(a) Plot the logarithm of number of hospital days (for the smokers) against number of cigarettes. Do you think a linear regression will adequately represent the relationship?

(b) Plot the logarithm of number of hospital days for smokers minus the logarithm of number of hospital days for the control group against number of cigarettes. Do you think a linear regression will adequately represent the relationship? Has subtraction of the control group means reduced the dispersion?

(c) Define $Y = \ln(\#\text{ days for smokers}) - \ln(\#\text{ days for nonsmokers})$ and $X = (\#\text{cigarettes})^2$. Fit the linear regression of Y on X. Make a test of significance to determine if the intercept can be set to zero. Depending on your results, give the regression equation, the standard errors of the estimates, and the summary analysis of variance.

1.22. Use the normal equations in 1.6 to show that

(a) $\sum X_i Y_i = \sum X_i \widehat{Y}_i$.

(b) $\sum X_i e_i = 0$.

(c) $\sum \widehat{Y}_i e_i = 0$. (Hint: use Exercise 1.8).

1.23 Consider the regression through the origin model in equation 1.39. Suppose $X_i \geq 0$. Define $\tilde{\beta}_1 = \sum Y_i / \sum X_i$ and $\widehat{\beta}_1 = \sum X_i Y_i / \sum X_i^2$.

(a) Show that $\tilde{\beta}_1$ and $\widehat{\beta}_1$ are unbiased for β_1.

(b) Compare the variances of $\tilde{\beta}_1$ and $\widehat{\beta}_1$.

2
INTRODUCTION TO MATRICES

Chapter 1 reviewed simple linear regression in algebraic notation and showed that the notation for models involving several variables is very cumbersome.

This chapter introduces matrix notation and all matrix operations that are used in this text. Matrix algebra greatly simplifies the presentation of regression and is used throughout the text. Sections 2.7 and 2.8 are not used until later in the text and can be omitted for now.

Matrix algebra is extremely helpful in multiple regression for simplifying notation and algebraic manipulations. You must be familiar with the basic operations of matrices in order to understand the regression results presented. A brief introduction to the key matrix operations is given in this chapter. You are referred to matrix algebra texts, for example, Searle (1982), Searle and Hausman (1970), or Stewart (1973), for more complete presentations of matrix algebra.

2.1 Basic Definitions

A **matrix** is a rectangular array of numbers arranged in orderly rows and columns. Matrices are denoted with boldface capital letters. The following

Matrix

are examples.

$$\boldsymbol{Z} = \begin{bmatrix} 1 & 2 \\ 6 & 4 \\ 5 & 7 \end{bmatrix} \qquad \boldsymbol{X} = \begin{bmatrix} 1 & 5 \\ 1 & 6 \\ 1 & 4 \\ 1 & 9 \\ 1 & 2 \\ 1 & 6 \end{bmatrix}$$

$$\boldsymbol{B} = \begin{bmatrix} 15 & 7 & -1 & 0 \\ 15 & 5 & -2 & 10 \end{bmatrix}.$$

Elements

The numbers that form a matrix are called the **elements** of the matrix. A general matrix could be denoted as

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

The subscripts on the elements denote the row and column, respectively, in which the element appears. For example, a_{23} is the element found in the second row and third column. The row number is always given first.

Order

The **order** of a matrix is its size given by the number of rows and columns. The first matrix given, $\boldsymbol{Z}$, is of order (3, 2). That is, $\boldsymbol{Z}$ is a 3×2 matrix, since it has three rows and two columns. Matrix $\boldsymbol{A}$ is an $m \times n$ matrix.

Rank

The **rank** of a matrix is defined as the number of linearly independent columns (or rows) in the matrix. Any subset of columns of a matrix are **linearly independent** if no column in the subset can be expressed as a linear combination of the others in the subset. The matrix

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 4 \\ 3 & 0 & 6 \\ 5 & 3 & 13 \end{bmatrix}$$

contains a linear dependency among its columns. The first column multiplied by two and added to the second column produces the third column. In fact, any one of the three columns of $\boldsymbol{A}$ can be written as a linear combination of the other two columns. On the other hand, any *two* columns of $\boldsymbol{A}$ are linearly independent since one cannot be produced as a multiple of the other. Thus, the rank of the matrix $\boldsymbol{A}$, denoted by $r(\boldsymbol{A})$, is two.

Full-Rank Matrices

If there are no linear dependencies among the columns of a matrix, the matrix is said to be of **full rank**, or **nonsingular**. If a matrix is *not* of full rank it is said to be **singular**. The number of linearly independent rows of a matrix will always equal the number of linearly independent columns. The linear dependency among the rows of $\boldsymbol{A}$ is shown by 9(row1)+ 7(row2) = 6(row3). The critical matrices in regression will almost always

have fewer columns than rows and, therefore, rank is more easily visualized by inspection of the columns.

The collection of all linear combinations of columns of $\boldsymbol{A}$ is called the **column space** of $\boldsymbol{A}$ or the space spanned by the columns of $\boldsymbol{A}$. **Column Space**

2.2 Special Types of Matrices

A **vector** is a matrix having only one row or one column, and is called a row or column vector, respectively. Although vectors are often designated with boldface lowercase letters, this convention is not followed rigorously in this text. A boldface capital letter is used to designate a data vector and a boldface Greek letter is used for vectors of parameters. Thus, for example, **Vector**

$$\boldsymbol{v} = \begin{pmatrix} 3 \\ 8 \\ 2 \\ 1 \end{pmatrix} \text{ is a } 4 \times 1 \text{ column vector.}$$

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 & \mu_2 & \mu_3 \end{pmatrix} \text{ is a } 1 \times 3 \text{ row vector.}$$

We usually define the vectors as column vectors but they need not be. A single number such as 4, -2.1, or 0 is called a **scalar**.

A **square matrix** has an equal number of rows and columns. **Square Matrix**

$$\boldsymbol{D} = \begin{bmatrix} 2 & 4 \\ 6 & 7 \end{bmatrix} \text{ is a } 2 \times 2 \text{ square matrix.}$$

A **diagonal matrix** is a square matrix in which all elements are zero except the elements on the main diagonal, the diagonal of elements, a_{11}, a_{22}, $\ldots$, a_{nn}, running from the upper left postion to the lower right position. **Diagonal Matrix**

$$\boldsymbol{A} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 8 \end{bmatrix} \text{ is a } 3 \times 3 \text{ diagonal matrix.}$$

An **identity matrix** is a diagonal matrix having all the diagonal elements equal to 1; such a matrix is denoted by $\boldsymbol{I}_n$. The subscript identifies the order of the matrix and is omitted when the order is clear from the context. **Identity Matrix**

$$\boldsymbol{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ is a } 3 \times 3 \text{ identity matrix.}$$

After matrix multiplication is discussed, it can be verified that multiplying any matrix by the identity matrix will not change the original matrix.

A **symmetric matrix** is a square matrix in which element a_{ij} equals element a_{ji} for all i and j. The elements form a symmetric pattern around the diagonal of the matrix. **Symmetric Matrix**

$$\boldsymbol{A} = \begin{bmatrix} 5 & -2 & 3 \\ -2 & 4 & -1 \\ 3 & -1 & 8 \end{bmatrix} \text{ is a } 3 \times 3 \text{ symmetric matrix.}$$

Note that the first row is identical to the first column, the second row is identical to the second column, and so on.

2.3 Matrix Operations

The **transpose** of a matrix $\boldsymbol{A}$, designated $\boldsymbol{A}'$, is the matrix obtained by using the rows of $\boldsymbol{A}$ as the columns of $\boldsymbol{A}'$. If **Transpose**

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 3 & 8 \\ 4 & 1 \\ 5 & 9 \end{bmatrix},$$

the transpose of $\boldsymbol{A}$ is

$$\boldsymbol{A}' = \begin{bmatrix} 1 & 3 & 4 & 5 \\ 2 & 8 & 1 & 9 \end{bmatrix}.$$

If a matrix $\boldsymbol{A}$ has order $m \times n$, its transpose $\boldsymbol{A}'$ has order $n \times m$. A symmetric matrix is equal to its transpose: $\boldsymbol{A}' = \boldsymbol{A}$.

Addition of two matrices is defined if and only if the matrices are of the same order. Then, addition (or subtraction) consists of adding (or subtracting) the corresponding elements of the two matrices. For example, **Addition**

$$\begin{bmatrix} 1 & 2 \\ 3 & 8 \end{bmatrix} + \begin{bmatrix} 7 & -6 \\ 8 & 2 \end{bmatrix} = \begin{bmatrix} 8 & -4 \\ 11 & 10 \end{bmatrix}.$$

Addition is commutative: $\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}$.

Multiplication of two matrices is defined if and only if the number of *columns* in the first matrix equals the number of *rows* in the second matrix. If $\boldsymbol{A}$ is of order $r \times s$ and $\boldsymbol{B}$ is of order $m \times n$, the matrix product $\boldsymbol{AB}$ exists only if $s = m$. The matrix product $\boldsymbol{BA}$ exists only if $r = n$. Multiplication is most easily defined by first considering the multiplication of a row vector times a column vector. Let $\boldsymbol{a}' = (\, a_1 \quad a_2 \quad a_3 \,)$ and $\boldsymbol{b}' = (\, b_1 \quad b_2 \quad b_3 \,)$. (Notice that both $\boldsymbol{a}$ and $\boldsymbol{b}$ are defined as column vectors.) Then, the product of $\boldsymbol{a}'$ and $\boldsymbol{b}$ is **Multiplication**

$$\begin{aligned} \boldsymbol{a}'\boldsymbol{b} &= (\, a_1 \quad a_2 \quad a_3 \,) \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \\ &= a_1 b_1 + a_2 b_2 + a_3 b_3. \end{aligned} \tag{2.1}$$

The result is a scalar equal to the sum of products of the corresponding elements. Let

$$\boldsymbol{a}' = \begin{pmatrix} 3 & 6 & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{b}' = \begin{pmatrix} 2 & 4 & 8 \end{pmatrix}.$$

The matrix product is

$$\boldsymbol{a}'\boldsymbol{b} = \begin{pmatrix} 3 & 6 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 4 \\ 8 \end{pmatrix} = 6 + 24 + 8 = 38.$$

Matrix multiplication is defined as a sequence of vector multiplications. Write

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \quad \text{as} \quad \boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1' \\ \boldsymbol{a}_2' \end{pmatrix},$$

where $\boldsymbol{a}_1' = \begin{pmatrix} a_{11} & a_{12} & a_{13} \end{pmatrix}$ and $\boldsymbol{a}_2' = \begin{pmatrix} a_{21} & a_{22} & a_{23} \end{pmatrix}$ are the 1×3 row vectors in $\boldsymbol{A}$. Similarly, write

$$\boldsymbol{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} \quad \text{as } \boldsymbol{B} = \begin{pmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 \end{pmatrix},$$

where $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ are the 3×1 column vectors in $\boldsymbol{B}$. Then the product of $\boldsymbol{A}$ and $\boldsymbol{B}$ is the 2×2 matrix

$$\boldsymbol{A}\boldsymbol{B} = \boldsymbol{C} = \begin{bmatrix} \boldsymbol{a}_1'\boldsymbol{b}_1 & \boldsymbol{a}_1'\boldsymbol{b}_2 \\ \boldsymbol{a}_2'\boldsymbol{b}_1 & \boldsymbol{a}_2'\boldsymbol{b}_2 \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}, \tag{2.2}$$

where

$$\begin{aligned} c_{11} &= \boldsymbol{a}_1'\boldsymbol{b}_1 = \sum_{j=1}^{3} a_{1j}b_{j1} = a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} \\ c_{12} &= \boldsymbol{a}_1'\boldsymbol{b}_2 = \sum_{j=1}^{3} a_{1j}b_{j2} = a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ c_{21} &= \boldsymbol{a}_2'\boldsymbol{b}_1 = \sum_{j=1}^{3} a_{2j}b_{j1} = a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} \\ c_{22} &= \boldsymbol{a}_2'\boldsymbol{b}_2 = \sum_{j=1}^{3} a_{2j}b_{j2} = a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32}. \end{aligned}$$

In general, element c_{ij} is obtained from the vector multiplication of the ith *row* vector from the first matrix and the jth *column* vector from the second matrix. The resulting matrix $\boldsymbol{C}$ has the number of rows equal to

the number of rows in $\boldsymbol{A}$ and number of columns equal to the number of columns in $\boldsymbol{B}$.

Example 2.1

Let

$$\boldsymbol{T} = \begin{bmatrix} 1 & 2 \\ 4 & 5 \\ 3 & 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{W} = \begin{pmatrix} -1 \\ 3 \end{pmatrix}.$$

The product $\boldsymbol{WT}$ is not defined since the number of columns in $\boldsymbol{W}$ is not equal to the number of rows in $\boldsymbol{T}$. The product $\boldsymbol{TW}$, however, is defined:

$$\begin{aligned} \boldsymbol{TW} &= \begin{bmatrix} 1 & 2 \\ 4 & 5 \\ 3 & 0 \end{bmatrix} \begin{pmatrix} -1 \\ 3 \end{pmatrix} \\ &= \begin{pmatrix} (1)(-1)+(2)(3) \\ (4)(-1)+(5)(3) \\ (3)(-1)+(0)(3) \end{pmatrix} = \begin{pmatrix} 5 \\ 11 \\ -3 \end{pmatrix}. \end{aligned}$$

The resulting matrix is of order 3×1 with the elements being determined by multiplication of the corresponding row vector from $\boldsymbol{T}$ with the column vector in $\boldsymbol{W}$. ■

Matrix multiplication is not commutative; $\boldsymbol{AB}$ does not necessarily equal $\boldsymbol{BA}$ even if both products exist. As for the matrices $\boldsymbol{W}$ and $\boldsymbol{T}$ in Example 2.1, the matrices are not of the proper order for multiplication to be defined in both ways. The first step in matrix multiplication is to verify that the matrices do conform (have the proper order) for multiplication.

The transpose of a product is equal to the product *in reverse order* of the transposes of the two matrices. That is,

$$(\boldsymbol{AB})' = \boldsymbol{B}'\boldsymbol{A}'. \tag{2.3}$$

The transpose of the product of $\boldsymbol{T}$ and $\boldsymbol{W}$ from Example 2.1 is

$$\begin{aligned} (\boldsymbol{TW})' = \boldsymbol{W}'\boldsymbol{T}' &= \begin{pmatrix} -1 & 3 \end{pmatrix} \begin{bmatrix} 1 & 4 & 3 \\ 2 & 5 & 0 \end{bmatrix} \\ &= \begin{pmatrix} 5 & 11 & -3 \end{pmatrix}. \end{aligned}$$

Scalar multiplication is the multiplication of a matrix by a single number. Every element in the matrix is multiplied by the scalar. Thus,

$$3 \begin{bmatrix} 2 & 1 & 7 \\ 3 & 5 & 9 \end{bmatrix} = \begin{bmatrix} 6 & 3 & 21 \\ 9 & 15 & 27 \end{bmatrix}.$$

Determinant

The **determinant** of a matrix is a scalar computed from the elements of

the matrix according to well-defined rules. Determinants are defined only for square matrices and are denoted by $|\boldsymbol{A}|$, where $\boldsymbol{A}$ is a square matrix. The determinant of a 1×1 matrix is the scalar itself. The determinant of a 2×2 matrix,

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

is defined as

$$|\boldsymbol{A}| \quad = \quad a_{11}a_{22} - a_{12}a_{21}. \tag{2.4}$$

For example, if

$$\boldsymbol{A} = \begin{bmatrix} 1 & 6 \\ -2 & 10 \end{bmatrix},$$

the determinant of $\boldsymbol{A}$ is

$$|\boldsymbol{A}| = (1)(10) - (6)(-2) = 22.$$

The determinants of higher-order matrices are obtained by expanding the determinants as linear functions of determinants of 2×2 submatrices. First, it is convenient to define the **minor** and the **cofactor** of an element in a matrix. Let $\boldsymbol{A}$ be a square matrix of order n. For any element a_{rs} in $\boldsymbol{A}$, a square matrix of order $(n-1)$ is formed by eliminating the row and column containing the element a_{rs}. Label this matrix $\boldsymbol{A}_{rs}$, with the subscripts designating the row and column eliminated from $\boldsymbol{A}$. Then $|\boldsymbol{A}_{rs}|$, the determinant of $\boldsymbol{A}_{rs}$, is called the **minor** of the element a_{rs}. The product $\theta_{rs} = (-1)^{r+s} |\boldsymbol{A}_{rs}|$ is called the **cofactor** of a_{rs}. Each element in a square matrix has its own minor and cofactor.

The determinant of a matrix of order n is expressed in terms of the elements of any row or column and their cofactors. Using row i for illustration, we can express the determinant of $\boldsymbol{A}$ as

$$|\boldsymbol{A}| \quad = \quad \sum_{j=1}^{n} a_{ij}\theta_{ij}, \tag{2.5}$$

where each θ_{ij} contains a determinant of order $(n-1)$. Thus, the determinant of order n is expanded as a function of determinants of one less order. Each of these determinants, in turn, is expanded as a linear function of determinants of order $(n-2)$. This substitution of determinants of one less order continues until $|\boldsymbol{A}|$ is expressed in terms of determinants of 2×2 submatrices of $\boldsymbol{A}$.

The first step of the expansion is illustrated for a 3×3 matrix $\boldsymbol{A}$. To compute the determinant of $\boldsymbol{A}$, choose any row or column of the matrix. For each element of the row or column chosen, compute the cofactor of the element. Then, if the ith row of $\boldsymbol{A}$ is used for the expansion,

$$|\boldsymbol{A}| = a_{i1}\theta_{i1} + a_{i2}\theta_{i2} + a_{i3}\theta_{i3}. \tag{2.6}$$

Example 2.2

For illustration, let

$$\boldsymbol{A} = \begin{bmatrix} 2 & 4 & 6 \\ 1 & 2 & 3 \\ 5 & 7 & 9 \end{bmatrix}$$

and use the first row for the expansion of $|\boldsymbol{A}|$. The cofactors of the elements in the first row are

$$\begin{aligned} \theta_{11} &= (-1)^{(1+1)} \begin{vmatrix} 2 & 3 \\ 7 & 9 \end{vmatrix} = (18 - 21) = -3, \\ \theta_{12} &= (-1)^{(1+2)} \begin{vmatrix} 1 & 3 \\ 5 & 9 \end{vmatrix} = -(9 - 15) = 6, \text{ and} \\ \theta_{13} &= (-1)^{(1+3)} \begin{vmatrix} 1 & 2 \\ 5 & 7 \end{vmatrix} = (7 - 10) = -3. \end{aligned}$$

Then, the determinant of $\boldsymbol{A}$ is

$$|\boldsymbol{A}| = 2(-3) + 4(6) + 6(-3) = 0$$

. ■

If the determinant of a matrix is zero, the matrix is **singular**, or it is not of full rank. Otherwise, the matrix is **nonsingular**. Thus, the matrix $\boldsymbol{A}$ in Example 2.2 is singular. The linear dependency is seen by noting that row 1 is equal to twice row 2. The determinants of larger matrices rapidly become difficult to compute and are obtained with the help of a computer.

Inverse of a Matrix

Division in the usual sense does not exist in matrix algebra. The concept is replaced by multiplication by the **inverse** of the matrix. The inverse of a matrix $\boldsymbol{A}$, designated by $\boldsymbol{A}^{-1}$, is defined as the matrix that gives the identity matrix when multiplied by $\boldsymbol{A}$. That is,

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}. \tag{2.7}$$

The inverse of a matrix may not exist. A matrix has a *unique inverse* if and only if the matrix is square and nonsingular. A matrix is nonsingular if and only if its determinant is not zero.

The inverse of a 2×2 matrix is easily computed. If

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

then

$$\boldsymbol{A}^{-1} = \frac{1}{|\boldsymbol{A}|} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}. \tag{2.8}$$

Note the rearrangement of the elements and the use of the determinant of $\boldsymbol{A}$ as the scalar divisor. For example, if

$$\boldsymbol{A} = \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix}, \quad \text{then} \quad \boldsymbol{A}^{-1} = \begin{bmatrix} \frac{2}{5} & -\frac{3}{5} \\ -\frac{1}{5} & \frac{4}{5} \end{bmatrix}.$$

That this is the inverse of $\boldsymbol{A}$ is verified by multiplication of $\boldsymbol{A}$ and $\boldsymbol{A}^{-1}$:

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \frac{2}{5} & -\frac{3}{5} \\ -\frac{1}{5} & \frac{4}{5} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The inverse of a matrix is obtained in general by (1) replacing every element of the matrix with its cofactor, (2) transposing the resulting matrix, and (3) dividing by the determinant of the original matrix, as illustrated in the next example.

Example 2.3

Consider the following matrix,

$$\boldsymbol{B} = \begin{bmatrix} 1 & 3 & 2 \\ 4 & 5 & 6 \\ 8 & 7 & 9 \end{bmatrix}.$$

The determinant of $\boldsymbol{B}$ is

$$\begin{aligned} |\boldsymbol{B}| &= 1\begin{vmatrix} 5 & 6 \\ 7 & 9 \end{vmatrix} - 3\begin{vmatrix} 4 & 6 \\ 8 & 9 \end{vmatrix} + 2\begin{vmatrix} 4 & 5 \\ 8 & 7 \end{vmatrix} \\ &= (45-42) - 3(36-48) + 2(28-40) \\ &= 15. \end{aligned}$$

The unique inverse of $\boldsymbol{B}$ exists since $|\boldsymbol{B}| \neq 0$. The cofactors for the elements of the first row of $\boldsymbol{B}$ were used in obtaining $|\boldsymbol{B}|$: $\theta_{11} = 3,\ \theta_{12} = 12,\ \theta_{13} = -12$. The remaining cofactors are:

$$\theta_{21} = -\begin{vmatrix} 3 & 2 \\ 7 & 9 \end{vmatrix} = -13 \quad \theta_{22} = \begin{vmatrix} 1 & 2 \\ 8 & 9 \end{vmatrix} = -7 \quad \theta_{23} = -\begin{vmatrix} 1 & 3 \\ 8 & 7 \end{vmatrix} = 17$$
$$\theta_{31} = \begin{vmatrix} 3 & 2 \\ 5 & 6 \end{vmatrix} = 8 \qquad \theta_{32} = -\begin{vmatrix} 1 & 2 \\ 4 & 6 \end{vmatrix} = 2 \quad \theta_{33} = \begin{vmatrix} 1 & 3 \\ 4 & 5 \end{vmatrix} = -7.$$

Thus, the matrix of cofactors is

$$\begin{bmatrix} 3 & 12 & -12 \\ -13 & -7 & 17 \\ 8 & 2 & -7 \end{bmatrix}$$

and the inverse of $\boldsymbol{B}$ is

$$\boldsymbol{B}^{-1} = \frac{1}{15}\begin{bmatrix} 3 & -13 & 8 \\ 12 & -7 & 2 \\ -12 & 17 & -7 \end{bmatrix}.$$

Notice that the matrix of cofactors has been transposed and divided by $|\boldsymbol{B}|$ to obtain $\boldsymbol{B}^{-1}$. It is left as an exercise to verify that this is the inverse of $\boldsymbol{B}$. As with the determinants, computers are used to find the inverses of larger matrices. ■

Inverse of a Diagonal Matrix

Note that if $\boldsymbol{A}$ is a diagonal nonsingular matrix, then $\boldsymbol{A}^{-1}$ is also a diagonal matrix where the diagonal elements of $\boldsymbol{A}^{-1}$ are the reciprocals of the diagonal elements of $\boldsymbol{A}$. That is, if

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ 0 & a_{22} & 0 & \cdots & 0 \\ 0 & 0 & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix},$$

where $a_{ii} \neq 0$, then

$$\boldsymbol{A} = \begin{bmatrix} a_{11}^{-1} & 0 & 0 & \cdots & 0 \\ 0 & a_{22}^{-1} & 0 & \cdots & 0 \\ 0 & 0 & a_{33}^{-1} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn}^{-1} \end{bmatrix}.$$

Also, if $\boldsymbol{A}$ and $\boldsymbol{B}$ are two nonsingular matrices, then

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{A}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B}^{-1} \end{bmatrix}.$$

2.4 Geometric Interpretations of Vectors

The elements of an $n \times 1$ vector can be thought of as the coordinates of a point in an n-dimensional coordinate system. The vector is represented in this n-space as the directional line connecting the origin of the coordinate system to the point specified by the elements. The direction of the vector is *from* the origin *to* the point; an arrowhead at the terminus indicates direction.

Vector Length

To illustrate, let $\boldsymbol{x}' = (3 \quad 2)$. This vector is of order two and is plotted in two-dimensional space as the line vector going from the origin (0, 0) to

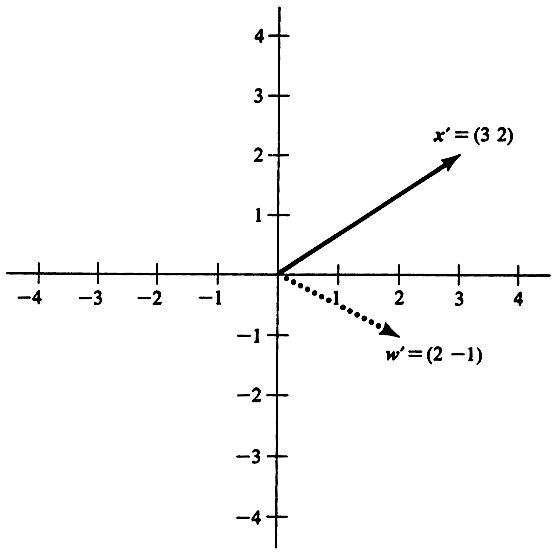

FIGURE 2.1. *The geometric representation of the vectors* $\boldsymbol{x}' = (3, 2)$ *and* $\boldsymbol{w}' = (2, -1)$ *in two-dimensional space.*

the point (3, 2) (see Figure 2.1). This can be viewed as the hypotenuse of a right triangle whose sides are of length 3 and 2, the elements of the vector $\boldsymbol{x}$. The length of $\boldsymbol{x}$ is then given by the Pythagorean theorem as the square root of the *sum of squares* of the elements of $\boldsymbol{x}$. Thus,

$$\text{length}(\boldsymbol{x}) = \sqrt{3^2 + 2^2} = \sqrt{13} = 3.61.$$

This result extends to the length of any vector regardless of its order. The sum of squares of the elements in a column vector $\boldsymbol{x}$ is given by (the matrix multiplication) $\boldsymbol{x}'\boldsymbol{x}$. Thus, the length of any vector $\boldsymbol{x}$ is

$$\text{length}(\boldsymbol{x}) = \sqrt{\boldsymbol{x}'\boldsymbol{x}}. \tag{2.9}$$

Space Defined by $\boldsymbol{x}$

Multiplication of $\boldsymbol{x}$ by a scalar defines another vector that falls precisely on the line formed by extending the vector $\boldsymbol{x}$ indefinitely in both directions. For example,

$$\boldsymbol{u}' = (-1)\boldsymbol{x}' = (\,-3 \quad -2\,)$$

falls on the extension of $\boldsymbol{x}$ in the negative direction. Any point on this indefinite extension of $\boldsymbol{x}$ in both directions can be "reached" by multiplication of $\boldsymbol{x}$ with an appropriate scalar. This set of points constitutes the **space** defined by $\boldsymbol{x}$, or the space **spanned** by $\boldsymbol{x}$. It is a one-dimensional subspace of the two-dimensional space in which the vectors are plotted. A single

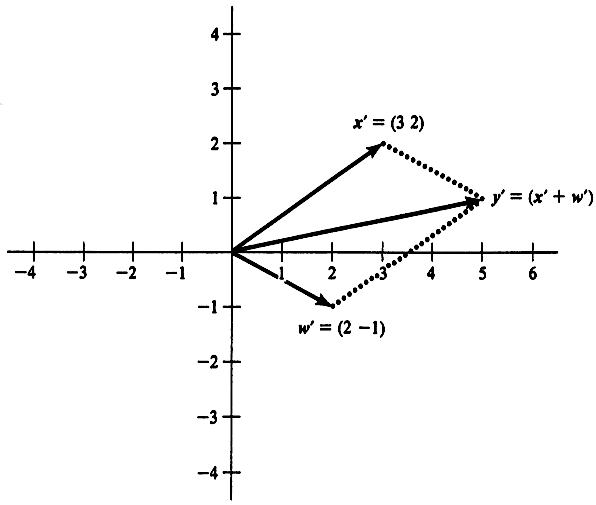

FIGURE 2.2. *Geometric representation of the sum of two vectors.*

vector of order n defines a one-dimensional subspace of the n-dimensional space in which the vector falls.

Linear Independence

The second vector $\boldsymbol{w}' = (2 \quad 1)$, shown in Figure 2.1 with a dotted line, defines another one-dimensional subspace. The two subspaces defined by $\boldsymbol{x}$ and $\boldsymbol{w}$ are disjoint subspaces (except for the common origin). The two vectors are said to be **linearly independent** since neither falls in the subspace defined by the other. This implies that one vector cannot be obtained by multiplication of the other vector by a scalar.

Two-Dimensional Subspace

If the two vectors are considered jointly, any point in the plane can be "reached" by an appropriate linear combination of the two vectors. For example, the sum of the two vectors gives the vector $\boldsymbol{y}$ (see Figure 2.2),

$$\boldsymbol{y}' = \boldsymbol{x}' + \boldsymbol{w}' = (3 \quad 2) + (2 \quad -1) = (5 \quad 1).$$

The two vectors $\boldsymbol{x}$ and $\boldsymbol{w}$ define, or span, the two-dimensional subspace represented by the plane in Figure 2.2. Any third vector of order 2 in this two-dimensional space *must* be a linear combination of $\boldsymbol{x}$ and $\boldsymbol{w}$. That is, there *must* be a linear dependency among any three vectors that fall on this plane.

Vector Addition

Geometrically, the vector $\boldsymbol{x}$ is added to $\boldsymbol{w}$ by moving $\boldsymbol{x}$, while maintaining its direction, until the base of $\boldsymbol{x}$ rests on the terminus of $\boldsymbol{w}$. The resultant vector $\boldsymbol{y}$ is the vector from the origin (0, 0) to the new terminus of $\boldsymbol{x}$. The same result is obtained by moving $\boldsymbol{w}$ along the vector $\boldsymbol{x}$. This is equivalent

to completing the parallelogram using the two original vectors as adjacent sides. The sum $\boldsymbol{y}$ is the diagonal of the parallelogram running from the origin to the opposite corner (see Figure 2.2). Subtraction of two vectors, say $\boldsymbol{w}' - \boldsymbol{x}'$, is most easily viewed as the addition of $\boldsymbol{w}'$ and $(-\boldsymbol{x}')$.

Three-Dimensional Subspace

Vectors of order 3 are considered briefly to show the more general behavior. Each vector of order 3 can be plotted in three-dimensional space; the elements of the vector define the endpoint of the vector. Each vector individually defines a one-dimensional subspace of the three-dimensional space. This subspace is formed by extending the vector indefinitely in both directions. Any *two* vectors define a two-dimensional subspace if the two vectors are **linearly independent**—that is, as long as the two vectors do not define the same subspace. The two-dimensional subspace defined by two vectors is the set of points in the *plane* defined by the origin and the endpoints of the two vectors. The two vectors defining the subspace and any linear combination of them lie in this plane.

A three-dimensional space contains an infinity of two-dimensional subspaces. These can be visualized by rotating the plane around the origin. Any third vector that does not fall in the original plane will, in conjunction with either of the first two vectors, define another plane. Any three linearly independent vectors, or any two planes, completely define, or span, the three-dimensional space. Any fourth vector in that three-dimensional subspace must be a linear function of the first three vectors. That is, any four vectors in a three-dimensional subspace *must* contain a linear dependency.

The general results are stated in the box:

1. Any vector of order n can be plotted in n-dimensional space and defines a one-dimensional subspace of the n-dimensional space.
2. Any p linearly independent vectors of order n, $p < n$, define a p-dimensional subspace.
3. Any $p + 1$ vectors in a p-dimensional subspace *must* contain a linear dependency.

Orthogonal Vectors

Two vectors $\boldsymbol{x}$ and $\boldsymbol{w}$ of the same order are **orthogonal** vectors if the vector product

$$\boldsymbol{x}'\boldsymbol{w} \quad = \quad \boldsymbol{w}'\boldsymbol{x} = 0. \tag{2.10}$$

If

$$\boldsymbol{x} = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 4 \end{pmatrix} \quad \text{and} \quad \boldsymbol{w} = \begin{pmatrix} 3 \\ 4 \\ -1 \\ -1 \end{pmatrix},$$

then $\boldsymbol{x}$ and $\boldsymbol{w}$ are orthogonal because

$$\boldsymbol{x}'\boldsymbol{w} = (1)(3) + (0)(4) + (-1)(-1) + (4)(-1) = 0.$$

Geometrically, two orthogonal vectors are perpendicular to each other or they form a right angle at the origin.

Linearly Dependent Vectors

Two **linearly dependent** vectors form angles of 0 or 180 degrees at the origin. All other angles reflect vectors that are neither orthogonal nor linearly dependent. In general, the cosine of the angle α between two (column) vectors $\boldsymbol{x}$ and $\boldsymbol{w}$ is

$$\cos(\alpha) = \frac{\boldsymbol{x}'\boldsymbol{w}}{\sqrt{\boldsymbol{x}'\boldsymbol{x}}\sqrt{\boldsymbol{w}'\boldsymbol{w}}}. \tag{2.11}$$

If the elements of each vector have mean zero, the *cosine* of the angle formed by two vectors is the **product moment correlation** between the two columns of data in the vectors. Thus, orthogonality of two such vectors corresponds to a zero correlation between the elements in the two vectors. If two such vectors are linearly dependent, the correlation coefficient between the elements of the two vectors will be either $+1.0$ or -1.0 depending on whether the vectors have the same or opposite directions.

2.5 Linear Equations and Solutions

A set of r linear equations in s unknowns is represented in matrix notation as $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$, where $\boldsymbol{x}$ is a vector of the s unknowns, $\boldsymbol{A}$ is the $r \times s$ matrix of known coefficients on the s unknowns, and $\boldsymbol{y}$ is the $r \times 1$ vector of known constants on the right-hand side of the equations.

A set of equations may have (1) no solution, (2) a unique solution, or (3) an infinite number of solutions. In order to have at least one solution, the equations must be **consistent**. This means that any linear dependencies among the rows of $\boldsymbol{A}$ must also exist among the corresponding elements of $\boldsymbol{y}$ (Searle and Hausman, 1970). For example, the equations

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 3 & 3 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 6 \\ 10 \\ 9 \end{pmatrix}$$

are **inconsistent** since the second row of $\boldsymbol{A}$ is twice the first row but the second element of $\boldsymbol{y}$ is *not* twice the first element. Since they are *not* consistent, there is no solution to this set of equations. Note that $\boldsymbol{x}' = (\,1 \quad 1 \quad 1\,)$ satisfies the first and third equations but not the second. If the second element of $\boldsymbol{y}$ were 12 instead of 10, the equations would be consistent and the solution $\boldsymbol{x}' = (\,1 \quad 1 \quad 1\,)$ would satisfy all three equations.

Consistent Equations

One method of determining if a set of equations is consistent is to compare the rank of $\boldsymbol{A}$ to the rank of the augmented matrix $[\boldsymbol{A}\ \boldsymbol{y}]$. The equations are consistent if and only if

$$r(\boldsymbol{A}) = r([\boldsymbol{A}\ \boldsymbol{y}]). \tag{2.12}$$

Rank can be determined by using elementary (row and column) operations to reduce the elements below the diagonal to zero. Operations such as addition of two rows, interchanging rows, and obtaining a scalar multiple of a row are called elementary row operations. (In a rectangular matrix, the diagonal is defined as the elements $a_{11}, a_{22}, \ldots, a_{dd}$, where d is the lesser of the number of rows and number of columns.) The number of rows with at least one nonzero element after reduction is the rank of the matrix.

Example 2.4

Elementary operations on

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 3 & 3 \end{bmatrix}$$

give

$$\boldsymbol{A}^* = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 0 & 0 \end{bmatrix}$$

so that $r(\boldsymbol{A}) = 2$. [The elementary operations to obtain $\boldsymbol{A}^*$ are (1) subtract 2 times row 1 from row 2, (2) subtract 3 times row 1 from row 3, and (3) interchange rows 2 and 3.] The same elementary operations, plus interchanging columns 3 and 4, on the augmented matrix

$$[\boldsymbol{A}\ \boldsymbol{y}] = \begin{bmatrix} 1 & 2 & 3 & 6 \\ 2 & 4 & 6 & 10 \\ 3 & 3 & 3 & 9 \end{bmatrix}$$

give

$$[\boldsymbol{A}\ \boldsymbol{y}]^* = \begin{bmatrix} 1 & 2 & 6 & 3 \\ 0 & -3 & -9 & -6 \\ 0 & 0 & -2 & 0 \end{bmatrix}.$$

Thus, $r([\boldsymbol{A}\ \boldsymbol{y}]) = 3$. Since $r([\boldsymbol{A}\ \boldsymbol{y}]) \neq r(\boldsymbol{A})$, the equations are not consistent and, therefore, they have no solution. ■

Unique Solution

Consistent equations either have a unique solution or an infinity of solutions. If $r(\boldsymbol{A})$ equals the number of unknowns, the solution is unique and is given by

1. $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{y}$, when $\boldsymbol{A}$ is square; or
2. $\boldsymbol{x} = \boldsymbol{A}_1^{-1}\boldsymbol{y}$, where $\boldsymbol{A}_1$ is a full rank submatrix of $\boldsymbol{A}$, when $\boldsymbol{A}$ is rectangular.

Example 2.5

The equations $\boldsymbol{Ax} = \boldsymbol{y}$ with

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 3 & 3 \\ 5 & 7 \end{bmatrix} \quad \text{and} \quad \boldsymbol{y} = \begin{pmatrix} 6 \\ 9 \\ 21 \end{pmatrix}$$

are consistent. (Proof of consistency is left as an exercise.) The rank of $\boldsymbol{A}$ equals the number of unknowns $[r(\boldsymbol{A}) = 2]$, so that the solution is unique. Any two linearly independent equations in the system of equations can be used to obtain the solution. Using the first two rows gives the full-rank equations

$$\begin{bmatrix} 1 & 2 \\ 3 & 3 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 6 \\ 9 \end{pmatrix}$$

with the solution

$$\begin{aligned} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \begin{bmatrix} 1 & 2 \\ 3 & 3 \end{bmatrix}^{-1} \begin{pmatrix} 6 \\ 9 \end{pmatrix} \\ &= \frac{1}{3} \begin{bmatrix} -3 & 2 \\ 3 & -1 \end{bmatrix} \begin{pmatrix} 6 \\ 9 \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \end{pmatrix}. \end{aligned}$$

Notice that the solution $\boldsymbol{x}' = (\,0 \quad 3\,)$ satisfies the third equation also. ■

Infinite Solutions

When $r(\boldsymbol{A})$ in a consistent set of equations is less than the number of unknowns, there is an infinity of solutions.

Example 2.6

Consider the equations $\boldsymbol{Ax} = \boldsymbol{y}$ with

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 3 & 3 \end{bmatrix} \quad \text{and } \boldsymbol{y} = \begin{pmatrix} 6 \\ 12 \\ 9 \end{pmatrix}.$$

The rank of $\boldsymbol{A}$ is $r(\boldsymbol{A}) = 2$ and elementary operations on the augmented matrix $[\boldsymbol{A}\ \boldsymbol{y}]$ give

$$[\boldsymbol{A}\ \boldsymbol{y}]^* = \begin{bmatrix} 1 & 2 & 3 & 6 \\ 0 & -3 & -6 & -18 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Thus, $r([\boldsymbol{A}\ \boldsymbol{y}]) = 2$, which equals $r(\boldsymbol{A})$, and the equations are consistent. However, $r(\boldsymbol{A})$ is less than the number of unknowns so that there is an

infinity of solutions. This infinity of solutions comes from the fact that one element of $\boldsymbol{x}$ can be chosen arbitrarily and the remaining two chosen so as to satisfy the set of equations. For example, if x_1 is chosen to be 1, the solution is $\boldsymbol{x}' = (1 \quad 1 \quad 1)$, whereas if x_1 is chosen to be 2, the solution is $\boldsymbol{x}' = (2 \quad -1 \quad 2)$. ■

Solutions Using Generalized Inverses

A more general method of finding a solution to a set of consistent equations involves the use of **generalized inverses**. There are several definitions of generalized inverses [see Searle (1971), Searle and Hausman (1970), and Rao (1973)]. An adequate definition for our purposes is the following (Searle and Hausman, 1970).

> A generalized inverse of $\boldsymbol{A}$ is any matrix $\boldsymbol{A}^-$ that satisfies the condition $\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A} = \boldsymbol{A}$.

($\boldsymbol{A}^-$ is used to denote a generalized inverse.) The generalized inverse is not unique (unless $\boldsymbol{A}$ is square and of full rank, in which case $\boldsymbol{A}^- = \boldsymbol{A}^{-1}$). A generalized inverse can be used to express a solution to a set of consistent equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ as $\boldsymbol{x} = \boldsymbol{A}^-\boldsymbol{y}$. This solution is unique only when $r(\boldsymbol{A})$ equals the number of unknowns in the set of equations. (The computer is used to obtain generalized inverses when needed.)

Example 2.7

For illustration, consider the set of consistent equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ where

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 3 & 3 \\ 5 & 7 \end{bmatrix} \quad \text{and} \quad \boldsymbol{y} = \begin{pmatrix} 6 \\ 9 \\ 21 \end{pmatrix}.$$

It has been shown that $r(\boldsymbol{A}) = 2$ which equals the number of unknowns so that the solution is unique. A generalized inverse of $\boldsymbol{A}$ is

$$\boldsymbol{A}^- = \frac{1}{18}\begin{bmatrix} -10 & 16 & -4 \\ 8 & -11 & 5 \end{bmatrix}$$

and the unique solution is given by

$$\boldsymbol{x} = \boldsymbol{A}^-\boldsymbol{y} = \begin{pmatrix} 0 \\ 3 \end{pmatrix}.$$

It is left as an exercise to verify the matrix multiplication of $\boldsymbol{A}^-\boldsymbol{y}$ and that $\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A} = \boldsymbol{A}$. ■

Example 2.8

For another illustration, consider again the consistent equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ from Example 2.6, where

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 3 & 3 \end{bmatrix} \quad \text{and } \boldsymbol{y} = \begin{pmatrix} 6 \\ 12 \\ 9 \end{pmatrix}.$$

This system has been shown to have an infinity of solutions. A generalized inverse of $\boldsymbol{A}$ is

$$\boldsymbol{A}^{-} = \begin{bmatrix} -\frac{1}{10} & -\frac{2}{10} & \frac{4}{9} \\ 0 & 0 & \frac{1}{9} \\ \frac{1}{10} & \frac{2}{10} & -\frac{2}{9} \end{bmatrix},$$

which gives the solution

$$\boldsymbol{x} = \boldsymbol{A}^{-}\boldsymbol{y} = (\,1 \quad 1 \quad 1\,)'.$$

This happens to agree with the first solution obtained in Example 2.6. Again, it is left as an exercise to verify that $\boldsymbol{x} = \boldsymbol{A}^{-}\boldsymbol{y}$ and $\boldsymbol{A}\boldsymbol{A}^{-}\boldsymbol{A} = \boldsymbol{A}$. A different generalized inverse of $\boldsymbol{A}$ may lead to another solution of the equations. ■

2.6 Orthogonal Transformations and Projections

The linear transformation of vector $\boldsymbol{x}$ to vector $\boldsymbol{y}$, both of order n, is written as $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, where $\boldsymbol{A}$ is the $n \times n$ matrix of coefficients effecting the transformation. The transformation is a one-to-one transformation only if $\boldsymbol{A}$ is nonsingular. Then, the inverse transformation of $\boldsymbol{y}$ to $\boldsymbol{x}$ is $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{y}$.

Orthogonal Transformations

A linear transformation is an **orthogonal** transformation if $\boldsymbol{A}\boldsymbol{A}' = \boldsymbol{I}$. This condition implies that the row vectors of $\boldsymbol{A}$ are orthogonal and of unit length. Orthogonal transformations maintain distances and angles between vectors. That is, the spatial relationships among the vectors are not changed with orthogonal transformations.

Example 2.9

For illustration, let $\boldsymbol{y}_1' = (\,3 \quad 10 \quad 20\,)$, $\boldsymbol{y}_2' = (\,6 \quad 14 \quad 21\,)$, and

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & 2 & -1 \end{bmatrix}.$$

Then

$$\boldsymbol{x}_1 = \boldsymbol{A}\boldsymbol{y}_1 = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & 2 & -1 \end{bmatrix} \begin{pmatrix} 3 \\ 10 \\ 20 \end{pmatrix} = \begin{pmatrix} 33 \\ 17 \\ -3 \end{pmatrix}$$

and

$$\boldsymbol{x}_2 = \boldsymbol{A}\boldsymbol{y}_2 = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & 2 & -1 \end{bmatrix} \begin{pmatrix} 6 \\ 14 \\ 21 \end{pmatrix} = \begin{pmatrix} 41 \\ 15 \\ 1 \end{pmatrix}$$

are linear transformations of $\boldsymbol{y}_1$ to $\boldsymbol{x}_1$ and $\boldsymbol{y}_2$ to $\boldsymbol{x}_2$. These are not orthogonal transformations because

$$\boldsymbol{A}\boldsymbol{A}' = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 6 \end{bmatrix} \neq \boldsymbol{I}.$$

The rows of $\boldsymbol{A}$ are mutually orthogonal (the off-diagonal elements are zero) but they do not have unit length. This can be made into an orthogonal transformation by scaling each row vector of $\boldsymbol{A}$ to have unit length by dividing each vector by its length. Thus,

$$\boldsymbol{x}_1^* = \boldsymbol{A}^*\boldsymbol{y}_1 = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \end{bmatrix} \boldsymbol{y}_1 = \begin{pmatrix} \frac{33}{\sqrt{3}} \\ \frac{17}{\sqrt{2}} \\ -\frac{3}{\sqrt{6}} \end{pmatrix}$$

and

$$\boldsymbol{x}_2^* = \boldsymbol{A}^*\boldsymbol{y}_2 = \begin{pmatrix} \frac{41}{\sqrt{3}} \\ \frac{15}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} \end{pmatrix}$$

are orthogonal transformations. It is left as an exercise to verify that the orthogonal transformation has maintained the distance between the two vectors; that is, verify that

$$(\boldsymbol{y}_1 - \boldsymbol{y}_2)'(\boldsymbol{y}_1 - \boldsymbol{y}_2) = (\boldsymbol{x}_1^* - \boldsymbol{x}_2^*)'(\boldsymbol{x}_1^* - \boldsymbol{x}_2^*) = 26.$$

[The squared distance between two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ is $(\boldsymbol{u} - \boldsymbol{v})'(\boldsymbol{u} - \boldsymbol{v})$.] ■

Projections

Projection of a vector onto a subspace is a special case of a transformation. (Projection is a key step in least squares.) The objective of a projection is to transform $\boldsymbol{y}$ in n-dimensional space to that vector $\widehat{\boldsymbol{y}}$ in a subspace such that $\widehat{\boldsymbol{y}}$ is as close to $\boldsymbol{y}$ as possible. A linear transformation of $\boldsymbol{y}$ to $\widehat{\boldsymbol{y}}$, $\widehat{\boldsymbol{y}} = \boldsymbol{P}\boldsymbol{y}$, is a **projection** if and only if $\boldsymbol{P}$ is idempotent and symmetric (Rao, 1973), in which case $\boldsymbol{P}$ is referred to as a projection matrix.

Idempotent Matrices

An **idempotent** matrix is a square matrix that remains unchanged when multiplied by itself. That is, the matrix $\boldsymbol{A}$ is idempotent if $\boldsymbol{A}\boldsymbol{A} = \boldsymbol{A}$. It can be verified that the rank of an idempotent matrix is equal to the sum of the elements on the diagonal (Searle, 1982; Searle and Hausman, 1970). This sum of elements on the diagonal of a square matrix is called the **trace** of

the matrix and is denoted by $\mathrm{tr}(\boldsymbol{A})$. Symmetry is not required for a matrix to be idempotent. However, all idempotent matrices with which we are concerned are symmetric.

The subspace of a projection is defined, or spanned, by the columns or rows of the projection matrix $\boldsymbol{P}$. If $\boldsymbol{P}$ is a projection matrix, $(\boldsymbol{I}-\boldsymbol{P})$ is also a projection matrix. However, since $\boldsymbol{P}$ and $(\boldsymbol{I}-\boldsymbol{P})$ are orthogonal matrices, the projection by $(\boldsymbol{I}-\boldsymbol{P})$ is onto the subspace *orthogonal* to that defined by $\boldsymbol{P}$. The rank of a projection matrix is the dimension of the subspace onto which it projects and, since the projection matrix is idempotent, the rank is equal to its trace.

Example 2.10

The matrix

$$\boldsymbol{A} = \frac{1}{6}\begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix}$$

is idempotent since

$$\begin{aligned} \boldsymbol{A}\boldsymbol{A} = \boldsymbol{A}^2 &= \frac{1}{6}\begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix} \frac{1}{6}\begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix} \\ &= \frac{1}{6}\begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix} = \boldsymbol{A}. \end{aligned}$$

The rank of $\boldsymbol{A}$ is given by

$$r(\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{A}) = \frac{1}{6}(5+2+5) = 2.$$

■

Since $\boldsymbol{A}$ is symmetric, it is also a projection matrix. Thus, the linear transformation

$$\widehat{\boldsymbol{y}} = \boldsymbol{A}\boldsymbol{y}_1 = \frac{1}{6}\begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix} \begin{pmatrix} 3 \\ 10 \\ 20 \end{pmatrix} = \begin{pmatrix} 2.5 \\ 11.0 \\ 19.5 \end{pmatrix}$$

is a projection of $\boldsymbol{y}_1 = (3 \quad 10 \quad 20)'$ onto the subspace defined by the columns of $\boldsymbol{A}$. The vector $\widehat{\boldsymbol{y}}$ is the unique vector in this subspace that is closest to $\boldsymbol{y}_1$. That is, $(\boldsymbol{y}_1 - \widehat{\boldsymbol{y}})'(\boldsymbol{y}_1 - \widehat{\boldsymbol{y}})$ is a minimum. Since $\boldsymbol{A}$ is a projection matrix, so is

$$\boldsymbol{I} - \boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{1}{6}\begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix} = \frac{1}{6}\begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}.$$

Then,

$$\boldsymbol{e} = (\boldsymbol{I} - \boldsymbol{A})\boldsymbol{y}_1 = \frac{1}{6}\begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}\begin{pmatrix} 3 \\ 10 \\ 20 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ -1 \\ \frac{1}{2} \end{pmatrix}$$

is a projection onto the subspace orthogonal to the subspace defined by $\boldsymbol{A}$. Note that $\widehat{\boldsymbol{y}}'\boldsymbol{e} = 0$ and $\widehat{\boldsymbol{y}} + \boldsymbol{e} = \boldsymbol{y}_1$. ■

2.7 Eigenvalues and Eigenvectors

Eigenvalues and eigenvectors of matrices are needed for some of the methods to be discussed, including principal component analysis, principal component regression, and assessment of the impact of collinearity (see Chapter 13). Determining the eigenvalues and eigenvectors of a matrix is a difficult computational problem and computers are used for all but the very simplest cases. However, the reader needs to develop an understanding of the eigenanalysis of a matrix.

The discussion of eigenanalysis is limited to real, symmetric, nonnegative definite matrices and, then, only key results are given. The reader is referred to other texts [such as Searle and Hausman (1970)] for more general discussions. In particular, Searle and Hausman (1970) show several important applications of eigenanalysis of asymmetric matrices. **Real matrices** do not contain any complex numbers as elements. **Symmetric, nonnegative definite matrices** are obtained from products of the type $\boldsymbol{B}'\boldsymbol{B}$ and, if used as defining matrices in quadratic forms (see Chapter 4), yield only zero or positive scalars.

Definitions

It can be shown that for a **real**, **symmetric** matrix $\boldsymbol{A}$ $(n \times n)$ there exists a set of n scalars λ_i, and n nonzero vectors $\boldsymbol{z}_i$, $i = 1, \ldots, n$, such that

$$\begin{aligned} \boldsymbol{A}\boldsymbol{z}_i &= \lambda_i \boldsymbol{z}_i, \\ \text{or} \quad \boldsymbol{A}\boldsymbol{z}_i - \lambda_i \boldsymbol{z}_i &= \boldsymbol{0}, \\ \text{or} \quad (\boldsymbol{A} - \lambda_i \boldsymbol{I})\boldsymbol{z}_i &= \boldsymbol{0}, \quad i = 1, \ldots, n. \end{aligned} \tag{2.13}$$

The λ_i are the n **eigenvalues** (characteristic roots or latent roots) of the matrix $\boldsymbol{A}$ and the $\boldsymbol{z}_i$ are the corresponding (column) **eigenvectors** (characteristic vectors or latent vectors).

Solution

There are nonzero solutions to equation 2.13 only if the matrix $(\boldsymbol{A} - \lambda_i \boldsymbol{I})$ is less than full rank—that is, only if the determinant of $(\boldsymbol{A} - \lambda_i \boldsymbol{I})$ is zero. The λ_i are obtained by solving the general determinantal equation

$$|\boldsymbol{A} - \lambda \boldsymbol{I}| = 0. \tag{2.14}$$

Since $\boldsymbol{A}$ is of order $n \times n$, the determinant of $(\boldsymbol{A} - \lambda \boldsymbol{I})$ is an nth degree polynomial in λ. Solving this equation gives the n values of λ, which are not necessarily distinct. Each value of λ is then used in turn in Equation 2.13 to find the companion eigenvector $\boldsymbol{z}_i$.

When the eigenvalues are distinct, the vector solution to Equation 2.13 is unique except for an arbitrary scale factor and sign. By convention, each eigenvector is defined to be the solution vector scaled to have unit length; that is, $\boldsymbol{z}_i'\boldsymbol{z}_i = 1$. Furthermore, the eigenvectors are mutually orthogonal; $\boldsymbol{z}_i'\boldsymbol{z}_j = 0$ when $i \neq j$. When the eigenvalues are not distinct, there is an additional degree of arbitrariness in defining the subsets of vectors corresponding to each subset of nondistinct eigenvalues. Nevertheless, the eigenvectors for each subset can be chosen so that they are mutually orthogonal as well as orthogonal to the eigenvectors of all other eigenvalues. Thus, if $\boldsymbol{Z} = (\, \boldsymbol{z}_1 \quad \boldsymbol{z}_2 \quad \cdots \quad \boldsymbol{z}_n \,)$ is the matrix of eigenvectors, then $\boldsymbol{Z}'\boldsymbol{Z} = \boldsymbol{I}$. This implies that $\boldsymbol{Z}'$ is the inverse of $\boldsymbol{Z}$ so that $\boldsymbol{Z}\boldsymbol{Z}' = \boldsymbol{I}$ as well.

Decomposition of a Matrix

Using $\boldsymbol{Z}$ and $\boldsymbol{L}$, defined as the diagonal matrix of the λ_i, we can write the initial equations $\boldsymbol{A}\boldsymbol{z}_i = \lambda_i \boldsymbol{z}_i$ as

$$\begin{aligned} \boldsymbol{A}\boldsymbol{Z} &= \boldsymbol{Z}\boldsymbol{L}, && (2.15) \\ \text{or} \quad \boldsymbol{Z}'\boldsymbol{A}\boldsymbol{Z} &= \boldsymbol{L}, && (2.16) \\ \text{or} \quad \boldsymbol{A} &= \boldsymbol{Z}\boldsymbol{L}\boldsymbol{Z}'. && (2.17) \end{aligned}$$

Equation 2.17 shows that a real symmetric matrix $\boldsymbol{A}$ can be transformed to a diagonal matrix by pre- and postmultiplying by $\boldsymbol{Z}'$ and $\boldsymbol{Z}$, respectively. Since $\boldsymbol{L}$ is a diagonal matrix, equation 2.17 shows that $\boldsymbol{A}$ can be expressed as the sum of matrices:

$$\boldsymbol{A} = \boldsymbol{Z}\boldsymbol{L}\boldsymbol{Z}' = \sum \lambda_i (\boldsymbol{z}_i \boldsymbol{z}_i'), \qquad (2.18)$$

where the summation is over the n eigenvalues and eigenvectors. Each term is an $n \times n$ matrix of rank 1 so that the sum can be viewed as a decomposition of the matrix $\boldsymbol{A}$ into n matrices that are mutually orthogonal. Some of these may be zero matrices if the corresponding λ_i are zero. The rank of $\boldsymbol{A}$ is revealed by the number of nonzero eigenvalues λ_i.

Example 2.11

For illustration, consider the matrix

$$\boldsymbol{A} = \begin{bmatrix} 10 & 3 \\ 3 & 8 \end{bmatrix}.$$

The eigenvalues of $\boldsymbol{A}$ are found by solving the determinantal equation (equation 2.14),

$$|(\boldsymbol{A} - \lambda \boldsymbol{I})| = \left| \begin{bmatrix} 10 - \lambda & 3 \\ 3 & 8 - \lambda \end{bmatrix} \right| = 0$$

or

$$(10-\lambda)(8-\lambda)-9 \quad = \quad \lambda^2-18\lambda+71=0.$$

The solutions to this quadratic (in λ) equation are

$$\lambda_1 = 12.16228 \quad \text{and} \quad \lambda_2 = 5.83772$$

arbitrarily ordered from largest to smallest. Thus, the matrix of eigenvalues of $\boldsymbol{A}$ is

$$\boldsymbol{L} = \begin{bmatrix} 12.16228 & 0 \\ 0 & 5.83772 \end{bmatrix}.$$

The eigenvector corresponding to $\lambda_1 = 12.16228$ is obtained by solving equation 2.13 for the elements of $\boldsymbol{z}_1$:

$$(\boldsymbol{A} - 12.16228\boldsymbol{I}) \begin{pmatrix} z_{11} \\ z_{21} \end{pmatrix} = 0$$

or

$$\begin{bmatrix} -2.162276 & 3 \\ 3 & -4.162276 \end{bmatrix} \begin{pmatrix} z_{11} \\ z_{21} \end{pmatrix} = 0.$$

Arbitrarily setting $z_{11} = 1$ and solving for z_{21}, using the first equation, gives $z_{21} = .720759$. Thus, the vector $\boldsymbol{z}_1' = (\,1 \quad .720759\,)$ satisfies the first equation (and it can be verified that it also satisfies the second equation). Rescaling this vector so it has unit length by dividing by

$$\text{length}(\boldsymbol{z}_1) = \sqrt{\boldsymbol{z}_1'\boldsymbol{z}_1} = \sqrt{1.5194935} = 1.232677$$

gives the first eigenvector

$$\boldsymbol{z}_1 = (\,.81124 \quad .58471\,)'.$$

The elements of $\boldsymbol{z}_2$ are found in the same manner to be

$$\boldsymbol{z}_2 = (\,-.58471 \quad .81124\,)'.$$

Thus, the matrix of eigenvectors for $\boldsymbol{A}$ is

$$\boldsymbol{Z} = \begin{bmatrix} .81124 & -.58471 \\ .58471 & .81124 \end{bmatrix}.$$

Notice that the first *column* of $\boldsymbol{Z}$ is the first eigenvector, and the second column is the second eigenvector. ■

Example 2.12

Continuing with Example 2.11, notice that the matrix $\boldsymbol{A}$ is of rank two because both eigenvalues are nonzero. The decomposition of $\boldsymbol{A}$ into two

orthogonal matrices each of rank one, $\boldsymbol{A} = \boldsymbol{A}_1 + \boldsymbol{A}_2$, equation 2.18, is given by

$$\begin{aligned}\boldsymbol{A}_1 = \lambda_1 \boldsymbol{z}_1 \boldsymbol{z}_1' &= 12.16228 \begin{pmatrix} .81124 \\ .58471 \end{pmatrix} (.81124 \quad .58471) \\ &= \begin{bmatrix} 8.0042 & 5.7691 \\ 5.7691 & 4.1581 \end{bmatrix}\end{aligned}$$

and

$$\boldsymbol{A}_2 = \lambda_2 \boldsymbol{z}_2 \boldsymbol{z}_2' = \begin{bmatrix} 1.9958 & -2.7691 \\ -2.7691 & 3.8419 \end{bmatrix}.$$

Since the two columns of $\boldsymbol{A}_1$ are multiples of the same vector $\boldsymbol{u}_1$, they are linearly dependent and, therefore, $r(\boldsymbol{A}_1) = 1$. Similarly, $r(\boldsymbol{A}_2) = 1$. Multiplication of $\boldsymbol{A}_1$ with $\boldsymbol{A}_2$ shows that the two matrices are orthogonal to each other: $\boldsymbol{A}_1 \boldsymbol{A}_2 = \boldsymbol{0}$, where $\boldsymbol{0}$ is a 2×2 matrix of zeros. Thus, the eigenanalysis has decomposed the rank-2 matrix $\boldsymbol{A}$ into two rank-1 matrices. It is left as an exercise to verify the multiplication and that $\boldsymbol{A}_1 + \boldsymbol{A}_2 = \boldsymbol{A}$. ■

Notice that the sum of the eigenvalues in Example 2.11, $\lambda_1 + \lambda_2 = 18$, is equal to $\text{tr}(\boldsymbol{A})$. This is a general result: the sum of the eigenvalues for any square symmetric matrix is equal to the trace of the matrix. Furthermore, the trace of each of the component rank-1 matrices is equal to its eigenvalue:

$$\text{tr}(\boldsymbol{A}_1) = \lambda_1 \quad \text{and} \quad \text{tr}(\boldsymbol{A}_2) = \lambda_2.$$

Note that for $\boldsymbol{A} = \boldsymbol{B}'\boldsymbol{B}$, we have

$$\boldsymbol{z}_i' \boldsymbol{A} \boldsymbol{z}_i = \lambda_i \boldsymbol{z}_i' \boldsymbol{z}_i$$

and

$$\begin{aligned}\lambda_i &= \frac{\boldsymbol{z}_i' \boldsymbol{A} \boldsymbol{z}_i}{\boldsymbol{z}_i' \boldsymbol{z}_i} = \frac{\boldsymbol{z}_i' \boldsymbol{B}' \boldsymbol{B} \boldsymbol{z}_i}{\boldsymbol{z}_i' \boldsymbol{z}_i} \\ &= \frac{\boldsymbol{c}_i' \boldsymbol{c}_i}{\boldsymbol{z}_i' \boldsymbol{z}_i},\end{aligned}$$

where $\boldsymbol{c}_i = \boldsymbol{B}\boldsymbol{z}_i$. Therefore, if $\boldsymbol{A} = \boldsymbol{B}'\boldsymbol{B}$ for some real matrix $\boldsymbol{B}$, then the eigenvalues of $\boldsymbol{A}$ are nonnegative. Symmetric matrices with nonnegative eigenvalues are called **nonnegative definite matrices**.

2.8 Singular Value Decomposition

The eigenanalysis, Section 2.7, applies to a square symmetric matrix. In this section, the eigenanalysis is used to develop a similar decomposition,

called the **singular value decomposition**, for a rectangular matrix. The singular value decomposition is then used to give the **principal component analysis**.

Singular Value Decomposition

Let $\boldsymbol{X}$ be an $n \times p$ matrix with $n > p$. Then $\boldsymbol{X}'\boldsymbol{X}$ is a square symmetric matrix of order $p \times p$. From Section 2.7, $\boldsymbol{X}'\boldsymbol{X}$ can be expressed in terms of its eigenvalues $\boldsymbol{L}$ and eigenvectors $\boldsymbol{Z}$ as

$$\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{Z}\boldsymbol{L}\boldsymbol{Z}'. \tag{2.19}$$

Here $\boldsymbol{L}$ is a diagonal matrix consisting of eigenvalues $\lambda_1, \ldots, \lambda_p$ of $\boldsymbol{X}'\boldsymbol{X}$. From Section 2.7, we know that $\lambda_1, \ldots, \lambda_p$ are nonnegative. Similarly, $\boldsymbol{X}\boldsymbol{X}'$ is a square symmetric matrix but of order $n \times n$. The rank of $\boldsymbol{X}\boldsymbol{X}'$ will be at most p so there will be at most p nonzero eigenvalues; they are in fact the same p eigenvalues obtained from $\boldsymbol{X}'\boldsymbol{X}$. In addition, $\boldsymbol{X}\boldsymbol{X}'$ will have at least $n - p$ eigenvalues that are zero. These $n - p$ eigenvalues and their vectors are dropped in the following. Denote with $\boldsymbol{U}$ the matrix of eigenvectors of $\boldsymbol{X}\boldsymbol{X}'$ that correspond to the p eigenvalues common to $\boldsymbol{X}'\boldsymbol{X}$. Each eigenvector $\boldsymbol{u}_i$ will be of order $n \times 1$. Then,

$$\boldsymbol{X}\boldsymbol{X}' = \boldsymbol{U}\boldsymbol{L}\boldsymbol{U}'. \tag{2.20}$$

Equations 2.19 and 2.20 jointly imply that the rectangular matrix $\boldsymbol{X}$ can be written as

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{L}^{1/2}\boldsymbol{Z}', \tag{2.21}$$

where $\boldsymbol{L}^{1/2}$ is the diagonal matrix of the positive square roots of the p eigenvalues of $\boldsymbol{X}'\boldsymbol{X}$. Thus, $\boldsymbol{L}^{1/2}\boldsymbol{L}^{1/2} = \boldsymbol{L}$. Equation 2.21 is the **singular value decomposition** of the rectangular matrix $\boldsymbol{X}$. The elements of $\boldsymbol{L}^{1/2}$, $\lambda_i^{1/2}$ are called the **singular values** and the *column* vectors in $\boldsymbol{U}$ and $\boldsymbol{Z}$ are the left and right singular vectors, respectively.

Since $\boldsymbol{L}^{1/2}$ is a diagonal matrix, the singular value decomposition expresses $\boldsymbol{X}$ as a sum of p rank-1 matrices,

$$\boldsymbol{X} = \sum \lambda_i^{1/2} \boldsymbol{u}_i \boldsymbol{z}_i', \tag{2.22}$$

where summation is over $i = 1, \ldots, p$. Furthermore, if the eigenvalues have been ranked from largest to smallest, the first of these matrices is the "best" rank-1 approximation to $\boldsymbol{X}$, the *sum* of the first two matrices is the "best" rank-2 approximation of $\boldsymbol{X}$, and so forth. These are "best" approximations in the least squares sense; that is, no other matrix (of the same rank) will give a better agreement with the original matrix $\boldsymbol{X}$ as measured by the sum of squared differences between the corresponding elements of $\boldsymbol{X}$ and the approximating matrix (Householder and Young, 1938). The goodness of fit of the approximation in each case is given by the ratio of the sum of the eigenvalues (squares of the singular values)

used in the approximation to the sum of *all* eigenvalues. Thus, the rank-1 approximation has a goodness of fit of $\lambda_1 / \sum \lambda_i$, the rank-2 approximation has a goodness of fit of $(\lambda_1 + \lambda_2) / \sum \lambda_i$, and so forth.

Recall that there is an arbitrariness of sign for the eigenvectors obtained from the eigenalysis of $\boldsymbol{X}'\boldsymbol{X}$ and $\boldsymbol{X}\boldsymbol{X}'$. Thus, care must be exercised in choice of sign for the eigenvectors in reconstructing $\boldsymbol{X}$ or lower-order approximations of $\boldsymbol{X}$ when the left and right eigenvectors have been obtained from eigenanalyses. This is not a problem when $\boldsymbol{U}$ and $\boldsymbol{Z}$ have been obtained directly from the singular value decomposition of $\boldsymbol{X}$.

Example 2.13

Singular value decomposition is illustrated using data on average minimum daily temperature X_1, average maximum daily temperature X_2, total rainfall X_3, and total growing degree days X_4, for six locations. The data were reported by Saeed and Francis (1984) to relate environmental conditions to cultivar by environment interactions in sorghum and are used with their kind permission. Each variable has been centered to have zero mean, and standardized to have unit sum of squares. (The centering and standardization are not necessary for a singular value decomposition. The centering removes the mean effect of each variable so that the dispersion about the mean is being analyzed. The standardization puts all variables on an equal basis and is desirable in most cases, particularly when the variables have different units of measure.) The $\boldsymbol{X}$ matrix is

$$\begin{aligned}\boldsymbol{X} &= (\boldsymbol{X}_1 \quad \boldsymbol{X}_2 \quad \boldsymbol{X}_3 \quad \boldsymbol{X}_4) \\ &= \begin{bmatrix} .178146 & -.523245 & .059117 & -.060996 \\ .449895 & -.209298 & .777976 & .301186 \\ -.147952 & .300866 & -.210455 & -.053411 \\ -.057369 & .065406 & .120598 & -.057203 \\ -.782003 & -.327028 & -.210455 & -.732264 \\ .359312 & .693299 & -.536780 & .602687 \end{bmatrix}.\end{aligned}$$

The singular value decomposition of $\boldsymbol{X}$ into $\boldsymbol{U}\boldsymbol{L}^{1/2}\boldsymbol{Z}'$ gives

$$\boldsymbol{U} = \begin{bmatrix} -.113995 & .308905 & -.810678 & .260088 \\ .251977 & .707512 & .339701 & -.319261 \\ .007580 & -.303203 & .277432 & .568364 \\ -.028067 & .027767 & .326626 & .357124 \\ -.735417 & -.234888 & .065551 & -.481125 \\ .617923 & -.506093 & -.198632 & -.385189 \end{bmatrix}$$

$$\boldsymbol{L}^{1/2} = \begin{bmatrix} 1.496896 & 0 & 0 & 0 \\ 0 & 1.244892 & 0 & 0 \\ 0 & 0 & .454086 & 0 \\ 0 & 0 & 0 & .057893 \end{bmatrix}$$

$$\boldsymbol{Z} = \begin{bmatrix} .595025 & .336131 & .383204 & .621382 \\ .451776 & .540753 & .657957 & .265663 \\ .004942 & .768694 & .639051 & .026450 \\ .664695 & .060922 & .108909 & .736619 \end{bmatrix}.$$

The columns of $\boldsymbol{U}$ and $\boldsymbol{Z}$ are the left and right singular vectors, respectively. The first column of $\boldsymbol{U}$, $\boldsymbol{u}_1$, the first column of $\boldsymbol{Z}$, $\boldsymbol{z}_1$, and the first singular value, $\lambda_1 = 1.496896$, give the best rank-1 approximation of $\boldsymbol{X}$,

$$\begin{aligned} \boldsymbol{A}_1 &= \lambda_1^{1/2} \boldsymbol{u}_1 \boldsymbol{z}_1' \\ &= (1.4969) \begin{pmatrix} -.1140 \\ .2520 \\ .0076 \\ -.0281 \\ -.7354 \\ .6179 \end{pmatrix} (.5950 \quad .4518 \quad .0049 \quad .6647) \\ &= \begin{bmatrix} -.101535 & -.077091 & -.000843 & -.113423 \\ .224434 & .170403 & .001864 & .250712 \\ .006752 & .005126 & .000056 & .007542 \\ -.024999 & -.018981 & -.000208 & -.027927 \\ -.655029 & -.497335 & -.005440 & -.731725 \\ .550378 & .417877 & .004571 & .614820 \end{bmatrix}. \end{aligned}$$

The goodness of fit of $\boldsymbol{A}_1$ to $\boldsymbol{X}$ is measured by

$$\frac{\lambda_1}{\sum \lambda_i} = \frac{(1.4969)^2}{4} = .56$$

or the sum of squares of the differences between the elements of $\boldsymbol{X}$ and $\boldsymbol{A}_1$, the lack of fit, is 44% of the total sum of squares of the elements in $\boldsymbol{X}$. This is not a very good approximation.

The rank-2 approximation to $\boldsymbol{X}$ is obtained by *adding* to $\boldsymbol{A}_1$ the matrix $\boldsymbol{A}_2 = \lambda_2^{1/2} \boldsymbol{u}_2 \boldsymbol{z}_2'$. This gives

$$\boldsymbol{A}_1 + \boldsymbol{A}_2 = \begin{bmatrix} .027725 & -.285040 & .295197 & -.089995 \\ .520490 & -.305880 & .678911 & .304370 \\ -.120122 & .209236 & -.290091 & -.015453 \\ -.013380 & -.037673 & .026363 & -.025821 \\ -.753317 & -.339213 & -.230214 & -.749539 \\ .338605 & .758568 & -.479730 & .576438 \end{bmatrix},$$

which has goodness of fit

$$\frac{\lambda_1 + \lambda_2}{\sum \lambda_i} = \frac{(1.4969)^2 + (1.2449)^2}{4} = .95.$$

In terms of approximating $\boldsymbol{X}$ with the rank-2 matrix $\boldsymbol{A}_1+\boldsymbol{A}_2$, the goodness of fit of .95 means that the sum of squares of the discrepancies between $\boldsymbol{X}$ and $(\boldsymbol{A}_1+\boldsymbol{A}_2)$ is 5% of the total sum of squares of all elements in $\boldsymbol{X}$. The sum of squares of all elements in $\boldsymbol{X}$ is $\sum\lambda_i$, the sum of squares of all elements in $(\boldsymbol{A}_1+\boldsymbol{A}_2)$ is $(\lambda_1+\lambda_2)$, and the sum of squares of all elements in $[\boldsymbol{X}-(\boldsymbol{A}_1+\boldsymbol{A}_2)]$ is $(\lambda_3+\lambda_4)$. In terms of the geometry of the data vectors, the goodness of fit of .95 means that 95% of the dispersion of the "cloud" of points in the original four-dimensional space is, in reality, contained in two dimensions, or the points in four-dimensional space very nearly fall on a plane. Only 5% of the dispersion is lost if the third and fourth dimensions are ignored.

Using all four singular values and their singular vectors gives the complete decomposition of $\boldsymbol{X}$ into four orthogonal rank-1 matrices. The sum of the four matrices equals $\boldsymbol{X}$, within the limits of rounding error. The analysis has shown, by the relatively small size of the third and fourth singular values, that the last two dimensions contain little of the dispersion and can safely be ignored in interpretation of the data. ■

Principal Component Analysis

The singular value decomposition is the first step in **principal component analysis**. Using the result $\boldsymbol{X}=\boldsymbol{U}\boldsymbol{L}^{1/2}\boldsymbol{Z}'$ and the property that $\boldsymbol{Z}'\boldsymbol{Z}=\boldsymbol{I}$, one can define the $n\times p$ matrix $\boldsymbol{W}$ as

$$\boldsymbol{W} \quad = \quad \boldsymbol{X}\boldsymbol{Z}=\boldsymbol{U}\boldsymbol{L}^{1/2}. \tag{2.23}$$

The first column of $\boldsymbol{Z}$ is the first of the right singular vectors of $\boldsymbol{X}$, or the first eigenvector of $\boldsymbol{X}'\boldsymbol{X}$. Thus, the coefficients in the first eigenvector define the particular linear function of the columns of $\boldsymbol{X}$ (of the original variables) that generates the first column of $\boldsymbol{W}$. The second column of $\boldsymbol{W}$ is obtained using the second eigenvector of $\boldsymbol{X}'\boldsymbol{X}$, and so on. Notice that $\boldsymbol{W}'\boldsymbol{W}=\boldsymbol{L}$. Thus, $\boldsymbol{W}$ is an $n\times p$ matrix that, unlike $\boldsymbol{X}$, has the property that all its columns are orthogonal. ($\boldsymbol{L}$ is a diagonal matrix so that all off-diagonal elements, the sums of products between columns of $\boldsymbol{W}$, are zero.) The sum of squares of the ith column of $\boldsymbol{W}$ is λ_i, the ith diagonal element of $\boldsymbol{L}$. Thus, if $\boldsymbol{X}$ is an $n\times p$ matrix of observations on p variables, each column of $\boldsymbol{W}$ is a new variable defined as a linear transformation of the original variables. The ith new variable has sum of squares λ_i and all are pairwise orthogonal. This analysis is called the **principal component analysis** of $\boldsymbol{X}$, and the columns of $\boldsymbol{W}$ are the **principal components** (sometimes called principal component scores).

Principal component analysis is used where the columns of $\boldsymbol{X}$ correspond to the observations on different variables. The transformation is to a set of orthogonal variables such that the first principal component accounts for the largest possible amount of the total dispersion, measured by λ_1, the second principal component accounts for the largest possible amount of the remaining dispersion λ_2, and so forth. The total dispersion is given by the

sum of all eigenvalues, which is equal to the sum of squares of the original variables; $\text{tr}(\boldsymbol{X}'\boldsymbol{X}) = \text{tr}(\boldsymbol{W}'\boldsymbol{W}) = \sum \lambda_i$.

Example 2.14

For the Saeed and Francis data, Example 2.13, each column of $\boldsymbol{Z}$ contains the coefficients that define one of the principal components as a linear function of the original variables. The first vector in $\boldsymbol{Z}$,

$$\boldsymbol{z}_1 = (\ .5950 \quad .4518 \quad .0049 \quad .6647\ )',$$

has similar first, second, and fourth coefficients with the third coefficient being near zero. Thus, the first principal component is essentially an average of the three temperature variables X_1, X_2, and X_4. The second column vector in $\boldsymbol{Z}$,

$$\boldsymbol{z}_2 = (\ .3361 \quad -.5408 \quad .7687 \quad .0609\ )',$$

gives heavy positive weight to X_3, heavy negative weight to X_2, and moderate positive weight to X_1. Thus, the second principal component will be large for those observations that have high rainfall X_3, and small difference between the maximum and minimum daily temperatures X_2 and X_1.

The third and fourth principal components account for only 5% of the total dispersion. This small amount of dispersion may be due more to random "noise" than to real patterns in the data. Consequently, the interpretation of these components may not be very meaningful. The third principal component will be large when there is high rainfall and large difference between the maximum and minimum daily temperatures,

$$\boldsymbol{z}_3 = (\ -.3832 \quad .6580 \quad .6391 \quad -.1089\ )'.$$

The variable degree days X_4 has little involvement in the second and third principal components; the fourth coefficient is relatively small. The fourth principal component is determined primarily by the difference between an average minimum daily temperature and degree days,

$$\boldsymbol{z}_4 = (\ .6214 \quad .2657 \quad -.0265 \quad -.7366\ )'. \qquad \blacksquare$$

The principal component vectors are obtained either by the multiplication $\boldsymbol{W} = \boldsymbol{U}\boldsymbol{L}^{1/2}$ or $\boldsymbol{W} = \boldsymbol{X}\boldsymbol{Z}$. The first is easier since it is the simple scalar multiplication of each column of $\boldsymbol{U}$ with the appropriate $\lambda_i^{1/2}$.

Example 2.15

The principal component vectors for the Saeed and Francis data of Ex-

ample 2.13 are (with some rounding)

$$\boldsymbol{W} = \begin{bmatrix} -.1706 & .3846 & -.3681 & .0151 \\ .3772 & .8808 & .1543 & -.0185 \\ .0113 & -.3775 & .1260 & .0329 \\ -.0420 & .0346 & .1483 & .0207 \\ -1.1008 & -.2924 & .0298 & -.0279 \\ .9250 & -.6300 & -.0902 & -.0223 \end{bmatrix}.$$

The sum of squares of the first principal component, the first column of $\boldsymbol{W}$, is $\lambda_1 = (1.4969)^2 = 2.2407$. Similarly, the sums of squares for the second, third, and fourth principal components are

$$\begin{aligned} \lambda_2 &= (1.2449)^2 = 1.5498 \\ \lambda_3 &= (.4541)^2 = .2062 \\ \lambda_4 &= (.0579)^2 = .0034. \end{aligned}$$

These sum to 4.0, the total sum of squares of the original three variables after they were standardized. The proportion of the total sum of squares accounted for by the first principal component is $\lambda_1 / \sum \lambda_i = 2.2407/4 = .56$ or 56%. The first two principal components account for $(\lambda_1 + \lambda_2)/4 = 3.79/4 = .95$ or 95% of the total sum of squares of the four original variables.

Each of the original data vectors in $\boldsymbol{X}$ was a vector in six-dimensional space and, together, the four vectors defined a four-dimensional subspace. These vectors were not orthogonal. The four vectors in $\boldsymbol{W}$, the principal component vectors, are linear functions of the original vectors and, as such, they fall in the *same* four-dimensional subspace. The principal component vectors, however, are orthogonal and defined such that the first principal component vector has the largest possible sum of squares. This means that the direction of the first principal component axis coincides with the major axis of the elipsoid of observations, Figure 2.3. Note that the "cloud" of observations, the data points, does *not* change; only the axes are being redefined. The second principal component has the largest possible sum of squares of all vectors orthogonal to the first, and so on. The fact that the first two principal components account for 95% of the sum of squares in this example shows that very little of the dispersion among the data points occurs in the third and fourth principal component dimensions. In other words, the variability among these six locations in average minimum and average maximum temperature, total rainfall, and total growing degree days, can be adequately described by considering *only* the two dimensions (or variables) defined by the first two principal components.

The plot of the first two principal components from the Saeed and Francis data, Figure 2.3, shows that locations 5 and 6 differ from each other primarily in the first principal component. This component was noted earlier to be mainly a temperature difference; location 6 is the warmer and has

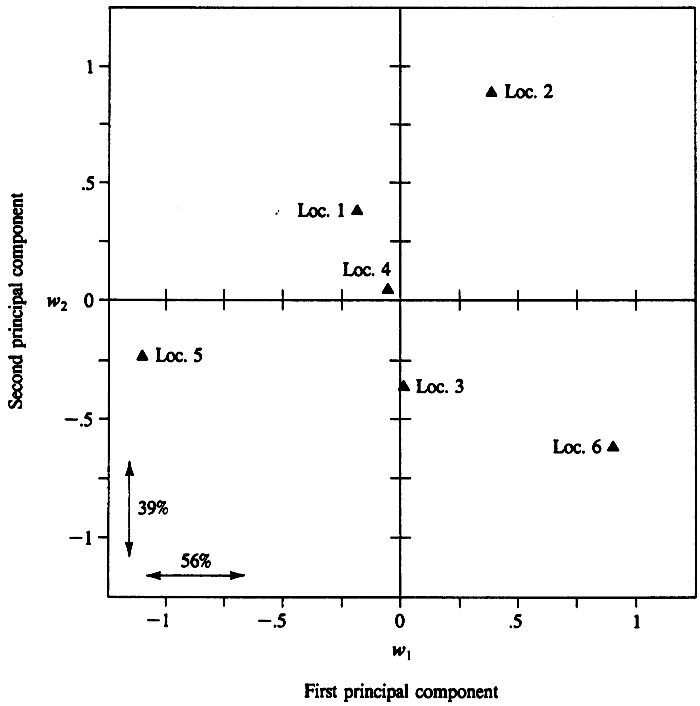

FIGURE 2.3. *The first two principal components of the Saeed and Francis (1984) data on average minimum temperature, average maximum temperature, total rainfall, and growing degree days for six locations. The first principal component primarily reflects average temperature. The second principal component is a measure of rainfall minus the spread between minimum and maximum temperature.*

the longer growing season. The other four locations differ primarily in the second principal component which reflects amount of rainfall and the difference in maximum and minimum temperature. Location 2 has the highest rainfall and tends to have a large difference in maximum and minimum daily temperature. Location 6 is also the lowest in the second principal component indicating a lower rainfall and small difference between the maximum and minimum temperature. Thus, location 6 appears to be a relatively hot, dry environment with somewhat limited diurnal temperature variation. ■

2.9 Summary

This chapter has presented the key matrix operations that are used in this text. The student must be able to use matrix notation and matrix operations. Of particular importance are

- the concepts of rank and the transpose of a matrix;
- the special types of matrices: square, symmetric, diagonal, identity, and idempotent;
- the elementary matrix operations of addition and multiplication; and
- the use of the inverse of a square symmetric matrix to solve a set of equations.

The geometry of vectors and projections is useful in understanding least squares principles. Eigenanalysis and singular value decomposition are used later in the text.

2.10 Exercises

2.1. Let

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

$$\boldsymbol{c}' = \begin{pmatrix} 1 & 2 & 0 \end{pmatrix}, \quad \text{and} \quad d = 2, \text{ a scalar.}$$

Perform the following operations, if possible. If the operation is not possible, explain why.

(a) $\boldsymbol{c}'\boldsymbol{A}$

(b) $\boldsymbol{A}'\boldsymbol{c}$

(c) $\boldsymbol{B}' + \boldsymbol{A}$

(d) $\boldsymbol{c}'\boldsymbol{B}$

(e) $\boldsymbol{A} - d$

(f) $(d\boldsymbol{B}' + \boldsymbol{A})$.

2.2. Find the rank of each of the following matrices. Which matrices are of full rank?

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \boldsymbol{B} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{C} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 \end{bmatrix}.$$

2.3. Use $\boldsymbol{B}$ in Exercise 2.2 to compute $\boldsymbol{D} = \boldsymbol{B}(\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'$. Determine whether $\boldsymbol{D}$ is idempotent. What is the rank of $\boldsymbol{D}$?

2.4. Find a_{ij} elements to make the following matrix symmetric. Can you choose a_{33} to make the matrix idempotent?

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & a_{13} & 4 \\ 2 & -1 & 0 & a_{24} \\ 6 & 0 & a_{33} & -2 \\ a_{41} & 8 & -2 & 3 \end{bmatrix}.$$

2.5. Verify that $\boldsymbol{A}$ and $\boldsymbol{B}$ are inverses of each other.

$$\boldsymbol{A} = \begin{bmatrix} 10 & 5 \\ 3 & 2 \end{bmatrix} \quad \boldsymbol{B} = \begin{bmatrix} \frac{2}{5} & -1 \\ -\frac{3}{5} & 2 \end{bmatrix}.$$

2.6. Find b_{41} such that $\boldsymbol{a}$ and $\boldsymbol{b}$ are orthogonal.

$$\boldsymbol{a} = \begin{pmatrix} 2 \\ 0 \\ -1 \\ 3 \end{pmatrix} \quad \boldsymbol{b} = \begin{pmatrix} 6 \\ -1 \\ 3 \\ b_{41} \end{pmatrix}.$$

2.7. Plot the following vectors on a two-dimensional coordinate system.

$$\boldsymbol{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \boldsymbol{v}_2 = \begin{pmatrix} 4 \\ 1 \end{pmatrix} \quad \boldsymbol{v}_3 = \begin{pmatrix} 1 \\ -4 \end{pmatrix}.$$

By inspection of the plot, which pairs of vectors appear to be orthogonal? Verify numerically that they are orthogonal *and* that all other

pairs in this set are *not* orthogonal. Explain from the geometry of the plot how you know there is a linear dependency among the three vectors.

2.8. The three vectors in Exercise 2.7 are linearly dependent. Find the linear function of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ that equals $\boldsymbol{v}_3$. Set the problem up as a system of linear equations to be solved. Let $\boldsymbol{V} = (\,\boldsymbol{v}_1 \quad \boldsymbol{v}_2\,)$, and let $\boldsymbol{x}' = (\,x_1 \quad x_2\,)$ be the vector of unknown coefficients. Then, $\boldsymbol{V}\boldsymbol{x} = \boldsymbol{v}_3$ is the system of equations to be solved for $\boldsymbol{x}$.

(a) Show that the system of equations is consistent.

(b) Show that there is a unique solution.

(c) Find the solution.

2.9. Expand the set of vectors in Exercise 2.7 to include a fourth vector, $\boldsymbol{v}_4' = (\,8 \quad 5\,)$. Reformulate Exercise 2.8 to include the fourth vector by including $\boldsymbol{v}_4$ in $\boldsymbol{V}$and an additional coefficient in $\boldsymbol{x}$. Is this system of equations consistent? Is the solution unique? Find a solution. If solutions are not unique, find another solution.

2.10. Use the determinant to determine which of the following matrices has a unique inverse.

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 4 & 10 \end{bmatrix} \quad \boldsymbol{B} = \begin{bmatrix} 4 & -1 \\ 0 & 6 \end{bmatrix} \quad \boldsymbol{C} = \begin{bmatrix} 6 & 3 \\ 4 & 2 \end{bmatrix}.$$

2.11. Given the following matrix,

$$\boldsymbol{A} = \begin{bmatrix} 3 & \sqrt{2} \\ \sqrt{2} & 2 \end{bmatrix},$$

(a) find the eigenvalues and eigenvectors of $\boldsymbol{A}$.

(b) What do your findings tell you about the rank of $\boldsymbol{A}$?

2.12. Given the following eigenvalues with their corresponding eigenvectors, and knowing that the original matrix was square and symmetric, reconstruct the original matrix.

$$\lambda_1 = 6 \quad \boldsymbol{z}_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\lambda_2 = 2 \quad \boldsymbol{z}_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

2.13. Find the inverse of the following matrix,

$$\boldsymbol{A} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 10 & 2 \\ 0 & 2 & 3 \end{bmatrix}.$$

2.14. Let

$$\boldsymbol{X} = \begin{bmatrix} 1 & .2 & 0 \\ 1 & .4 & 0 \\ 1 & .6 & 0 \\ 1 & .8 & 0 \\ 1 & .2 & .1 \\ 1 & .4 & .1 \\ 1 & .6 & .1 \\ 1 & .8 & .1 \end{bmatrix} \qquad \boldsymbol{Y} = \begin{pmatrix} 242 \\ 240 \\ 236 \\ 230 \\ 239 \\ 238 \\ 231 \\ 226 \end{pmatrix}.$$

(a) Compute $\boldsymbol{X}'\boldsymbol{X}$ and $\boldsymbol{X}'\boldsymbol{Y}$. Verify by separate calculations that the $(i, j) = (2, 2)$ element in $\boldsymbol{X}'\boldsymbol{X}$ is the sum of squares of column 2 in $\boldsymbol{X}$. Verify that the $(2, 3)$ element is the sum of products between columns 2 and 3 of $\boldsymbol{X}$. Identify the elements in $\boldsymbol{X}'\boldsymbol{Y}$ in terms of sums of squares or products of the columns of $\boldsymbol{X}$ and $\boldsymbol{Y}$.

(b) Is $\boldsymbol{X}$ of full column rank? What is the rank of $\boldsymbol{X}'\boldsymbol{X}$?

(c) Obtain $(\boldsymbol{X}'\boldsymbol{X})^{-1}$. What is the rank of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$? Verify by matrix multiplication that $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{I}$.

(d) Compute $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and verify by matrix multiplication that $\boldsymbol{P}$ is idempotent. Compute the trace $\mathrm{tr}(\boldsymbol{P})$. What is $r(\boldsymbol{P})$?

2.15. Use $\boldsymbol{X}$ as defined in Exercise 2.14.

(a) Find the singular value decomposition of $\boldsymbol{X}$. Explain what the singular values tell you about the rank of $\boldsymbol{X}$.

(b) Compute the rank-1 approximation of $\boldsymbol{X}$; call it $\boldsymbol{A}_1$. Use the singular values to state the "goodness of fit" of this rank-1 approximation.

(c) Use $\boldsymbol{A}_1$ to compute a rank-1 approximation of $\boldsymbol{X}'\boldsymbol{X}$; that is, compute $\boldsymbol{A}_1'\boldsymbol{A}_1$. Compare $\mathrm{tr}(\boldsymbol{A}_1'\boldsymbol{A}_1)$ with λ_1 and $\mathrm{tr}(\boldsymbol{X}'\boldsymbol{X})$.

2.16. Use $\boldsymbol{X}'\boldsymbol{X}$ as computed in Exercise 2.14.

(a) Compute the eigenanalysis of $\boldsymbol{X}'\boldsymbol{X}$. What is the relationship between the singular values of $\boldsymbol{X}$ obtained in Exercise 2.15 and the eigenvalues obtained for $\boldsymbol{X}'\boldsymbol{X}$?

(b) Use the results of the eigenanalysis to compute the rank-1 approximation of $\boldsymbol{X}'\boldsymbol{X}$. Compare this result to the approximation of $\boldsymbol{X}'\boldsymbol{X}$ obtained in Exercise 2.15.

(c) Show algebraically that they should be identical.

2.17. Verify that

$$\boldsymbol{A} = \frac{1}{15}\begin{bmatrix} 3 & -13 & 8 \\ 12 & -7 & 2 \\ -12 & 17 & -7 \end{bmatrix}$$

is the inverse of

$$\boldsymbol{B} = \begin{bmatrix} 1 & 3 & 2 \\ 4 & 5 & 6 \\ 8 & 7 & 9 \end{bmatrix}.$$

2.18. Show that the equations $\boldsymbol{Ax} = \boldsymbol{y}$ are consistent where

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 3 & 3 \\ 5 & 7 \end{bmatrix} \quad \text{and} \quad \boldsymbol{y} = \begin{pmatrix} 6 \\ 9 \\ 21 \end{pmatrix}.$$

2.19. Verify that

$$\boldsymbol{A}^{-} = \frac{1}{18}\begin{bmatrix} -10 & 16 & -4 \\ 8 & -11 & 5 \end{bmatrix}$$

is a generalized inverse of

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 3 & 3 \\ 5 & 7 \end{bmatrix}.$$

2.20. Verify that

$$\boldsymbol{A}^{-} = \begin{bmatrix} -\frac{1}{10} & -\frac{2}{10} & \frac{4}{9} \\ 0 & 0 & \frac{1}{9} \\ \frac{1}{10} & \frac{2}{10} & -\frac{2}{9} \end{bmatrix}$$

is a generalized inverse of

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 3 & 3 \end{bmatrix}.$$

2.21. Use the generalized inverse in Exercise 2.20 to obtain a solution to the equations $\boldsymbol{Ax} = \boldsymbol{y}$, where $\boldsymbol{A}$ is defined in Exercise 2.20 and $\boldsymbol{y} = (\,6 \quad 12 \quad 9\,)'$. Verify that the solution you obtained satisfies $\boldsymbol{Ax} = \boldsymbol{y}$.

2.22. The eigenanalysis of

$$\boldsymbol{A} = \begin{bmatrix} 10 & 3 \\ 3 & 8 \end{bmatrix}$$

in Section 2.7 gave

$$\boldsymbol{A}_1 = \begin{bmatrix} 8.0042 & 5.7691 \\ 5.7691 & 4.1581 \end{bmatrix} \quad \text{and} \quad \boldsymbol{A}_2 = \begin{bmatrix} 1.9958 & -2.7691 \\ -2.7691 & 3.8419 \end{bmatrix}.$$

Verify the multiplication of the eigenvectors to obtain $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$. Verify that $\boldsymbol{A}_1 + \boldsymbol{A}_2 = \boldsymbol{A}$, and that $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are orthogonal to each other.

2.23. In Section 2.6, a linear transformation of $\boldsymbol{y}_1 = (\,3 \quad 10 \quad 20\,)'$ to $\boldsymbol{x}_1 = (\,33 \quad 17 \quad -3\,)'$ and of $\boldsymbol{y}_2 = (\,6 \quad 14 \quad 21\,)'$ to $\boldsymbol{x}_2 = (\,41 \quad 15 \quad 1\,)'$ was made using the matrix

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & 2 & -1 \end{bmatrix}.$$

The vectors of $\boldsymbol{A}$ were then standardized so that $\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{I}$ to produce the *orthogonal* transformation of $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ to

$$\boldsymbol{x}_1^* = (\,33/\sqrt{3} \quad 17/\sqrt{2} \quad -3/\sqrt{6}\,)'$$

and

$$\boldsymbol{x}_2^* = (\,41/\sqrt{3} \quad 15/\sqrt{2} \quad 1/\sqrt{6}\,)',$$

respectively. Show that the squared distance between $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ is unchanged when the orthogonal transformation is made but not when the nonorthogonal transformation is made. That is, show that

$$(\boldsymbol{y}_1 - \boldsymbol{y}_2)'(\boldsymbol{y}_1 - \boldsymbol{y}_2) = (\boldsymbol{x}_1^* - \boldsymbol{x}_2^*)'(\boldsymbol{x}_1^* - \boldsymbol{x}_2^*)$$

but that

$$(\boldsymbol{y}_1 - \boldsymbol{y}_2)'(\boldsymbol{y}_1 - \boldsymbol{y}_2) \neq (\boldsymbol{x}_1 - \boldsymbol{x}_2)'(\boldsymbol{x}_1 - \boldsymbol{x}_2).$$

2.24. (a) Let $\boldsymbol{A}$ be an $m \times n$ matrix and $\boldsymbol{B}$ be an $n \times m$ matrix. Then show that $\text{tr}(\boldsymbol{AB}) = \text{tr}(\boldsymbol{BA})$.
(b) Use (a) to show that $\text{tr}(\boldsymbol{ABC}) = \text{tr}(\boldsymbol{BCA})$, where $\boldsymbol{C}$ is an $m \times m$ matrix.

2.25. Let $\boldsymbol{a}^*$ be an $m \times 1$ vector with $\boldsymbol{a}^{*\prime}\boldsymbol{a}^* > 0$. Define $\boldsymbol{a} = \boldsymbol{a}^*/(\boldsymbol{a}^{*\prime}\boldsymbol{a}^*)^{1/2}$ and $\boldsymbol{A} = \boldsymbol{a}\boldsymbol{a}'$. Show that $\boldsymbol{A}$ is a symmetric idempotent matrix of rank 1.

2.26. Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be two $m \times 1$ vectors that are orthogonal to each other. Define $\boldsymbol{A} = \boldsymbol{a}\boldsymbol{a}'$ and $\boldsymbol{B} = \boldsymbol{b}\boldsymbol{b}'$. Show that $\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{0}$, a matrix of zeros.

2.27. **Gram–Schmidt orthogonalization**. An orthogonal basis for a space spanned by some vectors can be obtained using the Gram–Schmidt orthogonalization procedure.

(a) Consider two linearly independent vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. Define $\boldsymbol{z}_1 = \boldsymbol{v}_1$ and $\boldsymbol{z}_2 = \boldsymbol{v}_2 - \boldsymbol{v}_1 c_{2.1}$, where $c_{2.1} = (\boldsymbol{v}_1'\boldsymbol{v}_2)/(\boldsymbol{v}_1'\boldsymbol{v}_1)$. Show that $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ are orthogonal. Also, show that $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ span the same space as $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$.

(b) Consider three linearly independent vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, and $\boldsymbol{v}_3$. Define $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ as in (a) and $\boldsymbol{z}_3 = \boldsymbol{v}_3 - c_{3.1}\boldsymbol{z}_1 - c_{3.2}\boldsymbol{z}_2$, where $c_{3.i} = (\boldsymbol{z}_i'\boldsymbol{v}_3)/(\boldsymbol{z}_i'\boldsymbol{z}_i)$, $i = 1,\ 2$. Show that $\boldsymbol{z}_1$, $\boldsymbol{z}_2$, and $\boldsymbol{z}_3$ are mutually orthogonal and span the same space as $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, and $\boldsymbol{v}_3$.

3

MULTIPLE REGRESSION IN MATRIX NOTATION

We have reviewed linear regression in algebraic notation and have introduced the matrix notation and operations needed to continue with the more complicated models.

This chapter presents the model, and develops the normal equations and solution to the normal equations for a general linear model involving any number of independent variables. The matrix formulation for the variances of linear functions is used to derive the measures of precision of the estimates.

Chapter 1 provided an introduction to multiple regression and suggested that a more convenient notation was needed. Chapter 2 familiarized you with matrix notation and operations with matrices. This chapter states multiple regression results in matrix notation. Developments in the chapter are for full rank models. Less than full rank models that use generalized inverses are discussed in Chapter 9.

3.1 The Model

The linear additive model for relating a dependent variable to p independent variables is

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \epsilon_i. \tag{3.1}$$