Woodward, J.A., Bonett, D.G., & Brecht, M-L. (1990).
Introduction to Linear Models and Experimental Design.
Harcourt Brace Jovanovich: Orlando, FL.

APPENDIX

A INTRODUCTION TO MATRIX NOTATION

Matrix notation simplifies the presentation of the linear statistical model and the associated estimators and test statistics. Using the matrix notation introduced in this section, we represent all of the models, estimators, and test statistics for the analysis of variance with a few general matrix equations. Although it requires effort to learn the matrix notation, the benefit is a great reduction in notational complexity.

A.1 Types of Matrices

We refer to single numbers by symbols such as a, b, μ, σ^2, and the like. Now we adopt a notation for referring to a rectangular array of such numbers, also called a *matrix*. We use lowercase letters in boldface, such as **x, y, a,** and **b**, to denote more than one number that has been arranged into a rectangular pattern with r rows and c columns. The number of rows need not equal the number of columns, but all rows are of the same length, and all columns are of the same length. A matrix **x** can be diagrammed in the following manner:

$$\underset{(3,\,4)}{\mathbf{x}} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \end{bmatrix}.$$

where the notation (3, 4) above the name of the matrix is called the *order* of the matrix and indicates the number of rows and the number of columns. The contents of the matrix are called elements; these have subscripts to indicate the number of the row (first subscript) and the number of the column (second subscript) for an element. We represent a matrix with r rows and c columns as

X_{52} = element in x matrix found in row 5, col 2

.. but what if double digits?

$$\text{order} \leftarrow \underset{\substack{\downarrow \\ \text{denotes matrix}}}{\underset{(r,\,c)}{\mathbf{x}}} = \underbrace{\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1c} \\ x_{21} & x_{22} & \cdots & x_{2c} \\ \vdots & & & \\ x_{r1} & x_{r2} & \cdots & x_{rc} \end{bmatrix}}_{\text{elements}} \Bigg\}\ \text{matrix}$$

When the numerical values of the elements are known, we specify the numbers in the representation of a matrix. For example, a 2×3 matrix **b** with known elements is diagrammed as

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

We distinguish six types of matrices, as follows.

Rectangular Matrix. In general, a rectangular matrix has r rows and c columns and $r \neq c$.

Square Matrix. A matrix that contains the same number of rows and columns is a square matrix (i.e., $r = c$). The matrix **a** is square because it has three rows and three columns:

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

principal diagonal.

The elements of a square matrix that have identical row and column subscripts form the *principal diagonal* of the matrix (i.e., a_{11}, a_{22}, a_{33}). A rectangular matrix ($r \neq c$) does not have a principal diagonal.

Vector. A matrix that has a single column (with more than one row) or a single row (with more than one column) is called a *vector*. The matrices **c** and **d** are vectors:

like in R you create vectors and bind them to form a df or tibble

$$\mathbf{c} = [1 \quad 3 \quad 2] \qquad \mathbf{d} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$$

In previous sections we referred to collections of random variables such as $Y_1, Y_2, \ldots, Y_n$. These random variables are represented in a matrix called a *random vector*

$$\underset{(n,\,1)}{\mathbf{Y}} = \left.\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}\right\}$$

functions (AKA random vars)

where the capital letters indicate that the elements are functions (i.e., random variables) rather than constants or mathematical variables.

Diagonal Matrix. A square matrix with all elements equal to zero except the principal diagonal is called a *diagonal matrix*. The matrix **d** is diagonal:

$$\mathbf{d} = \begin{bmatrix} d_{11} & 0 & 0 & 0 \\ 0 & d_{22} & 0 & 0 \\ 0 & 0 & d_{33} & 0 \\ 0 & 0 & 0 & d_{44} \end{bmatrix}$$

A diagonal matrix **c** with known elements can be diagrammed as

$$\mathbf{c} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Identity Matrix. The diagonal matrix with 1's along the principal diagonal is called the *identity matrix* and is represented by the capital letter **I**, even though it contains known constants. The identity matrix of order 4 is diagrammed as

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

The symbol for an identity matrix of order n is $\mathbf{I}_n$.

Null Matrix. A matrix consisting entirely of zeros is called the null matrix and is represented as **0**. The null matrix of order 2×3 is diagrammed as

$$\mathbf{0} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

A.2 Operations

Arithmetic operations on the elements of matrices are easy to represent by matrix notation because we represent the operations on the single symbols for the matrices, rather than by listing the operations on all of the elements separately. We discuss nine operations on matrices: equality, transpose, trace, addition, subtraction, multiplication, row echelon form, determinant, and inverse.

1 Equality

Two matrices of the same order are equal if all elements in corresponding positions are equal. For matrices **a** and **b**, the statement $\mathbf{a} = \mathbf{b}$ is true if $a_{ij} = b_{ij}$ for all combinations of i and j.

EXERCISE A.2.1

Which matrices are equal?

$$\mathbf{a} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 3 & 3 \end{bmatrix} \quad \mathbf{c} = \begin{bmatrix} 1 & 1 \\ 2 & 4 \\ 3 & 5 \end{bmatrix} \quad \mathbf{d} = \begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 3 & 3 \end{bmatrix}$$

(2,3) (3,2) (3,2) (3,2)

∴ b, c, & d ?

or just b & d ?

1 Transpose

The transpose of $\mathbf{a}$ is denoted as $\mathbf{a}'$ and represents the operation of exchanging the rows and columns of $\mathbf{a}$ to produce $\mathbf{a}'$. Thus the ith row of $\mathbf{a}$ becomes the ith column of $\mathbf{a}'$:

$$\mathbf{a} = \begin{bmatrix} 1 & 3 \\ 2 & 3 \\ 1 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{a}' = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 3 & 4 \end{bmatrix}$$

In general the ijth element of $\mathbf{a}$ is the jith element of $\mathbf{a}'$.

A square matrix is *symmetric* if it equals its transpose. The matrix $\mathbf{b}$ is symmetric:

$$\mathbf{b} = \begin{bmatrix} 1 & 3 & 2 \\ 3 & 1 & 4 \\ 2 & 4 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{b}' = \begin{bmatrix} 1 & 3 & 2 \\ 3 & 1 & 4 \\ 2 & 4 & 1 \end{bmatrix}$$

EXERCISE A.2.2

1. Form the transpose of each matrix:

$$\mathbf{a} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \mathbf{c} = \begin{bmatrix} 1 & 4 & 3 \\ 4 & 1 & 2 \\ 3 & 2 & 1 \end{bmatrix} \quad \mathbf{d} = [4 \quad 4 \quad 4 \quad 1]$$

$a' = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$

$b' = [1 \quad 2 \quad 3]$

$c' = \begin{bmatrix} 1 & 4 & 3 \\ 4 & 1 & 2 \\ 3 & 2 & 1 \end{bmatrix}$

$d' = \begin{bmatrix} 4 \\ 4 \\ 4 \\ 1 \end{bmatrix}$

2. Which of the above matrices are symmetric? c + c' are symmetric

3 Trace

The trace of a square $n \times n$ matrix $\mathbf{c}$ is the sum of the diagonal elements and is denoted as TR($\mathbf{c}$). Thus, for the 2×2 matrix $\mathbf{c}$,

$$\mathbf{c} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \end{bmatrix}$$

$$\text{TR}(\mathbf{c}) = 1 + 5 = 6$$

4 Addition and Subtraction

The operation of addition is defined on matrices of the same order and represents an element-by-element addition of elements. Thus

$$\mathbf{b} = \mathbf{a} + \mathbf{c}$$

is a convenient shorthand for

$$\begin{bmatrix} (2+1) & (3+6) \\ (1+9) & (4+7) \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 6 \\ 9 & 7 \end{bmatrix}$$

The matrix **b** is the sum of **a** and **c** and equals

$$\mathbf{b} = \begin{bmatrix} 3 & 9 \\ 10 & 11 \end{bmatrix}$$

which is the element-by-element sum of corresponding elements. Subtraction is defined analogously. If the matrix **b** is the sum of **a** and **c**, then $b_{ij} = a_{ij} + c_{ij}$. Likewise, if the matrix **b** equals **a** − **c**, then $b_{ij} = a_{ij} - c_{ij}$. The addition of matrices is commutative and associative. Because addition is *commutative*, **a** + **b** = **b** + **a**. That is,

$$\begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} + \begin{bmatrix} 5 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 6 \\ 5 \\ 5 \end{bmatrix}$$

Because addition is *associative*, for matrices of equal order we have

$$(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$$

That is, if $\mathbf{a} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, and $\mathbf{c} = \begin{bmatrix} 3 & 3 \\ 1 & 2 \end{bmatrix}$, then $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$:

$$\begin{bmatrix} (1+1) & (2+1) \\ (3+1) & (4+1) \end{bmatrix} + \begin{bmatrix} 3 & 3 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} (1+3) & (1+3) \\ (1+1) & (1+2) \end{bmatrix}$$

$$\begin{bmatrix} 5 & 6 \\ 5 & 7 \end{bmatrix} = \begin{bmatrix} 5 & 6 \\ 5 & 7 \end{bmatrix}$$

EXERCISE A.2.3

Given the matrices

$$\mathbf{a} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}, \mathbf{c} = \begin{bmatrix} 4 & 1 & 3 \\ 2 & 1 & 4 \end{bmatrix}, \text{and } \mathbf{d} = \begin{bmatrix} 3 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

perform the following operations:

1. **a** + **b**
2. **a** − **b**
3. **b** − **a**
4. **c** + **d**

1 $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$

2. $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 \\ -1 \\ 0 \end{bmatrix}$

3. $\begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$

4. $\begin{bmatrix} 4 & 1 & 3 \\ 2 & 1 & 4 \end{bmatrix} + \begin{bmatrix} 3 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$

$= \begin{bmatrix} 7 & 3 & 4 \\ 3 & 2 & 6 \end{bmatrix}$ 5

Multiplication

We discuss three types of multiplication for matrices: multiplication of a matrix by scalar, matrix product, and Kronecker product.

* Scalar = a number

5a **Multiplication of a Matrix by a Scalar.** A matrix can be multiplied by a single number, called a scalar (here denoted by c), using the following rule:

$$\mathbf{cb} = c\begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} cb_{11} & cb_{12} & cb_{13} \\ cb_{21} & cb_{22} & cb_{23} \\ cb_{31} & cb_{32} & cb_{33} \end{bmatrix}$$

Each element of the matrix is multiplied by this constant.

1. $\begin{bmatrix} 12 & 6 & 18 \\ 24 & 6 & 36 \end{bmatrix}$

2. Same as 1

3. $\begin{bmatrix} 6 \\ 6 \\ 6 \end{bmatrix}$

EXERCISE A.2.4

Let

$$a = 6, \quad \mathbf{b} = \begin{bmatrix} 2 & 1 & 3 \\ 4 & 1 & 6 \end{bmatrix}, \quad \text{and} \quad \mathbf{1}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

compute the following:

1. $\mathbf{a}\mathbf{b}$
2. $\mathbf{b}a$
3. $\mathbf{1}_3 a$

5b **Matrix Product.** Two matrices can be multiplied by forming the sum of cross products of all rows of the left matrix and all columns of the right matrix, so $\mathbf{ab} = \mathbf{c}$ is defined as

$$\begin{bmatrix} 1 & 3 & 4 \\ 2 & 1 & 3 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 4 & 2 \end{bmatrix} = \begin{bmatrix} (1)(1)+(3)(3)+(4)(4) & (1)(2)+(3)(1)+(4)(2) \\ (2)(1)+(1)(3)+(3)(4) & (2)(2)+(1)(1)+(3)(2) \end{bmatrix}$$

→ multiply then sum

$$= \begin{bmatrix} 26 & 13 \\ 17 & 11 \end{bmatrix}$$

In general, if $\mathbf{a}$ is r by s and $\mathbf{b}$ is s by t, then the ijth element of $\mathbf{c}$ is

$$c_{ij} = \sum_{k=1}^{s} a_{ik} b_{kj}$$

Matrices can be multiplied in this manner only if the number of columns of the left matrix equals the number of rows of the right matrix; such matrices are said to be *conformable* for multiplication. The dimensions of the resulting product equals the number of rows of the left matrix and the number of columns of the right matrix. Thus

$$\underset{(r,\,t)}{\mathbf{c}} = \underset{(r,\,s)}{\mathbf{a}} \; \underset{(s,\,t)}{\mathbf{b}}$$

$(2,\underbrace{3)\cdot(3}_{\text{inners cancel out}},2) = (2,2)$

Several illustrations may help explain the matrix product. Suppose that

① $\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}\begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix}$

② $\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}\begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix}$

③ $\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}\begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix}$

④ $\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}\begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix}$

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

$$\mathbf{1}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad \mathbf{1}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 6 \\ 5 \\ 4 \end{bmatrix} \quad \mathbf{c} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \quad \mathbf{d} = \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix}$$

$$\mathbf{e} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \quad \mathbf{f} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 9 \end{bmatrix}$$

Then

$$\mathbf{cd} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}\begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix} = \begin{bmatrix} (c_{11}d_{11} + c_{12}d_{21}) & (c_{11}d_{12} + c_{12}d_{22}) \\ (c_{21}d_{11} + c_{22}d_{21}) & (c_{21}d_{12} + c_{22}d_{22}) \end{bmatrix}$$

The product $\mathbf{1}_3'\mathbf{b}$ is

$$\mathbf{1}_3'\mathbf{b} = [1 \quad 1 \quad 1]\begin{bmatrix} 6 \\ 5 \\ 4 \end{bmatrix} = [(6 + 5 + 4)] = 15$$

which is the sum of the elements in **b**. The product $\mathbf{1}_2'\mathbf{c}$ is

$$\mathbf{1}_2'\mathbf{c} = [1 \quad 1]\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} = [(c_{11} + c_{21}) \quad (c_{12} + c_{22})]$$

which is the sum of the columns of **c** (note that $\mathbf{1}_3' = [1 \quad 1 \quad 1]$ and $\mathbf{1}_2' = [1 \quad 1]$, where the subscript of the 1 vector indicates its length.)

If we multiply $\mathbf{b}'\mathbf{b}$, we form the "sum of squares" of the elements in **b**, sometimes called the *inner product* of vectors:

$$\mathbf{b}'\mathbf{b} = [6 \quad 5 \quad 4]\begin{bmatrix} 6 \\ 5 \\ 4 \end{bmatrix} = [(6)(6) + (5)(5) + (4)(4)]$$

$$= [36 + 25 + 16] = 77$$

The *outer product* of these vectors is

$$\mathbf{bb}' = \begin{bmatrix} 6 \\ 5 \\ 4 \end{bmatrix}[6 \quad 5 \quad 4] = \begin{bmatrix} (6)(6) & (6)(5) & (6)(4) \\ (5)(6) & (5)(5) & (5)(4) \\ (4)(6) & (4)(5) & (4)(4) \end{bmatrix} = \begin{bmatrix} 36 & 30 & 24 \\ 30 & 25 & 20 \\ 24 & 20 & 16 \end{bmatrix}$$

Note that in multiplying $\mathbf{bb}'$, each row of **b** and each column of $\mathbf{b}'$ contains only one element, so the sum of cross products of row 1 in **b** and column 1 in $\mathbf{b}'$ is just (6)(6).

The product **ef** is

$$\mathbf{ef} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 9 \end{bmatrix} = \begin{bmatrix} 20 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 18 \end{bmatrix}$$

The matrix product (row-by-column rule) possesses certain general properties. The matrix product is *associative*; that is,

$$\mathbf{a(bc) = (ab)c}$$

provided that the matrices are conformable. The order in which the multiplications are carried out is irrelevant.

The matrix product does not possess the *commutative* property. Thus, in general,

$$\mathbf{ab} \neq \mathbf{ba}$$

For this reason, the left-hand matrix in a product is called the premultiplier and the right-hand matrix is called the postmultiplier. As an example, if $\mathbf{a} = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, then $\mathbf{ab} \neq \mathbf{ba}$ because

a * b ≠ b * a
pre-multiplier
post-multiplier

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

$$\begin{bmatrix} 7 & 10 \\ 6 & 10 \end{bmatrix} \neq \begin{bmatrix} 7 & 4 \\ 15 & 10 \end{bmatrix}$$

The three special cases in which the matrix product is commutative are $\mathbf{aI} = \mathbf{Ia}$, $\mathbf{a0} = \mathbf{0a}$, and $\mathbf{ab} = \mathbf{ba}$ if both **a** and **b** are diagonal, provided the matrices are conformable. It is important to note that $\mathbf{aI} = \mathbf{a}$; that is, **I** is the identity element of this type of matrix multiplication. This can be seen in the following illustration:

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

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

Matrix addition and matrix product possess the *distributive* property. That is,

$$\mathbf{a(b + c) = ab + ac}$$

provided that **b** and **c** are of the same order (i.e., can be added) and provided that **a**, **b**, and **c** are conformable for multiplication.

Note that $\mathbf{(abc)' = c'b'a'}$ so long as all of the matrices are conformable.

EXERCISE A.2.5

The matrices **a**, **b**, **c**, **d**, **e**, and $\mathbf{I}_3$ are defined as follows:

$$\mathbf{a} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 2 & 1 \\ 2 & 3 \\ 3 & 4 \end{bmatrix} \quad \mathbf{c} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \quad \mathbf{d} = \begin{bmatrix} 9 & 8 & 7 \\ 6 & 5 & 4 \\ 3 & 2 & 1 \end{bmatrix}$$

$$\mathbf{e} = [1 \quad 4 \quad 1] \quad \mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Compute the following:

1. $\mathbf{a'a}$ and $\mathbf{aa'}$
2. $\mathbf{a'b}$ and $\mathbf{b'a}$
3. $\mathbf{cd}$ and $\mathbf{dc}$
4. $\mathbf{dI}_3$ and $\mathbf{I}_3\mathbf{d}$
5. $\mathbf{e(c + d)}$
6. $\mathbf{ec + ed}$
7. $\mathbf{e(c - d)}$
8. $\mathbf{ec - ed}$
9. $\mathbf{cdb}$

EXERCISE A.2.6

Using the distributive property of matrix addition and matrix product, find alternate expressions for each of the following. (Assume all products are among conformable matrices):

1. $\mathbf{ax + bx} =$ x(a+b)
2. $\mathbf{xa + xb} =$ x(a+b)
3. $\mathbf{ax + a} =$ a(x+1)
4. $\mathbf{xa + a} =$ a(x+1)

5c **Kronecker Product.** The Kronecker product of matrices is defined as the product of each element of the left-hand matrix by the entire right-hand matrix. That is,

$$\underset{\mathbf{a}}{(p, q)} \otimes \underset{\mathbf{b}}{(m, n)} = \begin{bmatrix} a_{11}\mathbf{b} & \cdots & a_{1q}\mathbf{b} \\ \vdots & & \\ a_{p1}\mathbf{b} & \cdots & a_{pq}\mathbf{b} \end{bmatrix}$$

where the order of the product is pm by qn. As an example, consider

$$\mathbf{a} = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 2 & 3 \end{bmatrix} \text{ and } \mathbf{b} = [6 \quad 4]$$

1 [1 2 3

then

$$\mathbf{a} \otimes \mathbf{b} = \begin{bmatrix} 1[6 \quad 4] & 2[6 \quad 4] & 1[6 \quad 4] \\ 3[6 \quad 4] & 2[6 \quad 4] & 3[6 \quad 4] \end{bmatrix}$$

$$= \begin{bmatrix} 6 & 4 & 12 & 8 & 6 & 4 \\ 18 & 12 & 12 & 8 & 18 & 12 \end{bmatrix}$$

EXERCISE A.2.7

The matrices **a**, **b**, and **c** are defined as

$$\mathbf{a} = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{bmatrix} \qquad \mathbf{c} = \begin{bmatrix} 3 & 3 & 3 \\ 1 & 1 & 1 \end{bmatrix}$$

Compute the following:

1. $\mathbf{a} \otimes \mathbf{b}$, $\mathbf{a} \otimes \mathbf{c}$, $\mathbf{b} \otimes \mathbf{c}$
2. $\mathbf{a} \otimes (\mathbf{b} + \mathbf{c})$
3. $\mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c}$

Row Echelon Form and the Rank of a Matrix

Any matrix can be transformed into a unique form called *row echelon form.* The row echelon form of a matrix is useful for determining the number of linearly independent rows in a matrix; this number is called the row *rank* of the matrix. If a matrix contains at least one row that is a linear combination of other rows, then the rows of the matrix are said to be linearly dependent; otherwise, they are linearly independent. (This terminology should not be confused with the concept of stochastic independence discussed in Chapter 1.) We make direct use of the row echelon form of a matrix in Section 6.4, where we transform a modified hypothesis matrix for an incomplete design into row echelon form and delete all zero rows, thus retaining only linearly independent rows in the modified hypothesis matrix.

To compute the row echelon form of a matrix, we need to define three types of elementary matrices that are used to perform *elementary row operations* (ero) on a matrix:

- *Type I.* The interchange of any two distinct rows is a Type I elementary row operation. For the matrix

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

the interchange of rows 1 and 2 can be accomplished by premultiplying **a** by a Type I *elementary matrix* of the form

$$\mathbf{e}_{\mathrm{I}} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Thus $\mathbf{e}_{\mathrm{I}}\mathbf{a}$ interchanges rows 1 and 2. In general, to interchange rows r and s, $\mathbf{e}_{\mathrm{I}}$ is the identity matrix, except that it will have a 1 in the sth column of the rth row, and the rth column of the sth row. The diagonal elements in rows r and s are 0.

- *Type II.* The multiplication of a row by any nonzero number, denoted as λ, is a Type II elementary row operation. To multiply row 2 of **a** by λ, the elementary matrix of Type II can be used: for this example,

$$\mathbf{e}_{\mathrm{II}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

  In general, to multiply the rth row of **a** by λ, $\mathbf{e}_{\mathrm{II}}$ is a matrix with ones along the diagonal and λ in the rth diagonal position.

- *Type III.* The addition of a scalar multiple of some row to another row is a Type III elementary row operation. To add λ times row 3 to row 1, we can premultiply **a** by the elementary matrix

$$\mathbf{e}_{\mathrm{III}} = \begin{bmatrix} 1 & 0 & \lambda \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

  In general, to add λ times the sth row to the rth row, $\mathbf{e}_{\mathrm{III}}$ is a matrix with ones along the diagonal and λ in the rth row of the sth column.

Using the elementary row operations just defined, we can transform any matrix into a unique form called the row echelon form and denoted $\mathbf{a}_r$. A matrix $\mathbf{a}_r$ is in row echelon form if and only if

1. all nonzero rows are at the top of the matrix (if $\mathbf{a}_r$ has r nonzero rows, then the first r rows are nonzero);
2. the leading entries move to the right as we go down the rows of the matrix (a leading entry is the first nonzero element in a row); and
3. all leading entries are 1.
4. any column that contains a leading entry has all other entries equal to zero.

To transform a matrix to row echelon form, we apply an algorithm such as that in Figure A.1.

In the following example, we illustrate computation of the row echelon form as represented in the flow diagram of Figure A.1. With

$$\mathbf{a} = \begin{bmatrix} 0 & 0 & 3 & -1 \\ 0 & -1 & 4 & 7 \\ 0 & -1 & 7 & 6 \end{bmatrix}$$

set $m = 3$ (m is the number of rows in $\mathbf{a}$), and carry out the following:

Step 1

(a) and (b) $k = 1$ and $\mathbf{a}_1 = \mathbf{a}$.
(c) Does $\mathbf{a}_1 = \mathbf{0}$? No
(d) The first nonzero column of $\mathbf{a}_1$ is $p = 2$.
(e) The first row of $\mathbf{a}_1$ with nonzero entry in column 2 is row 2
(f) Put row 2 first using $\mathbf{e}_1\mathbf{a}$, where

$$\begin{bmatrix} 0 & -1 & 4 & 7 \\ 0 & 0 & 3 & -1 \\ 0 & -1 & 7 & 6 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 3 & -1 \\ 0 & -1 & 4 & 7 \\ 0 & -1 & 7 & 6 \end{bmatrix}$$

(g) Use ero II to make the first nonzero entry in row 1 equal to 1:

$$\begin{bmatrix} 0 & 1 & -4 & -7 \\ 0 & 0 & 3 & -1 \\ 0 & -1 & 7 & 6 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 & 4 & 7 \\ 0 & 0 & 3 & -1 \\ 0 & -1 & 7 & 6 \end{bmatrix}$$

(h) Reduce the remaining elements in column 2 to zero using ero III:

$$\begin{bmatrix} 0 & 1 & -4 & -7 \\ 0 & 0 & 3 & -1 \\ 0 & 0 & 3 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & -4 & -7 \\ 0 & 0 & 3 & -1 \\ 0 & -1 & 7 & 6 \end{bmatrix}$$

(i) Does $k = 1$ equal 3? No
(j) Partition $\mathbf{a}$

$$\mathbf{a} = \begin{bmatrix} [0 \quad 1 \quad -4 \quad -7] \\ \begin{bmatrix} 0 & 0 & 3 & -1 \\ 0 & 0 & 3 & -1 \end{bmatrix} \end{bmatrix}$$

where

$$\mathbf{a}_{k+1} = \mathbf{a}_2 = \begin{bmatrix} 0 & 0 & 3 & -1 \\ 0 & 0 & 3 & -1 \end{bmatrix}$$

$k = k + 1 = 2$

(k) Now $k = 2$; return to (c) and start step 2.

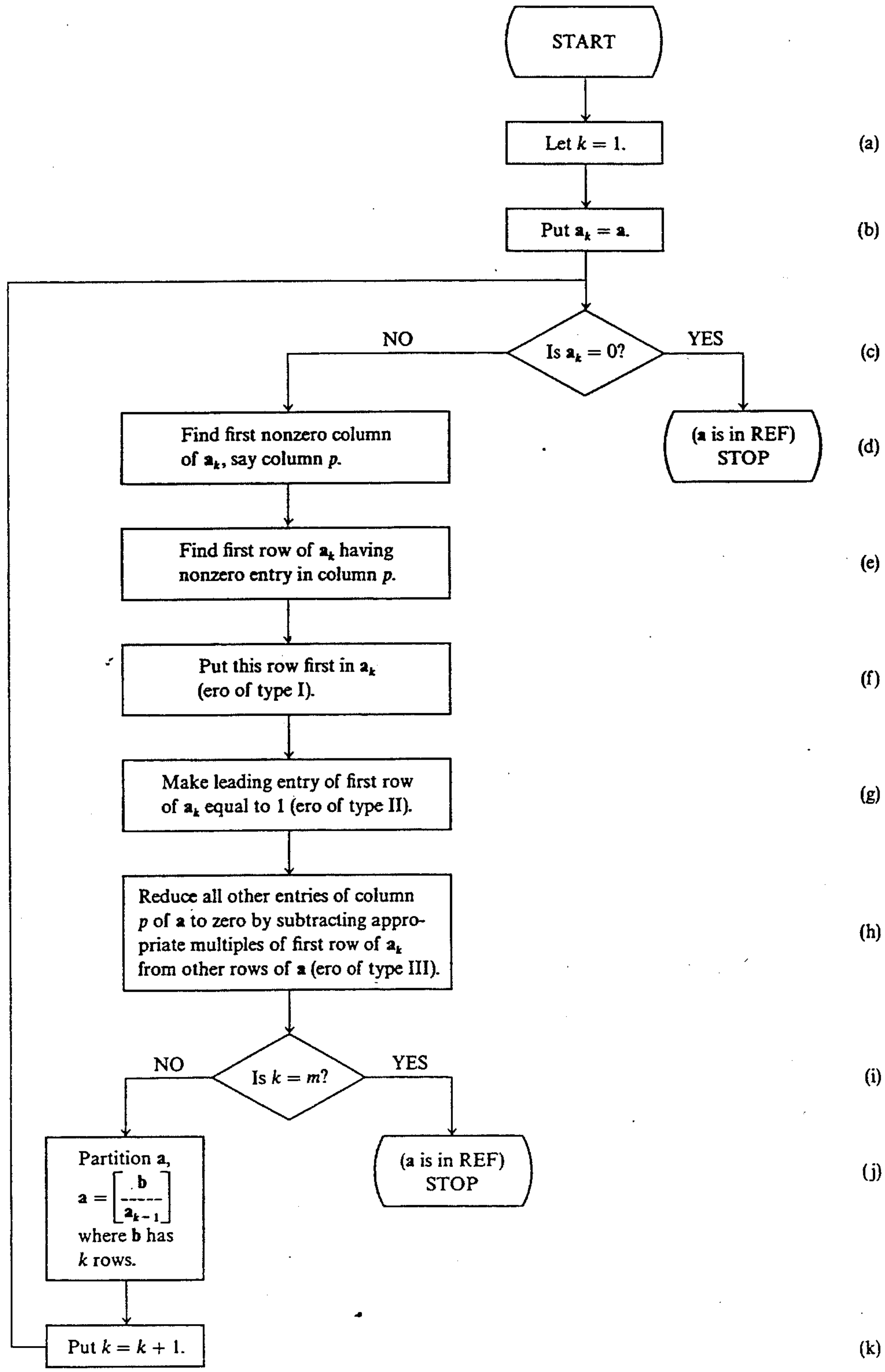


FIGURE A.1 Algorithm for transforming matrix to row echelon form. Note: ero denotes elementary row operations; REF denotes row echelon form. Adapted from Schneider and Barker (1968).

Step 2

(c) Does $\mathbf{a}_2 = 0$? No

(d) The first nonzero column of $\mathbf{a}_2$ is $p = 3$.

(e) The first row of $\mathbf{a}_2$ with nonzero entry in column 3 is row 1.

(f) This operation is not needed for this matrix at this step.

(g) Use ero II to make the first nonzero entry in row 1 (of $\mathbf{a}_2$) equal to 1:

$$\begin{bmatrix} 0 & 1 & -4 & -7 \\ 0 & 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 3 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & -4 & -7 \\ 0 & 0 & 3 & -1 \\ 0 & 0 & 3 & -1 \end{bmatrix}$$

(h) Reduce the remaining elements in column 2 to zero using ero III, first for row 1, column 3:

$$\begin{bmatrix} 0 & 1 & 0 & -\frac{25}{3} \\ 0 & 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 3 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & -4 & -7 \\ 0 & 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 3 & -1 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & 0 & -\frac{25}{3} \\ 0 & 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & -\frac{25}{3} \\ 0 & 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 3 & -1 \end{bmatrix}$$

(i) Does $k = 3$? No

(j) Partition $\mathbf{a}$

$$\mathbf{a} = \begin{bmatrix} \begin{bmatrix} 0 & 1 & 0 & -\frac{25}{3} \\ 0 & 0 & 1 & -\frac{1}{3} \end{bmatrix} \\ [0 \quad 0 \quad 0 \quad 0] \end{bmatrix}$$

where $\mathbf{a}_{k+1} = \mathbf{a}_3 = [0 \quad 0 \quad 0 \quad 0]$.

(k) Now $k = 3$; return to (c) and begin step 3.

Step 3

(c) Does $\mathbf{a}_3 = 0$? Yes

(d) Stop. $\mathbf{a}$ is now in row echelon form.

If we number the elementary matrices $\mathbf{e}_1$ through $\mathbf{e}_6$, we have

$$\mathbf{e}_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{a}_2 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{e}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

$$\mathbf{e}_4 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{e}_5 = \begin{bmatrix} 1 & 4 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{e}_6 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{bmatrix}$$

Then the row echelon form of $\mathbf{a}$ can be represented in the more compact form as

$$\mathbf{a}_r = (\mathbf{e}_6\mathbf{e}_5\mathbf{e}_4\mathbf{e}_3\mathbf{e}_2\mathbf{e}_1)\mathbf{a}$$

or

$$\mathbf{a}_r = \mathbf{e}\mathbf{a}$$

where

$$\mathbf{e} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 4 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{4}{3} & -1 & 0 \\ \frac{1}{3} & 0 & 0 \\ -1 & -1 & 1 \end{bmatrix}$$

The reader can verify that

$$\mathbf{a}_r = \begin{bmatrix} \frac{4}{3} & -1 & 0 \\ \frac{1}{3} & 0 & 0 \\ -1 & -1 & 1 \end{bmatrix} \mathbf{a}$$

The number of nonzero rows in $\mathbf{a}_r$ is equal to two. Thus the number of linearly independent rows in **a** is two. Equivalently, we can say that the *row rank* of the preceding matrix is two. The row echelon form of a matrix can be computed using the personal computer program MATRIX, which is included with this text. The number of linearly independent columns of a matrix is called the *column rank*; it can be determined by transforming the transpose of the matrix to row echelon form. The $m \times n$ matrix **c** is said to be of full row rank if the number of linearly independent rows equals m. The $m \times n$ matrix **c** is of full column rank if the number of linearly independent columns equals n.

The following properties of the rank of a matrix are important.

1. The row rank of a matrix equals the column rank. Thus we can speak of the rank of an $m \times n$ matrix **c** as a single number, denoted as $r = R(\mathbf{c})$ (Schneider and Barker, 1968, p. 152).
2. The rank of an $m \times n$ matrix **c** cannot exceed the smaller of m and n; that is, $R(\mathbf{c}) \leq \text{MIN}(m\ n)$. Because the row rank cannot exceed m, and the column rank cannot exceed n, this property follows from the first.
3. If the $n \times n$ matrix **c** is less than full rank (i.e., $R(\mathbf{c}) < n$), then the inverse $\mathbf{c}^{-1}$ does not exist. (The inverse $\mathbf{c}^{-1}$ will be defined later.) In this case, the matrix is said to be singular (Schneider and Barker, 1968, p. 88).
4. If **a** is $m \times n$ and of rank r_1, and **b** is $n \times b$ and of rank r_2, then the rank of the product **ab** cannot exceed the smaller of the ranks; that is, $R(\mathbf{ab}) \leq \text{MIN}(r_1, r_2)$ (Bhattacharya, Jain, and Naqpaul, 1983, p. 61).

5. If $\mathbf{a}$ is $n \times n$ and of full rank, and $\mathbf{b}$ is $n \times p$ and of rank r, then $R(\mathbf{ab}) = R(\mathbf{b})$ (Bhattacharya et al., 1983, p. 62).

We illustrate the first, fourth, and fifth properties. As an example of property 1, consider the 3×4 matrix

$$\mathbf{a} = \begin{bmatrix} 1 & 2 & 5 & 4 \\ 3 & 2 & 4 & 3 \\ 6 & 4 & 8 & 6 \end{bmatrix}$$

and its row echelon form

$$\mathbf{a}_r = \begin{bmatrix} 1 & 0 & -0.50 & -0.50 \\ 0 & 1 & 2.75 & 2.25 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The row rank of $\mathbf{a}$ is 2 because there are two nonzero rows in $\mathbf{a}_r$. Property 1 says that the number of linearly independent columns in $\mathbf{a}$ also equals 2 (that is, the column rank of $\mathbf{a}$ is 2). This can be verified by transforming $\mathbf{a}'$ to row echelon form, which is

$$(\mathbf{a}')_r = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and the column rank also is 2.

To illustrate property 4, consider the matrices $\mathbf{a}$ and $\mathbf{b}$ and their row echelon forms

$$\mathbf{a} = \begin{bmatrix} 1 & 2 & 1 & 2 \\ 3 & 6 & 4 & 1 \\ 2 & 4 & 2 & 4 \end{bmatrix} \qquad \mathbf{a}_r = \begin{bmatrix} 1 & 2 & 0 & 7 \\ 0 & 0 & 1 & -5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and

$$\mathbf{b} = \begin{bmatrix} 1 & 1 & 2 \\ 6 & 2 & 4 \\ 9 & 8 & 17 \\ 4 & 1 & 2 \end{bmatrix} \qquad \mathbf{b}_r = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Thus $R(\mathbf{a}) = r_1 = 2$ and $R(\mathbf{b}) = r_2 = 3$. Property 4 tells us that $R(\mathbf{ab})$ cannot exceed the smaller of the ranks, which is $r_1 = 2$. This can be seen in the following:

$$\mathbf{ab} = \begin{bmatrix} 30 & 15 & 31 \\ 79 & 48 & 100 \\ 60 & 30 & 62 \end{bmatrix} \qquad (\mathbf{ab})_r = \begin{bmatrix} 1 & 0 & -0.047 \\ 0 & 1 & 2.161 \\ 0 & 0 & 0 \end{bmatrix}$$

and the rank of $\mathbf{ab}$ does not exceed $r_1 = 2$.

Finally, property 5 is illustrated using the following matrices **a** and **b** and their row echelon forms $\mathbf{a}_r$ and $\mathbf{b}_r$:

$$\mathbf{a} = \begin{bmatrix} 2 & 1 & 3 \\ 5 & 1 & 4 \\ 2 & 3 & 4 \end{bmatrix} \qquad \mathbf{a}_r = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and

$$\mathbf{b} = \begin{bmatrix} 2 & 1 & 3 & 5 \\ 1 & 2 & 4 & 3 \\ 4 & 2 & 6 & 10 \end{bmatrix} \qquad \mathbf{b}_r = \begin{bmatrix} 1 & 0 & 0.67 & 2.33 \\ 0 & 1 & 1.67 & 0.33 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

As can be seen, **a** is of full rank because $R(\mathbf{a}) = 3$, and the rank of **b** is $R(\mathbf{b}) = 2$. Property 5 tells us that $R(\mathbf{ab}) = R(\mathbf{b}) = 2$. This can be verified as follows:

$$\mathbf{ab} = \begin{bmatrix} 17 & 10 & 28 & 43 \\ 27 & 15 & 43 & 68 \\ 23 & 16 & 42 & 59 \end{bmatrix} \qquad (\mathbf{ab})_r = \begin{bmatrix} 1 & 0 & 0.67 & 2.33 \\ 0 & 1 & 1.67 & 0.33 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Determinant

The determinant of a square matrix **a** is denoted as DET(**a**) or $|\mathbf{a}|$. For the 2×2 matrix,

$$\mathbf{a} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

the determinant is defined as

$$|\mathbf{a}| = a_{11}a_{22} - a_{12}a_{21}$$

For the 3×3 matrix,

$$\mathbf{b} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

the determinant is

$$|\mathbf{b}| = b_{11}b_{22}b_{33} + b_{12}b_{23}b_{31} + b_{13}b_{21}b_{32} - b_{13}b_{22}b_{31} - b_{11}b_{23}b_{32} - b_{12}b_{21}b_{33}$$

For larger matrices, this expanded form of the determinant would contain too many terms to list conveniently, but the form would be the same: a sum of products, in which the first is the product of all diagonal elements and in which all other products contain at least one element from below and above the diagonal. To compute the determinant of a large matrix it is convenient to use type III elementary row operations to transform the matrix to lower triangular form so that all elements above the principal diagonal are 0 (e.g., see Bock, 1975, p. 57). (Type III elementary row operations do not alter the determinant.) Then each product in the expanded form of the determinant, except the first one, contains at least one zero. Thus a determinant in lower triangular form can be computed simply as the product of the diagonal elements of the matrix.

Matrix Inverse

The inverse is an important operation on matrices. For the square matrix $\mathbf{a}$, the inverse is represented as $\mathbf{a}^{-1}$ and is the same order as $\mathbf{a}$. When the inverse of a matrix exists, it has the property that

$$\mathbf{a}\mathbf{a}^{-1} = \mathbf{a}^{-1}\mathbf{a} = \mathbf{I}$$

When the inverse of a matrix does not exist, the matrix is said to be *singular*, or less than full rank. The inverse of the square matrix is analogous to the reciprocal in scalar algebra. The constant c multiplied by its reciprocal c^{-1} equals 1, which is the identity element of scalar multiplication; thus

$$cc^{-1} = c^{-1}c = 1$$

If the constant c is 0, c^{-1} is undefined. By analogy, a singular matrix has the property that its inverse is undefined.

Because it is computationally difficult to calculate the inverse of a matrix, this is usually accomplished using a computer. The exceptions are the diagonal matrix and the square matrix of order 2.

To illustrate this calculation for a diagonal matrix, suppose first that

$$\mathbf{d} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

then

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

If $\mathbf{d}$ is $n \times n$, then the iith element of $\mathbf{d}^{-1}$ is d_{ii}^{-1}, for $i = 1$ to n.

EXERCISE A.2.8

Verify that $\mathbf{d}\mathbf{d}^{-1} = \mathbf{d}^{-1}\mathbf{d} = \mathbf{I}$ in the preceding illustration.

Now, suppose that

$$\mathbf{d} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

so that at least one of the diagonal elements is 0. Then **d** is singular because

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

is undefined (i.e., the second diagonal element $\frac{1}{0}$ is undefined).

To calculate the matrix inverse for the square matrix of order 2, suppose that **a** is the 2×2 matrix,

$$\mathbf{a} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

then

$$\mathbf{a}^{-1} = \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \times \frac{1}{(a_{11}a_{22} - a_{12}a_{21})}$$

or

$$\mathbf{a}^{-1} = \begin{bmatrix} \dfrac{a_{22}}{(a_{11}a_{22} - a_{12}a_{21})} & -\dfrac{a_{12}}{(a_{11}a_{22} - a_{12}a_{21})} \\ -\dfrac{a_{21}}{(a_{11}a_{22} - a_{12}a_{21})} & \dfrac{a_{11}}{(a_{11}a_{22} - a_{12}a_{21})} \end{bmatrix}$$

where the diagonal elements have been reversed and the signs of the off-diagonal elements are changed. As an example, suppose that

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

then

$$\mathbf{a}^{-1} = \begin{bmatrix} \dfrac{4}{(3)(4) - (2)(2)} & \dfrac{-2}{(3)(4) - (2)(2)} \\ \dfrac{-2}{(3)(4) - (2)(2)} & \dfrac{3}{(3)(4) - (2)(2)} \end{bmatrix} = \begin{bmatrix} 0.500 & -0.250 \\ -0.250 & 0.375 \end{bmatrix}$$

One way to compute the inverse of a matrix of order greater than 2 involves the row echelon form. It can be shown from the definition of row echelon form that the row echelon form of a square, full-rank matrix is the identity matrix. Thus we can write

$$\mathbf{a}_r = \mathbf{I} = \mathbf{ea}$$

where **a** is $m \times m$ and rank m. But there is only one matrix that can be multiplied times **a** to produce the identity matrix: the inverse of **a**. Thus, for the square, full-rank matrix, $\mathbf{e} = \mathbf{a}^{-1}$. The personal computer program MATRIX can be used to compute the inverse of the $m \times m$ matrix **a** by computing the row echelon form of the augmented $m \times 2m$ matrix $[\mathbf{a} : \mathbf{I}]$.

Then the product of each elementary matrix times $[\mathbf{a} : \mathbf{I}]$ yields

$$[\mathbf{e}_1\, \mathbf{a} : \mathbf{e}_1]$$

then

$$[\mathbf{e}_2\, \mathbf{e}_1\, \mathbf{a} : \mathbf{a}_2\, \mathbf{e}_1]$$

and finally

$$[\mathbf{e}\, \mathbf{a} : \mathbf{e}],$$

or

$$[\mathbf{I} : \mathbf{a}^{-1}]$$

because $\mathbf{ea} = \mathbf{I}$ and $\mathbf{e} = \mathbf{a}^{-1}$. The inverse of $\mathbf{a}$ occupies the rightmost $m \times m$ portion of the augmented matrix after transformation to row echelon form.

EXERCISE A.2.9

a. Verify that $\mathbf{aa}^{-1} = \mathbf{I}_2$ in the preceding example.

b. Compute the inverse of $\mathbf{b}$, where $\mathbf{b} = \begin{bmatrix} 6 & 1 \\ 1 & 3 \end{bmatrix}$.

c. Verify that $\mathbf{bb}^{-1} = \mathbf{I}_2$ in part (b).

The inverse is useful for solving matrix equations in a manner analogous to that used in scalar algebra. As an example, consider the scalar algebra equation

$$ax = c$$

where a and c are known constants and x is unknown. The solution for x is

$$(a^{-1})ax = a^{-1}c$$
$$(1)x = a^{-1}c$$

If $a = 5$ and $c = 10$, then

$$5x = 10$$
$$x = 10/5 = 2$$

A similar type of solution exists for an equation containing matrices, provided the matrix involved has an inverse. As an illustration, suppose we multiply $\mathbf{ax} = \mathbf{c}$ and obtain

$$\begin{bmatrix} 3 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 10 & 11 \\ 9 & 12 \end{bmatrix} = \begin{bmatrix} 48 & 57 \\ 56 & 70 \end{bmatrix}$$

As with the scalar equation, suppose we did not know the content of the matrix $\mathbf{x}$, so we could write only

$$\begin{bmatrix} 3 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} = \begin{bmatrix} 48 & 57 \\ 56 & 70 \end{bmatrix}$$

Here the matrix **x** is unknown, and we can use the matrix inverse to solve for **x** just as we solved for the scalar x in the scalar equation. We premultiply both sides of the equation by $\mathbf{a}^{-1}$, obtaining

$$\mathbf{a}^{-1}\mathbf{a}\mathbf{x} = \mathbf{a}^{-1}\mathbf{c}$$

but because $\mathbf{a}^{-1}\mathbf{a} = \mathbf{I}_2$, we have

$$\mathbf{I}_2\mathbf{x} = \mathbf{a}^{-1}\mathbf{c}$$

and because $\mathbf{I}_2\mathbf{x} = \mathbf{x}$, the solution for **x** is

$$\mathbf{x} = \mathbf{a}^{-1}\mathbf{c}$$

From the previous illustration, we know that

$$\mathbf{a}^{-1} = \begin{bmatrix} 0.500 & -0.250 \\ -0.250 & 0.375 \end{bmatrix}$$

and we obtain

$$\mathbf{x} = \begin{bmatrix} 0.500 & -0.250 \\ -0.250 & 0.375 \end{bmatrix}\begin{bmatrix} 48 & 57 \\ 56 & 70 \end{bmatrix} = \begin{bmatrix} 10 & 11 \\ 9 & 12 \end{bmatrix}$$

which is the correct result for the matrix **x**. This solution can be used provided that the matrix **a** is nonsingular (i.e., its inverse exists). A solution to the equation when **a** is singular and has no regular inverse requires the use of a matrix called the generalized inverse.

Finally, the following rule applies to conformable matrices: $(\mathbf{abc})^{-1} = \mathbf{c}^{-1}\mathbf{b}^{-1}\mathbf{a}^{-1}$

The generalized Inverse

Consider the matrix equation

$$\begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 28 \\ 14 \end{bmatrix}$$

where $\mathbf{a} = \begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, and $\mathbf{c} = \begin{bmatrix} 28 \\ 14 \end{bmatrix}$. If we try to apply the solution of the previous section,

$$\mathbf{x} = \mathbf{a}^{-1}\mathbf{c}$$

we would discover that we cannot solve for **x** because $\mathbf{a}^{-1}$ does not exist. The matrix **a** is singular, and $\mathbf{a}^{-1}$ cannot be computed.

EXERCISE A.2.10

Demonstrate that the inverse of $\mathbf{a} = \begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix}$ does not exist.

When $\mathbf{a}$ is singular, then the equation $\mathbf{ax} = \mathbf{c}$ may or may not have a solution; that is, one may or may not be able to find specific values of $\mathbf{x}$ for which the equation is true. Equations that have solutions are called *consistent*. For a consistent equation $\mathbf{ax} = \mathbf{c}$, a solution for $\mathbf{x}$ can be found using a *generalized inverse* of a singular matrix $\mathbf{a}$. A generalized inverse is symbolized as $\mathbf{a}^-$ and satisfies the four properties

$$\mathbf{aa}^-\mathbf{a} = \mathbf{a}$$

$$\mathbf{a}^-\mathbf{aa}^- = \mathbf{a}^-$$

$$\mathbf{a}^-\mathbf{a} = \mathbf{aa}^-$$

$$\mathbf{aa}^- = \mathbf{a}^-\mathbf{a}$$

Although this generalized inverse is unique, it does *not* have the property of the regular inverse, that is, $\mathbf{a}^-\mathbf{a} \neq \mathbf{I}$, where $\mathbf{a}$ is singular.

Using the generalized inverse, we find a solution to the equation $\mathbf{ax} = \mathbf{c}$ to be

$$\mathbf{x} = \mathbf{a}^-\mathbf{c}$$

but it must be recognized that this solution for $\mathbf{x}$ is not unique (as it is when $\mathbf{a}$ is nonsingular and the regular inverse can be used). That is, there is more than one solution.

The computation of the generalized inverse is complicated and must be carried out with a computer, except in the case of a diagonal matrix, for which the computation is simple and can be done by hand. Consider the singular diagonal matrix

$$\mathbf{d} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

A generalized inverse is simply

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

where the nonzero diagonal elements are inverted, and the zero diagonal elements remain zero.

A special type of inverse, called the **conditional** inverse, satisfies only the first property $\mathbf{a}^-\mathbf{aa}^- = \mathbf{a}^-$, and it is used in Chapter 8.

EXERCISE A.2.11

Demonstrate through computation that

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

is a generalized inverse of

$$\mathbf{d} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

Returning to our original equation

$$\begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 28 \\ 14 \end{bmatrix}$$

we use the personal computer program MATRIX to compute a generalized inverse of the matrix

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

and find that

$$\mathbf{a}^{-} = \begin{bmatrix} 0.08 & 0.04 \\ 0.16 & 0.08 \end{bmatrix}$$

A solution for $\mathbf{x}$ is $\mathbf{x} = \mathbf{a}^{-}\mathbf{c}$

$$\mathbf{x} = \begin{bmatrix} 0.08 & 0.04 \\ 0.16 & 0.08 \end{bmatrix} \begin{bmatrix} 28 \\ 14 \end{bmatrix} = \begin{bmatrix} 2.8 \\ 5.6 \end{bmatrix}$$

If these values for $\mathbf{x}$ are a solution to the equation, then it should be true that

$$\begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2.8 \\ 5.6 \end{bmatrix} = \begin{bmatrix} 28 \\ 14 \end{bmatrix}$$

EXERCISE A.2.12

1. Show that

$$\mathbf{a}^{-} = \begin{bmatrix} 0.08 & 0.04 \\ 0.16 & 0.08 \end{bmatrix}$$

is a ~~generalized~~ conditional inverse of

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

2. Demonstrate that

$$\mathbf{x} = \begin{bmatrix} 2.8 \\ 5.6 \end{bmatrix}$$

is a solution to the equation

$$\begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 28 \\ 14 \end{bmatrix}$$

3. Show that

$$\mathbf{x} = \begin{bmatrix} -1.2 \\ 7.6 \end{bmatrix}$$

also is a solution.

It is easy to prove that $\mathbf{a}^{-}\mathbf{c}$ is a solution to a consistent equation $\mathbf{ax} = \mathbf{c}$ when $\mathbf{a}$ is singular. Because the equation is consistent, a solution exists, and we can call this hypothetical solution $\mathbf{w}$. Then we can write

$$\mathbf{aw} = \mathbf{aw}$$

but because $\mathbf{aa}^{-}\mathbf{a} = \mathbf{a}$, we can change the left side to

$$(\mathbf{aa}^{-}\mathbf{a})\mathbf{w} = \mathbf{aw}$$

Because $\mathbf{w}$ is a solution, we know that $\mathbf{aw} = \mathbf{c}$, and we can substitute $\mathbf{c}$ for each $(\mathbf{aw})$ and obtain

$$\mathbf{aa}^{-}(\mathbf{c}) = (\mathbf{c})$$

Because $\mathbf{a}(\mathbf{a}^{-}\mathbf{c}) = \mathbf{c}$, this shows that $\mathbf{a}^{-}\mathbf{c}$ is a solution since any matrix multiplied by $\mathbf{a}$ that yields $\mathbf{c}$ is a solution.

The equation $\mathbf{ax} = \mathbf{c}$ has a solution if and only if $\mathbf{aa}^{-}\mathbf{c} = \mathbf{c}$. If this condition holds (and the equation is consistent), then the matrix $\mathbf{a}^{-}\mathbf{c}$ is just one of the possible solutions. Other solutions $\mathbf{s}$ can be determined from the expression

$$\mathbf{s} = \mathbf{a}^{-}\mathbf{c} + (\mathbf{I} - \mathbf{a}^{-}\mathbf{a})\mathbf{z}$$

where $\mathbf{z}$ is an arbitrary vector. For the equation in which $\mathbf{a}$ is $m \times n$, $\mathbf{x}$ is $n \times 1$, and $\mathbf{c}$ is $m \times 1$, $\mathbf{z}$ is an arbitrary $n \times 1$ vector.

As an example, we found earlier that one solution to the equation $\mathbf{ax} = \mathbf{c}$,

$$\begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 28 \\ 14 \end{bmatrix}$$

is

$$\mathbf{a}^{-}\mathbf{c} = \begin{bmatrix} 2.8 \\ 5.6 \end{bmatrix}$$

We can determine other solutions using

$$\mathbf{s} = \mathbf{a}^{-}\mathbf{c} + (\mathbf{I} - \mathbf{a}^{-}\mathbf{a})\mathbf{z}$$

We know that

$$\mathbf{a}^{-}\mathbf{c} = \begin{bmatrix} 2.8 \\ 5.6 \end{bmatrix}$$

and we determine that

$$(\mathbf{I} - \mathbf{a}^{-}\mathbf{a}) = \begin{bmatrix} 0.8 & -0.4 \\ -0.4 & 0.2 \end{bmatrix}$$

Then by arbitrarily selecting $\mathbf{z}$ to be

$$\mathbf{z} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

we find that another solution is

$$\begin{aligned} \mathbf{s} &= \begin{bmatrix} 2.8 \\ 5.6 \end{bmatrix} + \begin{bmatrix} 0.8 & -0.4 \\ -0.4 & 0.2 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 2.8 \\ 5.6 \end{bmatrix} + \begin{bmatrix} 2.8 \\ -1.4 \end{bmatrix} \\ &= \begin{bmatrix} 5.6 \\ 4.2 \end{bmatrix} \end{aligned}$$

The reader can verify that $\begin{bmatrix} 5.6 \\ 4.2 \end{bmatrix}$ also is a solution to the equation $\mathbf{ax} = \mathbf{c}$.

EXERCISE A.2.13

The matrices $\mathbf{b}$ and $\mathbf{c}$ are defined as

$$\mathbf{b} = \begin{bmatrix} 6 & 1 \\ 1 & 3 \end{bmatrix} \quad \text{and} \quad \mathbf{c} = \begin{bmatrix} 8 & 9 \\ 12 & 10 \end{bmatrix}$$

Find $\mathbf{x}$ in the matrix equation $\mathbf{bx} = \mathbf{c}$. (Note: $\mathbf{b}^{-1}$ was computed in Exercise A.2.9.)

EXERCISE A.2.14

Assuming all operations to be defined, solve the following matrix equations for the matrix $\mathbf{x}$:

1. $\mathbf{ax} - \mathbf{c} = \mathbf{d}$
2. $\mathbf{ax} + \mathbf{c} = \mathbf{dx} - \mathbf{c}$
3. $\mathbf{ax} + \mathbf{x} = \mathbf{d}$

A.3 Applications

An important reason for studying matrix notation is the simplicity it allows when complicated operations are performed on large sets of numbers. In this section, some of the statistical concepts represented in Chapters 1 and 2 in scalar notation are presented in matrix notation.

Expected Values, Variances, and Covariances. Let $\mathbf{Y}$ be a random vector containing n random variables,

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

The expected value of $\mathbf{Y}$ is

$$E(\mathbf{Y}) = \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_n) \end{bmatrix} = \begin{bmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{bmatrix} = \boldsymbol{\mu}$$

Note that when all $E(Y_i)$ have the same expected value, we could write

$$E(\mathbf{Y}) = \boldsymbol{\mu} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \mu = \mathbf{1}_n \mu$$

The deviation random vector $\boldsymbol{\varepsilon} = \mathbf{Y} - \boldsymbol{\mu}$ represents the vector of random variables:

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} (Y_1 - \mu) \\ (Y_2 - \mu) \\ \vdots \\ (Y_n - \mu) \end{bmatrix}$$

All the variances and covariances of these n random variables can be represented in one simple matrix equation:

$$E[\varepsilon\varepsilon'] = E\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} [\varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_n]$$

$$= E\begin{bmatrix} \varepsilon_1^2 & \varepsilon_1\varepsilon_2 & \cdots & \varepsilon_1\varepsilon_n \\ \varepsilon_2\varepsilon_1 & \varepsilon_2^2 & \cdots & \varepsilon_2\varepsilon_n \\ \vdots & & & \\ \varepsilon_n\varepsilon_1 & \varepsilon_n\varepsilon_2 & \cdots & \varepsilon_n^2 \end{bmatrix}$$

$$= E\begin{bmatrix} (Y_1-\mu)^2 & (Y_1-\mu)(Y_2-\mu) & \cdots & (Y_1-\mu)(Y_n-\mu) \\ (Y_2-\mu)(Y_1-\mu) & (Y_2-\mu)^2 & \cdots & (Y_2-\mu)(Y_n-\mu) \\ \vdots & & & \\ (Y_n-\mu)(Y_1-\mu) & (Y_n-\mu)(Y_2-\mu) & \cdots & (Y_n-\mu)^2 \end{bmatrix}$$

$$= \begin{bmatrix} E(Y_1-\mu)^2 & E(Y_1-\mu)(Y_2-\mu) & \cdots & E(Y_1-\mu)(Y_n-\mu) \\ E(Y_2-\mu)(Y_1-\mu) & E(Y_2-\mu)^2 & \cdots & E(Y_2-\mu)(Y_n-\mu) \\ \vdots & & & \\ E(Y_n-\mu)(Y_1-\mu) & E(Y_n-\mu)(Y_2-\mu) & \cdots & E(Y_n-\mu)^2 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & & & \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix}$$

Recall that

$$\begin{aligned} \sigma_1^2 &= E(Y_1-\mu)^2 \\ &= E(Y_1^2) - E(Y_1)^2 \end{aligned}$$

and

$$\begin{aligned} \sigma_{12} &= E(Y_1-\mu)(Y_2-\mu) \\ &= E(Y_1Y_2) - E(Y_1)E(Y_2) \end{aligned}$$

The matrix expression $E(\varepsilon\varepsilon')$ represents an $n \times n$ matrix of all variances and covariances among the n variables. This matrix is called the variance covariance matrix, or more briefly the covariance matrix, and is represented by the symbol Σ (capital sigma). Thus

$$\Sigma_\varepsilon = E(\varepsilon\varepsilon')$$

represents the $n \times n$ covariance matrix of the deviation random variables $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$. Note that the variances of the n random variables are on the principal diagonal of Σ_ε. Also note that $\Sigma_\varepsilon = \Sigma'_\varepsilon$, that is, the covariance matrix is symmetric.

Estimators. Recall the sampling of n cases from the population. On the sample space of each sampling event, we define a random variable Y and represent these n random variables as

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

The estimator of μ in scalar notation is

$$\hat{\mu} = \sum_{i=1}^{n} Y_i(1/n)$$

The estimator of μ in matrix notation is

$$\hat{\mu} = \mathbf{1}'_n \mathbf{Y}(1/n)$$

where $\mathbf{1}'_n = [1 \quad 1 \quad \cdots \quad 1]$ is a row of n ones and $(1/n)$ is a constant.

In matrix notation, the estimate $\bar{\mu}$ is

$$\bar{\mu} = \mathbf{1}'_n \mathbf{y}(1/n)$$

where the data are

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Suppose that y is

$$\mathbf{y} = \begin{bmatrix} 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{bmatrix}$$

then $n = 5$ and

$$\bar{\mu} = [1 \quad 1 \quad 1 \quad 1 \quad 1] \begin{bmatrix} 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{bmatrix} (1/n) = (30)(\tfrac{1}{5}) = 6$$

The estimate of μ is the 1×1 matrix containing the number 6.

The estimator for the variance is represented in scalar notation as

$$\hat{\sigma}^2 = \sum_{i=1}^{n} (Y_i - \hat{\mu})^2 (1/n)$$

In matrix notation we can define the random vector $\hat{\boldsymbol{\varepsilon}}$ as

$$\hat{\boldsymbol{\varepsilon}} = (\mathbf{Y} - \mathbf{1}'_n \hat{\mu}) = (\mathbf{Y} - \hat{\boldsymbol{\mu}})$$

because

$$\hat{\boldsymbol{\mu}} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \hat{\mu} = \begin{bmatrix} \hat{\mu} \\ \hat{\mu} \\ \vdots \\ \hat{\mu} \end{bmatrix}.$$

Then the estimator of the variance in matrix notation is

$$\hat{\sigma}^2 = \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}(1/n)$$

The estimate $\bar{\sigma}^2$ is a function of the data and is

$$\bar{\sigma}^2 = \bar{\varepsilon}'\bar{\varepsilon}(1/n) = (\mathrm{y} - \bar{\mu})'(\mathrm{y} - \bar{\mu})(1/n)$$

From the above illustration we have

$$\bar{\varepsilon} = \begin{bmatrix} 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} 6 = \begin{bmatrix} 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{bmatrix} - \begin{bmatrix} 6 \\ 6 \\ 6 \\ 6 \\ 6 \end{bmatrix} = \begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}$$

and

$$\bar{\sigma}^2 = \bar{\varepsilon}'\bar{\varepsilon}(1/n)$$

$$= [-2 \quad -1 \quad 0 \quad 1 \quad 2] \begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{bmatrix} (\tfrac{1}{5}) = 2$$

EXERCISE A.3.1

The following data resulted from sampling six subjects from the population and measuring their IQ:

$$\mathrm{y} = \begin{bmatrix} 98 \\ 120 \\ 115 \\ 100 \\ 110 \\ 120 \end{bmatrix}$$

Compute $\bar{\mu}$ and $\bar{\sigma}^2$, using both scalar and matrix notation.

We now derive the unconstrained and constrained least-squares estimates of β in the general linear model $\mathrm{Y} = \mathrm{x}\beta + \varepsilon$. The unconstrained estimate of β minimizes the function

$$LS(\beta) = (\mathrm{y} - \mathrm{x}\beta)'(\mathrm{y} - \mathrm{x}\beta)$$

The value of $U(\beta)$ is at a minimum when the derivative of $U(\beta)$ with respect to β equals zero. We compute this derivative, set the result equal to 0, and solve for the value of β using some of the matrix operations described in Section A.2. The solution is denoted as $\bar{\beta}$. The derivative of $U(\beta)$ with respect to β is

$$\frac{\partial LS(\beta)}{\partial \beta'} = 2\beta'\mathrm{x}'\mathrm{x} - 2\mathrm{y}'\mathrm{x}$$

Setting this result equal to 0 and solving for $\boldsymbol{\beta}$ gives

$$\begin{aligned}
2\boldsymbol{\beta}'\mathbf{x}'\mathbf{x} - 2\mathbf{y}'\mathbf{x} &= \mathbf{0} \\
2\boldsymbol{\beta}'\mathbf{x}'\mathbf{x} &= 2\mathbf{y}'\mathbf{x} \\
\boldsymbol{\beta}'\mathbf{x}'\mathbf{x} &= \mathbf{y}'\mathbf{x} \\
\boldsymbol{\beta}'(\mathbf{x}'\mathbf{x})(\mathbf{x}'\mathbf{x})^{-1} &= \mathbf{y}'\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1} \\
\bar{\boldsymbol{\beta}}' &= \mathbf{y}'\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}
\end{aligned}$$

and taking the transpose we obtain

$$\bar{\boldsymbol{\beta}} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}$$

Thus the value of $\boldsymbol{\beta}$ that minimizes the least-squares function is $(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}$, and this is denoted as $\bar{\boldsymbol{\beta}}$.

The constrained least-squares estimate of $\boldsymbol{\beta}$ minimizes the function

$$U(\boldsymbol{\beta}, \boldsymbol{\lambda}) = (\mathbf{y} - \mathbf{x}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{x}\boldsymbol{\beta}) - \boldsymbol{\lambda}'(\mathbf{a} - \mathbf{h}\boldsymbol{\beta})$$

where $\boldsymbol{\lambda}$ is a vector of Lagrange multipliers. To obtain the constrained least-squares estimate of $\boldsymbol{\beta}$, we take the derivative of $U(\boldsymbol{\beta}, \boldsymbol{\lambda})$ with respect to both $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$, set each result equal to 0, first solve for $\boldsymbol{\lambda}$, and then solve for $\boldsymbol{\beta}$. The derivative of $U(\boldsymbol{\beta}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{\beta}$ is

$$\frac{\partial US(\boldsymbol{\beta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta}'} = 2\boldsymbol{\beta}'(\mathbf{x}'\mathbf{x}) - 2\mathbf{y}'\mathbf{x} + \boldsymbol{\lambda}'\mathbf{h}$$

and the derivative of $U(\boldsymbol{\beta}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{\lambda}$ is

$$\frac{\partial U(\boldsymbol{\beta}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}'} = -\mathbf{a}' + \boldsymbol{\beta}'\mathbf{h}'$$

Setting $\partial U(\boldsymbol{\beta}, \boldsymbol{\lambda})/\partial \boldsymbol{\beta}'$ to 0 and postmultiplying both sides of the equation by $(\mathbf{x}'\mathbf{x})^{-1}\mathbf{h}'$ gives

$$\begin{aligned}
\mathbf{0} &= 2\boldsymbol{\beta}'(\mathbf{x}'\mathbf{x})(\mathbf{x}'\mathbf{x})^{-1}\mathbf{h}' - 2\mathbf{y}'\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{h}' + \boldsymbol{\lambda}'\mathbf{h}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{h}' \\
&= 2\boldsymbol{\beta}'\mathbf{h}' - 2\mathbf{y}'\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{h}' + \boldsymbol{\lambda}'\mathbf{h}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{h}'
\end{aligned}$$

Solving for $\boldsymbol{\lambda}'$, we obtain

$$\begin{aligned}
\boldsymbol{\lambda}'\mathbf{h}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{h}' &= 2\mathbf{y}'\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{h}' - 2\boldsymbol{\beta}'\mathbf{h}' \\
&= 2\bar{\boldsymbol{\beta}}'\mathbf{h}' - 2\boldsymbol{\beta}'\mathbf{h}' \\
&= 2(\boldsymbol{\beta}'\mathbf{h}' - \mathbf{a}')
\end{aligned}$$

Postmultiplying both sides by $(\mathbf{h}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{h}')^{-1}$ yields

$$\boldsymbol{\lambda}'\mathbf{h}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{h}'(\mathbf{h}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{h}')^{-1} = 2\left(\boldsymbol{\beta}'\mathbf{h}' - \mathbf{a}'\right)\left(\mathbf{h}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{h}'\right)^{-1}$$

$$\boldsymbol{\lambda}' = 2\left(\boldsymbol{\beta}'\mathbf{h}' - \mathbf{a}'\right)\left(\mathbf{h}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{h}'\right)^{-1}$$

Substituting this expression for $\boldsymbol{\lambda}'$ in $\partial U(\boldsymbol{\beta}, \boldsymbol{\lambda})/\partial\boldsymbol{\beta}'$, setting the result to 0, and solving for $\boldsymbol{\beta}$ gives

$$0 = 2\boldsymbol{\beta}'(\mathbf{x}'\mathbf{x}) - 2\mathbf{y}'\mathbf{x} + 2\left(\boldsymbol{\beta}'\mathbf{h}' - \mathbf{a}'\right)\left(\mathbf{h}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{h}'\right)^{-1}\mathbf{h}$$

$$2\boldsymbol{\beta}'(\mathbf{x}'\mathbf{x}) = 2\mathbf{y}'\mathbf{x} - 2\left(\boldsymbol{\beta}'\mathbf{h}' - \mathbf{a}'\right)\left(\mathbf{h}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{h}'\right)^{-1}\mathbf{h}$$

Postmultiplying both sides by $\frac{1}{2}(\mathbf{x}'\mathbf{x})^{-1}$ yields

$$\boldsymbol{\beta}'(\mathbf{x}'\mathbf{x})(\mathbf{x}'\mathbf{x})^{-1} = \mathbf{y}'\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1} - \left(\boldsymbol{\beta}'\mathbf{h}' - \mathbf{a}'\right)\left(\mathbf{h}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{h}'\right)^{-1}\mathbf{h}(\mathbf{x}'\mathbf{x})^{-1}$$

Solving for $\boldsymbol{\beta}$ yields the constrained estimate

$$\bar{\boldsymbol{\beta}}^{*\prime} = \bar{\boldsymbol{\beta}}' - \left(\bar{\boldsymbol{\beta}}'\mathbf{h}' - \mathbf{a}'\right)\left(\mathbf{h}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{h}'\right)^{-1}\mathbf{h}(\mathbf{x}'\mathbf{x})^{-1}$$

Taking the transpose of both sides yields

$$\bar{\boldsymbol{\beta}}^{*} = \bar{\boldsymbol{\beta}} - (\mathbf{x}'\mathbf{x})^{-1}\mathbf{h}'\left(\mathbf{h}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{h}'\right)^{-1}(\mathbf{h}\bar{\boldsymbol{\beta}} - \mathbf{a})$$

which is the contrained least-squares estimate of $\boldsymbol{\beta}$ subject to the linear constraints $\mathbf{h}\boldsymbol{\beta} = \mathbf{a}$.

7.3.1 (1) $\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \\ Y_7 \\ Y_8 \\ Y_9 \\ Y_{10} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \varepsilon_{10} \end{bmatrix}$

7.3.2 Let $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix}$, $\gamma = \begin{bmatrix} \mu_1 & \mu_2 \\ \mu_3 & \mu_4 \end{bmatrix}$, $\mathbf{h}_b = [1 \quad -1]$,

and $\mathbf{h}_p = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Then $\mathbf{h}_b \otimes \mathbf{h}_p =$

$\begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}$, so $(\mathbf{h}_b \otimes \mathbf{h}_p)\boldsymbol{\mu} = \mathbf{0}$ equals

$\begin{bmatrix} \mu_1 - \mu_3 \\ \mu_2 - \mu_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Similarly, $\mathbf{h}_b \gamma \mathbf{h}_p = \mathbf{0}$ is

$[1 \quad -1] \begin{bmatrix} \mu_1 & \mu_2 \\ \mu_3 & \mu_4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, which equals

$\begin{bmatrix} \mu_1 - \mu_3 \\ \mu_2 - \mu_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

7.3.3 (1) Unrestricted

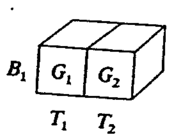
(2) No, there were no within-subject factor levels presented in random order for each subject.

CHAPTER 8

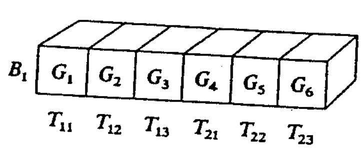
8.1.1 (1)

		t	b	p
For Example	8.1.1	4	2	2
	8.1.2	12	6	2
	8.1.3	12	4	3
	8.1.4	8	2	4
	8.1.5	24	6	4

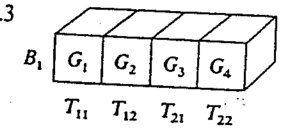
(2) For Example 8.1.1



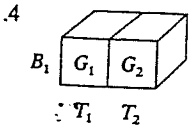
8.1.2



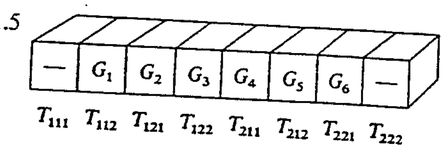
8.1.3



8.1.4



8.1.5



APPENDIX A

A.2.1 $\mathbf{b} = \mathbf{d}$

A.2.2 (1) $\mathbf{a}' = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$, $\quad \mathbf{b}' = [1 \quad 2 \quad 3]$,

$\mathbf{c}' = \begin{bmatrix} 1 & 4 & 3 \\ 4 & 1 & 2 \\ 3 & 2 & 1 \end{bmatrix}$, $\quad \mathbf{d}' = \begin{bmatrix} 4 \\ 4 \\ 4 \\ 1 \end{bmatrix}$

(2) $\mathbf{c}$

A.2.3 (1) $\mathbf{a} + \mathbf{b} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$

(2) $\mathbf{a} - \mathbf{b} = \begin{bmatrix} -2 \\ -1 \\ 0 \end{bmatrix}$

(3) $\mathbf{b} - \mathbf{a} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$

(4) $\mathbf{c} + \mathbf{d} = \begin{bmatrix} 7 & 3 & 4 \\ 3 & 2 & 6 \end{bmatrix}$

A.2.4 (1) $a\mathbf{b} = \begin{bmatrix} 12 & 6 & 18 \\ 24 & 6 & 36 \end{bmatrix}$

(2) Same as **(1)**

(3) $\mathbf{1}_3 \mathbf{a} = \begin{bmatrix} 6 \\ 6 \\ 6 \end{bmatrix}$

A.2.5 **(1)** $a'a = 14, aa' = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}$

(2) $a'b = [15 \quad 19], \qquad b'a = \begin{bmatrix} 15 \\ 19 \end{bmatrix}$

(3) $cd = \begin{bmatrix} 30 & 24 & 18 \\ 84 & 69 & 54 \\ 138 & 114 & 90 \end{bmatrix},$

$dc = \begin{bmatrix} 90 & 114 & 138 \\ 54 & 69 & 84 \\ 18 & 24 & 30 \end{bmatrix}$

(4) $dI_3 = d, I_3 d = d$

(5) $e(c + d) = [60 \quad 60 \quad 60]$

(6) $ec + ed = [24 \quad 30 \quad 36] + [36 \quad 30 \quad 24]$
$= [60 \quad 60 \quad 60]$

(7) $e(c - d) = [-12 \quad 0 \quad 12]$

(8) $ec - ed = [24 \quad 30 \quad 36] - [36 \quad 30 \quad 24]$
$= [-12 \quad 0 \quad 12]$

(9) $cdb = \begin{bmatrix} 162 & 174 \\ 468 & 507 \\ 774 & 840 \end{bmatrix}$

A.2.6 **(1)** $(a + b)x$ **(3)** $a(x + I)$
(2) $x(a + b)$ **(4)** $(x + I)a$

A.2.7 **(1)** $a \otimes b = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \\ 3 & 6 & 9 \\ 3 & 9 & 6 \\ 1 & 2 & 3 \\ 1 & 3 & 2 \end{bmatrix},$

$a \otimes c = \begin{bmatrix} 3 & 3 & 3 \\ 1 & 1 & 1 \\ 9 & 9 & 9 \\ 3 & 3 & 3 \\ 3 & 3 & 3 \\ 1 & 1 & 1 \end{bmatrix},$

$b \otimes c = \begin{bmatrix} 3 & 3 & 3 & 6 & 6 & 6 & 9 & 9 & 9 \\ 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 3 \\ 3 & 3 & 3 & 9 & 9 & 9 & 6 & 6 & 6 \\ 1 & 1 & 1 & 3 & 3 & 3 & 2 & 2 & 2 \end{bmatrix}$

(2) $a \otimes (b + c) \begin{bmatrix} 4 & 5 & 6 \\ 2 & 4 & 3 \\ 12 & 15 & 18 \\ 6 & 12 & 9 \\ 4 & 5 & 6 \\ 2 & 4 & 3 \end{bmatrix}$

$a \otimes b \otimes c = \begin{bmatrix} 3 & 3 & 3 & 6 & 6 & 6 & 9 & 9 & 9 \\ 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 3 \\ 3 & 3 & 3 & 9 & 9 & 9 & 6 & 6 & 6 \\ 1 & 1 & 1 & 3 & 3 & 3 & 2 & 2 & 2 \\ 9 & 9 & 9 & 18 & 18 & 18 & 27 & 27 & 27 \\ 3 & 3 & 3 & 6 & 6 & 6 & 9 & 9 & 9 \\ 9 & 9 & 9 & 27 & 27 & 27 & 18 & 18 & 18 \\ 3 & 3 & 3 & 9 & 9 & 9 & 6 & 6 & 6 \\ 3 & 3 & 3 & 6 & 6 & 6 & 9 & 9 & 9 \\ 1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 3 \\ 3 & 3 & 3 & 9 & 9 & 9 & 6 & 6 & 6 \\ 1 & 1 & 1 & 3 & 3 & 3 & 2 & 2 & 2 \end{bmatrix}$

A.2.8 $\begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{4} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

A.2.9 **(1)** $\begin{bmatrix} 3 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 0.5 & -0.25 \\ -0.25 & 0.375 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

(2) $b^{-1} = \begin{bmatrix} 0.1765 & -0.0588 \\ -0.0588 & 0.3529 \end{bmatrix}$

(3) $b^{-1}b = I$

A.2.10

$a^{-1} = \begin{bmatrix} 2 & -4 \\ -1 & 2 \end{bmatrix} \cdot \frac{1}{0}$ which is undefined

A.2.11 $dd^{-}d = d, \qquad d^{-}dd^{-} = d^{-}, \qquad d^{-}d = dd^{-},$
$dd^{-} = d^{-}d$

A.2.12 **(1)** $aa^{-}a = a,$ $\cancel{a^{-}aa^{-} = a^{-}},$
$\cancel{a^{-}a = aa^{-}},$ $\cancel{aa^{-} = a^{-}a}$

(2) $\begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2.8 \\ 5.6 \end{bmatrix} = \begin{bmatrix} 28 \\ 14 \end{bmatrix}$

(3) $\begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -1.2 \\ 7.6 \end{bmatrix} = \begin{bmatrix} 28 \\ 14 \end{bmatrix}$

A.2.13 $x = b^{-1}c = \begin{bmatrix} 0.7064 & 1.0005 \\ 3.7644 & 2.9998 \end{bmatrix}$

A.2.14 **(1)** $x = a^{-1}(d + c)$ **(2)** $x = -2(a - d)^{-1}c$
(3) $x = (a + I)^{-1}d$

A.3.1 $\bar{\mu} = 110.5, \bar{\sigma}^2 = 77.9166$