5. Matrix Algebra
A Prelude to Multiple Regression

Matrices are rectangular arrays of numbers and are denoted using boldface (mostly capital) symbols.

Example: a 2×2 matrix (always #rows $\times$ #columns)

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

Example: a 4×2 matrix $\mathbf{B}$, and a 2×3 matrix $\mathbf{C}$

$$\mathbf{B} = \begin{bmatrix} 4 & 6 \\ 1 & 10 \\ 5 & 7 \\ 12 & 2 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1 & 1 & 4 \\ 2 & 4 & 3 \end{bmatrix}$$

In general, an $r \times c$ matrix is given by

$$\mathbf{A}_{r \times c} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1c} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2c} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{ic} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rj} & \cdots & a_{rc} \end{bmatrix}$$

or in abbreviated form

$$\mathbf{A}_{r \times c} = [a_{ij}], \quad i = 1, 2, \ldots, r, \ j = 1, 2, \ldots, c$$

1st subscript gives row#, 2nd subscript gives column#

Where is a_{79} or a_{44} ?

A matrix $\mathbf{A}$ is called **square**, if it has the same # of rows and columns ($r = c$).

Example:

$$\mathbf{A}_{2 \times 2} = \begin{bmatrix} 2.7 & 7.0 \\ 1.4 & 3.4 \end{bmatrix}$$

Matrices having either 1 row ($r = 1$) or 1 column ($c = 1$) are called **vectors**.

Example:
column vector $\mathbf{A}$ ($c = 1$) and row vector $\mathbf{C}'$ ($r = 1$)

$$\mathbf{A} = \begin{bmatrix} 4 \\ 7 \\ 13 \end{bmatrix}, \quad \mathbf{C}' = \begin{bmatrix} c_1 & c_2 & c_3 & c_4 \end{bmatrix}$$

Row vectors always have the prime!

Transpose: $\mathbf{A}'$ is the transpose of $\mathbf{A}$ where

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 5 & 6 \\ 2 & 4 & 3 & 7 \\ 10 & 0 & 1 & 2 \end{bmatrix}, \quad \mathbf{A}' = \begin{bmatrix} 3 & 2 & 10 \\ 1 & 4 & 0 \\ 5 & 3 & 1 \\ 6 & 7 & 2 \end{bmatrix}$$

$\mathbf{A}'$ is obtained by interchanging columns & rows of $\mathbf{A}$

a_{ij} is the typical element of $\mathbf{A}$
a'_{ij} is the typical element of $\mathbf{A}'$

$$a_{ij} = a'_{ji} \qquad (a_{12} = a'_{21})$$

Equality of Matrices: Two matrices $\mathbf{A}$ and $\mathbf{B}$ are said to be **equal** if they are of the same dimension and all corresponding elements are equal.

$\mathbf{A}_{r\times c} = \mathbf{B}_{r\times c}$ means
$a_{ij} = b_{ij}$, $i = 1, \dots, r$, $j = 1, \dots, c$.

Addition and Subtraction: To add or subtract matrices they must be of the same dimension. The result is another matrix of this dimension. If

$$\mathbf{A}_{3\times 2} = \begin{bmatrix} 4 & 6 \\ 1 & 10 \\ 5 & 7 \end{bmatrix}, \qquad \mathbf{B}_{3\times 2} = \begin{bmatrix} 2 & 3 \\ 0 & 1 \\ 7 & 5 \end{bmatrix},$$

then its sum and its difference is calculated elementwise

$$\mathbf{C} = \mathbf{A} + \mathbf{B} = \begin{bmatrix} 4+2 & 6+3 \\ 1+0 & 10+1 \\ 5+7 & 7+5 \end{bmatrix} = \begin{bmatrix} 6 & 9 \\ 1 & 11 \\ 12 & 12 \end{bmatrix}$$

$$\mathbf{D} = \mathbf{A} - \mathbf{B} = \begin{bmatrix} 4-2 & 6-3 \\ 1-0 & 10-1 \\ 5-7 & 7-5 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 1 & 9 \\ -2 & 2 \end{bmatrix}.$$

Regression Analysis

Remember, we had $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ and wrote the SLR as

$$Y_i = E(Y_i) + \epsilon_i, \qquad i = 1, 2, \ldots, n.$$

Now we are able to write the above model as

$$\mathbf{Y}_{n\times 1} = E(\mathbf{Y}_{n\times 1}) + \boldsymbol{\epsilon}_{n\times 1}$$

with the $n \times 1$ column vectors

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad E(\mathbf{Y}) = \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_n) \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Matrix Multiplication:

(1) by a scalar, which is a (1×1) matrix. Let

$$\mathbf{A} = \begin{bmatrix} 5 & 2 \\ 3 & 4 \\ 1 & 7 \end{bmatrix}$$

If the scalar is 3, then $3 \times \mathbf{A} = \mathbf{A} + \mathbf{A} + \mathbf{A}$ or

$$3 \times \mathbf{A} = \begin{bmatrix} 3\times 5 & 3\times 2 \\ 3\times 3 & 3\times 4 \\ 3\times 1 & 3\times 7 \end{bmatrix} = \begin{bmatrix} 15 & 6 \\ 9 & 12 \\ 3 & 21 \end{bmatrix}$$

Generally, if λ denotes the scalar, we get

$$\lambda \times \mathbf{A} = \begin{bmatrix} 5\lambda & 2\lambda \\ 3\lambda & 4\lambda \\ \lambda & 7\lambda \end{bmatrix} = \mathbf{A} \times \lambda$$

We can also factor out a common factor, e.g.

$$\begin{bmatrix} 15 & 5 \\ 10 & 0 \end{bmatrix} = 5 \times \begin{bmatrix} 3 & 1 \\ 2 & 0 \end{bmatrix}$$

(2) by a matrix: we write the product of two matrices $\mathbf{A}$ and $\mathbf{B}$ as $\mathbf{AB}$. For $\mathbf{AB}$ to exist, the #col's of $\mathbf{A}$ must be the same as the #rows of $\mathbf{B}$.

$$\mathbf{A}_{3\times 2} = \begin{bmatrix} 2 & 5 \\ 4 & 1 \\ 3 & 2 \end{bmatrix}, \qquad \mathbf{B}_{2\times 3} = \begin{bmatrix} 4 & 6 & -1 \\ 0 & 5 & 8 \end{bmatrix}$$

Let $\mathbf{C} = \mathbf{AB}$. You get c_{ij} by taking the inner product of the ith row of $\mathbf{A}$ and the jth column of $\mathbf{B}$, that is

$$c_{ij} = \sum_{k=1}^{\text{\#col's in } \mathbf{A}} a_{ik}b_{kj}$$

Since $i = 1, \ldots, \#\text{rows in } \mathbf{A}$, $j = 1, \ldots, \#\text{col's in } \mathbf{B}$ the resulting matrix $\mathbf{C}$ has dimension:

(#rows in $\mathbf{A}$)$\times$(#col's in $\mathbf{B}$).

For $\mathbf{C}$ to exist, (#col's in $\mathbf{A}$)=(#rows in $\mathbf{B}$).

Hence, for $\mathbf{A}_{3\times 2}\mathbf{B}_{2\times 3}$ we get the 3×3 matrix

$$\mathbf{C} = \begin{bmatrix} 2\times 4 + 5\times 0 & 2\times 6 + 5\times 5 & 2\times(-1) + 5\times 8 \\ 4\times 4 + 1\times 0 & 4\times 6 + 1\times 5 & 4\times(-1) + 1\times 8 \\ 3\times 4 + 2\times 0 & 3\times 6 + 2\times 5 & 3\times(-1) + 2\times 8 \end{bmatrix} = \begin{bmatrix} 8 & 37 & 38 \\ 16 & 29 & 4 \\ 12 & 28 & 13 \end{bmatrix}$$

Note, this is different to $\mathbf{D}_{2\times 2} = \mathbf{B}_{2\times 3}\mathbf{A}_{3\times 2}$ which gives the 2×2 matrix

$$\mathbf{D} = \begin{bmatrix} 4\times 2 + 6\times 4 - 1\times 3 & 4\times 5 + 6\times 1 - 1\times 2 \\ 0\times 2 + 5\times 4 + 8\times 3 & 0\times 5 + 5\times 1 + 8\times 2 \end{bmatrix} = \begin{bmatrix} 29 & 24 \\ 44 & 21 \end{bmatrix}$$

For $\mathbf{AB}$ we say, $\mathbf{B}$ is premultiplied by $\mathbf{A}$ or $\mathbf{A}$ is postmultiplied by $\mathbf{B}$.

Regression Analysis

Remember our SLR with all means on the straight line

$$E(Y_i) = \beta_0 + \beta_1 X_i, \qquad i = 1, 2, \ldots, n$$

With

$$\mathbf{X}_{n\times 2} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}, \quad \boldsymbol{\beta}_{2\times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

we get the mean model

$$E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_1 \\ \beta_0 + \beta_1 X_2 \\ \vdots \\ \beta_0 + \beta_1 X_n \end{bmatrix}$$

Thus we rewrite the SLR as

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

Important Matrices in Regression:

$$\begin{aligned} \mathbf{Y}'\mathbf{Y} &= \begin{bmatrix} Y_1 & Y_2 & \ldots & Y_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \\ &= \sum_{i=1}^{n} Y_i^2 \\ \mathbf{X}'\mathbf{X} &= \begin{bmatrix} 1 & 1 & \ldots & 1 \\ X_1 & X_2 & \ldots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \\ &= \begin{bmatrix} n & \sum_i X_i \\ \sum_i X_i & \sum_i X_i^2 \end{bmatrix} \\ \mathbf{X}'\mathbf{Y} &= \begin{bmatrix} 1 & 1 & \ldots & 1 \\ X_1 & X_2 & \ldots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \\ &= \begin{bmatrix} \sum_i Y_i \\ \sum_i X_i Y_i \end{bmatrix} \end{aligned}$$

Special Types of Matrices:

Symmetric Matrix, if $\mathbf{A} = \mathbf{A}'$, $\mathbf{A}$ is said to be symmetric, e.g.

$$\mathbf{A} = \begin{bmatrix} 2 & 5 & 8 \\ 5 & 1 & 3 \\ 8 & 3 & 2 \end{bmatrix}, \qquad \mathbf{A}' = \begin{bmatrix} 2 & 5 & 8 \\ 5 & 1 & 3 \\ 8 & 3 & 2 \end{bmatrix}$$

A symmetric matrix necessarily is square! Any product like $\mathbf{Z}'\mathbf{Z}$ is symmetric.

Diagonal Matrix is a square matrix whose off-diagonal elements are all zeros, e.g.

$$\mathbf{A} = \begin{bmatrix} 7 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} b_{11} & 0 & 0 & 0 \\ 0 & b_{22} & 0 & 0 \\ 0 & 0 & b_{33} & 0 \\ 0 & 0 & 0 & b_{44} \end{bmatrix}$$

Identity Matrix $\mathbf{I}$ is a diagonal matrix whose elements are all 1s, e.g. $\mathbf{B}$ above with $b_{ii} = 1$, $i = 1, 2, 3, 4$.
Pre- and postmultiplying by $\mathbf{I}$ does not change a matrix, $\mathbf{A} = \mathbf{IA} = \mathbf{AI}$.

Vector and matrix with all elements Unity

A column vector with all elements 1 is denoted by $\mathbf{1}$, a square matrix with all elements 1 is denoted by $\mathbf{J}$,

$$\mathbf{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \mathbf{J} = \begin{bmatrix} 1 & \ldots & 1 \\ \vdots & & \vdots \\ 1 & \ldots & 1 \end{bmatrix}$$

Note that for an $n \times 1$ vector $\mathbf{1}$ we obtain

$$\mathbf{1}'\mathbf{1} = \begin{bmatrix} 1 & \ldots & 1 \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = n$$

and

$$\mathbf{1}\mathbf{1}' = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} 1 & \ldots & 1 \end{bmatrix} = \begin{bmatrix} 1 & \ldots & 1 \\ \vdots & & \vdots \\ 1 & \ldots & 1 \end{bmatrix} = \mathbf{J}_{n \times n}$$

Zero vector A column vector with all elements 0

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

Linear Dependence and Rank of Matrix

Consider the following matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 4 & 1 \\ 2 & 2 & 6 & 6 \\ 3 & 4 & 10 & 1 \end{bmatrix}$$

Think of $\mathbf{A}$ as being made up of 4 column vectors

$$\mathbf{A} = \left[\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix} \begin{bmatrix} 4 \\ 6 \\ 10 \end{bmatrix} \begin{bmatrix} 1 \\ 6 \\ 1 \end{bmatrix} \right]$$

The third column is 2 times the first plus the second

$$\begin{bmatrix} 4 \\ 6 \\ 10 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix}$$

We say the columns of $\mathbf{A}$ are **linearly dependent** (or $\mathbf{A}$ is **singular**). When no such relationships exist, $\mathbf{A}$'s columns are said to be **linearly independent**.

The **rank of a matrix** is the number of linearly independent columns (in the example, its rank is 3).

Inverse of a Matrix

Q: What's the inverse of a number (6)?

A: Its reciprocal (1/6)!

A number multiplied by its inverse always equals 1

Generally, for the inverse $1/x$ of a scalar x

$$x\frac{1}{x} = \frac{1}{x}x = x^{-1}x = xx^{-1} = 1$$

In matrix algebra, the inverse of $\mathbf{A}$ is the matrix $\mathbf{A}^{-1}$, for which

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

In order for $\mathbf{A}$ to have an inverse:
- $\mathbf{A}$ must be square,
- col's of $\mathbf{A}$ must be linearly independent.

Example: Inverse of a matrix

$$\mathbf{A}_{2\times2} = \begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix}, \qquad \mathbf{A}_{2\times2}^{-1} = \begin{bmatrix} -0.1 & 0.4 \\ 0.3 & -0.2 \end{bmatrix}$$

$$\mathbf{A}\mathbf{A}^{-1} = \begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} -0.1 & 0.4 \\ 0.3 & -0.2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{A}^{-1}\mathbf{A} = \begin{bmatrix} -0.1 & 0.4 \\ 0.3 & -0.2 \end{bmatrix} \begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Example: Inverse of a diagonal matrix

$$\mathbf{D} = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix}, \quad \mathbf{D}^{-1} = \begin{bmatrix} 1/d_1 & 0 & 0 \\ 0 & 1/d_2 & 0 \\ 0 & 0 & 1/d_3 \end{bmatrix}$$

e.g.,

$$\mathbf{D} = \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix}, \quad \mathbf{D}^{-1} = \begin{bmatrix} 1/3 & 0 \\ 0 & 1/4 \end{bmatrix}$$

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

Finding the Inverse: The 2×2 case

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad \mathbf{A}^{-1} = \frac{1}{D} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

where $D = ad - bc$ denotes the **determinant** of $\mathbf{A}$. If $\mathbf{A}$ is singular then $D = 0$ and no inverse would exist.

$$\begin{aligned} \mathbf{A}\mathbf{A}^{-1} &= \begin{bmatrix} a & b \\ c & d \end{bmatrix} \frac{1}{D} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \\ &= \frac{1}{D} \begin{bmatrix} ad - bc & -ab + ba \\ cd - dc & -cb + da \end{bmatrix} = \frac{1}{D} \begin{bmatrix} D & 0 \\ 0 & D \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I} \end{aligned}$$

Example:

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

Determinant $D = ad - bc = 2 \times 1 - 4 \times 3 = -10$

$$\mathbf{A}^{-1} = -\frac{1}{10} \begin{bmatrix} 1 & -4 \\ -3 & 2 \end{bmatrix} = \begin{bmatrix} -0.1 & 0.4 \\ 0.3 & -0.2 \end{bmatrix}$$

Regression Analysis

Principal inverse matrix in regression is the inverse of

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} n & \sum_i X_i \\ \sum_i X_i & \sum_i X_i^2 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Its determinant is

$$\begin{aligned} D &= n\sum_i X_i^2 - \Big(\sum_i X_i\Big)^2 = n\left(\sum_i X_i^2 - \frac{1}{n}\big(n\overline{X}\big)^2\right) \\ &= n\left(\sum_i X_i^2 - n\overline{X}^2\right) = n\left(\sum_i (X_i - \overline{X})^2\right) \\ &= nS_{XX} \neq 0. \end{aligned}$$

Thus

$$\begin{aligned} (\mathbf{X}'\mathbf{X})^{-1} &= \frac{1}{nS_{XX}}\begin{bmatrix} \sum_i X_i^2 & -\sum_i X_i \\ -\sum_i X_i & n \end{bmatrix} \\ &= \frac{1}{S_{XX}}\begin{bmatrix} \frac{1}{n}\sum_i X_i^2 & -\overline{X} \\ -\overline{X} & 1 \end{bmatrix} \end{aligned}$$

Uses of Inverse Matrix

- In ordinary algebra, we solve an equation of the type

$$5y = 20$$

by multiplying both sides by the inverse of 5

$$\frac{1}{5}(5y) = \frac{1}{5}20$$

and obtain $y = 4$.

- System of equations:

$$\begin{aligned} 2y_1 + 4y_2 &= 20 \\ 3y_1 + y_2 &= 10 \end{aligned}$$

With matrix algebra we rewrite this system as

$$\begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 20 \\ 10 \end{bmatrix}$$

Thus, we have to solve

$$\mathbf{AY} = \mathbf{C}$$

$$\mathbf{AY} = \mathbf{C}$$

Premultiplying with the inverse $\mathbf{A}^{-1}$ gives

$$\begin{aligned} \mathbf{A}^{-1}\mathbf{AY} &= \mathbf{A}^{-1}\mathbf{C} \\ \mathbf{Y} &= \mathbf{A}^{-1}\mathbf{C} \end{aligned}$$

The solution of these equations then is

$$\begin{aligned} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} &= \begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 20 \\ 10 \end{bmatrix} \\ &= \begin{bmatrix} -0.1 & 0.4 \\ 0.3 & -0.2 \end{bmatrix} \begin{bmatrix} 20 \\ 10 \end{bmatrix} \\ &= \begin{bmatrix} 2 \\ 4 \end{bmatrix} \end{aligned}$$

Some Basic Matrix Facts

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

Random Vectors and Matrices

A random vector is a vector of random variables, e.g. $\mathbf{Y} = \begin{bmatrix} Y_1 & Y_2 & \dots & Y_n \end{bmatrix}'$

The **expected value** of $\mathbf{Y}$ is a vector, denoted by $E(\mathbf{Y}) = \begin{bmatrix} E(Y_1) & E(Y_2) & \dots & E(Y_n) \end{bmatrix}'$.

Regression Example:

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}; \quad E(\boldsymbol{\epsilon}) = \begin{bmatrix} E(\epsilon_1) \\ E(\epsilon_2) \\ \vdots \\ E(\epsilon_n) \end{bmatrix} = \mathbf{0}_{n\times 1}$$

The usual rules for expectation still work:
Suppose $\mathbf{V}$ and $\mathbf{W}$ are random vectors and $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are matrices of constants. Then

$$E(\mathbf{AV} + \mathbf{BW} + \mathbf{C}) = \mathbf{A}E(\mathbf{V}) + \mathbf{B}E(\mathbf{W}) + \mathbf{C}$$

Regression Example: Find $E(\mathbf{Y}) = E(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon})$

$$E(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) = E(\mathbf{X}\boldsymbol{\beta}) + E(\boldsymbol{\epsilon}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{0} = \mathbf{X}\boldsymbol{\beta}$$

Variance-Covariance Matrix of a Random Vector

For a random vector $\mathbf{Z}_{n\times 1}$ define $Var(\mathbf{Z}) =$

$$\begin{bmatrix} Var(Z_1) & Cov(Z_1, Z_2) & \dots & Cov(Z_1, Z_n) \\ Cov(Z_2, Z_1) & Var(Z_2) & \dots & Cov(Z_2, Z_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(Z_n, Z_1) & Cov(Z_n, Z_2) & \dots & Var(Z_n) \end{bmatrix}$$

where $Cov(Z_i, Z_j) = E\Big[(Z_i - E(Z_i))(Z_j - E(Z_j))\Big] = Cov(Z_j, Z_i)$. It is a symmetric ($n \times n$) matrix.

If Z_i and Z_j are independent, then $Cov(Z_i, Z_j) = 0$.

Regression Example: because we assumed n independent random errors ϵ_i, each with the same variance σ^2, we have

$$Var(\boldsymbol{\epsilon}) = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}_{n\times n}$$

Rules for a Variance-Covariance Matrix

Remember: if V is a r.v. and a, b are constant terms, then

$$Var(aV + b) = Var(aV) = a^2 Var(V)$$

Suppose now that $\mathbf{V}$ is a random vector and $\mathbf{A}$, $\mathbf{B}$ are matrices of constants. Then

$$Var(\mathbf{AV} + \mathbf{B}) = Var(\mathbf{AV}) = \mathbf{A}Var(\mathbf{V})\mathbf{A}'$$

Regression Analysis: Find $Var(\mathbf{Y}) = Var(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon})$

$$Var(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) = Var(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}_{n\times n}$$

Off-diagonal elements are zero because the ϵ_i's, and hence the Y_i's, are independent.

SLR in Matrix Terms

Now we can write the SLR in matrix terms compactly as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

and we assume that

- $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$
- $\boldsymbol{\beta}$ and σ^2 are unknown parameters
- $\mathbf{X}$ is a constant matrix

Consequences: $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ and $Var(\mathbf{Y}) = \sigma^2 \mathbf{I}$.

In the next step we define the Least Squares (LS) estimators (b_0, b_1) using matrix notation.

Normal Equations: Remember the LS criterion

$$Q = \sum_{i=1}^{n} (Y_i - (\beta_0 + \beta_1 X_i))^2 = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

Recall that when we take derivatives of Q w.r.t. β_0 and β_1 and set the resulting equations equal to zero, we get the normal equations

$$\begin{aligned} nb_0 + \quad n\overline{X}b_1 &= n\overline{Y} \\ n\overline{X}b_0 + \sum_i X_i^2 b_1 &= \sum_i X_i Y_i \end{aligned}$$

Let's write these equations in matrix form

$$\begin{bmatrix} n & n\overline{X} \\ n\overline{X} & \sum_i X_i^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} n\overline{Y} \\ \sum_i X_i Y_i \end{bmatrix}$$

But with $\mathbf{b}_{2\times 1} = (b_0 \ \ b_1)'$, this is exactly equivalent to

$$(\mathbf{X}'\mathbf{X})\mathbf{b} = (\mathbf{X}'\mathbf{Y})$$

Premultiplying with the inverse $(\mathbf{X}'\mathbf{X})^{-1}$ gives

$$\begin{aligned} (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})\mathbf{b} &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y}) \\ \mathbf{b} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \end{aligned}$$

Fitted Values and Residuals

Remember $\hat{Y}_i = b_0 + b_1 X_i$. Because

$$\begin{bmatrix} \hat{Y}_1 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} b_0 + b_1 X_1 \\ \vdots \\ b_0 + b_1 X_n \end{bmatrix}$$

we write the vector of fitted values as

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{b}$$

With $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ we get

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{b} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

We can express this by using the $(n \times n)$ **Hat Matrix**

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

(it puts the hat on $\mathbf{Y}$) as

$$\hat{\mathbf{Y}} = \mathbf{H}\mathbf{Y}.$$

$\mathbf{H}$ is **symmetric** $(\mathbf{H} = \mathbf{H}')$ & **idempotent** $(\mathbf{H}\mathbf{H} = \mathbf{H})$

Symmetric:

$$\begin{aligned} \mathbf{H}' &= \Big(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Big)' \overset{9.}{=} \mathbf{X}\Big((\mathbf{X}'\mathbf{X})^{-1}\Big)'\mathbf{X}' \\ &\overset{8.}{=} \mathbf{X}\Big((\mathbf{X}'\mathbf{X})'\Big)^{-1}\mathbf{X}' \overset{5.}{=} \mathbf{X}\Big(\mathbf{X}'\mathbf{X}\Big)^{-1}\mathbf{X}' = \mathbf{H} \end{aligned}$$

Idempotent: because $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{I}$ we have

$$\begin{aligned} \mathbf{HH} &= \Big(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Big)\Big(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Big) \\ &= \mathbf{XI}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{H} \end{aligned}$$

With these results we get $(\mathbf{HX} = \mathbf{X})$ $(\mathbf{HIH} = \mathbf{H})$

$$\begin{aligned} E(\hat{\mathbf{Y}}) &= E(\mathbf{HY}) = \mathbf{HX}\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta} \\ Var(\hat{\mathbf{Y}}) &= Var(\mathbf{HY}) = \mathbf{H}\ \sigma^2\mathbf{I}\ \mathbf{H} = \sigma^2\mathbf{H}. \end{aligned}$$

Residuals: $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{IY} - \mathbf{HY} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$.

Like $\mathbf{H}$, also $\mathbf{I} - \mathbf{H}$ is symmetric and idempotent.

$$\begin{aligned} E(\mathbf{e}) &= (\mathbf{I} - \mathbf{H})E(\mathbf{Y}) = \mathbf{0} \\ Var(\mathbf{e}) &= Var((\mathbf{I} - \mathbf{H})\mathbf{Y}) = \sigma^2(\mathbf{I} - \mathbf{H}) \end{aligned}$$

Inferences in Regression Analysis

Distribution of LS Estimates

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{CY} = \begin{bmatrix} c_{11} & \dots & c_{1n} \\ c_{21} & \dots & c_{2n} \end{bmatrix} \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

with $\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ a $2 \times n$ matrix of constants. Thus, each element of $\mathbf{b}$ is a linear combination of independent normals, Y_i's, and therefore a normal r.v.

$$\begin{aligned} E(\mathbf{b}) &= E\Big((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\Big) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{Y}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{I}\boldsymbol{\beta} = \boldsymbol{\beta} \\ Var(\mathbf{b}) &= Var\Big((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\Big) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\ Var(\mathbf{Y})\ \Big((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Big)' \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\ \sigma^2\mathbf{I}\ \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2\mathbf{I}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

With the previous result we have

$$Var(\mathbf{b}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} = \frac{\sigma^2}{S_{XX}} \begin{bmatrix} \frac{1}{n}\sum_i X_i^2 & -\overline{X} \\ -\overline{X} & 1 \end{bmatrix}$$

Its estimator is

$$\widehat{Var}(\mathbf{b}) = \frac{MSE}{S_{XX}} \begin{bmatrix} \frac{1}{n}\sum_i X_i^2 & -\overline{X} \\ -\overline{X} & 1 \end{bmatrix}$$

As covariance/correlation between b_0 and b_1 we get

$$\begin{aligned} Cov(b_0, b_1) &= -\frac{\sigma^2}{S_{XX}}\overline{X} \\ Corr(b_0, b_1) &= \frac{Cov(b_0, b_1)}{\sqrt{Var(b_0)Var(b_1)}} = \frac{-\overline{X}}{\sqrt{\frac{1}{n}\sum_i X_i^2}} \end{aligned}$$

b_0, b_1 are not independent! Together we have

$$\mathbf{b} \sim N\Big(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\Big)$$

This is used to construct CI's and tests regarding $\boldsymbol{\beta}$ as before.

Estimate the Mean of the Response at X_h

Recall our estimate for $E(Y_h) = \beta_0 + \beta_1 X_h$ is

$$\hat{Y}_h = b_0 + b_1 X_h = \mathbf{X}_h'\mathbf{b}$$

where $\mathbf{X}_h' = (1 \;\; X_h)$. The fitted value is a normal r.v. with mean and variance

$$\begin{aligned} E(\hat{Y}_h) &= E(\mathbf{X}_h'\mathbf{b}) = \mathbf{X}_h'E(\mathbf{b}) \\ &= \mathbf{X}_h'\boldsymbol{\beta} \\ Var(\hat{Y}_h) &= Var(\mathbf{X}_h'\mathbf{b}) = \mathbf{X}_h'Var(\mathbf{b})\mathbf{X}_h \\ &= \mathbf{X}_h'\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h \\ &= \sigma^2\mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h \end{aligned}$$

Thus,

$$\begin{aligned} \frac{\hat{Y}_h - \mathbf{X}_h'\boldsymbol{\beta}}{\sqrt{\sigma^2 \cdot \mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h}} &\sim N(0,1) \\ \frac{\hat{Y}_h - \mathbf{X}_h'\boldsymbol{\beta}}{\sqrt{MSE \cdot \mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h}} &\sim t(n-2) \end{aligned}$$

What is $\mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h$?

$$\begin{aligned}
&= \left[\begin{array}{cc} 1 & X_h \end{array}\right] \frac{1}{S_{XX}} \left[\begin{array}{cc} \frac{1}{n}\sum_i X_i^2 & -\overline{X} \\ -\overline{X} & 1 \end{array}\right] \left[\begin{array}{c} 1 \\ X_h \end{array}\right] \\
&= \frac{1}{S_{XX}} \left[\begin{array}{cc} \frac{1}{n}\sum_i X_i^2 - \overline{X}X_h & -\overline{X} + X_h \end{array}\right] \left[\begin{array}{c} 1 \\ X_h \end{array}\right] \\
&= \frac{1}{S_{XX}} \left(\frac{1}{n}\sum_i X_i^2 - \overline{X}X_h - \overline{X}X_h + X_h^2\right) \\
&= \frac{1}{S_{XX}} \left(\frac{1}{n}\Big(S_{XX} + n\overline{X}^2\Big) - 2\overline{X}X_h + X_h^2\right) \\
&= \frac{1}{n} + \frac{1}{S_{XX}}(X_h - \overline{X})^2
\end{aligned}$$

by applying $S_{XX} = \sum_i X_i^2 - n\overline{X}^2$.

Matrix Algebra with R: Whiskey Example

```
> one <- rep(1,10); age <- c(0,.5,1,2,3,4,5,6,7,8)
> y <- c(104.6, 104.1, 104.4, 105.0, 106.0,
+        106.8, 107.7, 108.7, 110.6, 112.1)
> X <- matrix(c(one, age), ncol=2)
```



```
> XtX <- t(X) %*% X; XtX
     [,1]   [,2]
[1,] 10.0  36.50
[2,] 36.5 204.25
> solve(XtX)
            [,1]        [,2]
[1,]  0.28757480 -0.05139036
[2,] -0.05139036  0.01407955
> b <- solve(XtX) %*% t(X)%*%y; b
            [,1]
[1,] 103.5131644
[2,]   0.9552974
> H <- X %*% solve(XtX) %*% t(X)
> e <- y - H %*% y; SSE <- t(e) %*% e; SSE
         [,1]
[1,] 3.503069
> as.numeric(SSE/8) * solve(XtX)
            [,1]         [,2]
[1,]  0.12592431 -0.022502997
[2,] -0.02250300  0.006165205
> summary(lm(y ~ age))
Coefficients:
             Estimate Std.Error t value Pr(>|t|)
(Intercept)  103.51316  0.35486 291.70  < 2e-16 ***
age            0.95530  0.07852  12.17 1.93e-06 ***
```