4 Matrix Algebra: A Brief Introduction

In the previous chapter, we learned the algebraic results that form the foundation for the study of factor analysis and structural equation modeling. These results, powerful as they are, are somewhat cumbersome to apply in more complicated systems involving large numbers of variables. Matrix algebra provides us with a new mathematical notation that is ideally suited for developing results involving linear combinations and transformations. Once we have developed a few basic results and learned how to use them, we will be ready to derive the fundamental equations of factor analysis and structural modeling.

4.1 Basic Terminology

Definition (Matrix) A **matrix** is defined as an ordered array of numbers, of dimensions p,q.

Our standard notation for a matrix $\boldsymbol{A}$ of order p, q will be:

$$ {}_p\boldsymbol{A}_q $$

There are numerous other notations. For example, one might indicate a matrix of order p, q as $\boldsymbol{A}$ $(p \times q)$. Frequently, we shall refer to such a matrix as "a $p \times q$ matrix $\boldsymbol{A}$."

On occasion, we shall refer explicitly to the *elements* of a matrix (i.e., the numbers or random variables in the array). In this case, we use the following notation to indicate that "$\boldsymbol{A}$ is a matrix with elements a_{ij}".

$$\boldsymbol{A} = \{a_{ij}\}$$

When we refer to element a_{ij}, the *first* subscript will refer to the *row position* of the elements in the array. The *second* subscript (regardless of which letter is used in this position) will refer to the column position. Hence, a typical matrix ${}_p\boldsymbol{A}_q$ will be of the form:

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1q} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2q} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3q} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{p1} & a_{p2} & a_{p3} & \cdots & a_{pq} \end{bmatrix}$$

We shall generally use bold italic capital letters to indicate matrices, and employ lower case letters to signify elements of an array, except where clarity dictates. In particular, we may find it convenient, when referring to matrices of random variables, to refer to the elements with capital letters, so we can distinguish notationally between constants and random variables.

Definition (Column Vector) A **column vector** of numbers or random variables will be a matrix of order $p \times 1$. We will, in general, indicate column vectors with the following notation:

$${}_p\boldsymbol{x}_1$$

Definition (Row Vector) A **row vector** of numbers or random variables will be a matrix of order $1 \times q$. We will, in general, indicate row vectors with the following notation:

$${}_1\boldsymbol{x}'_q$$

A column vector with all elements equal to one will be symbolized as either $\boldsymbol{j}$ or $\boldsymbol{1}$.

Note how we reserve the use of italic boldface for matrices and vectors.

4.1.1 Special Matrices

We will refer occasionally to special types of matrices by name. For any ${}_p\boldsymbol{A}_q$,

(a) If $p \neq q$, $\boldsymbol{A}$ is a *rectangular* matrix.

(b) If $p = q$, $\boldsymbol{A}$ is a *square* matrix.

(c) In a *square* matrix $\boldsymbol{A}$, the elements a_{ii}, $i = 1,\ p$ define the *diagonal* of the matrix.

(d) A square matrix $\boldsymbol{A}$ is *lower triangular* if $a_{ij} = 0$ for $i < j$.

(e) A square matrix $\boldsymbol{A}$ is *upper triangular* if $a_{ij} = 0$ for $i > j$.

(f) A square matrix $\boldsymbol{A}$ is a *diagonal matrix* if $a_{ij} = 0$ for $i \neq j$

(g) A square matrix $\boldsymbol{A}$ is a *scalar matrix* if it is a diagonal matrix and all diagonal elements are equal.

(h) An *identity matrix* is a scalar matrix with diagonal elements equal to one. We use the notation $\boldsymbol{I}_p$ to denote a $p \times p$ identity matrix.

(i) $\boldsymbol{0}$, a matrix composed entirely of zeros, is called a *null matrix*.

(j) A square matrix $\boldsymbol{A}$ is *symmetric* if $a_{ij} = a_{ji}$ for all i, j.

(k) A 1×1 matrix is a *scalar*.

Example 4.1.1 (Special Matrices) Some examples follow:

(a) A rectangular matrix

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix}$$

(b) A square matrix

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

(c) A lower triangular matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 3 & 0 & 0 \\ 4 & 5 & 6 & 0 \\ 7 & 8 & 9 & 10 \end{bmatrix}$$

(d) An upper triangular matrix

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 5 & 6 & 7 \\ 0 & 0 & 8 & 9 \\ 0 & 0 & 0 & 10 \end{bmatrix}$$

(e) A diagonal matrix

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 7 \end{bmatrix}$$

(f) A scalar matrix

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

(g) A symmetric matrix

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

4.1.2 Partitioning of Matrices

In many theoretical discussions of matrices, it will be useful to conceive of a matrix as being composed of sub-matrices. When we do this, we will "partition" the matrix symbolically by breaking it down into its components. The components can be either matrices or scalars. Here is a simple example.

Example 4.1.2 (A Simple Partitioned Matrix) In discussions of simple multiple regression, where there is one criterion variable Y and p predictor variables that can be placed in a vector $\boldsymbol{x}$, it is common to refer to the correlation matrix of the entire set of variables using partitioned notation, as follows:

$$\boldsymbol{R} = \begin{bmatrix} 1 & \boldsymbol{r}'_{Y\boldsymbol{x}} \\ \boldsymbol{r}_{\boldsymbol{x}Y} & \boldsymbol{R}_{\boldsymbol{xx}} \end{bmatrix}$$

In the above notation, the scalar 1 in the upper-left partition refers to the correlation of the Y variable with itself. The row vector $\boldsymbol{r}'_{Y\boldsymbol{x}}$ refers to the set of correlations between the variable Y and the set of p random variables in $\boldsymbol{x}$. $\boldsymbol{R}_{\boldsymbol{xx}}$ is the $p \times p$ matrix of correlations of the predictor variables. We will refer to the "order" of the "partitioned form" as the number of rows and columns in the partitioning, which is distinct from the number of rows and columns in the matrix being represented. For example, suppose there were $p = 5$ predictor variables in Example 4.1.2. Then the matrix $\boldsymbol{R}$ is a 6×6 matrix, but the example shows a "2×2 partitioned form."

When matrices are partitioned properly, it is understood that "pieces" that appear to the left or right of other pieces have the same number of rows, and pieces that appear above or below other pieces have the same number of columns. So, in the above example, $\boldsymbol{R}_{\boldsymbol{xx}}$, appearing to the right of the $p \times 1$ column vector $\boldsymbol{R}_{\boldsymbol{x}Y}$, must have p rows, and since it appears below the $1 \times p$ row vector $\boldsymbol{r}'_{Y\boldsymbol{x}}$, it must have p columns. Hence, it must be a $p \times p$ matrix.

4.2 Some Matrix Operations

In this section, we review the fundamental operations on matrices.

4.2.1 Matrix (and Vector) Addition and Subtraction

For the addition and subtraction operations to be defined for two matrices $\boldsymbol{A}$, $\boldsymbol{B}$, they must be *conformable*.

Definition (Conformability for Addition and Subtraction) Two matrices are *conformable for addition and subtraction* if and only if they are of the same order.

Definition (Matrix Addition) Let $\boldsymbol{A} = \{a_{ij}\}$ and $\boldsymbol{B} = \{b_{ij}\}$. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be conformable. The sum $\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{C}$ is defined as:

$$\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B} = \{c_{ij}\} = \{a_{ij} + b_{ij}\}$$

Definition (Matrix Subtraction) Let $\boldsymbol{A} = \{a_{ij}\}$ and $\boldsymbol{B} = \{b_{ij}\}$. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be conformable. The sum $\boldsymbol{A} - \boldsymbol{B} = \boldsymbol{C}$ is defined as:

$$\boldsymbol{C} = \boldsymbol{A} - \boldsymbol{B} = \{c_{ij}\} = \{a_{ij} - b_{ij}\}$$

Matrix addition and subtraction are natural, intuitive extensions to scalar addition and subtraction. One simply adds elements in the same position.

Example 4.2.1 (Matrix Addition and Subtraction)

Let $\boldsymbol{A} = \begin{bmatrix} 1 & 4 & 5 \\ 2 & 3 & 4 \\ 4 & 4 & 0 \end{bmatrix}$, and $\boldsymbol{B} = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 3 & 2 \end{bmatrix}$.

Find $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}$ and $\boldsymbol{D} = \boldsymbol{A} - \boldsymbol{B}$.

Solution $\boldsymbol{C} = \begin{bmatrix} 4 & 6 & 6 \\ 4 & 6 & 5 \\ 5 & 7 & 2 \end{bmatrix}$ $\boldsymbol{D} = \begin{bmatrix} -2 & 2 & 4 \\ 0 & 0 & 3 \\ 3 & 1 & -2 \end{bmatrix}$

Definition (Matrix Equality) Two matrices are equal if and only if they are of the same row and column order, and have all elements equal.

For matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$, properties of matrix addition include:

(a) Associativity
$$\boldsymbol{A} + (\boldsymbol{B} + \boldsymbol{C}) = (\boldsymbol{A} + \boldsymbol{B}) + \boldsymbol{C}$$

(b) Commutativity
$$\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}$$

(c) There exists a "neutral element" for addition, i.e., the null matrix $\mathbf{0}$, such that $\boldsymbol{A} + \mathbf{0} = \boldsymbol{A}$.

(d) There exist inverse elements for addition, in the sense that for any matrix $\boldsymbol{A}$, there exists a matrix $-\boldsymbol{A}$, such that $\boldsymbol{A} + -\boldsymbol{A} = \mathbf{0}$.

These properties mimic those of scalar addition and subtraction. Consequently, there is little "negative transfer" involved in generalizing from the scalar to the matrix operations.

4.2.2 Scalar Multiples and Scalar Products

In the previous section, we examined some matrix operations, addition and subtraction, that operate very much like their scalar algebraic counterparts. In this section, we begin to see a divergence between matrix algebra and scalar algebra.

Definition (Scalar Multiplication) Given a matrix $\boldsymbol{A} = \{a_{ij}\}$, and a scalar c. Then $\boldsymbol{B} = c\boldsymbol{A} = \{ca_{ij}\}$ is called a *scalar multiple* of $\boldsymbol{A}$.

Scalar multiples are not to be confused with *scalar products*, which will be defined subsequently. Scalar multiplication is a simple idea — multiply a matrix by a scalar, and you simply multiply every element of the matrix by the scalar.

Example 4.2.2 (Scalar Multiple) Let $\boldsymbol{A} = \begin{bmatrix} 2 & -1 \\ 3 & 4 \end{bmatrix}$. Then $2\boldsymbol{A} = \begin{bmatrix} 4 & -2 \\ 6 & 8 \end{bmatrix}$.

For matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, and scalars a and b, scalar multiplication has the following mathematical properties:

(a) $(a+b)\boldsymbol{A} = a\boldsymbol{A} + b\boldsymbol{A}$

(b) $a(\boldsymbol{A}+\boldsymbol{B}) = a\boldsymbol{A} + a\boldsymbol{B}$

(c) $a(b\boldsymbol{A}) = (ab)\boldsymbol{A}$

(d) $a\boldsymbol{A} = \boldsymbol{A}a$

Definition (Scalar Product) Given row vector ${}_1\boldsymbol{a}'_p$ and ${}_p\boldsymbol{b}_1$. Let $\boldsymbol{a}' = \{a_i\}$ and $\boldsymbol{b} = \{b_i\}$. The *scalar product* $\boldsymbol{a}'\boldsymbol{b}$ is defined as

$$\boldsymbol{a}'\boldsymbol{b} = \sum_{i=1}^{p} a_i b_i$$

The scalar product is simply the sum of cross products of the elements of the two vectors.

Example 4.2.3 (Scalar Product) Let $\boldsymbol{a}' = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$. Let $\boldsymbol{b} = \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}$.

Then $\boldsymbol{a}'\boldsymbol{b} = 11$.

4.2.3 Matrix Multiplication

Matrix multiplication is an operation with properties quite different from its scalar counterpart. To begin with, *order matters* in matrix multiplication. That is, the matrix product $\boldsymbol{AB}$ need not be the same as the matrix product $\boldsymbol{BA}$. Indeed, the matrix product $\boldsymbol{AB}$ might be well-defined, while the product $\boldsymbol{BA}$ might not exist. To begin with, we establish the fundamental property of *conformability*.

Definition (Conformability for Matrix Multiplication) ${}_p\boldsymbol{A}_q$ and ${}_r\boldsymbol{B}_s$ are *conformable for matrix multiplication* if and only if $q = r$.

The matrix multiplication operation is defined as follows.

Definition (Matrix Multiplication) Let ${}_p\boldsymbol{A}_q = \{a_{ij}\}$ and ${}_q\boldsymbol{B}_s = \{b_{jk}\}$. Then ${}_p\boldsymbol{C}_s = \boldsymbol{AB} = \{c_{ik}\}$ where:

$$c_{ik} = \sum_{j=1}^{q} a_{ij} b_{jk}$$

Example 4.2.4 (The Row by Column Method) The meaning of the formal definition of matrix multiplication might not be obvious at first glance. Indeed, there are several ways of thinking about matrix multiplication. The first way, which I call the *row by column approach*, works as follows. Visualize ${}_p\boldsymbol{A}_q$ as a set of p *row vectors* and ${}_q\boldsymbol{B}_s$ as a set of s *column vectors*. Then if $\boldsymbol{C} = \boldsymbol{AB}$, element c_{ik} of $\boldsymbol{C}$ is the scalar product (i.e., the sum of cross products) of the ith row of $\boldsymbol{A}$ with the kth column of $\boldsymbol{B}$.

For example, let $\boldsymbol{A} = \begin{bmatrix} 2 & 4 & 6 \\ 5 & 7 & 1 \\ 2 & 3 & 5 \end{bmatrix}$, and let $\boldsymbol{B} = \begin{bmatrix} 4 & 1 \\ 0 & 2 \\ 5 & 1 \end{bmatrix}$.

Then $\boldsymbol{C} = \boldsymbol{AB} = \begin{bmatrix} 38 & 16 \\ 25 & 20 \\ 33 & 13 \end{bmatrix}$.

Consider element c_{21}, which has a value of 25. This element, which is in the second row and first column of $\boldsymbol{C}$, is computed by taking the sum of cross products of the elements of the second row of $\boldsymbol{A}$ with the first column of $\boldsymbol{B}$. That is, $(5 \times 4) + (7 \times 0) + (1 \times 5) = 25$.

The following are some key properties of matrix multiplication:

(a) Associativity.

$$(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$$

(b) Not generally commutative. That is, often $\boldsymbol{AB} \neq \boldsymbol{BA}$.

(c) Distributive over addition and subtraction.

$$\boldsymbol{C}(\boldsymbol{A}+\boldsymbol{B}) = \boldsymbol{C}\boldsymbol{A}+\boldsymbol{C}\boldsymbol{B}$$

(d) Assuming it is conformable, the identity matrix $\boldsymbol{I}$ functions like the number 1, that is ${}_p\boldsymbol{A}_q\boldsymbol{I}_q = \boldsymbol{A}$, and ${}_p\boldsymbol{I}_p\boldsymbol{A}_q = \boldsymbol{A}$.

(e) $\boldsymbol{AB} = \boldsymbol{0}$ does not necessarily imply that either $\boldsymbol{A} = \boldsymbol{0}$ or $\boldsymbol{B} = \boldsymbol{0}$.

Several of the above results are surprising, and result in negative transfer for beginning students as they attempt to reduce matrix algebra expressions.

Example 4.2.5 (A Null Matrix Product) The following example shows that one can, indeed, obtain a null matrix as the product of two non-null matrices.

Let $\boldsymbol{a}' = \begin{bmatrix} 6 & 2 & 2 \end{bmatrix}$, and let $\boldsymbol{B} = \begin{bmatrix} -8 & 12 & 12 \\ 12 & -40 & 4 \\ 12 & 4 & -40 \end{bmatrix}$.

Then $\boldsymbol{a}'\boldsymbol{B} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$.

Definition (Pre-multiplication and Post-multiplication) When we talk about the "product of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$," it is important to remember that $\boldsymbol{AB}$ and $\boldsymbol{BA}$ are usually not the same. Consequently, it is common to use the terms "pre-multiplication" and "post-multiplication." When we say "$\boldsymbol{A}$ is post-multiplied by $\boldsymbol{B}$," or "$\boldsymbol{B}$ is pre-multiplied by $\boldsymbol{A}$," we are referring to the product $\boldsymbol{AB}$. When we say "$\boldsymbol{B}$ is post-multiplied by $\boldsymbol{A}$," or "$\boldsymbol{A}$ is pre-multiplied by $\boldsymbol{B}$," we are referring to the product $\boldsymbol{BA}$.

4.2.4 Matrix Transposition

"Transposing" a matrix is an operation which plays a very important role in multivariate statistical theory. The operation, in essence, switches the rows and columns of a matrix.

Definition (Transpose of a Matrix) Let ${}_p\boldsymbol{A}_q = \{a_{ij}\}$. Then the *transpose* of $\boldsymbol{A}$, denoted $\boldsymbol{A}'$, is defined as

$$ {}_q\boldsymbol{A}'_p = \{a_{ji}\} $$

Comment. The above notation may seem cryptic to the beginner, so some clarification may be useful. The typical element of $\boldsymbol{A}$ is a_{ij}. This means that, in the i,j position of matrix $\boldsymbol{A}$ is found element a_{ij}. Suppose $\boldsymbol{A}$ is of order 3×3. Then it can be written as

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

$\boldsymbol{A}'$ is a matrix with typical element a_{ji}. This means that, in the i,j position of matrix $\boldsymbol{A}'$ is found element a_{ji} of the original matrix $\boldsymbol{A}$. For example, element 2,1 of $\boldsymbol{A}'$ is element 1,2 of the original matrix. So, in terms of the elements of the original $\boldsymbol{A}$, the transpose has the representation

$$ \boldsymbol{A}' = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{22} & a_{33} \end{bmatrix} $$

Studying the above element-wise representation, you can see that transposition does not change the diagonal elements of a matrix, and constructs columns of the transpose from the rows of the original matrix.

Example 4.2.6 (Matrix Transposition)

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

Here are some frequently used properties of the matrix transposition operation:

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

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

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

(d) $(\boldsymbol{A}\boldsymbol{B})' = \boldsymbol{B}'\boldsymbol{A}'$

(e) A square matrix $\boldsymbol{A}$ is symmetric if and only if $\boldsymbol{A} = \boldsymbol{A}'$.

You should study the properties of matrix transposition carefully, as they are used frequently in reducing and simplifying matrix algebra expressions.

4.2.5 Linear Combinations of Matrix Rows and Columns

We have already discussed the "row by column" conceptualization of matrix multiplication. However, there are some other ways of conceptualizing matrix multiplication that are particularly useful in the field of multivariate statistics. To begin with, we need to enhance our understanding of the way matrix multiplication and transposition works with partitioned matrices.

Definition (Multiplication and Transposition of Partitioned Matrices) Being able to transpose and multiply a partitioned matrix is a skill that is important for understanding the key equations of structural equation modeling. Assuming that the matrices are partitioned properly, the rules are quite simple:

(a) To transpose a partitioned matrix, treat the sub-matrices in the partition as though they were elements of a matrix, but transpose each sub-matrix. The transpose of a $p \times q$ partitioned form will be a $q \times p$ partitioned form.

(b) To multiply partitioned matrices, treat the sub-matrices as though they were elements of a matrix. The product of $p \times q$ and $q \times r$ partitioned forms is a $p \times r$ partitioned form.

Some examples will illustrate the above definition.

Example 4.2.7 (Transposing a Partitioned Matrix) Suppose $\boldsymbol{A}$ is partitioned as

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{D} \\ \boldsymbol{E} & \boldsymbol{F} \\ \boldsymbol{G} & \boldsymbol{H} \end{bmatrix}$$

Then

$$\boldsymbol{A}' = \begin{bmatrix} \boldsymbol{C}' & \boldsymbol{E}' & \boldsymbol{G}' \\ \boldsymbol{D}' & \boldsymbol{F}' & \boldsymbol{H}' \end{bmatrix}$$

Example 4.2.8 (Product of two Partitioned Matrices)

Notice how a 1×2 partitioned form post-multiplied by a 2×1 partitioned form yields a 1×1 partitioned form $\boldsymbol{XG} + \boldsymbol{YH}$.

Suppose $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{X} & \boldsymbol{Y} \end{bmatrix}$ and $\boldsymbol{B} = \begin{bmatrix} \boldsymbol{G} \\ \boldsymbol{H} \end{bmatrix}$.

Then $\boldsymbol{AB} = \begin{bmatrix} \boldsymbol{XG} + \boldsymbol{YH} \end{bmatrix}$.

Example 4.2.9 (Linearly Combining Columns of a Matrix) Consider an $n \times p$ matrix $\boldsymbol{X}$, containing the scores of n persons on p variables. One can conceptualize the matrix $\boldsymbol{X}$ as a set of p column vectors. In "partitioned matrix form," we can represent $\boldsymbol{X}$ as

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \boldsymbol{x}_3 & \cdots & \boldsymbol{x}_p \end{bmatrix}$$

Now suppose one were to post-multiply $\boldsymbol{X}$ with a $p \times 1$ vector $\boldsymbol{b}$. The product is a $n \times 1$ column vector $\boldsymbol{y}$. Utilizing the above results on multiplication of partitioned matrices, it is easy to see that the result can be written as follows.

$$
\begin{aligned}
\boldsymbol{y} &= \boldsymbol{X}\boldsymbol{b} \\
&= \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \boldsymbol{x}_3 & \cdots & \boldsymbol{x}_p \\ & & & & \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_p \end{bmatrix} \\
&= b_1\boldsymbol{x}_1 + b_2\boldsymbol{x}_2 + b_3\boldsymbol{x}_3 + \cdots + b_p\boldsymbol{x}_p
\end{aligned}
$$

The above example illustrates a very general principle in matrix algebra. If you wish to linearly combine a set of variables that are in the columns of a matrix $\boldsymbol{X}$, simply place the desired linear weights in a vector $\boldsymbol{b}$, and post-multiply $\boldsymbol{X}$ by $\boldsymbol{b}$. We illustrate this below with a simple numerical example.

Example 4.2.10 (Computing Difference Scores) Suppose the matrix $\boldsymbol{X}$ consists of a set of scores on two variables, and you wish to compute the difference scores on the variables. Simply apply the linear weight $+1$ to the first column, and -1 to the second column, by post-multiplying with a vector containing those weights. Specifically

$$
\begin{aligned}
\boldsymbol{y} &= \boldsymbol{X}\boldsymbol{b} \\
&= \begin{bmatrix} 80 & 70 \\ 77 & 79 \\ 64 & 64 \end{bmatrix} \begin{bmatrix} +1 \\ -1 \end{bmatrix} \\
&= +1 \begin{bmatrix} 80 \\ 77 \\ 64 \end{bmatrix} - 1 \begin{bmatrix} 70 \\ 79 \\ 64 \end{bmatrix} \\
&= \begin{bmatrix} 10 \\ -2 \\ 0 \end{bmatrix}
\end{aligned}
$$

Example 4.2.11 (Computing Course Grades) Suppose the matrix $\boldsymbol{X}$ contained grades on two exams, a mid-term and a final, and that the final exam is counted twice as much as the mid-term in determining the course grade. In that case, course grades might be computed using the linear weights $+1/3$ and $+2/3$. Specifically

$$\begin{aligned}
\boldsymbol{y} &= \boldsymbol{X}\boldsymbol{b} \\
&= \begin{bmatrix} 80 & 70 \\ 77 & 79 \\ 64 & 64 \end{bmatrix} \begin{bmatrix} +1/3 \\ +2/3 \end{bmatrix} \\
&= +1/3 \begin{bmatrix} 80 \\ 77 \\ 64 \end{bmatrix} + 2/3 \begin{bmatrix} 70 \\ 79 \\ 64 \end{bmatrix} \\
&= \begin{bmatrix} 73\frac{1}{3} \\ 78\frac{1}{3} \\ 64 \end{bmatrix}
\end{aligned}$$

In the preceding examples, we linearly combined columns of a matrix, using post-multiplication by a vector of linear weights. Of course, one can perform an analogous operation on the rows of a matrix by pre-multiplication with a row vector of linear weights.

Example 4.2.12 (Linearly Combining Rows of a Matrix) Suppose we view the $p \times q$ matrix $\boldsymbol{X}$ as being composed of p row vectors. If we pre-multiply $\boldsymbol{X}$ with a $1 \times p$ row vector $\boldsymbol{b}'$, the elements of $\boldsymbol{b}'$ are linear weights applied to the rows of $\boldsymbol{X}$. Specifically,

$$\begin{aligned}
\boldsymbol{y}' &= \boldsymbol{b}'\boldsymbol{X} \\
&= \begin{bmatrix} b_1 & b_2 & \cdots & b_p \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1' \\ \boldsymbol{x}_2' \\ \vdots \\ \boldsymbol{x}_p' \end{bmatrix} \\
&= b_1\boldsymbol{x}_1' + b_2\boldsymbol{x}_2' + \ldots + b_p\boldsymbol{x}_p'
\end{aligned}$$

4.2.6 Sets of Linear Combinations

There is, of course, no need to restrict oneself to a single linear combination of the rows and columns of a matrix. To create more than one linear combination, simply add columns (or rows) to the post-multiplying (or pre-multiplying) matrix!

Example 4.2.13 (Taking the Sum and Difference of Two Columns)
Suppose the matrix $\boldsymbol{X}$ consists of a set of scores on two variables, and you wish to compute both the sum and the difference scores on the variables. In this case, we post-multiply by two vectors of linear weights, creating two linear combinations. Specifically

$$\begin{aligned}
\boldsymbol{y} &= \boldsymbol{XB} \\
&= \begin{bmatrix} 80 & 70 \\ 77 & 79 \\ 64 & 64 \end{bmatrix} \begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix} \\
&= \begin{bmatrix} 150 & 10 \\ 156 & -2 \\ 128 & 0 \end{bmatrix}
\end{aligned}$$

Linear Combinations can be used to perform a number of basic operations on matrices. For example, you can

(a) Extract a column of a matrix;

(b) Extract a row of a matrix;

(c) Exchange rows and/or columns of a matrix;

(d) Re-scale the rows and/or columns of a matrix by multiplication or division.

(e) Extract a particular single element from a matrix.

See if you can figure out, for yourself, *before* scanning the examples below, how to perform the above 5 operations with linear combinations.

Example 4.2.14 (Extracting a Column of a Matrix) This is perhaps the simplest case of a linear combination. Simply multiply the desired column by the linear weight +1, and make the linear weights 0 for all other columns. So, for example,

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

We can give such a matrix a formal name! (See definition below.)

Definition (Selection Vector) The **selection vector** $\boldsymbol{s}_{[i]}$ is a vector with all elements zero except the ith element, which is 1.

Example 4.2.15 (Extracting a Row of a Matrix) Just as we can extract the ith column of a matrix by post-multiplication by $\boldsymbol{s}_{[i]}$, we can extract the ith row by pre-multiplication by $\boldsymbol{s}'_{[i]}$. In this case, we extract the second row by pre-multiplication by $\boldsymbol{s}'_{[2]}$.

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

Example 4.2.16 (Exchanging Columns of a Matrix) We can use two strategically selected linear combinations to exchange columns of a matrix. Suppose we wish to exchange columns 1 and 2 of the matrix from the previous example. Consider a post-multiplying matrix of linear weights $\boldsymbol{B}$. We simply make its first column $\boldsymbol{s}_{[2]}$, thereby selecting the second column of the matrix to be the first column of the new matrix. We then make the second column of $\boldsymbol{B}$ the selection vector $\boldsymbol{s}_{[1]}$, placing the first column of the old matrix in the second column position of the new matrix. This is illustrated below.

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

Comment. Exchanging rows of a matrix is an obvious generalization. Simply pre-multiply by appropriately chosen (row) selection vectors.

Example 4.2.17 (Rescaling Rows and Columns of a Matrix) This is accomplished by post-multiplication and/or pre-multiplication by appropriately configured diagonal matrices. Consider the following example.

$$\begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 12 \\ 4 & 15 \\ 6 & 18 \end{bmatrix}$$

Example 4.2.18 (Selecting a Specific Element of a Matrix) By selecting the column of a matrix with post-multiplication, and the row with pre-multiplication, one may, using two appropriately chosen selection vectors, "pick out" any element of an array. For example, if we wished to select element $\boldsymbol{X}_{12}$, we would pre-multiply by $\boldsymbol{s}_{[1]}$ and post-multiply by $\boldsymbol{s}_{[2]}$. For example,

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

4.3 Matrix Algebra of Some Sample Statistics

In this section, we show how matrix algebra can be used to express some common statistical formulas in a succinct way that allows us to derive some important results in multivariate analysis.

4.3.1 The Data Matrix

Suppose we wish to discuss a set of sample data representing scores for N people on p variables. We can represent the people in rows and the variables in columns, or vice-versa. Placing the variables in columns seems like a more natural way to do things for the modern computer user, as most computer files for standard statistical software represent the "cases" as rows, and the variables as columns. Ultimately, we will develop the ability to work with both notational variations, but for the time being, we'll work with our data in "column form," i.e., with the variables in columns. Consequently, our standard notation for a data matrix is ${}_N\boldsymbol{X}_p$.

4.3.2 Converting to Deviation Scores

Suppose $\boldsymbol{x}$ is an $n \times 1$ vector of scores for n people on a single variable. We wish to transform the scores in $\boldsymbol{x}$ to *deviation score form*. (In general, we will find this a source of considerable convenience.) To accomplish the deviation score transformation, the arithmetic mean $\overline{x}_\bullet$, must be subtracted from each score in $\boldsymbol{x}$.

Let $\boldsymbol{1}$ be a $n \times 1$ vector of ones. We will refer to such a vector on occasion as a "summing vector," for the following reason. Consider any vector x, for example a 3×1 column vector with the numbers 1, 2, 3. If we compute $\boldsymbol{1}'\boldsymbol{x}$, we are taking the sum of cross-products of a set of 1's with the numbers in x. In summation notation,

$$\boldsymbol{1}'\boldsymbol{x} = \sum_{i=1}^{n} 1_i x_i = \sum_{i=1}^{n} x_i$$

So $\boldsymbol{1}'\boldsymbol{x}$ is how we express "the sum of the x's" in matrix notation.

Consequently,

$$\overline{x}_\bullet = (1/n)\boldsymbol{1}'\boldsymbol{x}$$

To transform $\boldsymbol{x}$ to deviation score form, we need to subtract $\overline{x}_\bullet$ from every element of $\boldsymbol{x}$. We can easily construct a vector with every element equal to $\overline{x}_\bullet$ by simply multiplying the scalar $\overline{x}_\bullet$ by a summing vector. Consequently, if we denote the vector of deviation scores as $\boldsymbol{x}^*$, we have

$$\begin{aligned}
\boldsymbol{x}^* &= \boldsymbol{x} - \boldsymbol{1}\overline{x}_\bullet \\
&= \boldsymbol{x} - \boldsymbol{1}\left(\frac{\boldsymbol{1}'\boldsymbol{x}}{n}\right) && (4.1) \\
&= \boldsymbol{x} - \frac{\boldsymbol{1}\boldsymbol{1}'}{n}\boldsymbol{x} \\
&= \boldsymbol{x} - \left(\frac{\boldsymbol{1}\boldsymbol{1}'}{n}\right)\boldsymbol{x} \\
&= \left(\boldsymbol{I} - \frac{\boldsymbol{1}\boldsymbol{1}'}{n}\right)\boldsymbol{x} && (4.2) \\
&= (\boldsymbol{I} - \boldsymbol{P})\,\boldsymbol{x} && (4.3) \\
\boldsymbol{x}^* &= \boldsymbol{Q}\boldsymbol{x} && (4.4)
\end{aligned}$$

where

$$\boldsymbol{Q} = \boldsymbol{I} - \boldsymbol{P}$$

and

$$\boldsymbol{P} = \frac{\boldsymbol{1}\boldsymbol{1}'}{n}$$

A number of points need to be made about the above derivation:

(a) You should study the above derivation carefully, making certain you understand all steps.

(b) You should carefully verify that the matrix $\boldsymbol{1}\boldsymbol{1}'$ is an $n \times n$ matrix of 1's, so the expression $\boldsymbol{1}\boldsymbol{1}'/n$ is an $n \times n$ matrix with each element equal to $1/n$(Division of matrix by a non-zero scalar is a special case of a scalar multiple, and is perfectly legal).

(c) Since $\boldsymbol{x}$ can be converted from raw score form to deviation score form by pre-multiplication with a single matrix, it follows that any *particular* deviation score can be computed with one pass through a list of numbers.

(d) We would probably never want to compute deviation scores in practice using the above formula, as it would be inefficient. However, the formula does allow us to see some interesting things that are difficult to see using scalar notation (more about that later).

(e) If one were, for some reason, to write a computer program using Equation 4.4, one would not need (or want) to save the matrix $\boldsymbol{Q}$, for several reasons. First, it can be very large! Second, no matter how large n is, the elements of $\boldsymbol{Q}$ take on only two distinct values. Diagonal elements of $\boldsymbol{Q}$ are always equal to $(n-1)/n$, and off-diagonal elements are always equal to $-1/n$. In general, there would be no need to store the numbers.

Example 4.3.1 (The Deviation Score Projection Operator) Any vector of N raw scores can be converted into deviation score form by pre-multiplication by a "projection operator" $\boldsymbol{Q}$. Diagonal elements of $\boldsymbol{Q}$ are always equal to $(n-1)/n$, and off-diagonal elements are always equal to $-1/n$. Suppose we have the vector

$$\boldsymbol{x} = \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix}$$

Construct a projection operator $\boldsymbol{Q}$ such that $\boldsymbol{Qx}$ will be in deviation score form.

Solution We have

$$\begin{aligned} \boldsymbol{Qx} &= \begin{bmatrix} 2/3 & -1/3 & -1/3 \\ -1/3 & 2/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} \end{aligned}$$

Example 4.3.2 (Computing the ith Deviation Score) An implication of the preceding result is that one can compute the ith deviation score as a single linear combination of the n scores in a list. For example, the 3rd deviation score in a list of 3 is computed as $[dx]_3 = -1/3x_1 - 1/3x_2 + 2/3x_3$.

Let us now investigate the properties of the matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ that accomplish this transformation. First, we should establish an additional definition and result.

Definition (Idempotent Matrix) A matrix $\boldsymbol{C}$ is *idempotent* if $\boldsymbol{C}^2 = \boldsymbol{CC} = \boldsymbol{C}$.

Theorem 4.3.3 *If $\boldsymbol{C}$ is idempotent and $\boldsymbol{I}$ is a conformable identity matrix, then $\boldsymbol{I} - \boldsymbol{C}$ is also idempotent.*

Proof *To prove the result, we need merely show that $(\boldsymbol{I} - \boldsymbol{C})^2 = (\boldsymbol{I} - \boldsymbol{C})$. This is straightforward.*

$$
\begin{aligned}
(\boldsymbol{I} - \boldsymbol{C})^2 &= (\boldsymbol{I} - \boldsymbol{C})(\boldsymbol{I} - \boldsymbol{C}) \\
&= \boldsymbol{I}^2 - \boldsymbol{CI} - \boldsymbol{IC} + \boldsymbol{C}^2 \\
&= \boldsymbol{I} - \boldsymbol{C} - \boldsymbol{C} + \boldsymbol{C} \\
&= \boldsymbol{I} - \boldsymbol{C}
\end{aligned}
$$

Recall that $\boldsymbol{P}$ is an $n \times n$ symmetric matrix with each element equal to $1/n$.

$\boldsymbol{P}$ is also idempotent, since:

$$
\begin{aligned}
\boldsymbol{PP} &= \frac{\boldsymbol{11}'}{n}\frac{\boldsymbol{11}'}{n} \\
&= \frac{\boldsymbol{11}'\boldsymbol{11}'}{n^2} \\
&= \frac{\boldsymbol{1}\left(\boldsymbol{1}'\boldsymbol{1}\right)\boldsymbol{1}'}{n^2} \\
&= \frac{\boldsymbol{1}\left(n\right)\boldsymbol{1}'}{n^2} \\
&= \frac{\boldsymbol{11}'\left(n\right)}{n^2} \\
&= \boldsymbol{11}'\frac{n}{n^2} \\
&= \frac{\boldsymbol{11}'}{n} \\
&= \boldsymbol{P}
\end{aligned}
$$

The above derivation demonstrates some principles that are generally useful in reducing simple statistical formulas in matrix form:

(a) Scalars can be "moved through" matrices to any position in the expression that is convenient.

(b) Any expression of the form $\boldsymbol{x}'\boldsymbol{y}$ is a scalar product, and hence it is a scalar, and can be moved intact through other matrices in the expression. So, for example, we recognized that $\boldsymbol{1}'\boldsymbol{1}$ is a scalar and can be reduced and eliminated in the above derivation.

You may easily verify the following properties:

(a) The matrix $\boldsymbol{Q} = \boldsymbol{I} - \boldsymbol{P}$ is also symmetric and idempotent. (Hint: Use Theorem 35.)

(b) $\boldsymbol{Q}\mathbf{1} = \mathbf{0}$ (Hint: First prove that $\boldsymbol{P}\mathbf{1} = \boldsymbol{1}$.)

4.3.3 The Sample Variance and Covariance

Since the sample variance S_X^2 is defined as in Definition **??** as the sum of squared deviations divided by $n-1$, it is easy to see that, if scores in a vector $\boldsymbol{x}$ are in deviation score form, then the sum of squared deviations is simply $\boldsymbol{x}^{*\prime}\boldsymbol{x}^{*}$, and the sample variance may be written

$$S_X^2 = 1/(n-1)\boldsymbol{x}^{*\prime}\boldsymbol{x}^{*} \tag{4.5}$$

If $\boldsymbol{x}$ is not in deviation score form, we may use the $\boldsymbol{Q}$ operator to convert it into deviation score form first. Hence, in general,

$$\begin{aligned} S_X^2 &= 1/(n-1)\boldsymbol{x}^{*\prime}\boldsymbol{x}^{*} \\ &= 1/(n-1)\boldsymbol{Qx}'\boldsymbol{Qx} \\ &= 1/(n-1)\boldsymbol{x}'\boldsymbol{Q}'\boldsymbol{Qx}, \end{aligned}$$

since the transpose of a product of two matrices is the product of their transposes in reverse order.

The expression can be reduced further. Since $\boldsymbol{Q}$ is symmetric, it follows immediately that $\boldsymbol{Q}' = \boldsymbol{Q}$, and (remembering also that $\boldsymbol{Q}$ is idempotent) that $\boldsymbol{Q}'\boldsymbol{Q} = \boldsymbol{Q}$. Hence

$$S_X^2 = 1/(n-1)\ \boldsymbol{x}'\boldsymbol{Qx}$$

As an obvious generalization of the above, we write the matrix form for the covariance between two vectors of scores $\boldsymbol{x}$ and $\boldsymbol{y}$ as

$$S_{XY} = 1/(n-1)\boldsymbol{x}'\boldsymbol{Qy}$$

Some times a surprising result is staring us right in the face, if we are only able to see it. Notice that the sum of cross products of deviation scores can be computed as

$$\begin{aligned} \boldsymbol{x}^{*\prime}\boldsymbol{y}^{*} &= (\boldsymbol{Qx})'(\boldsymbol{Qy}) \\ &= \boldsymbol{x}'\boldsymbol{Q}'\boldsymbol{Qy} \\ &= \boldsymbol{x}'\boldsymbol{Qy} \\ &= \boldsymbol{x}'(\boldsymbol{Qy}) \\ &= \boldsymbol{x}'\boldsymbol{y}^{*} \\ &= \boldsymbol{y}'\boldsymbol{x}^{*} \end{aligned}$$

Because products of the form $\boldsymbol{QQ}$ or $\boldsymbol{QQ'}$ can be collapsed into a single $\boldsymbol{Q}$, when computing the sum of cross products of deviation scores of two variables, one variable can be left in raw score form and the sum of cross products will remain the same! This surprising result is somewhat harder to see (and prove) using summation algebra.

In what follows, we will generally assume, unless explicitly stated otherwise, that our data matrices have been transformed to deviation score form. (The $\boldsymbol{Q}$ operator discussed above will accomplish this simultaneously for the case of scores of n subjects on several, say p, variates.) For example, consider a data matrix ${}_n\boldsymbol{X}_p$, whose p columns are the scores of n subjects on p different variables. If the columns of $\boldsymbol{X}$ are in raw score form, the matrix $\boldsymbol{X}^* = \boldsymbol{QX}$ will have p columns of deviation scores.

We shall concentrate on results in the case where $\boldsymbol{X}$ is in "column variate form," i.e., is an $n \times p$ matrix. Equivalent results may be developed for "row variate form" $p \times n$ data matrices which have the N scores on p variables arranged in p rows. The choice of whether to use row or column variate representations is arbitrary, and varies in books and articles. One must, ultimately, be equally fluent with either notation, although modern computer software tends to emphasize column variate form.

4.3.4 The Variance-Covariance Matrix

Consider the case in which we have n scores on p variables. We define the *variance-covariance matrix* $\boldsymbol{S_{xx}}$ to be a symmetric $p \times p$ matrix with element s_{ij} equal to the covariance between variable i and variable j. Naturally, the ith diagonal element of this matrix contains the covariance of variable i with itself, i.e., its variance. As a generalization of our results for a single vector of scores, the variance-covariance matrix may be written as follows. First, for raw scores in column variate form:

$$\boldsymbol{S_{xx}} = 1/(n-1)\boldsymbol{X'QX}$$

We obtain a further simplification if $\boldsymbol{X}$ is in deviation score form. In that case, we have:

$$\boldsymbol{S_{xx}} = 1/(n-1)\boldsymbol{X'X}$$

Note that some authors use the terms "variance-covariance matrix" and "covariance matrix" interchangeably.

4.3.5 The Correlation Matrix

For p variables in the data matrix X, the *correlation matrix* $\boldsymbol{R_{xx}}$ is a $p \times p$ symmetric matrix with typical element r_{ij} equal to the correlation between variables i and j. Of course, the diagonal elements of this matrix represent the correlation of a variable with itself, and are all equal to 1. Recall that all of the elements of the variance-covariance matrix $\boldsymbol{S_{xx}}$ are covariances, since the variances are covariances of variables with themselves. We know that, in order to convert s_{ij} (the covariance between variables i and j) to a correlation, we simply "standardize" it by dividing by the product of the standard deviations of variables i and j. This is very easy to accomplish in matrix notation.

Specifically, let $\boldsymbol{D_{xx}} = \operatorname{diag}(\boldsymbol{S_{xx}})$ be a diagonal matrix with ith diagonal element equal to the variance of the ith variable in $\boldsymbol{X}$. Then let $\boldsymbol{D}^{1/2}$ be a diagonal matrix with elements equal to standard deviations, and $\boldsymbol{D}^{-1/2}$ be a diagonal matrix with ith diagonal element equal to $1/s_i$, where s_i is the standard deviation of the ith variable. Then we may quickly verify that the correlation matrix is computed as:

$$\boldsymbol{R_{xx}} = \boldsymbol{D}^{-1/2}\boldsymbol{S_{xx}}\boldsymbol{D}^{-1/2}$$

4.3.6 The Covariance Matrix

Given ${}_N\boldsymbol{X}_m$ and ${}_N\boldsymbol{Y}_p$, two data matrices in deviation score form. The *covariance matrix* $\boldsymbol{S_{xy}}$ is a $m \times p$ matrix with element s_{ij} equal to the covariance between the ith variable in $\boldsymbol{X}$ and the jth variable in $\boldsymbol{Y}$. $\boldsymbol{S_{xy}}$ is computed as

$$\boldsymbol{S_{xy}} = 1/(n-1)\boldsymbol{X}'\boldsymbol{Y}$$

4.4 Variance of a Linear Combination

In an earlier section, we developed a summation algebra expression for evaluating the variance of a linear combination of variables. In this section, we derive the same result using matrix algebra.

We first note the following result.

Theorem 4.4.1 *Given $\mathbf{X}$, a data matrix in column variate deviation score form. For any linear composite $\boldsymbol{y} = \mathbf{X}\boldsymbol{b}$, $\boldsymbol{y}$ will also be in deviation score form.*

Proof *The variables in $\mathbf{X}$ are in deviation score form if and only if the sum of scores in each column is zero, i.e., $\mathbf{1}'\mathbf{X} = \mathbf{0}'$. But if $\mathbf{1}'\mathbf{X} = \mathbf{0}'$, then for any linear combination $\boldsymbol{y} = \mathbf{X}\boldsymbol{b}$, we have, immediately,*

$$\begin{aligned} \mathbf{1}'\boldsymbol{y} &= \mathbf{1}'\mathbf{X}\boldsymbol{b} \\ &= \left(\mathbf{1}'\mathbf{X}\right)\boldsymbol{b} \\ &= \mathbf{0}'\boldsymbol{b} \\ &= 0 \end{aligned}$$

Since, for any $\boldsymbol{b}$, the linear combination scores in $\boldsymbol{y}$ sum to zero, it must be in deviation score form.

We now give a result that is one of the cornerstones of multivariate statistics.

Theorem 4.4.2 (Variance of a Linear Combination) *Given X, a set of N deviation scores on p variables in column variate form, having variance-covariance matrix $\mathbf{S}_{xx}$. The variance of any linear combination $\boldsymbol{y} = \mathbf{X}\boldsymbol{b}$ may be computed as*

$$S_y^2 = \boldsymbol{b}'\mathbf{S}_{xx}\boldsymbol{b} \tag{4.6}$$

Proof *Suppose* $\boldsymbol{X}$ *is in deviation score form. Then, by Theorem 4.4.1, so must* $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b}$*, for any* $\boldsymbol{b}$*. From Equation 4.5, we know that*

$$\begin{aligned} S_y^2 &= 1/(N-1)\ \boldsymbol{y}'\boldsymbol{y} \\ &= 1/(N-1)\ (\boldsymbol{X}\boldsymbol{b})'(\boldsymbol{X}\boldsymbol{b}) \\ &= 1/(N-1)\ \boldsymbol{b}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} \\ &= \boldsymbol{b}'\ \left[1/(N-1)\ \boldsymbol{X}'\boldsymbol{X}\right]\boldsymbol{b} \\ &= \boldsymbol{b}'\boldsymbol{S_{xx}}\boldsymbol{b} \end{aligned}$$

This is a very useful result, as it allows to variance of a linear composite to be computed directly from the variance-covariance matrix of the original variables. This result may be extended immediately to obtain the variance-covariance *matrix* of a set of linear composites in a matrix $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B}$. The proof is not given as, it is a straightforward generalization of the previous proof.

Theorem 4.4.3 (Variance-Covariance Matrix of Several Linear Combinations) *Given X, a set of N deviation scores on p variables in column variate form, having variance-covariance matrix* $\boldsymbol{S_{xx}}$*. The variance-covariance matrix of any set of linear combinations* $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B}$ *may be computed as*

$$\boldsymbol{S_{YY}} = \boldsymbol{B}'\boldsymbol{S_{xx}}\boldsymbol{B} \tag{4.7}$$

In a similar manner, we may prove the following:

Theorem 4.4.4 (Covariance Matrix of Two Sets of Linear Combinations) *Given* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$*, two sets of n deviation scores on p and q variables in column variate form, having covariance matrix* $\boldsymbol{S_{xy}}$*. The covariance matrix of any two sets of linear combinations* $\boldsymbol{W} = \boldsymbol{X}\boldsymbol{B}$ *and* $\boldsymbol{M} = \boldsymbol{Y}\boldsymbol{C}$ *may be computed as*

$$\boldsymbol{S_{wm}} = \boldsymbol{B}'\boldsymbol{S_{xy}}\boldsymbol{C} \tag{4.8}$$

4.5 Trace of a Square Matrix

Traces play an important role in regression analysis and other fields of statistics.

Definition (Trace of a Square Matrix) The trace of a square matrix $\boldsymbol{A}$, denoted $\mathrm{Tr}\,(\boldsymbol{A})$, is defined as the sum of its diagonal elements, i.e.,

$$\mathrm{Tr}\,(\boldsymbol{A}) = \sum_{i=1}^{N} a_{ii}$$

We may verify that the trace has the following properties:

(a) $\mathrm{Tr}\,(\boldsymbol{A}+\boldsymbol{B}) = \mathrm{Tr}\,(\boldsymbol{A}) + \mathrm{Tr}\,(\boldsymbol{B})$

(b) $\mathrm{Tr}\,(\boldsymbol{A}) = \mathrm{Tr}\,(\boldsymbol{A}')$

(c) $\mathrm{Tr}\,(c\boldsymbol{A}) = c\,\mathrm{Tr}\,(\boldsymbol{A})$

(d) $\mathrm{Tr}\,(\boldsymbol{A}'\boldsymbol{B}) = \sum_i \sum_j a_{ij} b_{ij}$

(e) $\mathrm{Tr}\,(\boldsymbol{E}'\boldsymbol{E}) = \sum_i \sum_j e_{ij}^2$

(f) The "cyclic permutation rule":

$$\mathrm{Tr}\,(\boldsymbol{ABC}) = \mathrm{Tr}\,(\boldsymbol{CAB}) = \mathrm{Tr}\,(\boldsymbol{BCA})$$

4.6 Random Vectors and Random Matrices

In this section, we extend our results on linear combinations of variables to *random vector* notation. The generalization is straightforward.

Definition (Random Vector) A *random vector* $\boldsymbol{\xi}$ is a vector whose elements are random variables.

One (informal) way of thinking of a random variable is that it is a process that generates numbers according to some law. An analogous way of thinking of a random vector is that it produces a vector of numbers according to some law.

Definition (Random Matrix) A *random matrix* $\boldsymbol{\Psi}$ is a matrix whose elements are random variables.

Definition (Expected Value of a Random Vector) The expected value of a random vector (or matrix) is a vector (or matrix) whose elements are the expected values of the individual random variables that are the elements of the random vector.

Example 4.6.1 (Expected Value of a Random Vector) Suppose, for example, we have two random variables x and y, and their expected values are 0 and 2, respectively. If we put these variables into a vector $\boldsymbol{\xi}$, it follows that

$$\mathcal{E}(\boldsymbol{\xi}) = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

Definition (Variance-Covariance Matrix of a Random Vector) Given a random vector $\boldsymbol{\xi}$ with expected value $\boldsymbol{\mu}$, the variance-covariance matrix $\boldsymbol{\Sigma}_{\xi\xi}$ is defined as

$$\begin{aligned} \boldsymbol{\Sigma}_{\xi\xi} &= \mathcal{E}(\boldsymbol{\xi}-\boldsymbol{\mu})(\boldsymbol{\xi}-\boldsymbol{\mu})' && (4.9) \\ &= \mathcal{E}(\boldsymbol{\xi}\boldsymbol{\xi}') - \boldsymbol{\mu}\boldsymbol{\mu}' && (4.10) \end{aligned}$$

If $\boldsymbol{\xi}$ is a deviation score random vector, then

$$\boldsymbol{\Sigma}_{\xi\xi} = \mathcal{E}(\boldsymbol{\xi}\boldsymbol{\xi}')$$

The preceding definition is frequently is confusing to beginners. Let's "concretize" it a bit by giving an example with just two variables.

Example 4.6.2 (Variance-Covariance Matrix) Suppose

$$\boldsymbol{\xi} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

and

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

Note that $\boldsymbol{\xi}$ contains random variables, while $\boldsymbol{\mu}$ contains constants. Computing $\mathcal{E}(\boldsymbol{\xi}\boldsymbol{\xi}')$, we find

$$\begin{aligned} \mathcal{E}\left(\boldsymbol{\xi}\boldsymbol{\xi}'\right) &= \mathcal{E}\left(\begin{bmatrix} x_1 x_2 \end{bmatrix}\begin{bmatrix} x_1 & x_2 \end{bmatrix}\right) \\ &= \mathcal{E}\left(\begin{bmatrix} x_1^2 & x_1 x_2 \\ x_2 x_1 & x_2^2 \end{bmatrix}\right) \\ &= \begin{bmatrix} \mathcal{E}(x_1^2) & \mathcal{E}(x_1 x_2) \\ \mathcal{E}(x_2 x_1) & \mathcal{E}(x_2^2) \end{bmatrix} \end{aligned} \tag{4.11}$$

In a similar vein, we find that

$$\begin{aligned} \boldsymbol{\mu}\boldsymbol{\mu}' &= \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}\begin{bmatrix} \mu_1 & \mu_2 \end{bmatrix} \\ &= \begin{bmatrix} \mu_1^2 & \mu_1\mu_2 \\ \mu_2\mu_1 & \mu_2^2 \end{bmatrix} \end{aligned} \tag{4.12}$$

Subtracting Equation 4.12 from Equation 4.11, and recalling Equation 3.2, we find

$$
\begin{aligned}
\mathcal{E}(\boldsymbol{\xi}\boldsymbol{\xi}') - \boldsymbol{\mu}\boldsymbol{\mu}' &= \begin{bmatrix} \mathcal{E}(x_1^2) - \mu_1^2 & \mathcal{E}(x_1 x_2) - \mu_1\mu_2 \\ \mathcal{E}(x_2 x_1) - \mu_2\mu_1 & \mathcal{E}(x_2^2) - \mu_2^2 \end{bmatrix} \\
&= \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}
\end{aligned}
$$

Definition (Covariance Matrix for Two Random Vectors) Given two random vectors $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$, their covariance matrix $\boldsymbol{\Sigma}_{\xi\eta}$ is defined as

$$
\begin{aligned}
\boldsymbol{\Sigma}_{\xi\eta} &= \mathcal{E}\left(\boldsymbol{\xi}\boldsymbol{\eta}'\right) - \mathcal{E}(\boldsymbol{\xi})\mathcal{E}(\boldsymbol{\eta}') \qquad (4.13) \\
&= \mathcal{E}\left(\boldsymbol{\xi}\boldsymbol{\eta}'\right) - \mathcal{E}(\boldsymbol{\xi})\mathcal{E}(\boldsymbol{\eta})' \qquad (4.14)
\end{aligned}
$$

We now present some key results involving the "expected value algebra" of random matrices and vectors.

As a generalization of results we presented in scalar algebra, we find that, for a matrix of constants $\boldsymbol{B}$, and a random vector $\boldsymbol{x}$,

$$
\mathcal{E}\left(\boldsymbol{B}'\boldsymbol{x}\right) = \boldsymbol{B}'\mathcal{E}(\boldsymbol{x}) = \boldsymbol{B}'\boldsymbol{\mu}
$$

For random vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, we find

$$
\mathcal{E}\left(\boldsymbol{x} + \boldsymbol{y}\right) = \mathcal{E}(\boldsymbol{x}) + \mathcal{E}(\boldsymbol{y})
$$

The result obviously generalizes to the expected value of the difference of random vectors.

Some key implications of the preceding two results, which are especially useful for reducing matrix algebra expressions, are the following:

(a) The expected value operator distributes over addition and/or subtraction of random vectors and matrices.

(b) The parentheses of an expected value operator can be "moved through" multiplied matrices or vectors of constants from both the left and right of any term, until the first random vector or matrix is encountered.

(c) $\mathcal{E}(\boldsymbol{x}') = (\mathcal{E}(\boldsymbol{x}))'$

Example 4.6.3 (Expected Value Algebra) As an example of Result 4.6, we reduce the following expression. Suppose the Greek letters are random vectors with zero expected value, and the matrices contain constants. Then

$$\begin{aligned} \mathcal{E}\left(\boldsymbol{A}'\boldsymbol{B}'\boldsymbol{\eta}\boldsymbol{\xi}'\boldsymbol{C}\right) &= \boldsymbol{A}'\boldsymbol{B}'\mathcal{E}\left(\boldsymbol{\eta}\boldsymbol{\xi}'\right)\boldsymbol{C} \\ &= \boldsymbol{A}'\boldsymbol{B}'\boldsymbol{\Sigma}_{\eta\xi}\boldsymbol{C} \end{aligned}$$

Theorem 4.6.4 (Variance-Covariance Matrix of Linear Combinations) *Given $\boldsymbol{x}$, a random vector with p variables, having variance-covariance matrix $\boldsymbol{\Sigma}_{xx}$. The variance-covariance matrix of any set of linear combinations $\boldsymbol{y} = \boldsymbol{B}'\boldsymbol{x}$ may be computed as*

$$\boldsymbol{\Sigma}_{yy} = \boldsymbol{B}'\boldsymbol{\Sigma}_{xx}\boldsymbol{B} \tag{4.15}$$

In a similar manner, we may prove the following:

Theorem 4.6.5 (Covariance Matrix of Two Sets of Linear Combinations) *Given $\boldsymbol{x}$ and $\boldsymbol{y}$, two random vectors with p and q variables having covariance matrix $\boldsymbol{\Sigma}_{xy}$. The covariance matrix of any two sets of linear combinations $\boldsymbol{w} = \boldsymbol{B}'\boldsymbol{x}$ and $\boldsymbol{m} = \boldsymbol{C}'\boldsymbol{y}$ may be computed as*

$$\boldsymbol{\Sigma}_{wm} = \boldsymbol{B}'\boldsymbol{\Sigma}_{xy}\boldsymbol{C} \tag{4.16}$$

4.7 Inverse of a Square Matrix

For an $n \times n$ square matrix $\boldsymbol{A}$, the *inverse* of $\boldsymbol{A}$, bmA^{-1}, exists if and only if $\boldsymbol{A}$ is of full rank, i.e., if and only if no column of A is a linear combination of the others. $\boldsymbol{A}^{-1}$ is the unique matrix that satisfies

$$\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$$

If a square matrix $\boldsymbol{A}$ has an inverse, we say that $\boldsymbol{A}$ is "invertible," "nonsingular," and "of full rank." If the transpose of a matrix is its inverse, and vice versa, i.e., $\boldsymbol{A}\boldsymbol{A}' = \boldsymbol{A}'\boldsymbol{A} = \boldsymbol{I}$, we say that the matrix $\boldsymbol{A}$ is *orthogonal*.

$\boldsymbol{A}^{-1}$ has the following properties:

(a) $(\boldsymbol{A}')^{-1} = (\boldsymbol{A}^{-1})'$

(b) If $\boldsymbol{A} = \boldsymbol{A}'$, then $\boldsymbol{A}^{-1} = (\boldsymbol{A}^{-1})'$

(c) The inverse of the product of several invertible square matrices is the product of their inverses in reverse order. For example

$$(\boldsymbol{ABC})^{-1} = \boldsymbol{C}^{-1}\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$$

(d) For nonzero scalar c, $(c\boldsymbol{A})^{-1} = (1/c)\boldsymbol{A}^{-1}$

(e) For diagonal matrix $\boldsymbol{D}$, $\boldsymbol{D}^{-1}$ is a diagonal matrix with diagonal elements equal to the reciprocal of the corresponding diagonal elements of $\boldsymbol{D}$.

4.8 Bilinear and Quadratic Forms, Rank, and Linear Dependence

In this section, we discuss a number of terms related to the *rank* of a matrix, a concept that is closely related to the dimensionality of a linear model.

Definition (Bilinear Form) For a matrix $\boldsymbol{A}$, a *bilinear form* is a scalar expression of the form

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

for vectors $\boldsymbol{b}$ and bmc.

Definition (Quadratic Form) A *quadratic form* is a scalar expression of the form

$$\boldsymbol{b}'\boldsymbol{A}\boldsymbol{b}$$

for vector $\boldsymbol{b}$.

Definition (Positive Definite Matrix) A matrix $\boldsymbol{A}$ is *positive definite* if for any non-null $\boldsymbol{b}$, the quadratic form $\boldsymbol{b}'\boldsymbol{A}\boldsymbol{b}$ is greater than zero. Similarly, we say that $\boldsymbol{A}$ is *positive semidefinite* if for any non-null $\boldsymbol{b}$, the quadratic form $\boldsymbol{b}'\boldsymbol{A}\boldsymbol{b}$ is greater than or equal to zero.

Definition (Linear Independence and Dependence) A set of p-component vectors $\boldsymbol{x}_i$, $i = 1, \dots, p$ is said to be *linearly independent* if there exists no (non-zero) set of linear weights b_i such that

$$\sum_{i=1}^{p} b_i \boldsymbol{x}_i = \boldsymbol{0}$$

If a set of vectors is not linearly independent, then the vectors are *linearly dependent.*

Example 4.8.1 (Linear Dependence) Let $\boldsymbol{x}' = \begin{bmatrix} 1 & -1 & 2 \end{bmatrix}$, $\boldsymbol{y}' = \begin{bmatrix} 2 & 0 & -1 \end{bmatrix}$, and $\boldsymbol{z}' = \begin{bmatrix} 0 & -2 & 5 \end{bmatrix}$. Then $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$ are linearly dependent because $\boldsymbol{z} - 2\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{0}$.

Definition (Rank of a Matrix) The *rank* of a matrix $\boldsymbol{A}$, denoted "Rank $(\boldsymbol{A})$," is the maximal number of linearly independent rows or columns in $\boldsymbol{A}$.

Comment. In some theoretical discussions, reference is made to *row rank* and *column rank.* The *row rank* of a matrix is equal to the number of linearly independent rows in the matrix. The *column rank* of a matrix is equal to its number of linearly independent columns. For any matrix $\boldsymbol{A}$, row rank and column rank are equal. If the row rank of a matrix is equal to its number of rows, we say the matrix is of *full row rank*, and if its row rank is less than its number of rows, we say it is of *deficient row rank*. Analogous terminology holds for columns. If a matrix is square, we refer to it as *full rank* or *deficient rank*, depending on whether or not it has the maximum possible number of linearly independent rows and columns.

Here are some key properties of rank. For any matrix $\boldsymbol{A}$, we have

(a) $\text{Rank}\left(\boldsymbol{A}'\right) = \text{Rank}\left(\boldsymbol{A}\right)$.

(b) $\text{Rank}\left(\boldsymbol{A}'\boldsymbol{A}\right) = \text{Rank}\left(\boldsymbol{A}\boldsymbol{A}'\right) = \text{Rank}\left(\boldsymbol{A}\right)$.

(c) For any conformable matrix $\boldsymbol{B}$, $\text{Rank}\left(\boldsymbol{AB}\right) \leq \min\left(\text{Rank}\left(\boldsymbol{A}\right), \text{Rank}\left(\boldsymbol{B}\right)\right)$.

(d) For any nonsingular, square matrix $\boldsymbol{B}$, $\text{Rank}\left(\boldsymbol{AB}\right) = \text{Rank}\left(\boldsymbol{A}\right)$.

4.9 Determinant of a Square Matrix

In this section, we define the *determinant* of a square matrix, and explore its properties. The determinant of a matrix $\boldsymbol{A}$ is a scalar function that is zero if and only if a matrix is of deficient rank. This fact is sufficient information about the determinant to allow the reader to continue through much of the remainder of this book. The remainder of this section is presented primarily for mathematical completeness, and may be omitted on first reading.

Definition (Determinant) The *determinant* of a square matrix $\boldsymbol{A}$, denoted $\text{Det}\left(A\right)$ or $|\boldsymbol{A}|$, is defined as follows:

(a) For $N \times N$ matrix $\boldsymbol{A}$, form all possible products of N elements of the matrix such that no two elements in any product are from the same row or column of A. There are $N!$ such products available. For example, when $N = 3$, the required products are the 6 quantities, $a_{11}a_{22}a_{33}$, $a_{12}a_{23}a_{31}$, $a_{13}a_{21}a_{32}$, $a_{13}a_{22}a_{31}$, $a_{11}a_{23}a_{32}$, $a_{12}a_{21}a_{33}$.

(b) Within each product, arrange the factors so that the row subscripts are in natural order $1, 2, \ldots, N$, as in the example above.

(c) Then examine the order in which the column subscripts appear. Specifically, note how many times a larger number precedes a smaller number in the sequence of column subscripts for the product. For product p_i, call this number the "number of inversions" k_i.

(d) The determinant is computed as

$$|\boldsymbol{A}| = \sum_{i=1}^{N!} p_i(-1)^{k_i} \tag{4.17}$$

Example 4.9.1 (Some Simple Determinants) Applying the above formula to a 3×3 matrix, we obtain

$$\begin{aligned} |A| &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ &\quad -a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} \end{aligned} \tag{4.18}$$

For a 2×2 matrix, we have

$$|A| = a_{11}a_{22} - a_{12}a_{21} \tag{4.19}$$

Clearly, for $N > 3$, calculating the determinant using the basic formula would be exceedingly tedious. However, if we define M_{ij} as the determinant of the matrix of order $(N-1) \times (N-1)$ obtained by crossing out the ith row and jth column of $\boldsymbol{A}$, we find that, for any row i,

$$|A| = \sum_{j=1}^{N} a_{ij}M_{ij}(-1)^{i+j} \tag{4.20}$$

M_{ij} is referred to as a *minor* of $\boldsymbol{A}$. If we define the *cofactor* $C_{ij} = (-1)^{i+j}M_{ij}$, then we find that for any row i,

$$|A| = \sum_{j=1}^{N} a_{ij}C_{ij}, \tag{4.21}$$

Also, for any column j,

$$|A| = \sum_{i=1}^{N} a_{ij}C_{ij} \tag{4.22}$$

Here are some important properties of determinants. For any $N \times N$ matrix $\boldsymbol{A}$,

(a) The determinant of a diagonal matrix or a triangular matrix is the product of its diagonal elements.

(b) If the elements of a single row or column of $\boldsymbol{A}$ multiplied by the scalar c, the determinant of the new matrix is equal to $c|\boldsymbol{A}|$. If every element of $\boldsymbol{A}$ is multiplied by c, then $|c\boldsymbol{A}| = cN|\boldsymbol{A}|$.

(c) If two columns (or rows) of $\boldsymbol{A}$ are interchanged, the sign of $|\boldsymbol{A}|$ is reversed.

(d) If two columns or two rows of a matrix are equal, then $|\boldsymbol{A}| = 0$.

(e) The determinant of a matrix is unchanged if a multiple of some column is added to another column. A similar property holds for rows.

(f) If all elements of a row or column of $\boldsymbol{A}$ are zero, then $|\boldsymbol{A}| = 0$.

(g) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are both $N \times N$, then $|\boldsymbol{AB}| = |\boldsymbol{A}||\boldsymbol{B}|$.

(h) The sum of the products of the elements of a given row of a square matrix with the corresponding cofactors of a different row is equal to zero. A similar result holds for columns.

4.10 Eigenvalues and Eigenvectors of a Square Matrix

The *eigenvalues* and *eigenvectors* of a square matrix play a key role in some important operations in statistics. In particular, they are intimately connected with the determination of the rank of a matrix, and the "factoring" of a matrix into a product of matrices.

Definition (Eigenvalues and Eigenvectors) For a square matrix $\boldsymbol{A}$, a scalar c and a vector $\boldsymbol{v}$ are an *eigenvalue* and associated *eigenvector*, respectively, if and only if they satisfy the equation,

$$\boldsymbol{Av} = c\boldsymbol{v} \tag{4.23}$$

Comment. There are infinitely many solutions to Equation 4.23 unless some identification constraint is placed on the size of vector $\boldsymbol{v}$. For example for any c and $\boldsymbol{v}$ satisfying the equation, $c/2$ and $2\boldsymbol{v}$ must also satisfy the same equation. Consequently in eigenvectors are assumed to be "normalized," i.e., satisfy the constraint that $\boldsymbol{v}'\boldsymbol{v} = 1$. Eigenvalues c_i are roots to the determinantal equation

$$|\boldsymbol{A} - c\boldsymbol{I}| = 0 \tag{4.24}$$

Here are some key properties of eigenvalues and eigenvectors. For $n \times n$ matrix $\boldsymbol{A}$ with eigenvalues c_i and associated eigenvectors $\boldsymbol{v}_i$,

(a)

$$\text{Tr}\,(\boldsymbol{A}) = \sum_{i=1}^{n} c_i$$

(b)

$$|\boldsymbol{A}| = \prod_{i=1}^{n} c_i$$

(c) Eigenvalues of a symmetric matrix with real elements are all real.

(d) Eigenvalues of a positive definite matrix are all positive.

(e) If a $n \times n$ symmetric matrix $\boldsymbol{A}$ is positive semidefinite and of rank r, it has exactly r positive eigenvalues and $p - r$ zero eigenvalues.

(f) The nonzero eigenvalues of the product $\boldsymbol{AB}$ are equal to the nonzero eigenvalues of $\boldsymbol{BA}$. Hence the traces of $\boldsymbol{AB}$ and $\boldsymbol{BA}$ are equal.

(g) The characteristic roots of a diagonal matrix are its diagonal elements.

(h) The scalar multiple $b\boldsymbol{A}$ has eigenvalue bc_i with eigenvector $\boldsymbol{v}_i$. (Proof: $\boldsymbol{A}\boldsymbol{v}_i = c_i\boldsymbol{v}_i$ implies immediately that $(b\boldsymbol{A})\boldsymbol{v}_i = (bc_i)\boldsymbol{v}_i$.)

(i) Adding a constant b to every diagonal element of $\boldsymbol{A}$ creates a matrix $\boldsymbol{A} + b\boldsymbol{I}$ with eigenvalues $c_i + b$ and associated eigenvectors $\boldsymbol{v}_i$. (Proof: $(\boldsymbol{A} + b\boldsymbol{I})\boldsymbol{v}_i = \boldsymbol{A}\boldsymbol{v}_i + b\boldsymbol{v}_i = c_i\boldsymbol{v}_i + b\boldsymbol{v}_i = (c_i + b)\boldsymbol{v}_i$).

(j) $\boldsymbol{A}^m$ has c_i^m as an eigenvalue, and $\boldsymbol{v}_i$ as its eigenvector. Proof: Consider $\boldsymbol{A}^2\boldsymbol{v}_i = \boldsymbol{A}(\boldsymbol{A}\boldsymbol{v}_i) = \boldsymbol{A}(\boldsymbol{c}_i\boldsymbol{v}_i) = c_i(\boldsymbol{A}\boldsymbol{v}_i) = c_ic_i\boldsymbol{v}_i = c_i^2\boldsymbol{v}_i$. The general case follows by induction.

(k) $\boldsymbol{A}^{-1}$, if it exists, has $1/c_i$ as an eigenvalue, and $\boldsymbol{v}_i$ as its eigenvector. Proof: $\boldsymbol{A}\boldsymbol{v}_i = c_i\boldsymbol{v}_i = \boldsymbol{v}_ic$. $\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{v}_i = \boldsymbol{v}_i = \boldsymbol{A}^{-1}\boldsymbol{v}_ic_i$. $\boldsymbol{v}_i = \boldsymbol{A}^{-1}\boldsymbol{v}_ic_i = c_i\boldsymbol{A}^{-1}\boldsymbol{v}_i$. So $(1/c_i)\boldsymbol{v}_i = \boldsymbol{A}^{-1}\boldsymbol{v}_i$.

(l) For symmetric $\boldsymbol{A}$, for distinct eigenvalues c_i, c_j with associated eigenvectors $\boldsymbol{v}_i$, $\boldsymbol{v}_j$, we have $\boldsymbol{v}_i'\boldsymbol{v}_j = 0$. Proof: $\boldsymbol{A}\boldsymbol{v}_i = c_i\boldsymbol{v}_i$, and $\boldsymbol{A}\boldsymbol{v}_j = c_j\boldsymbol{v}_j$. So $\boldsymbol{v}_j'\boldsymbol{A}\boldsymbol{v}_i = c_i\boldsymbol{v}_j'\boldsymbol{v}_i$ and $\boldsymbol{v}_i'\boldsymbol{A}\boldsymbol{v}_j = c_j\boldsymbol{v}_i'\boldsymbol{v}_j$. But, since a bilinear form is a scalar, it is equal to its transpose, and, remembering that $\boldsymbol{A} = \boldsymbol{A}'$, $\boldsymbol{v}_j'\boldsymbol{A}\boldsymbol{v}_i = \boldsymbol{v}_i'\boldsymbol{A}\boldsymbol{v}_j$. So $c_i\boldsymbol{v}_j'\boldsymbol{v}_i = c_j\boldsymbol{v}_i'\boldsymbol{v}_j = c_j\boldsymbol{v}_j'\boldsymbol{v}_i$. If c_i and c_j are different, this implies $\boldsymbol{v}_j'\boldsymbol{v}_i = 0$.

(m) For any real, symmetric $\boldsymbol{A}$, there exists a $\boldsymbol{V}$ such that $\boldsymbol{V}'\boldsymbol{A}\boldsymbol{V} = \boldsymbol{D}$,where $\boldsymbol{D}$ is diagonal. Moreover, any real,symmetric matrix $\boldsymbol{A}$ can be written as $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}'$,where $\boldsymbol{V}$ contains the eigenvectors $\boldsymbol{v}_i$ of $\boldsymbol{A}$ in order in its columns, and D contains the eigenvalues c_i of $\boldsymbol{A}$ in the ith diagonal position.

4.11 Applications of Eigenvalues and Eigenvectors

Eigenvalues and eigenvectors have widespread practical application in multivariate statistics. In this section, we demonstrate a few such applications. First, we deal with the notion of *matrix factorization.*

Definition (Powers of a Diagonal Matrix) Diagonal matrices act much more like scalars than most matrices do. For example, we can define fractional powers of diagonal matrices, as well as positive powers. Specifically, if diagonal matrix $\boldsymbol{D}$ has diagonal elements d_i, the matrix $\boldsymbol{D}^x$ has elements d_i^x. If x is negative, it is assumed $\boldsymbol{D}$ is positive definite. With this definition, the powers of $\boldsymbol{D}$ behave essentially like scalars. For example, $\boldsymbol{D}^{1/2}\boldsymbol{D}^{1/2} = \boldsymbol{D}$.

Example 4.11.1 (Powers of a Diagonal Matrix) Suppose we have

$$\boldsymbol{D} = \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix}$$

Then

$$\boldsymbol{D}^{1/2} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

Example 4.11.2 (Matrix Factorization) Suppose you have a variance-covariance matrix $\boldsymbol{\Sigma}$ for some statistical population. Assuming $\boldsymbol{\Sigma}$ is positive semidefinite, then (from Result 4.10), it can be written in the form $\boldsymbol{\Sigma} = \boldsymbol{VDV}' = \boldsymbol{FF}'$, where $\boldsymbol{F} = \boldsymbol{VD}^{1/2}$. $\boldsymbol{F}$ is called a "Gram-factor of $\boldsymbol{\Sigma}$."

Comment. Gram-factors are not, in general, uniquely defined. For example, suppose $\boldsymbol{\Sigma} = \boldsymbol{FF}'$. Then consider any orthogonal matrix $\boldsymbol{T}$, such that $\boldsymbol{TT}' = \boldsymbol{T}'\boldsymbol{T} = \boldsymbol{I}$. There are infinitely many orthogonal matrices of order 2×2 and higher. Then for any such matrix $\boldsymbol{T}$, we have $\boldsymbol{\Sigma} = \boldsymbol{FTT}'\boldsymbol{F}' = \boldsymbol{F}^*\boldsymbol{F}^{*\prime}$, where $\boldsymbol{F}^* = \boldsymbol{FT}$.

Gram-factors have some significant applications. For example, in the field of random number generation, it is relatively easy to generate pseudo-random numbers that mimic p variables that are independent with zero mean and unit variance. But suppose we wish to mimic p variables that are not independent, but have variance-covariance matrix $\boldsymbol{\Sigma}$? The following example describes one method for doing this.

Example 4.11.3 (Simulating Nonindependent Random Numbers) Given $p \times 1$ random vector $\boldsymbol{x}$ having variance-covariance matrix $\boldsymbol{I}$. Let $\boldsymbol{F}$ be a Gram-factor of $\boldsymbol{\Sigma} = \boldsymbol{FF}'$. Then $\boldsymbol{y} = \boldsymbol{Fx}$ will have variance-covariance matrix $\boldsymbol{\Sigma}$.

In certain intermediate and advanced derivations in matrix algebra, reference is made to "symmetric powers" of a symmetric matrix $\boldsymbol{A}$, in particular the "symmetric square root" of $\boldsymbol{A}$, a symmetric matrix which, when multiplied by itself, yields $\boldsymbol{A}$.

Example 4.11.4 (Symmetric Powers of a Symmetric Matrix) When investigating properties of eigenvalues and eigenvectors, we pointed out that, for distinct eigenvalues of a symmetric matrix $\boldsymbol{A}$, the associated eigenvectors are orthogonal. Since the eigenvectors are normalized to have a sum of squares equal to 1, it follows that if we place the eigenvectors in a matrix $\boldsymbol{V}$, this matrix will be orthogonal, i.e. $\boldsymbol{V}\boldsymbol{V}' = \boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}$. This fact allows us to create "symmetric powers" of a symmetric matrix very efficiently if we know the eigenvectors. For example, suppose you wish to create a symmetric matrix $\boldsymbol{A}^{1/2}$ such that $\boldsymbol{A}^{1/2}\boldsymbol{A}^{1/2} = \boldsymbol{A}$. Let diagonal matrix $\boldsymbol{D}$ contain the eigenvalues of $\boldsymbol{A}$ in proper order. Then $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}'$, and it is easy to verify that $\boldsymbol{A}^{1/2} = \boldsymbol{V}\boldsymbol{D}^{1/2}\boldsymbol{V}'$ has the required properties. To prove that $\boldsymbol{A}^{1/2}$ is symmetric, we need simply show that it is equal to its transpose, which is trivial (so long as you recall that any diagonal matrix is symmetric, and that the transpose of a product of several matrices is the product of the transposes in reverse order). That $\boldsymbol{A}^{1/2}\boldsymbol{A}^{1/2} = \boldsymbol{A}$ follows immediately by substitution, i.e.,

$$
\begin{aligned}
\boldsymbol{A}^{1/2}\boldsymbol{A}^{1/2} &= \boldsymbol{V}\boldsymbol{D}^{1/2}\boldsymbol{V}'\boldsymbol{V}\boldsymbol{D}^{1/2}\boldsymbol{V}' \\
&= \boldsymbol{V}\boldsymbol{D}^{1/2}\left[\boldsymbol{V}'\boldsymbol{V}\right]\boldsymbol{D}^{1/2}\boldsymbol{V}' \\
&= \boldsymbol{V}\boldsymbol{D}^{1/2}\left[\boldsymbol{I}\right]\boldsymbol{D}^{1/2}\boldsymbol{V}' \\
&= \boldsymbol{V}\boldsymbol{D}^{1/2}\boldsymbol{D}^{1/2}\boldsymbol{V}' \\
&= \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}'
\end{aligned}
$$