2 Review of matrix algebra

2.1 Introduction

Before we begin our discussion of the statistical models and methods, we review elements of matrix algebra that will be quite useful in streamlining our presentation and representing data. Here, we will note some basic results and operations. Further results and definitions will be discussed as we need them throughout the course. Many useful facts here are stated systematically in this chapter; thus, this chapter will serve as a reference for later developments using matrix notation.

2.2 Matrix notation

MATRIX: A rectangular array of numbers, e.g.

$$\boldsymbol{A} = \begin{pmatrix} 3 & 5 & 7 & 8 \\ 1 & 2 & 3 & 7 \end{pmatrix}$$

As is standard, we will use boldface capital letters to denote an entire matrix.

DIMENSION: A matrix with r rows and c columns is said to be of **dimension** $(r \times c)$.

It is customary to refer generically to the elements of a matrix by using 2 subscripts, e.g.

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \end{pmatrix}$$

$a_{11} = 3$, $a_{12} = 5$, etc. In general, for a matrix with r rows and c columns, $\boldsymbol{A}$, the element of $\boldsymbol{A}$ in the ith row and the jth column is denoted as a_{ij}, where $i = 1, \ldots, r$ and $j = 1, \ldots, c$.

VECTOR: A column vector is a matrix with only one column, e.g.

$$\boldsymbol{a} = \begin{pmatrix} 2 \\ 0 \\ 3 \\ -2 \end{pmatrix}$$

A row vector is matrix with only one row, e.g.

$$\boldsymbol{b} = \begin{pmatrix} 1, & 3, & -5 \end{pmatrix}$$

It is worth noting some special cases of matrices.

SQUARE MATRIX: A matrix with $r = c$, that is, with the same number of rows and columns is called a **square matrix**. If a matrix $\boldsymbol{A}$ is square, the elements a_{ii} are said to lie on the (principal) **diagonal** of $\boldsymbol{A}$. For example,

$$\boldsymbol{A} = \begin{pmatrix} 4 & 0 & 7 \\ 9 & -1 & 3 \\ -8 & 4 & 5 \end{pmatrix}.$$

SYMMETRIC MATRIX: A square matrix $\boldsymbol{A}$ is called **symmetric** if $a_{ij} = a_{ji}$ for all values of i and j. The term symmetric refers to the fact that such a matrix "reflects" across its diagonal, e.g.

$$\boldsymbol{A} = \begin{pmatrix} 3 & 5 & 7 \\ 5 & 1 & 4 \\ 7 & 4 & 8 \end{pmatrix}$$

Symmetric matrices turn out to be quite important in formulating statistical models for all types of data!

IDENTITY MATRIX: An important special case of a square, symmetric matrix is the **identity** matrix – a square matrix with 1's on diagonal, 0's elsewhere, e.g.

$$\boldsymbol{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

As we will see shortly, the identity matrix functions the same way as "1" does in the real number system.

TRANSPOSE: The **transpose** of any $(r \times c)$ $\boldsymbol{A}$ matrix is the $(c \times r)$ matrix denoted as $\boldsymbol{A}'$ such that a_{ij} is replaced by a_{ji} everywhere. That is, the transpose of $\boldsymbol{A}$ is the matrix found by "flipping" the matrix around, e.g.

$$\boldsymbol{A} = \begin{pmatrix} 3 & 5 & 7 & 8 \\ 1 & 2 & 3 & 7 \end{pmatrix}, \quad \boldsymbol{A}' = \begin{pmatrix} 3 & 1 \\ 5 & 2 \\ 7 & 3 \\ 8 & 7 \end{pmatrix}$$

A fundamental property of a symmetric matrix is that the matrix and its transpose are the **same**; i.e., if $\boldsymbol{A}$ is symmetric then $\boldsymbol{A} = \boldsymbol{A}'$. (Try it on the symmetric matrix above.)

2.3 Matrix operations

The world of matrices can be thought of as an extension of the world of real (scalar) numbers. Just as we add, subtract, multiply, and divide real numbers, we can do the same in with matrices. It turns out that these operations make the expression of complicated calculations easy to talk about and express, hiding all the details!

MATRIX ADDITION AND SUBTRACTION: Adding or subtracting two matrices are operations that are defined **element-by-element**. That is, to add to matrices, add their corresponding elements, e.g.

$$\boldsymbol{A} = \begin{pmatrix} 5 & 0 \\ -3 & 2 \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} 6 & 4 \\ 2 & -1 \end{pmatrix}$$

$$\boldsymbol{A} + \boldsymbol{B} = \begin{pmatrix} 11 & 4 \\ -1 & 1 \end{pmatrix}, \quad \boldsymbol{A} - \boldsymbol{B} = \begin{pmatrix} -1 & -4 \\ -5 & 3 \end{pmatrix}$$

Note that these operations only make sense if the two matrices have the **same dimension** – the operations are not defined otherwise.

MULTIPLICATION BY A CONSTANT: The effect of multiplying a matrix $\boldsymbol{A}$ of any dimension by a real number (scalar) b, say, is to multiply each element in $\boldsymbol{A}$ by b. This is easy to see by considering that this is just equivalent to adding $\boldsymbol{A}$ to itself b times. E.g.

$$3 \begin{pmatrix} 5 & -2 \\ 6 & 4 \end{pmatrix} = \begin{pmatrix} 15 & -6 \\ 18 & 12 \end{pmatrix}.$$

GENERAL FACTS:

- $\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}$, $b(\boldsymbol{A} + \boldsymbol{B}) = b\boldsymbol{A} + b\boldsymbol{B}$
- $(\boldsymbol{A} + \boldsymbol{B})' = \boldsymbol{A}' + \boldsymbol{B}'$, $(b\boldsymbol{A})' = b\boldsymbol{A}'$

MATRIX MULTIPLICATION: This operation is a bit tricky, but as we will see in a moment, it proves most powerful for expressing a whole series of calculations in a very simple way.

- *Order matters*
- Number of columns of first matrix *must* = Number of rows of second matrix, e.g.

$$\boldsymbol{A} = \begin{pmatrix} 1 & 3 & 5 \\ -2 & -1 & 2 \end{pmatrix} \quad \boldsymbol{B} = \begin{pmatrix} 2 & 3 \\ 0 & 5 \\ 1 & -2 \end{pmatrix}$$

$$\boldsymbol{AB} = \begin{pmatrix} 7 & 8 \\ -2 & -15 \end{pmatrix}$$

  E.g. $(1)(2) + (3)(0) + (5)(1) = 7$ for the $(1,1)$ element.

- Two matrices satisfying these requirements are said to **conform** to multiplication.
- Formally, if $\boldsymbol{A}$ is $(r \times c)$ and $\boldsymbol{B}$ is $(c \times q)$, then $\boldsymbol{AB}$ is a $(r \times q)$ matrix with (i,j)th element $\sum_{k=1}^{c} a_{ik} b_{kj}$.

  Here, we say that $\boldsymbol{A}$ is **postmultiplied** by $\boldsymbol{B}$ and, equivalently, that $\boldsymbol{B}$ is **premultiplied** by $\boldsymbol{A}$.

EXAMPLE: Consider a **simple linear regression** model: suppose that we have n pairs $(x_1, Y_1), \ldots, (x_n, Y_n)$, and we believe that, except for a random deviation, the relationship between the **covariate** x and the response Y follows a straight line. That is, for $j = 1, \ldots, n$, we have

$$Y_j = \beta_0 + \beta_1 x_j + \epsilon_j,$$

where ϵ_j is a random deviation representing the amount by which the actual observed response Y_j deviates from the exact straight line relationship. Defining

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix},$$

we may express the model succinctly as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{2.1}$$

SPECIAL CASE: Multiplying vectors. With a row vector premultiplying a column vector, the result is a **scalar** (remember, a (1×1) matrix is just a real number!), e.g.

$$\boldsymbol{a}\,\boldsymbol{b} = \begin{pmatrix} 1, & 3, & -5, & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 0 \\ 3 \\ -2 \end{pmatrix} = -15$$

i.e. $(1)(2) + (3)(0) + (-5)(3) + (1)(-2) = -15$

With a column vector premultiplying a row vector, the result is a **matrix**. e.g.

$$\boldsymbol{bc} = \begin{pmatrix} 2 \\ 0 \\ 3 \\ -2 \end{pmatrix} \begin{pmatrix} 3 & -1 & 2 \end{pmatrix} = \begin{pmatrix} 6 & -2 & 4 \\ 0 & 0 & 0 \\ 9 & -3 & 6 \\ -6 & 2 & -4 \end{pmatrix}$$

MULTIPLICATION BY AN IDENTITY MATRIX: Multiplying **any** matrix by an identity matrix of appropriate dimension gives back the **same** matrix, e.g.

$$\boldsymbol{I}\,\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 5 \\ -2 & -1 & 2 \end{pmatrix} = \boldsymbol{A}$$

GENERAL FACTS:

- $\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC}$, $(\boldsymbol{A} + \boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC} + \boldsymbol{BC}$
- For any matrix $\boldsymbol{A}$, $\boldsymbol{A}'\boldsymbol{A}$ will be a *square* matrix.
- The **transpose** of a matrix product – if $\boldsymbol{A}$ and $\boldsymbol{B}$ conform to multiplication, then the transpose of their product

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

These latter results may be proved generically, but you may convince yourself by working them out for the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ given above.

LINEAR DEPENDENCE: This characteristic of a matrix is extremely important in that it describes the nature and extent of the information contained in the matrix. Consider the matrix

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

Refer to the columns as $\boldsymbol{c}_1$, $\boldsymbol{c}_2$, $\boldsymbol{c}_3$. Note that

$$2\boldsymbol{c}_1 + -\boldsymbol{c}_2 + -\boldsymbol{c}_3 = \boldsymbol{0},$$

where $\boldsymbol{0}$ is a column of zeros (in this case, a (3×1) vector). Because the 3 columns of $\boldsymbol{A}$ may be **combined** in a **linear** function to yield a vector of nothing but zeros, clearly, there is some kind of relationship, or **dependence**, among the information in the columns. Put another way, it seems as though there is some **duplication** of information in the columns.

In general, we say that k columns $\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_k$ of a matrix are **linearly dependent** if there exists a set of scalar values $\lambda_1, \ldots, \lambda_k$ such that

$$\lambda_1\boldsymbol{c}_1 + \cdots + \lambda_k\boldsymbol{c}_k = \boldsymbol{0}, \tag{2.2}$$

and at least one of the λ_j's is not equal to 0.

Linear dependence implies that each column vector is a combination of the others, e.g.,

$$\boldsymbol{c}_k = -(\lambda_1\boldsymbol{c}_1 + \cdots + \lambda_{k-1}\boldsymbol{c}_{k-1})/\lambda_k.$$

The implication is that all of the "information" in the matrix is contained in a subset of the columns – if we know any $(k-1)$ columns, we know them all. This formalizes our notion of "duplication" of information.

If, on the other hand, the only set of λ_j values we can come up with to satisfy (2.2) is a set of all zeros, then it must be that there is **no relationship** among the columns, e.g. they are "independent" in the sense of containing no overlap of information. The formal term is **linearly independent**.

RANK OF A MATRIX: The **rank** of a matrix is the maximum number of linearly independent columns that may be selected from the columns of the matrix. It is sort of a measure of the extent of "duplication of information" in the matrix. The rank of a matrix may be equivalently defined as the number of linearly independent **rows** (by turning the matrix on its side). The rank determined either way is the same.

Thus, the largest that the rank of a matrix can be is the minimum of r and c. The smallest rank may be is 1, in which case there is one column such that all other columns are direct multiples.

In the above, the rank of the matrix $\boldsymbol{A}$ is 2. To see this, eliminate one of the columns (we have already seen that the three columns are linearly dependent, so we can get the third from the other two). Now try to find a new linear combination of the remaining columns that has some λ_j not equal to 0. If this can not be done – stop and declare the rank to be the number of remaining columns.

FULL RANK: A matrix is said to be of **full rank** if its rank is **equal to** the minimum of r and c.

FACT: If $\boldsymbol{X}$ is a $(r \times c)$ matrix with rank k, then $\boldsymbol{X'X}$ also has rank k. Note, of course, that $\boldsymbol{X'X}$ is a square matrix of dimension $(c \times c)$. If $k = c$, then $\boldsymbol{X'X}$ is of full rank.

INVERSE OF A MATRIX: This is related to the matrix version of "division" – the inverse of a matrix may be thought of in way similar to a "reciprocal" in the world of real numbers.

- The notion of an inverse is only defined for **square** matrices, for reasons that will be clear below.
- The **inverse** of the square matrix $\boldsymbol{A}$ is denoted by $\boldsymbol{A}^{-1}$ and is the square matrix satisfying

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

  where $\boldsymbol{I}$ is an identity matrix of the same dimension.

- We sometimes write $\boldsymbol{I}_k$ when $\boldsymbol{I}$ is $(k \times k)$ when it is important to note explicitly the dimension.

Thus, the inverse of a matrix is like the analog of the reciprocal for scalars. Recall that if b is a scalar and $b = 0$, then the reciprocal of b, $1/b$ **does not exist** – it is not defined in this case. Similarly, there are matrices that "act like zero" for which no inverse is defined. Consequently, inverse is only defined when it exists.

Computing the inverse of a matrix is best done on a computer, where the intricate formulæ for matrices of general dimension are usually built in to software packages. Only in simple cases is an analytic expression obtained easily (see the next page).

A technical condition that an inverse of the matrix $\boldsymbol{A}$ exist is that the columns of $\boldsymbol{A}$ are linearly independent. This is related to the following. if and only if

DETERMINANT: When is a matrix "like zero?" The **determinant** of a square matrix is a **scalar** number that in some sense summarizes how "zero-like" a matrix is.

The determinant of a (2×2) matrix is defined as follows. Let

$$\boldsymbol{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Then the determinant of $\boldsymbol{A}$ is given by

$$|\boldsymbol{A}| = ad - bc.$$

The notation $|\boldsymbol{A}|$ means "determinant of;" this may also be written as $\det(\boldsymbol{A})$. Determinant is also defined for larger matrices, although the calculations become tedious (but are usually part of any decent software package).

The inverse of a matrix is related to the determinant. In the special case of a (2×2) matrix like $\boldsymbol{A}$ above, it may be shown that

$$\boldsymbol{A}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Inverse for matrices of larger dimension is also defined in terms of the determinant, but the expressions are complicated.

GENERAL FACTS:

- If a square matrix is not of full rank, then it will have determinant equal to 0. For example, for the (2×2) matrix above, suppose that the columns are **linearly dependent** with $a = 2b$ and $c = 2d$. Then note that
$$|\boldsymbol{A}| = ad - bc = 2bd - 2bd = 0.$$
- Thus, note that if a matrix is not of full rank, its inverse does not exist. In the case of a (2×2) matrix, note that the inverse formula requires division by $(ad - bc)$, which would be equal to zero.

EXAMPLE:

$$\boldsymbol{A} = \begin{pmatrix} 5 & 0 \\ -3 & 2 \end{pmatrix}, \quad |\boldsymbol{A}| = (5)(2) - (0)(-3) = 10$$

$$\boldsymbol{A}^{-1} = \frac{1}{10}\begin{pmatrix} 2 & 0 \\ 3 & 5 \end{pmatrix} = \begin{pmatrix} 1/5 & 0 \\ 3/10 & 1/2 \end{pmatrix}$$

Verify that $\boldsymbol{A}\,\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$.

ADDITIONAL FACTS: Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be square matrices of the same dimension whose inverses exist.

- $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$, $(\boldsymbol{A}^{-1})' = (\boldsymbol{A}')^{-1}$.
- If $\boldsymbol{A}$ is a **diagonal** matrix, that is, a matrix that has non-zero elements only on its diagonal, with 0's everywhere else, then its inverse is nothing more than a diagonal matrix whose diagonal elements are the **reciprocals** of the original diagonal elements, e.g., if

  $$\boldsymbol{A} = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -4 \end{pmatrix}, \quad \boldsymbol{A}^{-1} = \begin{pmatrix} 1/5 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & -1/4 \end{pmatrix}.$$

  Note that an identity matrix is just a diagonal matrix whose inverse is itself, just as 1/1=1.
- $|\boldsymbol{A}| = |\boldsymbol{A}'|$
- If each element of a row or column of $\boldsymbol{A}$ is zero, then $|\boldsymbol{A}| = 0$.
- If $\boldsymbol{A}$ has any rows or columns identical, then $|\boldsymbol{A}| = 0$.
- $|\boldsymbol{A}| = 1/|\boldsymbol{A}^{-1}|$
- $|\boldsymbol{AB}| = |\boldsymbol{A}||\boldsymbol{B}|$
- If b is a scalar, then $|b\boldsymbol{A}| = b^k|\boldsymbol{A}|$, where k is the dimension of $\boldsymbol{A}$.
- $(\boldsymbol{A}+\boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}(\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1})^{-1}\boldsymbol{A}^{-1}$
- If $\boldsymbol{A}$ is a **diagonal** matrix, then $|\boldsymbol{A}|$ is equal to the product of the diagonal elements, i.e.

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix} \Rightarrow |\boldsymbol{A}| = a_{11}a_{22}\cdots a_{nn}.$$

USE OF INVERSE – SOLVING SIMULTANEOUS EQUATIONS: Suppose we have a set of simultaneous equations with unknown values x, y, and z, e.g.

$$\begin{array}{rcrcrcr} x & - & y & + & z & = & 2 \\ 2x & + & y & & & = & 7 \\ 3x & + & y & + & z & = & \text{-}5. \end{array}$$

We may write this system succinctly in matrix notation as $\boldsymbol{Aa} = \boldsymbol{b}$, where

$$\boldsymbol{A} = \begin{pmatrix} 1 & -1 & 1 \\ 2 & 1 & 0 \\ 3 & 1 & 1 \end{pmatrix}, \quad \boldsymbol{a} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} 2 \\ 7 \\ -5 \end{pmatrix}.$$

Then, provided $\boldsymbol{A}^{-1}$ exists, we may write the solution as

$$\boldsymbol{a} = \boldsymbol{A}^{-1}\boldsymbol{b}.$$

Note that if $\boldsymbol{b} = \boldsymbol{0}$, then the above shows that if $\boldsymbol{A}$ has an inverse, then it must be that $\boldsymbol{a} = \boldsymbol{0}$. More formally, a square matrix $\boldsymbol{A}$ is said to be **nonsingular** if $\boldsymbol{Aa} = \boldsymbol{0}$ implies $\boldsymbol{a} = \boldsymbol{0}$. Otherwise, the matrix is said to be **singular**.

Equivalently, a square matrix is **nonsingular** if it is of **full rank**.

For a square matrix $\boldsymbol{A}$, the following are equivalent:

- $\boldsymbol{A}$ is nonsingular
- $|\boldsymbol{A}| \neq 0$
- $\boldsymbol{A}^{-1}$ exists

We will see that matrix notation is incredibly useful for summarizing models and methods for longitudinal data. As is true more generally in statistics, the concepts of rank and singularity are very important. Matrices in statistical models that are singular generally reflect a **problem** – most often, they reflect that there is not sufficient information available to learn about certain aspects of the model. We will see this in action later in the course.

EXAMPLE: Returning to the matrix representation of the simple linear regression model, it is possible to use these operations to streamline the statement of how to calculate the least squares estimators of β_0 and β_1. Recall that the least squares estimators $\widehat{\beta}_0$ and $\widehat{\beta}_1$ for the intercept and slope minimize the **sum of squared deviations**

$$\sum_{j=1}^{n}(Y_j - \beta_0 - x_j\beta_1)^2$$

and are given by

$$\widehat{\beta}_1 = \frac{S_{XY}}{S_{XX}}, \quad \widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1\overline{x},$$

where

$$S_{XY} = \sum_{j=1}^{n}(Y_j - \overline{Y})(x_j - \overline{x}) = \sum_{j=1}^{n} x_j Y_j - \frac{(\sum_{j=1}^{n} x_j)(\sum_{j=1}^{n} Y_j)}{n}, \quad \overline{Y} = n^{-1}\sum_{j=1}^{n} Y_j, \quad \overline{x} = n^{-1}\sum_{j=1}^{n} x_j$$

$$S_{XX} = \sum_{j=1}^{n}(x_j - \overline{x})^2 = \sum_{j=1}^{n} x_j^2 - \frac{(\sum_{j=1}^{n} x_j)^2}{n}, \quad S_{YY} = \sum_{j=1}^{n}(Y_j - \overline{Y})^2 = \sum_{j=1}^{n} Y_j^2 - \frac{(\sum_{j=1}^{n} Y_j)^2}{n},$$

We may summarize these calculations succinctly in matrix notation: the sum of squared deviations may be written as

$$(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}),$$

and, letting $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_0, \widehat{\beta}_1)'$, the least squares estimator for $\boldsymbol{\beta}$ may be written

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}.$$

Verify that, with $\boldsymbol{X}$ and $\boldsymbol{Y}$ defined as in (2.1), this matrix equation gives the usual estimators above.

CONVENTION: Here, we have referred to $\widehat{\beta}_0$ and $\widehat{\beta}_1$ as **estimators**, and have written them in terms of the **random variables** Y_j. The term **estimator** refers to the generic function of random variables one would use to learn about **parameters** like β_0 or β_1. The term **estimate** refers to the actual numerical values obtained by applying the estimator to data; e.g., $y_1, \ldots, y_n$ in this case.

We will see later that matrix notation is more generally useful for summarizing models for longitudinal data and the calculations required to fit them; the simple linear regression model above is a simple example.

TRACE OF A MATRIX: Defining this quantity allows a streamlined representation of many complex calculations. If $\boldsymbol{A}$ is a $(k \times k)$ square matrix, then define the **trace** of $\boldsymbol{A}$, $\text{tr}(\boldsymbol{A})$, to be the sum of the diagonal elements; i.e.

$$\text{tr}(\boldsymbol{A}) = \sum_{i=1}^{k} a_{ii}.$$

If $\boldsymbol{A}$ and $\boldsymbol{B}$ are both square with dimension k, then

- $\text{tr}(\boldsymbol{A}) = \text{tr}(\boldsymbol{A}')$, $\text{tr}(b\boldsymbol{A}) = b\text{tr}(\boldsymbol{A})$
- $\text{tr}(\boldsymbol{A} + \boldsymbol{B}) = \text{tr}(\boldsymbol{A}) + \text{tr}(\boldsymbol{B})$, $\text{tr}(\boldsymbol{AB}) = \text{tr}(\boldsymbol{BA})$

QUADRATIC FORMS: The following form arises quite often. Suppose $\boldsymbol{A}$ is a square, **symmetric** matrix of dimension k, and $\boldsymbol{x}$ is a $(k \times 1)$ column vector. Then

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$$

is called a **quadratic form**. It may be shown that

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \sum_{i=1}^{k}\sum_{j=1}^{k} a_{ij}x_i x_j.$$

Note that this sum will involve both **squared** terms x_i^2 and **cross-product** terms $x_i x_j$, which forms the basis for the name **quadratic**.

A quadratic form thus takes on **scalar** values. Depending on the value, the quadratic form and the matrix $\boldsymbol{A}$ may be classified. With $\boldsymbol{x} \neq \boldsymbol{0}$,

- If $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geq 0$, the quadratic form and the matrix $\boldsymbol{A}$ are said to be **nonnegative definite**
- If $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$, the quadratic form and the matrix $\boldsymbol{A}$ are said to be **positive definite**. If $\boldsymbol{A}$ is positive definite, then it is symmetric and nonsingular (so its inverse exists).

EXAMPLE: The sum of squared deviations that is minimized to obtain the least squares estimators in regression is a quadratic form with $\boldsymbol{A} = \boldsymbol{I}$,

$$(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{I}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}).$$

Note that this is strictly greater than 0 by definition, because it equals

$$\sum_{j=1}^{n}(Y_j - \beta_0 - x_j\beta_1)^2,$$

which is a sum of squared quantities, all of which must be positive (assuming that not all deviations are identically equal to zero, in which case the problem is rather nonsensical).

FACT: $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \text{tr}(\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}')$; this may be verified by simply multiplying out each side. (Try it for the sum of squared deviations above.)