

Statistical Methods II

Derek L. Sonderegger

2017-09-26

Contents

Chapter 1

Matrix Theory

Almost all of the calculations done in classical statistics require formulas with large number of subscripts and many different sums. In this chapter we will develop the mathematical machinery to write these formulas in a simple compact formula using *matrices*.

1.1 Types of Matrices

We will first introduce the idea behind a matrix and give several special types of matrices that we will encounter.

1.1.1 Scalars

To begin, we first define a *scalar*. A scalar is just a single number, either real or complex. The key is that a scalar is just a single number. For example, 6 is a scalar, as is -3 . By convention, variable names for scalars will be lower case and not in bold typeface.

Examples could be $a = 5$, $b = \sqrt{3}$, or $\sigma = 2$.

1.1.2 Vectors

A vector is collection of scalars, arranged as a row or column. Our convention will be that a vector will be a lower cased letter but written in a bold type. In other branches of mathematics is common to put a bar over the variable name to denote that it is a vector, but in statistics, we have already used a bar to denote a mean.

Examples of column vectors could be

$$\mathbf{a} = \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 2 \\ 8 \\ 3 \\ 4 \\ 1 \end{bmatrix}$$

and examples of row vectors are

$$\mathbf{c} = \begin{bmatrix} 8 & 10 & 43 & -22 \end{bmatrix}$$

$$\mathbf{d} = \begin{bmatrix} -1 & 5 & 2 \end{bmatrix}$$

To denote a specific entry in the vector, we will use a subscript. For example, the second element of \mathbf{d} is $d_2 = 5$. Notice, that we do not bold this symbol because the second element of the vector is the scalar value 5.

1.1.3 Matrix

Just as a vector is a collection of scalars, a matrix can be viewed as a collection of vectors (all of the same length). We will denote matrices with bold capitalized letters. In general, I try to use letters at the end of the alphabet for matrices. Likewise, I try to use symmetric letters to denote symmetric matrices.

For example, the following is a matrix with two rows and three columns

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

and there is no requirement that the number of rows be equal, less than, or greater than the number of columns. In denoting the size of the matrix, we first refer to the number of rows and then the number of columns. Thus \mathbf{W} is a 2×3 matrix and it sometimes is helpful to remind ourselves of this by writing $\mathbf{W}_{2 \times 3}$.

To pick out a particular element of a matrix, I will again use a subscripting notation, always with the row number first and then column. Notice the notational shift to lowercase, non-bold font.

$$w_{1,2} = 2 \quad \text{and} \quad w_{2,3} = 6$$

There are times I will wish to refer to a particular row or column of a matrix and we will use the following notation

$$\mathbf{w}_{1,\cdot} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$$

is the first row of the matrix \mathbf{W} . The second column of matrix \mathbf{W} is

$$\mathbf{w}_{\cdot,2} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$$

1.1.4 Square Matrices

A square matrix is a matrix with the same number of rows as columns. The following are square

$$\mathbf{Z} = \begin{bmatrix} 3 & 6 \\ 8 & 10 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 2 \\ 3 & 2 & 1 \end{bmatrix}$$

1.1.5 Symmetric Matrices

In statistics we are often interested in square matrices where the i, j element is the same as the j, i element. For example, $x_{1,2} = x_{2,1}$ in the above matrix \mathbf{X} .

Consider a matrix \mathbf{D} that contains the distance from four towns to each of the other four towns. Let $d_{i,j}$ be the distance from town i to town j . It only makes sense that the distance doesn't matter which direction you are traveling, and we should therefore require that $d_{i,j} = d_{j,i}$.

In this example, it is the values $d_{i,i}$ represent the distance from a town to itself, which should be zero. It turns out that we are often interested in the terms $d_{i,i}$ and I will refer to those terms as the *main diagonal* of matrix \mathbf{D} .

Symmetric matrices play a large role in statistics because matrices that represent the covariances between random variables must be symmetric because $Cov(Y, Z) = Cov(Z, Y)$.

1.1.6 Diagonal Matrices

A square matrix that has zero entries in every location except the main diagonal is called a diagonal matrix. Here are two examples:

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

Sometimes to make matrix more clear, I will replace the 0 with a dot to emphasize the non-zero components.

$$\mathbf{R} = \begin{bmatrix} 1 & \cdot & \cdot & \cdot \\ \cdot & 2 & \cdot & \cdot \\ \cdot & \cdot & 2 & \cdot \\ \cdot & \cdot & \cdot & 3 \end{bmatrix}$$

1.1.7 Identity Matrices

A diagonal matrix with main diagonal values exactly 1 is called the identity matrix. The 3×3 identity matrix is denoted I_3 .

$$\mathbf{I}_3 = \begin{bmatrix} 1 & \cdot & \cdot \\ \cdot & 1 & \cdot \\ \cdot & \cdot & 1 \end{bmatrix}$$

1.2 Operations on Matrices

1.2.1 Transpose

The simplest operation on a square matrix matrix is called *transpose*. It is defined as $\mathbf{M} = \mathbf{W}^T$ if and only if $m_{i,j} = w_{j,i}$.

$$\mathbf{Z} = \begin{bmatrix} 1 & 6 \\ 8 & 3 \end{bmatrix} \quad \mathbf{Z}^T = \begin{bmatrix} 1 & 8 \\ 6 & 3 \end{bmatrix}$$

$$\mathbf{M} = \begin{bmatrix} 3 & 1 & 2 \\ 9 & 4 & 5 \\ 8 & 7 & 6 \end{bmatrix} \quad \mathbf{M}^T = \begin{bmatrix} 3 & 9 & 8 \\ 1 & 4 & 7 \\ 2 & 5 & 6 \end{bmatrix}$$

We can think of this as swapping all elements about the main diagonal. Alternatively we could think about the transpose as making the first row become the first column, the second row become the second column, etc. In this fashion we could define the transpose of a non-square matrix.

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

$$\mathbf{W}^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

1.2.2 Addition and Subtraction

Addition and subtraction are performed *element-wise*. This means that two matrices or vectors can only be added or subtracted if their dimensions match.

$$\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} + \begin{bmatrix} 5 \\ 6 \\ 7 \\ 8 \end{bmatrix} = \begin{bmatrix} 6 \\ 8 \\ 10 \\ 12 \end{bmatrix}$$

$$\begin{bmatrix} 5 & 8 \\ 2 & 4 \\ 11 & 15 \end{bmatrix} - \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & -6 \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ -1 & 0 \\ 6 & 21 \end{bmatrix}$$

1.2.3 Multiplication

Multiplication is the operation that is vastly different for matrices and vectors than it is for scalars. There is a great deal of mathematical theory that suggests a useful way to define multiplication. What is presented below is referred to as the *dot-product* of vectors in calculus, and is referred to as the standard *inner-product* in linear algebra.

1.2.4 Vector Multiplication

We first define multiplication for a row and column vector. For this multiplication to be defined, both vectors must be the same length. The product is the sum of the element-wise multiplications.

$$\begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 5 \\ 6 \\ 7 \\ 8 \end{bmatrix} = (1 \cdot 5) + (2 \cdot 6) + (3 \cdot 7) + (4 \cdot 8) = 5 + 12 + 21 + 32 = 70$$

1.2.5 Matrix Multiplication

Matrix multiplication is just a sequence of vector multiplications. If \mathbf{X} is a $m \times n$ matrix and \mathbf{W} is $n \times p$ matrix then $\mathbf{Z} = \mathbf{X}\mathbf{W}$ is a $m \times p$ matrix where $z_{i,j} = \mathbf{x}_{i,\cdot} \cdot \mathbf{w}_{\cdot,j}$ where $\mathbf{x}_{i,\cdot}$ is the i th column of \mathbf{X} and $\mathbf{w}_{\cdot,j}$ is the j th column of \mathbf{W} . For example, let

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix} \quad \mathbf{W} = \begin{bmatrix} 13 & 14 \\ 15 & 16 \\ 17 & 18 \\ 19 & 20 \end{bmatrix}$$

so \mathbf{X} is 3×4 (which we remind ourselves by adding a 3×4 subscript to \mathbf{X} as $\mathbf{X}_{3 \times 4}$) and \mathbf{W} is $\mathbf{W}_{4 \times 2}$. Because the *inner* dimensions match for this multiplication, then $\mathbf{Z}_{3 \times 2} = \mathbf{X}_{3 \times 4} \mathbf{W}_{4 \times 2}$ is defined where

$$\begin{aligned} z_{1,1} &= \mathbf{x}_{1,\cdot} \cdot \mathbf{w}_{\cdot,1} \\ &= (1 \cdot 13) + (2 \cdot 15) + (3 \cdot 17) + (4 \cdot 19) = 170 \end{aligned}$$

and similarly

$$\begin{aligned} z_{2,1} &= \mathbf{x}_{2,\cdot} \mathbf{w}_{\cdot,1} \\ &= (5 \cdot 13) + (6 \cdot 15) + (7 \cdot 17) + (8 \cdot 19) = 426 \end{aligned}$$

so that

$$\mathbf{Z} = \begin{bmatrix} 170 & 180 \\ 426 & 452 \\ 682 & 724 \end{bmatrix}$$

For another example, we note that

$$\begin{aligned} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 2 \\ 1 & 2 \end{bmatrix} &= \begin{bmatrix} 1+4+3 & 2+4+6 \\ 2+6+4 & 4+6+8 \end{bmatrix} \\ &= \begin{bmatrix} 8 & 12 \\ 12 & 18 \end{bmatrix} \end{aligned}$$

Notice that this definition of multiplication means that the order matters. Above, we calculated $\mathbf{X}_{3 \times 4} \mathbf{W}_{4 \times 2}$ but we cannot reverse the order because the inner dimensions do not match up.

1.2.6 Scalar times a Matrix

Strictly speaking, we are not allowed to multiply a matrix by a scalar because the dimensions do not match. However, it is often notationally convenient. So we define $a\mathbf{X}$ to be the *element-wise* multiplication of each element of \mathbf{X} by the scalar a . Because this is just a notational convenience, the mathematical theory about inner-products does not apply to this operation.

$$5 \begin{bmatrix} 4 & 5 \\ 7 & 6 \\ 9 & 10 \end{bmatrix} = \begin{bmatrix} 20 & 25 \\ 35 & 30 \\ 45 & 50 \end{bmatrix}$$

Because of this definition, it is clear that $a\mathbf{X} = \mathbf{X}a$ and the order does not matter. Thus when mixing scalar multiplication with matrices, it is acceptable to reorder scalars, but not matrices.

1.2.7 Determinant

The determinant is defined only for square matrices and can be thought of as the matrix equivalent of the absolute value or magnitude (i.e. $|-6| = 6$). The determinant gives a measure of the multi-dimensional size of a matrix (say the matrix \mathbf{A}) and as such is denoted $\det(\mathbf{A})$ or $|\mathbf{A}|$. Generally this is a very tedious thing to calculate by hand and for completeness sake, we will give a definition and small examples.

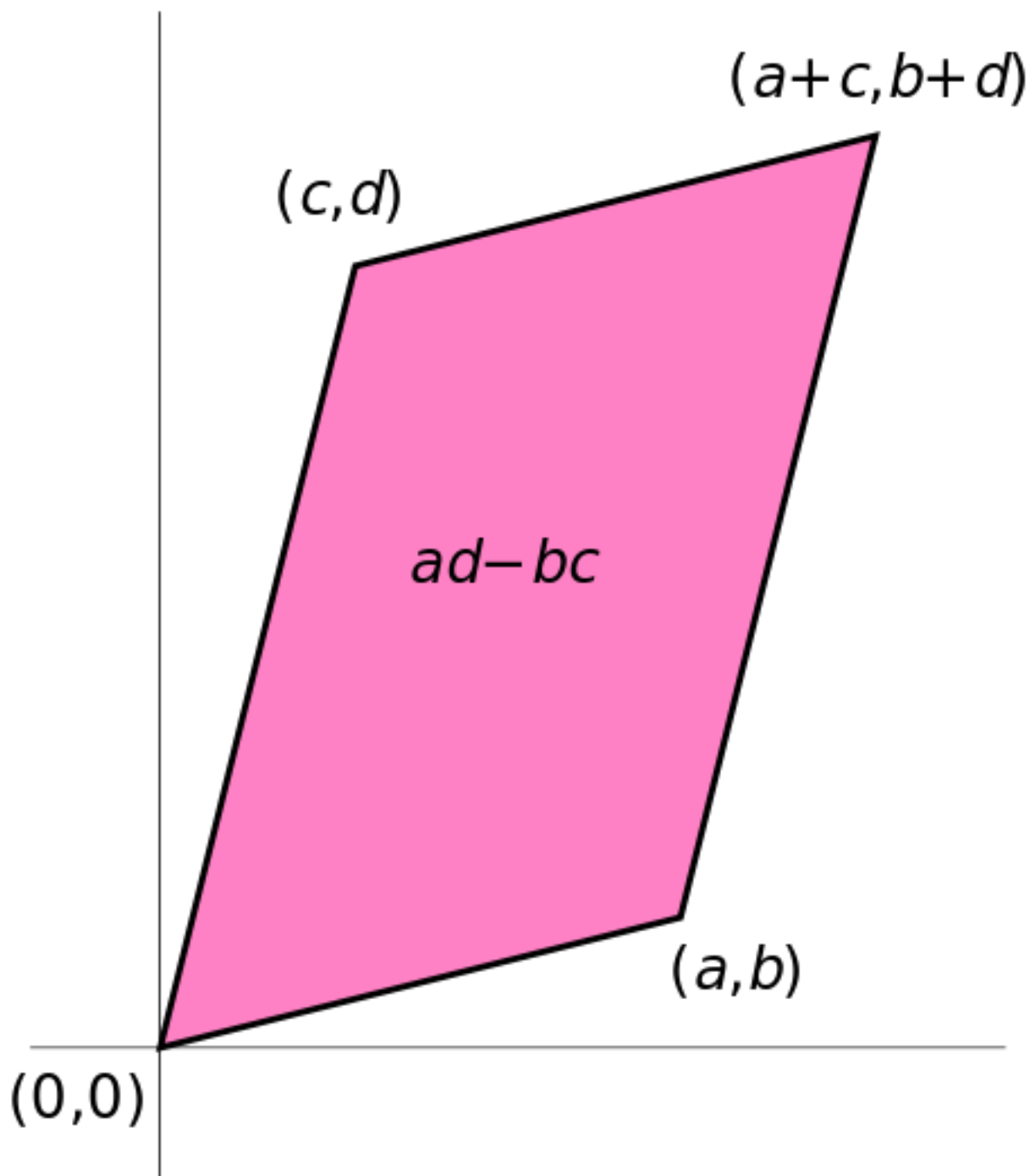
For a 2×2 matrix

$$\begin{vmatrix} a & c \\ b & d \end{vmatrix} = ad - cb$$

So a simple example of a determinant is

$$\begin{vmatrix} 5 & 2 \\ 3 & 10 \end{vmatrix} = 50 - 6 = 44$$

The determinant can be thought of as the area of the parallelogram created by the row or column vectors of the matrix.



1.2.8 Inverse

In regular algebra, we are often interested in solving equations such as

$$5x = 15$$

for x . To do so, we multiply each side of the equation by the inverse of 5, which is $1/5$.

$$\begin{aligned} 5x &= 15 \\ \frac{1}{5} \cdot 5 \cdot x &= \frac{1}{5} \cdot 15 \\ 1 \cdot x &= 3 \\ x &= 3 \end{aligned}$$

For scalars, we know that the inverse of scalar a is the value that when multiplied by a is 1. That is we see to find a^{-1} such that $aa^{-1} = 1$.

In the matrix case, I am interested in finding \mathbf{A}^{-1} such that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ and $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$. For both of these multiplications to be defined, \mathbf{A} must be a square matrix and so the inverse is only defined for square matrices.

For a 2×2 matrix

$$\mathbf{W} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

the inverse is given by:

$$\mathbf{W}^{-1} = \frac{1}{\det \mathbf{W}} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

For example, suppose

$$\mathbf{W} = \begin{bmatrix} 1 & 2 \\ 5 & 3 \end{bmatrix}$$

then $\det W = 3 - 10 = -7$ and

$$\begin{aligned} \mathbf{W}^{-1} &= \frac{1}{-7} \begin{bmatrix} 3 & -2 \\ -5 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -\frac{3}{7} & \frac{2}{7} \\ \frac{5}{7} & -\frac{1}{7} \end{bmatrix} \end{aligned}$$

and thus

$$\begin{aligned} \mathbf{W}\mathbf{W}^{-1} &= \begin{bmatrix} 1 & 2 \\ 5 & 3 \end{bmatrix} \begin{bmatrix} -\frac{3}{7} & \frac{2}{7} \\ \frac{5}{7} & -\frac{1}{7} \end{bmatrix} \\ &= \begin{bmatrix} -\frac{3}{7} + \frac{10}{7} & \frac{2}{7} - \frac{2}{7} \\ -\frac{15}{7} + \frac{15}{7} & \frac{10}{7} - \frac{3}{7} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}_2 \end{aligned}$$

Not every square matrix has an inverse. If the determinant of the matrix (which we think of as some measure of the magnitude or *size* of the matrix) is zero, then the formula would require us to divide by zero. Just as we cannot find the inverse of zero (i.e. solve $0x = 1$ for x), a matrix with zero determinate is said to have no inverse.

1.3 Exercises

Consider the following matrices:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 6 & 5 & 4 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 6 & 4 & 3 \\ 8 & 7 & 6 \end{bmatrix} \quad \mathbf{c} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \mathbf{d} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \quad \mathbf{E} = \begin{bmatrix} 1 & 2 \\ 2 & 6 \end{bmatrix}$$

1. Find \mathbf{Bc}
2. Find \mathbf{AB}^T
3. Find $\mathbf{c}^T \mathbf{d}$
4. Find \mathbf{cd}^T
5. Confirm that $\mathbf{E}^{-1} = \begin{bmatrix} 3 & -1 \\ -1 & 1/2 \end{bmatrix}$ is the inverse of \mathbf{E} by calculating $\mathbf{EE}^{-1} = \mathbf{I}$.

Chapter 2

Parameter Estimation

We have previously looked at ANOVA and regression models and, in many ways, they felt very similar. In this chapter we will introduce the theory that allows us to understand both models as a particular flavor of a larger class of models known as *linear models*.

First we clarify what a linear model is. A linear model is a model where the data (which we will denote using roman letters as \mathbf{x} and \mathbf{y}) and parameters of interest (which we denote using Greek letters such as α and β) interact only via addition and multiplication. The following are linear models:

| Model | Formula |
|--------------------|---------------------------------------------------------------------------------------------|
| ANOVA | $y_{ij} = \mu + \tau_i + \epsilon_{ij}$ |
| Simple Regression | $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ |
| Quadratic Term | $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i$ |
| General Regression | $y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_p x_{i,p} + \epsilon_i$ |

Notice in the Quadratic model, the square is not a parameter and we can consider x_i^2 as just another column of data. This leads to the second example of multiple regression where we just add more slopes for other covariates where the p th covariate is denoted $\mathbf{x}_{\cdot,p}$ and might be some transformation (such as x^2 or $\log x$) of another column of data. The critical point is that the transformation to the data \mathbf{x} does not depend on a parameter. Thus the following is *not* a linear model

$$y_i = \beta_0 + \beta_1 x_i^\alpha + \epsilon_i$$

2.1 Simple Regression

We would like to represent all linear models in a similar compact matrix representation. This will allow us to make the transition between simple and multiple regression (and ANCOVA) painlessly.

To begin, we think about how to write the simple regression model using matrices and vectors that correspond to the data and the parameters. Notice we have