1 Introduction to Matrices

In this section, important definitions and results from matrix algebra that are useful in regression analysis are introduced. While all statements below regarding the columns of matrices can also be said of rows, in regression applications we will typically be focusing on the columns.

A **matrix** is a rectangular array of numbers. The **order** or **dimension** of the matrix is the number of rows and columns that make up the matrix. The **rank** of a matrix is the number of linearly independent columns (or rows) in the matrix.

A subset of columns is said to be **linearly independent** if no column in the subset can be written as a linear combination of the other columns in the subset. A matrix is **full rank (nonsingular)** if there are no linear dependencies among its columns. The matrix is **singular** if lineardependencies exist.

The **column space** of a matrix is the collection of all linear combinations of the columns of a matrix.

The following are important types of matrices in regression:

Vector – Matrix with one row or column

Square Matrix – Matrix where number of rows equals number of columns

Diagonal Matrix – Square matrix where all elements off main diagonal are 0

Identity Matrix – Diagonal matrix with 1's everywhere on main diagonal

Symmetric Matrix – Matrix where element $a_{ij} = a_{ji}\ \forall i, j$

Scalar – A single ordinary number

The **transpose** of a matrix is the matrix generated by interchanging the rows and columns of the matrix. If the original matrix is $\boldsymbol{A}$, then its transpose is labelled $\boldsymbol{A'}$. For example:

$$\mathbf{A} = \begin{bmatrix} 2 & 4 & 7 \\ 1 & 7 & 2 \end{bmatrix} \quad \Rightarrow \quad \mathbf{A}' = \begin{bmatrix} 2 & 1 \\ 4 & 7 \\ 7 & 2 \end{bmatrix}$$

Matrix addition (subtraction) can be performed on two matrices as long as they are of equal order (dimension). The new matrix is obtained by elementwise addition (subtraction) of the two matrices. For example:

$$\mathbf{A} = \begin{bmatrix} 2 & 4 & 7 \\ 1 & 7 & 2 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 & 3 & 0 \\ 2 & 4 & 8 \end{bmatrix} \quad \Rightarrow \quad \mathbf{A} + \mathbf{B} = \begin{bmatrix} 3 & 7 & 7 \\ 3 & 11 & 10 \end{bmatrix}$$

Matrix multiplication can be performed on two matrices as long as the number of columns of the first matrix equals the number of rows of the second matrix. The resulting has the same number of rows as the first matrix and the same number of columns as the second matrix. If

$\mathbf{C} = \mathbf{AB}$ and $\mathbf{A}$ has s columns and $\mathbf{B}$ has s rows, the element in the i^{th} row and j^{th} column of $\mathbf{C}$, which we denote c_{ij} is obtained as follows (with similar definitions for a_{ij} and b_{ij}):

$$c_{ij} \quad = \quad a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots a_{is}b_{sj} \quad = \quad \sum_{k=1}^{s} a_{ik}b_{kj}$$

For example:

$$\mathbf{A} = \begin{bmatrix} 2 & 4 & 7 \\ 1 & 7 & 2 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 & 5 & 6 \\ 2 & 0 & 1 \\ 3 & 3 & 3 \end{bmatrix} \quad \Rightarrow$$

$$\mathbf{C} = \mathbf{AB} = \begin{bmatrix} 2(1)+4(2)+7(3) & 2(5)+4(0)+7(3) & 2(6)+4(1)+7(3) \\ 1(1)+7(2)+2(3) & 1(5)+7(0)+2(3) & 1(6)+7(1)+2(3) \end{bmatrix} = \begin{bmatrix} 31 & 31 & 37 \\ 21 & 11 & 19 \end{bmatrix}$$

Note that $\mathbf{C}$ has the same number of rows as $\mathbf{A}$ and the same number of columns as $\mathbf{B}$. Note that in general $\mathbf{AB} \neq \mathbf{BA}$; in fact, the second matrix may not exist due to dimensions of matrices. However, the following equality does hold: $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$.

Scalar Multiplication can be performed between any scalar and any matrix. Each element of the matrix is multiplied by the scalar. For example:

$$\mathbf{A} = \begin{bmatrix} 2 & 4 & 7 \\ 1 & 7 & 2 \end{bmatrix} \quad \Rightarrow \quad 2\mathbf{A} = \begin{bmatrix} 4 & 8 & 14 \\ 2 & 14 & 4 \end{bmatrix}$$

The **determinant** is scalar computed from the elements of a matrix via well–defined (although rather painful) rules. Determinants only exist for square matrices. The determinant of a matrix $\mathbf{A}$ is denoted as $|\mathbf{A}|$.

For a scalar (a 1×1 matrix): $|\mathbf{A}| = \mathbf{A}$.
For a 2×2 matrix: $|\mathbf{A}| = a_{11}a_{22} - a_{12}a_{21}$.
For $n \times n$ matrices ($n > 2$):

1. $\mathbf{A_{rs}} \equiv (n-1) \times (n-1)$ matrix with row r and column s removed from $\mathbf{A}$
2. $|\mathbf{A_{rs}}| \equiv$ the **minor** of element a_{rs}
3. $\theta_{rs} = (-1)^{r+s}|\mathbf{A_{rs}}| \equiv$ the **cofactor** of element a_{rs}
4. The determinant is obtained by summing the product of the elements and cofactors for any row or column of $\mathbf{A}$. By using row i of $\mathbf{A}$, we get $|\mathbf{A}| = \sum_{j=1}^{n} a_{ij}\theta_{ij}$

Example – Determinant of a 3×3 matrix
We compute the determinant of a 3×3 matrix, making use of its first row.

$$\mathbf{A} \quad = \quad \begin{bmatrix} 10 & 5 & 2 \\ 6 & 8 & 0 \\ 2 & 5 & 1 \end{bmatrix}$$

$$a_{11} = 10 \quad \mathbf{A_{11}} = \begin{bmatrix} 8 & 0 \\ 5 & 1 \end{bmatrix} \quad |\mathbf{A_{11}}| = 8(1) - 0(5) = 8 \quad \theta_{11} = (-1)^{1+1}(8) = 8$$

$$a_{12} = 5 \quad \mathbf{A_{12}} = \begin{bmatrix} 6 & 0 \\ 2 & 1 \end{bmatrix} \quad |\mathbf{A_{11}}| = 6(1) - 0(2) = 6 \quad \theta_{12} = (-1)^{1+2}(6) = -6$$

$$a_{13} = 2 \quad \mathbf{A_{13}} = \begin{bmatrix} 6 & 8 \\ 2 & 5 \end{bmatrix} \quad |\mathbf{A_{13}}| = 6(5) - 8(2) = 14 \quad \theta_{13} = (-1)^{1+3}(14) = 14$$

Then the determinant of $\mathbf{A}$ is:

$$|\mathbf{A}| \quad = \quad \sum_{j=1}^{n} a_{1j}\theta_{1j} \quad = \quad 10(8) + 5(-6) + 2(14) \quad = \quad 78$$

Note that we would have computed 78 regardless of which row and column we used.

An important result in linear algebra states that if $|\mathbf{A}| = 0$, then $\mathbf{A}$ is singular, otherwise $\mathbf{A}$ is nonsingular (full rank).

The **inverse** of a square matrix $\mathbf{A}$, denoted $\mathbf{A^{-1}}$, is a matrix such that $\mathbf{A^{-1}A} = \mathbf{I} = \mathbf{AA^{-1}}$ where $\mathbf{I}$ is the identity matrix of the same dimension as $\mathbf{A}$. A unique inverse exists if $\mathbf{A}$ is square and full rank.

The identity matrix, when multiplied by any matrix (such that matrix multiplication exists) returns the same matrix. That is: $\mathbf{AI} = \mathbf{A}$ and $\mathbf{IA} = \mathbf{A}$, as long as the dimensions of the matrices conform to matrix multiplication.

For a scalar (a 1×1 matrix): $\mathbf{A^{-1}}$=$1/\mathbf{A}$.

For a 2×2 matrix: $\mathbf{A^{-1}} = \frac{1}{|\mathbf{A}|} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$.

For $n \times n$ matrices ($n > 2$):

1. Replace each element with its cofactor (θ_{rs})
2. Transpose the resulting matrix
3. Divide each element by the determinant of the original matrix

Example – Inverse of a 3×3 matrix

We compute the inverse of a 3×3 matrix (the same matrix as before).

$$\mathbf{A} \quad = \quad \begin{bmatrix} 10 & 5 & 2 \\ 6 & 8 & 0 \\ 2 & 5 & 1 \end{bmatrix} \qquad |\mathbf{A}| = 78$$

$$|\mathbf{A_{11}}| = 8 \qquad |\mathbf{A_{12}}| = 6 \qquad |\mathbf{A_{13}}| = 14$$

$$|\mathbf{A_{21}}| = -5 \qquad |\mathbf{A_{22}}| = 6 \qquad |\mathbf{A_{23}}| = 40$$
$$|\mathbf{A_{31}}| = -16 \qquad |\mathbf{A_{32}}| = -12 \qquad |\mathbf{A_{33}}| = 50$$

$$\theta_{11} = 8 \quad \theta_{12} = -6 \quad \theta_{13} = 14 \quad \theta_{21} = 5 \quad \theta_{22} = 6 \quad \theta_{23} = -40 \quad \theta_{31} = -16 \quad \theta_{32} = 12 \quad \theta_{33} = 50$$

$$\mathbf{A^{-1}} = \frac{1}{78}\begin{bmatrix} 8 & 5 & -16 \\ -6 & 6 & 12 \\ 14 & -40 & 50 \end{bmatrix}$$

As a check:

$$\mathbf{A^{-1}A} = \frac{1}{78}\begin{bmatrix} 8 & 5 & -16 \\ -6 & 6 & 12 \\ 14 & -40 & 50 \end{bmatrix}\begin{bmatrix} 10 & 5 & 2 \\ 6 & 8 & 0 \\ 2 & 5 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \mathbf{I_3}$$

To obtain the inverse of a diagonal matrix, simply compute the recipocal of each diagonal element.

The following results are very useful for matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and scalar λ, as long as the matrices' dimensions are conformable to the operations in use:

1. $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
2. $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$
3. $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$
4. $\mathbf{C}(\mathbf{A} + \mathbf{B}) = \mathbf{CA} + \mathbf{CB}$
5. $\lambda(\mathbf{A} + \mathbf{B}) = \lambda\mathbf{A} + \lambda\mathbf{B}$
6. $(\mathbf{A}')' = \mathbf{A}$
7. $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$
8. $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$
9. $(\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}'$
10. $(\mathbf{AB})^{\mathbf{-1}} = \mathbf{B^{-1}A^{-1}}$
11. $(\mathbf{ABC})^{\mathbf{-1}} = \mathbf{C^{-1}B^{-1}A^{-1}}$
12. $(\mathbf{A^{-1}})^{\mathbf{-1}} = \mathbf{A}$
13. $(\mathbf{A}')^{\mathbf{-1}} = (\mathbf{A^{-1}})'$

The length of a column vector $\mathbf{x}$ and the distance between two column vectors $\mathbf{u}$ and $\mathbf{v}$ are:

$$l(\mathbf{x}) = \sqrt{\mathbf{x}'\mathbf{x}} \qquad l((\mathbf{u} - \mathbf{v})) = \sqrt{(\mathbf{u} - \mathbf{v})'(\mathbf{u} - \mathbf{v})}$$

Vectors $\mathbf{x}$ and $\mathbf{w}$ are **orthogonal** if $\mathbf{x}'\mathbf{w} = 0$.

1.1 Linear Equations and Solutions

Suppose we have a system of r linear equations in s unknown variables. We can write this in matrix notation as:

$$\mathbf{Ax} = \mathbf{y}$$

where $\mathbf{x}$ is a $s \times 1$ vector of s unknowns; $\mathbf{A}$ is a $r \times s$ matrix of known coefficients of the s unknowns; and $\mathbf{y}$ is a $r \times 1$ vector of known constants on the right hand sides of the equations. This set of equations may have:

- No solution
- A unique solution
- An infinite number of solutions

A set of linear equations is **consistent** if any linear dependencies among rows of $\mathbf{A}$ also appear in the rows of $\mathbf{y}$. For example, the following system is inconsistent:

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

This is inconsistent because the coefficients in the second row of $\mathbf{A}$ are twice those in the first row, but the element in the second row of $\mathbf{y}$ is not twice the element in the first row. There will be no solution to this system of equations.

A set of equations is consistent if $r(\mathbf{A}) = r([\mathbf{Ay}])$ where $[\mathbf{Ay}]$ is the augmented matrix $[\mathbf{A}|\mathbf{y}]$. When $r(\mathbf{A})$ equals the number of unknowns, and A is square:

$$\mathbf{x} = \mathbf{A^{-1}y}$$

1.2 Projection Matrices

The goal of regression is to transform a n-dimensional column vector $\mathbf{Y}$ onto a vector $\hat{\mathbf{Y}}$ in a subspace (such as a straight line in 2-dimensional space) such that $\hat{\mathbf{Y}}$ is as close to $\mathbf{Y}$ as possible. Linear transformation of $\mathbf{Y}$ to $\hat{\mathbf{Y}}$, $\hat{\mathbf{Y}} = \mathbf{PY}$ is said to be a **projection** iff $\mathbf{P}$ is idempotent and symmetric, in which case $\mathbf{P}$ is said to be a **projection matrix**.

A square matrix $\mathbf{A}$ is **idempotent** if $\mathbf{AA} = \mathbf{A}$. If $\mathbf{A}$ is idempotent, then:

$$r(\mathbf{A}) = \sum_{i=1}^{n} a_{ii} = tr(\mathbf{A})$$

where $tr(\mathbf{A})$ is the **trace** of $\mathbf{A}$. The subspace of a projection is defined, or spanned, by the columns or rows of the projection matrix $\mathbf{P}$.

$\hat{\mathbf{Y}} = \mathbf{PY}$ is the vector in the subspace spanned by $\mathbf{P}$ that is closest to $\mathbf{Y}$ in distance. That is:

$$SS(\text{RESIDUAL}) = (\mathbf{Y} - \hat{\mathbf{Y}})'(\mathbf{Y} - \hat{\mathbf{Y}})$$

is at a minimum. Further:

$$\mathbf{e} \quad = \quad (\mathbf{I} - \mathbf{P})\mathbf{Y}$$

is a projection onto a subspace orthogonal to the subspace defined by $\mathbf{P}$.

$$\hat{\mathbf{Y}}'\mathbf{e} = (\mathbf{PY})'(\mathbf{I} - \mathbf{P})\mathbf{Y} = \mathbf{Y}'\mathbf{P}'(\mathbf{I} - \mathbf{P})\mathbf{Y} = \mathbf{Y}'\mathbf{P}(\mathbf{I} - \mathbf{P})\mathbf{Y} = \mathbf{Y}'(\mathbf{P} - \mathbf{P})\mathbf{Y} = 0$$

$$\hat{\mathbf{Y}} + \mathbf{e} = \mathbf{PY} + (\mathbf{I} - \mathbf{P})\mathbf{Y} = \mathbf{Y}$$

1.3 Vector Differentiation

Let f be a function of $\mathbf{x} = [x_1, \ldots, x_p]'$. We define:

$$\frac{df}{d\mathbf{x}} \quad = \quad \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_p} \end{bmatrix}$$

From this, we get for $p \times 1$ vector $\mathbf{a}$ and $p \times p$ symmetric matrix $\mathbf{A}$:

$$\frac{d(\mathbf{a}'\mathbf{x})}{d\mathbf{x}} = \mathbf{a} \qquad\qquad \frac{d(\mathbf{x}'\mathbf{A}\mathbf{x})}{d\mathbf{x}} = 2\mathbf{A}\mathbf{x}$$

"Proof" – Consider $p = 3$:

$$\mathbf{a}'\mathbf{x} = a_1x_1 + a_2x_2 + a_3x_3 \quad \frac{d(\mathbf{a}'\mathbf{x})}{dx_i} = a_i \quad \Rightarrow \frac{d(\mathbf{a}'\mathbf{x})}{d\mathbf{x}} = \mathbf{a}$$

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \begin{bmatrix} x_1a_{11} + x_2a_{21} + x_3a_{31} & x_1a_{12} + x_2a_{22} + x_3a_{32} & x_1a_{13} + x_2a_{23} + x_3a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} =$$

$$= x_1^2a_{11} + x_1x_2a_{21} + x_1x_3a_{31} + x_1x_2a_{12} + x_2^2a_{22} + x_2x_3a_{32} + x_1x_3a_{13} + x_2x_3a_{23} + x_3^2a_{33}$$

$$\Rightarrow \frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial x_i} = 2a_{ii}x_i + 2\sum_{j \neq i} a_{ij}x_j \qquad (a_{ij} = a_{ji})$$

$$\Rightarrow \quad \frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial x_1} \\ \frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial x_2} \\ \frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 2a_{11}x_1 + 2a_{12}x_2 + 2a_{13}x_3 \\ 2a_{21}x_1 + 2a_{22}x_2 + 2a_{23}x_3 \\ 2a_{31}x_1 + 2a_{32}x_2 + 2a_{33}x_3 \end{bmatrix} = 2\mathbf{A}\mathbf{x}$$