed as $(\Sigma Y_i)^2/n$.] The abbreviated table takes the following form:

Source		df	SS	MS	
$SS(b_1\|b_0)$		1	$\mathbf{b}'\mathbf{X}'\mathbf{Y} - n\overline{Y}^2$		
Residual	Lack of fit	$n - 2 - n_e$	$\mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y} - SS(pe)$	MS_L	s^2
	Pure error	n_e	$SS(pe)$	s_e^2	
Total, corrected		$n - 1$	$\mathbf{Y}'\mathbf{Y} - n\overline{Y}^2$		

The tests for lack of fit and (if there is no lack of fit) for $H_0:\beta_1 = 0$ versus $H_1:\beta_1 \neq 0$ are performed as described in Chapters 1 and 2. An additional measure of the regression is provided by the ratio

$$R^2 = \frac{(\mathbf{b}'\mathbf{X}'\mathbf{Y} - n\overline{Y}^2)}{(\mathbf{Y}'\mathbf{Y} - n\overline{Y}^2)}$$

4. If no lack of fit is shown, so that s^2 can be used as an estimate of σ^2, $(\mathbf{X}'\mathbf{X})^{-1}s^2$ will provide estimates of $V(b_0)$, $V(b_1)$, and $\mathrm{cov}(b_0, b_1)$ and enable individual coefficients to be tested or other calculations made as in Chapter 1.

5. The following quantities can be found:

$$\text{The vector of fitted values: } \hat{\mathbf{Y}} = \mathbf{X}\mathbf{b};$$

$$\text{A prediction of } Y \text{ at } X_0\text{: } \hat{Y}_0 = \mathbf{X}_0'\mathbf{b} = \mathbf{b}'\mathbf{X}_0;$$

$$\text{with variance: } V(\hat{Y}_0) = \mathbf{X}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0\sigma^2;$$

where $\mathbf{X}_0' = (1, X_0)$.

4.7. THE GENERAL REGRESSION SITUATION

We have seen how the problem of fitting a straight line by least squares can be handled through the use of matrices. This approach is important for the following reason. If we wish to fit *any* model linear in parameters β_0, β_1, β_2, . . . , by least squares, the calculations necessary are of exactly the same form (in matrix terms) as those for the straight line involving only two parameters β_0 and β_1. Only the sizes of the matrices and the numbers of certain degrees of freedom change. The mechanics of calculation, however, increase sharply with the number of parameters. Thus while the formulas are easy to remember, the use of a computer is essential. Even when few parameters are involved, or when the data arise from a designed experiment that provides an $\mathbf{X}'\mathbf{X}$ matrix of simple or patterned form, computer evaluation is preferable.

We deal with the general regression situation in Chapter 5.

EXERCISES FOR CHAPTER 4

A. In this question we define

$$\mathbf{A} = \begin{bmatrix} 4 & 1 \\ 3 & -2 \\ 1 & 7 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 3 & -1 & 2 \\ -2 & 3 & 1 \\ 1 & 4 & 1 \end{bmatrix}.$$

Are the statements below true or false?

1. $\mathbf{B} + \mathbf{C} = \begin{bmatrix} 5 & 0 & 2 \\ -1 & 5 & 1 \\ 1 & 4 & 1 \end{bmatrix}.$

2. $\mathbf{AC} = \begin{bmatrix} 7 & 9 & 12 \\ 14 & 21 & 7 \end{bmatrix}.$

3. $\mathbf{AB} = \begin{bmatrix} 9 & 6 \\ 4 & -1 \\ 9 & 15 \end{bmatrix}.$

4. $\mathbf{B}^{-1} = \dfrac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$

5. $\mathbf{A}^{-1} = \dfrac{1}{11}\begin{bmatrix} 2 & 1 & 0 \\ 3 & -4 & 0 \end{bmatrix}.$

6. $(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{ABB}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$

B. We define

$$\mathbf{A} = \begin{bmatrix} 4 & 0 & 3 \\ 0 & 4 & 0 \\ 3 & 0 & 2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} -1 & 1 \\ 2 & 3 \\ 3 & 2 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix}.$$

Calculate the matrices below, or say that it is impossible to do so, if it is impossible.

1. $\mathbf{B} + \mathbf{C}$.

2. $\mathbf{BB}'$.

3. $\mathbf{A} + \mathbf{B}'\mathbf{B}$.

4. $\mathbf{BC}$.

5. $\mathbf{AA}^{-1}\mathbf{BC}$.
6. $\mathbf{CB}'$.
7. $\mathbf{CAB}$.
8. $\mathbf{BC}^{-1}$, where $\mathbf{C}^{-1} = \begin{bmatrix} -2 & 3 \\ 3 & -4 \end{bmatrix}$.
9. $\mathbf{A}^{-1}$.
10. $\mathbf{A}'\mathbf{A}(\mathbf{A}')^{-1}\mathbf{A}^{-1}$.

C. We define

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \\ -1 & 1 & 0 \end{bmatrix}.$$

Are the results below true or false? If false, say (briefly) why.

1. $\mathbf{Ab} = [0, 1, 2]$.
2. $\mathbf{A}'\mathbf{C} = \begin{bmatrix} 3 & 3 \\ 5 & 6 \\ 7 & 9 \end{bmatrix}$.
3. $\mathbf{AD} = \begin{bmatrix} 1 & 4 & 1 \\ 6 & 8 & 1 \end{bmatrix}$.
4. $\mathbf{C}^{-1} = \begin{bmatrix} 1 & 0 \\ -\frac{2}{3} & \frac{1}{3} \end{bmatrix}$.
5. $\mathbf{bC} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$.
6. $\mathbf{A} - \mathbf{D} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 3 \\ -1 & 1 & 0 \end{bmatrix}$.
7. $\mathbf{C}^{-1}\mathbf{CCC}^{-1}\mathbf{C}^{-1}\mathbf{CCC}^{-1}\mathbf{C} = \mathbf{C}$.
8. $\mathbf{D}^{-1} = \frac{1}{3}\begin{bmatrix} 0 & 1 & -2 \\ 0 & 1 & 1 \\ 3 & -2 & 1 \end{bmatrix}$.
9. $\mathbf{b}'\mathbf{Cb} = 2$.

D. Using the matrix development throughout, fit to the data below, the model $Y = \beta_0 + \beta_1 X + \epsilon$, and so obtain b_0 and b_1. Plot the data and the fitted line. Find the fitted values and the residuals correct to one decimal place. Evaluate the analysis of variance table, and test for lack of fit. Find the variance–covariance matrix of the estimated parameters, and the matrix expression for $V(\hat{Y})$. Hence find $V(\hat{Y})$ when $X = 65$, and construct a 95% confidence interval for $E(Y|X = 65)$.

X:	30	40	50	80	30	40	60	70	70	70	30	80	70	70
Y:	13	17	20	29	12	15	22	25	23	27	15	27	24	26

E. Using the matrix development, fit a straight line $Y = \beta_0 + \beta_1 X + \epsilon$ to the data below, and provide a complete analysis (i.e., all relevant details of Chapter 4).

X	Y
1	4.2
1	3.8
2	3.0
3	2.3
4	1.8
4	2.0
4	2.2
5	2.0
6	2.5
6	2.7

F. Using the data below, go through the following steps:

1. Fit a straight line model $Y = \beta_0 + \beta_1 X + \epsilon$, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma^2)$.
2. Plot the data and your fitted line.
3. Obtain the basic ANOVA table.
4. Add additional details on the ANOVA table.
5. Test for lack of fit.
6. Carry out the usual F-test for overall regression. Is this test valid?
7. Find, via the usual calculations, a 95% confidence interval for the true mean value of Y at $X_0 = \sqrt{122} + 6$. Is it valid?
8. Evaluate the fitted values and residuals. Plot each e_i versus its corresponding $\hat{Y}_i$.
9. Write down the form of the variance–covariance matrix of the b's in terms of σ^2.
10. Evaluate R^2.
11. Evaluate r_{XY}^2. What is its relationship to R^2, for this model?
12. What are your overall conclusions?

	X	Y	XY
	0	−2	0
	2	0	0
	2	2	4
	5	1	5
	5	3	15
	9	1	9
	9	0	0
	9	0	0
	9	1	9
	10	−1	−10
Sum	60	5	32
Sum of Squares	482	21	528

Note: When using a calculator, it is best to work with integers and simple fractions as long as you can. Convert to decimals only at the last possible moment.