743- Regression and Time Series

Mamikon S. Ginovyan

The Multiple Regression Model

Elements of Matrix Algebra

Definition:

A matrix is a rectangular array of real numbers:

$$A = A_{mn} = \left\| a_{kj} \right\|_{k=\overline{1,m},\, j=\overline{1,n}} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

is $(m \times n)$- **rectangular matrix**, where $\boldsymbol{m}$ is the number of **rows**, and $\boldsymbol{n}$ is the number of **columns**.

If $\boldsymbol{m} = \boldsymbol{n}$, then $A = \left\| a_{ki} \right\|_{k,j=\overline{1,n}}$ is called $(n \times n)$ - **square matrix**.

Elements of Matrix Algebra

- **Examples.**

$$A_{23} = \begin{bmatrix} 6 & 0 & 1 \\ 3 & -2 & 1 \end{bmatrix}$$ is a (2×3) -matrix, $\boldsymbol{m} = 2$, $\boldsymbol{n} = 3$.

$$A_{32} = \begin{bmatrix} 2 & 1 \\ 1 & -1 \\ 0 & 4 \end{bmatrix}$$ is a (3×2) -matrix, $\boldsymbol{m} = 2$, $\boldsymbol{n} = 3$.

$$A_{1n} = X_n = [x_1 \;\; x_2 \;\; \ldots \;\; x_n]$$ is a $(1 \times n)$ -matrix.

Elements of Matrix Algebra

- **Addition of Matrices.**

If $A = \|a_{kj}\|_{k=\overline{1,m},\, j=\overline{1,n}}$ and $B = \|b_{kj}\|_{k=\overline{1,m},\, j=\overline{1,n}}$,

then $A + B = C = \|c_{kj}\|_{k=\overline{1,m},\, j=\overline{1,n}}$,

where $c_{kj} = a_{kj} + b_{kj}$.

Similarly, can be defined $A + B + C + \cdots$

Elements of Matrix Algebra

- **Example.**

Let $A_{23} = \begin{bmatrix} 6 & 0 & 1 \\ 3 & -2 & 1 \end{bmatrix}$, and $B_{23} = \begin{bmatrix} 1 & -1 & 3 \\ 2 & 2 & -5 \end{bmatrix}$,

then

$$C_{23} = A + B = \begin{bmatrix} 7 & -1 & 4 \\ 5 & 0 & -4 \end{bmatrix}.$$

Elements of Matrix Algebra

- **Multiplication of a Matrix by a Real Number.**

Let $A_{mn} = \left\| a_{kj} \right\|_{k=\overline{1,m},\, j=\overline{1,n}}$, and $l \in R$,

then $l\, A_{mn} = \left\| l\, a_{kj} \right\|_{k=\overline{1,m},\, j=\overline{1,n}}$.

Elements of Matrix Algebra

- **Example.**

If $A_{23} = \begin{bmatrix} 6 & 0 & 1 \\ 3 & -2 & 7 \end{bmatrix}$ and $l = -2,$

then

$$l\,A_{23} = \begin{bmatrix} 6(-2) & 0(-2) & 1(-2) \\ 3(-2) & -2(-2) & 7(-2) \end{bmatrix} = \begin{bmatrix} -12 & 0 & -2 \\ -6 & 4 & -14 \end{bmatrix}.$$

Elements of Matrix Algebra

- **Matrix Multiplication.**

Let $A_{mp} = \|a_{kj}\|_{k=\overline{1,m},\, j=\overline{1,p}}$ and $B_{pn} = \|a_{kj}\|_{k=\overline{1,p},\, j=\overline{1,n}}$,

then

$$C_{mn} = A_{mp} B_{pn} = \|c_{kj}\|_{k=\overline{1,m},\, j=\overline{1,n}}$$

is a $(m \times n)$ -rectangular matrix with elements:

$$c_{kj} = \sum_{i=1}^{p} a_{ki} b_{ij}.$$

Elements of Matrix Algebra

Remark 1.

$AB \neq BA$ (in general), moreover $\boldsymbol{BA}$ may be **undefined.**

- **Example 1.**

Let $A_{32} = \begin{bmatrix} 2 & 1 \\ 1 & -1 \\ 0 & 4 \end{bmatrix}$ and $B_{23} = \begin{bmatrix} 4 & -1 & -1 \\ 2 & 0 & 2 \end{bmatrix}$.

Then

Elements of Matrix Algebra

1) $$C_{33} = A_{32}B_{23} = \begin{bmatrix} 2 & 1 \\ 1 & -1 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} 4 & -1 & -1 \\ 2 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 10 & -2 & 0 \\ 2 & -1 & -3 \\ 8 & 0 & 8 \end{bmatrix}$$

is a (3×3) -matrix, while

2) $$C_{22} = B_{23}A_{32} = \begin{bmatrix} 4 & -1 & -1 \\ 2 & 0 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & -1 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 7 & 1 \\ 4 & 10 \end{bmatrix}$$

is a (2×2) -matrix.

Elements of Matrix Algebra

- **Example 2.**

(a) $$A_{13}B_{32} = \begin{bmatrix} 2 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 4 & 3 \end{bmatrix} = C_{12},$$

(b) $B_{32}A_{13}$ is **undefined** because of the dimensions of ***A*** and ***B***,

(c) $$A_{14}B_{41} = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} = [30] = C_{11},$$

(d) $B_{41}A_{14} = C_{44}$ is (4×4)-square matrix.

Elements of Matrix Algebra

- **The Identify Matrix**

Definition 1.

The matrix $I_n = \|d_{kj}\|_{kj=\overline{1,n}} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$, where

$d_{kj} = \begin{cases} 1, & k = j \\ 0, & k \neq j \end{cases}$ is called $(n \times n)$ **-identity matrix.**

Definition 2.

The matrix $O = \|a_{kj}\|$, for which $a_{kj} = 0$, for all $k = \overline{1,m}$ and $j = \overline{1,n}$ is called **O-matrix or zero-matrix.**

Elements of Matrix Algebra

Definition 3.

Let $A_{mn} = \left\| a_{kj} \right\|_{k=\overline{1,m},\, j=\overline{1,n}}$ and $B_{mn} = \left\| b_{kj} \right\|_{k=\overline{1,m},\, j=\overline{1,n}}$, then

$A_{mn} = B_{mn} \Leftrightarrow a_{kj} = b_{kj}$ for all $k = \overline{1,m}$ and $j = \overline{1,n}$.

- **Properties of O and I matrices.**

1) $\boldsymbol{A} + \boldsymbol{O} = \boldsymbol{O} + \boldsymbol{A} = \boldsymbol{A}$;

2) $\boldsymbol{IA} = \boldsymbol{AI} = \boldsymbol{A}$.

Elements of Matrix Algebra

- **The Inverse Matrix**

Definition 4.

Let $A_{mn} = \left\| a_{kj} \right\|_{k=\overline{1,m},\, j=\overline{1,n}}$ be $(n \times n)$ -square matrix.

If a matrix, denoted by A_n^{-1}, can be found such that

$$A_n A_n^{-1} = A_n^{-1} A_n = I_n,$$

then A_n^{-1} is called the **inverse** of A_n.

Remark.

If $A_n^{-1} = \left\| b_{kj} \right\|_{kj=\overline{1,n}}$, then $\sum_{i=1}^{n} a_{ki} b_{ij} = d_{kj}$.

Elements of Matrix Algebra

Note: Let A and B be two matrices whose inverses exist. Let $C = AB$. Then the inverse of the matrix C exists and

$$C^{-1} = B^{-1}A^{-1}.$$

The Woodbury Theorem:

$$\left(A + BCD\right)^{-1} = A^{-1} - A^{-1}B\left[C^{-1} + DA^{-1}B\right]^{-1} DA^{-1}$$

where the inverses

$$A^{-1}, C^{-1} \ \text{ and } \ \left[C^{-1} + DA^{-1}B\right]^{-1} \ \text{ exist.}$$

Elements of Matrix Algebra

Note: The **Woodbury Theorem** can be used to find the inverse of some pattern matrices:

For Example:

$$\begin{bmatrix} b & a & \cdots & a \\ a & b & \cdots & a \\ \vdots & \vdots & \ddots & \vdots \\ a & a & \cdots & b \end{bmatrix}^{-1} = \begin{bmatrix} c & d & \cdots & d \\ d & c & \cdots & d \\ \vdots & \vdots & \ddots & \vdots \\ d & d & \cdots & c \end{bmatrix}$$

where:

$$d = -\frac{a}{(b-a)\big(b+a(n-1)\big)} \quad \text{and} \quad c = \frac{1}{b-a}\left[\frac{b+a(n-2)}{b+a(n-1)}\right].$$

Elements of Matrix Algebra

Example- Note 1: For $n = 2$

$$d = -\frac{a}{(b-a)(b+a)} = -\frac{a}{b^2 - a^2}$$

and $$c = \frac{1}{b-a}\left[\frac{b}{b+a}\right] = \frac{b}{b^2 - a^2}$$

Thus $$\begin{bmatrix} b & a \\ a & b \end{bmatrix}^{-1} = \frac{1}{b^2 - a^2}\begin{bmatrix} b & -a \\ -a & b \end{bmatrix}$$

Elements of Matrix Algebra

Example- Note 2: For special case $a = 0$, we have

$$\begin{bmatrix} b & 0 & \cdots & 0 \\ 0 & b & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b \end{bmatrix}^{-1} = \begin{bmatrix} 1/b & 0 & \cdots & 0 \\ 0 & 1/b & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/b \end{bmatrix}$$

Since in this case

$$d = -\frac{a}{(b-a)\left(b+a(n-1)\right)} = 0$$

$$c = \frac{1}{b-a}\left[\frac{b+a(n-2)}{b+a(n-1)}\right] = \frac{1}{b}.$$

Elements of Matrix Algebra

- **The Transpose of a Matrix.**

Definition 5.

Let $A_{mn} = \left\| a_{kj} \right\|_{k=\overline{1,m},\, j=\overline{1,n}}$.

The **transpose** of A, denoted by A', is defined to be a matrix obtained from A by **interchanging corresponding rows and columns** of A, that is first with first, second by second, and so on.

$$\text{Thus } A' = A'_{nm} = \left\| a_{jk} \right\|_{j=\overline{1,n},\, k=\overline{1,m}}.$$

Property.

The transpose of product: $(ABC)' = C'B'A'$.

Elements of Matrix Algebra

- **Example 1.**

If $A_{32} = \begin{bmatrix} 2 & 0 \\ 1 & 1 \\ 4 & 3 \end{bmatrix}$, then $A'_{23} = \begin{bmatrix} 2 & 1 & 4 \\ 0 & 1 & 3 \end{bmatrix}$.

- **Example 2.**

If $X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, then $X' = [x_1 \quad x_2 \quad x_3]$,

and $X'X = x_1^2 + x_2^2 + x_3^2$.

Elements of Matrix Algebra

- **A matrix Expression for a system of Linear Equations.**

Consider the systems of $\boldsymbol{n}$ equations

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n \end{cases} \qquad (1)$$

Elements of Matrix Algebra

- **A matrix Expression for a system of Linear Equations.**

Denoting by

$$A_n = \begin{bmatrix} a_{11} a_{12} \cdots a_{1n} \\ a_{21} a_{22} \cdots a_{2n} \\ \cdots \cdots \cdots \cdots \cdots \\ a_{n1} a_{n2} \cdots a_{nn} \end{bmatrix}; \quad X = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} \text{ and } B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}, \quad (2)$$

and using matrix operations we obtain that (1) is equivalent to (**matrix form** of (1)):

$$(1) \quad \Leftrightarrow \quad A_n X = B \qquad (3)$$

Elements of Matrix Algebra

If A_n has an inverse A_n^{-1}, then the solution of (3) (and hence (1)) is given by

$$X = A_n^{-1} B. \qquad (4)$$

Thus, to solve the system (1) follow the **steps**.

Step 1. Specify the matrices $\boldsymbol{A}$, $\boldsymbol{X}$ and $\boldsymbol{B}$ as in (2).

Step 2. Write (1) in equivalent matrix form (3).

Step 3. Find the inverse A^{-1} of $\boldsymbol{A}$.

Step 4. Multiply A^{-1} by $\boldsymbol{B}$, to get $X = A^{-1} B$.

Elements of Matrix Algebra

- **Example.**

Solve the system of equations $\begin{cases} 2x_1 + x_2 = 5 \\ x_1 - x_2 = 1 \end{cases}$

(observe that $x_1 = 2, x_2 = 1$).

- **Matrix - Solution.**

We have

$$A = \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}, \quad X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad B = \begin{bmatrix} 5 \\ 1 \end{bmatrix}.$$

Elements of Matrix Algebra

- **Matrix - Solution.**

For inverse A^{-1} we have

$$A^{-1} = \begin{bmatrix} 1/3 & 1/3 \\ 1/3 & -2/3 \end{bmatrix}.$$

So,

$$X = A^{-1}B = \begin{bmatrix} 1/3 & 1/3 \\ 1/3 & -2/3 \end{bmatrix} \begin{bmatrix} 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

and hence the solution is $x_1 = 2, x_2 = 1$.

The Multiple Regression Model

1. The Model.

Definition:

In multiple regression, the objective is to construct a probabilistic model that relates a **dependent variable** $(\,Y\,)$ to **more than one** (say $k \geq 2$) **independent** or **predictor variables**, which we denote by $x_1, \mathbf{L}\ , x_k$.

Thus, the **general additive (linear) regression model equation** is

$$Y = b_0 + b_1 x_1 + \mathbf{L} + b_k x_k + e = b_0 + \sum_{i=1}^{k} b_i x_i + e, \qquad (1)$$

where $b_i, i = 0, 1, \mathbf{L}\ , k$ are **unknown parameters** to be estimated, and e is the **error (random)** term.

The Model

If we make ***n*** independent observations on $Y: y_1, \ldots, y_n$,
then by (1) these observations we can write as

$$Y_i = b_0 + b_1 x_{i1} + \ldots + b_k x_{ik} + e_i, i = 1, 2, \ldots, n, \qquad (2)$$

x_{ij} stands for ***i*** -th observation on explanatory variable x_j,

$$i = \overline{1, n}, \quad j = \overline{1, k}.$$

Remark.

For $k = 1$, $b_0 = a$, $b_1 = b$, $x_1 = x$,

we obtain the **simple linear model:**

$$Y = b_0 + b_1 x_1 = a + b\, x.$$

Model Assumptions

1. The **relationship** between $\boldsymbol{Y}$ and $\underline{x} = (x_1, \ldots, x_k)$ is **linear** and is given by (1).

2. $x_j,\ j = \overline{1,k}$ $\ (k < n)$ are **non-random**, and **no exact linear relationship** exists between two or more x_i's.

3. $E(\varepsilon_i) = 0$ for all $i = \overline{1,n}$.

4. $Cov(\varepsilon_i, \varepsilon_j) = E[\varepsilon_i \varepsilon_j] = \sigma^2 \delta_{ij} = \begin{cases} \sigma^2, i = j \\ 0, i \neq j \end{cases}, \quad i, j = \overline{1,n}.$

5. $\varepsilon_i \sim iid - N(0, \sigma^2), \quad i = \overline{1,n}.$

Matrix form representation of the model

In multiple regression it is convenient to use the **matrix form** representation of the model.

Denote

$$b_0, b_1, \ldots, b_k$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \mathbf{L} \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 x_{11} x_{12} \ \mathbf{L} \ x_{1k} \\ 1 x_{21} x_{22} \ \mathbf{L} \ x_{2k} \\ \mathbf{L} \ \mathbf{L} \ \mathbf{L} \ \mathbf{L} \ \mathbf{L} \\ 1 x_{n1} x_{n2} \ \mathbf{L} \ x_{nk} \end{bmatrix}, \quad b = \begin{bmatrix} b_0 \\ b_1 \\ \mathbf{L} \\ b_k \end{bmatrix}, \quad and \quad e = \begin{bmatrix} e_1 \\ e_2 \\ \mathbf{L} \\ e_n \end{bmatrix}, \qquad (3)$$

Matrix form representation of the model

Remark 1.

Observe that $Y = Y_{n,1}$ is $(n \times 1)$ -matrix;

$X = X_{k+1,n}$ is $(k+1) \times n$ -matrix;

$b = b_{k+1,1}$ is $(k+1) \times 1$ -matrix;

$e = e_{n,1}$ is $(n \times 1)$ -matrix.

The **model equations** (2) we can write in the **equivalent matrix form**:

$$Y = Xb + e. \qquad (4)$$

Matrix form representation of the model

Remark 2.

For **matrix form**, the **model assumptions** are

1. The relationship between $\boldsymbol{Y}$ and $\boldsymbol{X}$ is given by (4)
2. Rank $\boldsymbol{X} = \boldsymbol{k} < \boldsymbol{n}$
3. $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon') = \sigma^2 I.$
4. $\varepsilon \sim N_n(0, \sigma^2 I)$, where $\boldsymbol{I}$ is $(n \times n)$ -identity matrix:

$$I = I_n = \left\| \delta_{kj} \right\|_{kj=\overline{1,n}} = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}.$$

Matrix form representation of the model

Remark 3.

The **data** consists of the following points:

$$\left\{\begin{matrix} x_{11}, x_{12}, \mathbf{L}\ , x_{1k}, y_1 \\ \mathbf{L}\ \mathbf{L}\ \mathbf{L}\ \mathbf{L}\ \mathbf{L}\ \mathbf{L}\ \mathbf{L}\ \mathbf{L} \\ x_{n1}, x_{n2}, \mathbf{L}\ , x_{nk}, y_n \end{matrix}\right\} = \begin{cases} x_{i1}, \mathbf{L}\ , x_{ik}, y_i \\ i = 1, n \end{cases}.$$

Remark 4.

$$E[Y \mid X] = X b \ = b_0 + b_1 x_1 + \mathbf{L} + b_k x_k .$$

Least-squares Estimation

The **objective** is to find a **vector of parameters** $\hat{b}$ which **minimizes the SSE** (= **Sum of Squares of Errors**) defined by

$$SSE = \sum_{i=1}^{n} \hat{e}_i^{\,2} = \hat{e}\,'\hat{e} \quad = \textbf{Residual Sum of Squares,} \qquad (1)$$

where

$$\hat{e} = Y - \hat{Y} \quad = \textbf{Regression Residuals} \qquad (2)$$

and

$$\hat{Y} = X\,\hat{b} \quad = \textbf{Fitted (= predicted) values of } \boldsymbol{Y} \qquad (3)$$

Least-squares Estimation

Substituting (2) and (3) into (1), we get

$$SSE = \hat{e}'\hat{e} = (Y - X\hat{b})'(Y - X\hat{b})$$
$$= YY' - Y'X\hat{b} - \hat{b}'X'Y + \hat{b}'X'X\hat{b}$$
$$= Y'Y - 2\hat{b}'X'Y + \hat{b}'X'X\hat{b} \qquad (4)$$

(since $\hat{b}'X'Y = Y'X\hat{b}$ = scalar).

Least-squares Estimation

Taking derivative w.r.t. $\hat{b}$ in (4) we obtain the **normal equation** to determine $\hat{b}$:

$$\frac{\partial SSE}{\partial \hat{b}} = -2X'Y + 2X'X\hat{b} = 0. \qquad (5)$$

Solving (5) for $\hat{b}$ we obtain the

Least-squares Estimators:

$$\hat{b} = (X'X)^{-1}(X'Y). \qquad (6)$$

Least-squares Estimation

Remark 1.
The matrix $X'X$, called the **cross-product matrix**, is guaranteed to have **inverse** $(X'X)^{-1}$, because by model assumptions $Rank\,\boldsymbol{X} = k$.

Remark 2.
The value of $\hat{b}$ given by (6) is a **minimum point** since $(X'X)$ is **positive definite matrix**, and by (5)

$$\frac{\partial^2 SSE}{\partial \hat{b}^2} = 2(X'X).$$

Least-squares Estimation

Remark 3.

To calculate $\hat{b}$ follow the steps:

Step 1. Compute $b = X'Y$.

Step 2. Calculate $(X'X)^{-1}$ (in general it will be given).

Step 3. Compute $\hat{b} = (X'X)^{-1}b$.

Least-squares Estimation

Remark 4.

The **estimated regression function** is given by equation:

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

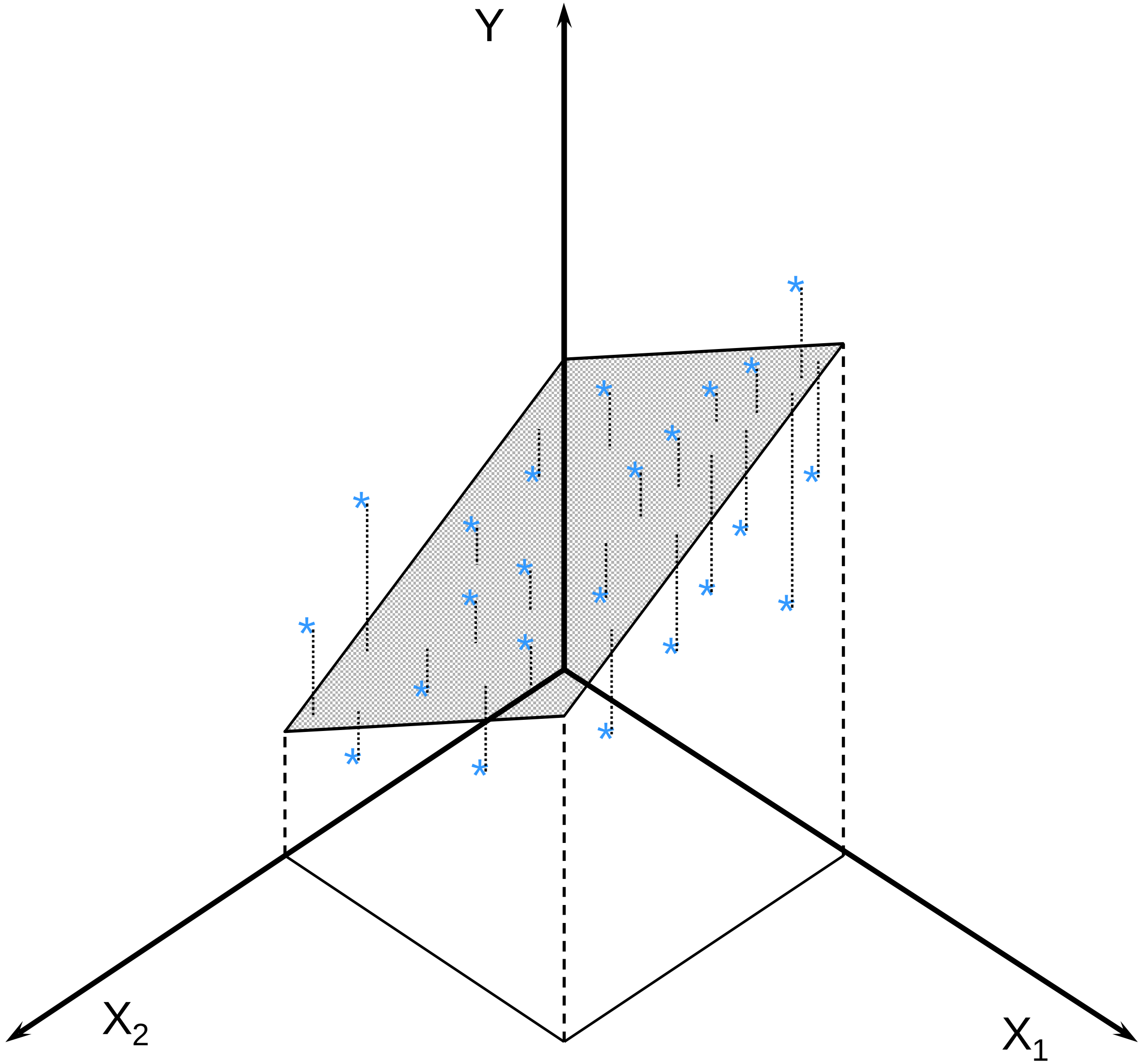
or, equivalently,

$$\hat{Y}_i = \hat{b}_0 + \hat{b}_1 x_{i1} + \cdots + \hat{b}_k x_{ik}, \quad i = 1, 2, \ldots, n,$$

where $\hat{b}_0, \hat{b}_1, \ldots, \hat{b}_k$ are the **least-squares estimators** for **regression unknown parameters** $b_0, b_1, \ldots, b_k$.

The resulting function fits a hyperplane to our sample data.

Example: Regression Surface (Hyperplane) for Two Independent Variables



A point estimator for s^2

As in the simple linear regression model (***k*** = 1), a point estimator for s^2 is the ***MSE*** equal to ***SSE*** divided by the **degrees of freedom (df)**.

Since in this case $\boldsymbol{df = n - (k+1)}$, we have

$$\hat{s}^2 = s^2 = MSE = \frac{SSE}{n-(k+1)}$$

$$= \frac{\hat{e}'\hat{e}}{n-(k+1)} = \frac{Y'Y - \hat{b}'X'Y}{n-(k+1)} . \tag{7}$$

Properties of the point estimators

1. $\hat{b}_i$ is an **unbiased** estimator for $b_i, i = \overline{0,k}$, that is,

$$E(\hat{b}_i) = b_i, \quad i = 0, 1, \mathbf{L}, k \quad \Leftrightarrow \quad E(\hat{b}) = b.$$

2. $Var(\hat{b}_i) = c_{ii} s^2, \quad i = \overline{0,k} \quad \Leftrightarrow \quad Var(\hat{b}) = s^2 (X'X)^{-1},$

where $C = (X'X)^{-1} = \left\| c_{ij} \right\|_{i,j=\overline{0,k}}$.

3. $Cov(\hat{b}_i, \hat{b}_j) = c_{ij} s^2, \quad i, j = \overline{0,k}.$

Properties of the point estimators

4. s^2, given by (7) is an **unbiased** estimator for s^2 :

$$E[s^2] = s^2.$$

5. **Gauss-Markov Theorem**.

$\hat{b}$ is the **best linear unbiased estimator (BLUE)** of b in the class of all unbiased estimators.

6. $$\hat{b}_i \sim N(b_i, c_{ii}s^2), \quad i = \overline{0, n}$$

$$\Leftrightarrow \quad \hat{b} \sim N_k(m, s^2(X'X)^{-1}).$$

Properties of the point estimators

7. $\hat{e}'\hat{e}/s^2 \sim c^2(n-k-1).$

8. $\frac{n-(k+1)}{s^2}s^2 \sim c^2(n-k-1).$

9. The statistics s^2 and $\hat{b}_i,\ i=\overline{0,k}$ are **independent** RV's.

Model Utility

First observe that if a regression model gives a **small** value of s^2 (the estimator for s^2), then the model will **accurately predict** individual $\boldsymbol{y}$-values.

For this reason, s^2 is one **measure of the usefulness**, or **utility**, of a regression model.

Now we are going to consider other ways to **assess the utility** of a regression model.

The multiple coefficient of determination

- **Notation and Explanations**

1) The total sum of squares = $SST = \sum_{i=1}^{n} (Y_i - \overline{Y})^2 = Y'Y$

is a measure of **total variation** in the observed ***y*** -values.

2) The $SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \hat{e}'\hat{e}$

is a measure of **unexplained variation** in the data.

The multiple coefficient of determination

3) The **Regression Sum of Squares**

$$= SSR = \sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2 = \hat{b}' X' X \hat{b}$$

is a measure of **explained (by the model) variation** in the data.

4) It can be shown that the following **ANOVA Identity for Regression** holds:

$$SST = SSE + SSR \qquad (1)$$

The multiple coefficient of determination

Definition 1:

The **Multiple Coefficient of Determination**, R^2, is defined to be

$$R^2 = \frac{SSR}{SST} = \frac{\hat{b}'X'X\hat{b}}{Y'Y}, \qquad (2)$$

that is, R^2 is the **proportion** of the total variation in the ***n*** observed values of the dependent variable (***Y***) that is **explained by the model relationship**.

The multiple coefficient of determination

Remark 1.

From (1) and (2) we have

$$R^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST}.$$

Remark 2.

$R = \sqrt{R^2}$ is called **multiple correlation coefficient**.

The multiple coefficient of determination

Remark 3.

Proof of (1). (Assume that $\overline{Y} = E[Y] = 0$).

We have $\hat{Y} = X\hat{b}$ and $Y = X\hat{b} + \hat{e}$.

Hence

$$SST = Y'Y = (X\hat{b} + \hat{e})'(X\hat{b} + \hat{e})$$
$$= \hat{b}'X'X\hat{b} + \hat{e}'X\hat{b} + \hat{b}'X'\hat{e} + \hat{e}'\hat{e}$$
$$= \hat{b}'X'X\hat{b} + \hat{e}'\hat{e} \qquad (\text{since} \quad X'\hat{e} = \hat{e}'X = 0)$$
$$= SSR + SSE.$$

The multiple coefficient of determination

If $\bar{Y} \neq 0$, we consider $y_i = Y_i - \bar{Y}$ and $y'y = Y'Y - n\bar{Y}^2$.
In this case

$$R^2 = \frac{SSR}{SST} = \frac{\hat{b}'X'X\hat{b} - n\bar{Y}^2}{y'y}.$$

Remark 4.
Unfortunately, there is a potential problem with R^2: Its value can be inflated by including predictors in the model that are relatively **unimportant**, yielding **overestimating** the importance of the dependent variables.
To avoid this overestimating we consider the

adjusted coefficient of determination.

The multiple coefficient of determination

Definition 2:

The **adjusted multiple coefficient of determination** (adjusted R^2) is defined to be

$$R_a^2 = 1 - \frac{MSE}{MST} = 1 - \frac{SSE/[n-(k+1)]}{SST/[n-1]} = 1 - \frac{n-1}{n-k-1} \cdot \frac{SSE}{SST}.$$

- It is clear that $R_a^2 \leq R^2$, and in fact will be much smaller when k (= the number of predictors) is large relative to n (= the number of observations) ($k < n$).

Remark: A value of R_a^2 much smaller than R^2 is a **warning flag** that the chosen model has too many predictors relative to the amount of data.

A model Utility Test

Another way to **<u>assess the utility of a regression model</u>** is to test the significance of the regression relationship between dependent ($\boldsymbol{y}$) and the independent ($\boldsymbol{x}$) variables.
For the multiple linear regression model, we test the null hypothesis

$$H_0 : b_1 = b_2 = \cdots = b_k = 0, \qquad (1)$$

which says that **<u>there is no useful relationship</u>** between $\boldsymbol{y}$ and any of the $\boldsymbol{k}$ predictors $(x_i, i = \overline{1, k})$, that is,

A model Utility Test

none of the independent variables $x_i, i = \overline{1, k}$
is significantly related to y
(the **regression relationship is not significant**),

versus the alternative hypothesis

$$H_a : \text{ at least one of } b_i, i = \overline{1, k} \text{ is not } 0, \qquad (2)$$

which says that **at least one** of the independent variables
is significantly related to y
(the **regression relationship is significant**).

A model Utility Test

An *F*-test for the complete multiple linear regression model:
Assume that the model assumptions are satisfied and we want to test the hypotheses (1) vs. (2):

$$H_0 : b_1 = b_2 = \dots = b_k = 0$$

vs.

$$H_a : \textbf{at least one} \text{ of } \quad b_i, i = \overline{1, k} \quad \text{is not } 0.$$

As a test statistic we consider the statistic

$$f = \frac{R^2 / k}{(1 - R^2) / (n - k - 1)} = \frac{SSR / k}{SSE / (n - k - 1)} = \frac{MSR}{MSE}, \tag{3}$$

where $SSR = SST - SSE$.

A model Utility Test

It can be shown that under $\boldsymbol{H_0}$ the RV defined by (3) has $F_{k,(n-k-1)}$ -distribution.

Indeed, the result follows from the following **facts**.

1) $SST = SSE + SSR \Rightarrow \dfrac{SST}{s^2} = \dfrac{SST}{s^2} + \dfrac{SSE}{s^2}.$

2) ***SSE*** and ***SSR*** are **independent.**

3) $\dfrac{SST}{s^2} \sim c^2(n-1).$

A model Utility Test

4) $\dfrac{SSE}{s^{2}} \sim c^{2}(n-k-1).$

5) If $X \sim c^{2}(n),\ \ Y \sim c^{2}(m),$ and $\boldsymbol{X}$ and $\boldsymbol{Y}$ are **independent**,

then $X+Y \sim c^{2}(n+m).$

From 1) - 5) we conclude that $\dfrac{SSR}{s^{2}}$ has c^{2}-distribution with

$$df = (n-1)-(n-k-1) = n-1-n+k+1 = k.$$

A model Utility Test

Thus,

6) $\frac{SSR}{s^2} \sim c^2(k),$ and hence

$$f = \frac{SSR/k}{SSE/(n-k-1)} = \frac{SSR/(s^2 k)}{SSE/[s^2(n-k-1)]} \sim F_{k,n-k-1}.$$

Now, for given **significance level α** , we denote by $F_a = F_{a,k,n-k-1}$ the **α -upper percentile** (= **Critical Value**) of ***F*** –distribution:

$$P(f \geq F_a) = a.$$

This **Critical Value**) can be found from **F-Table.**

A model Utility Test

For the corresponding ***P*** **–value** we have

$$\boldsymbol{P}\text{–value} = P(f \geq f(obs)),$$

where $\boldsymbol{f(obs)}$ is the **observed value** of the **test statistic** f given by (3).

Therefore we can set up the
Decision Rule:

- **Reject** $\boldsymbol{H_0}$ if $f(obs) \geq F_a$ (or P –value $<$ α).

- **Do not reject** $\boldsymbol{H_0}$ if $f(obs) < F_a$ (or $P - value \geq a$).

A model Utility Test

- **An *F*-test for a reduced model.**

In some situations we are interested in reduced model involving only $\boldsymbol{m}$ (out of $\boldsymbol{k}$ ($\boldsymbol{m} < \boldsymbol{k}$)) independent variables.
That is, if the **complete model** is

Model $\quad C : Y = b_0 + b_1 x_1 + \mathbf{L} + b_m x_m + b_{m+1} x_{m+1} + \mathbf{L} + b_k x_k + e\,.$

then the **reduced model** is

Model $\quad R : Y = b_0 + b_1 x_1 + \mathbf{L} + b_m x_m + e\,.$

To test the significance of the reduced model parameters we test the null hypothesis

$$H_0(R) : b_{m+1} = b_{m+2} = \mathbf{L}\ b_k = 0. \qquad (1)$$

A model Utility Test

The *F* -test can be carried out as in the complete model.

Denoting by ***SSE(R)*** and ***SSE(C)*** the ***SSE***'s for **reduced** and **complete** models, respectively, we have

$$SSE(R) = SSE(C) + [SSE(R) - SSE(C)] \qquad (2)$$

Observe that $SSE(C) \leq SSE(R)$.

A model Utility Test

It can be shown that

$$c_1^2 = \frac{SSE(R)}{s^2} \sim c^2(n-m-1),$$

$$c_2^2 = \frac{SSE(C)}{s^2} \sim c^2(n-k-1), \quad \text{hence by (2)}$$

$$c_3^2 = \frac{SSE(R)-SSE(C)}{s^2} \sim c^2(k-m).$$

A model Utility Test

Hence to test $H_0(R)$ given by (1) we can use the **test statistic**

$$f = \frac{c_3^2/(k-m)}{c_2^2/(n-k-2)} \sim F_{(k-m),(n-k-1)} \quad \text{under} \quad H_0(R).$$

The **Decision Rule** is the same as in complete model.

Remark.

If $\boldsymbol{m} = 0$, we obtain the complete model.

ANOVA Table for Multiple Regression Model

The following **ANOVA table** is useful for computations.

Source of Variation	Sum of Squares	Degree of Freedom	Mean Squares	F-Ratio
Regression	SSR	k	$MSR = \dfrac{SSR}{k}$	$f = \dfrac{MSR}{MSE}$
Error	SSE	$n - k - 1$	$MSE = \dfrac{SSR}{n-k-1}$	
Total	SST	$n - 1$		

ANOVA Table for Multiple Regression Model

The following method for calculating ***SSR*** and ***SSE*** is useful.

Step1.

Calculate

$$A = \sum_{i=1}^{n} Y_i, \quad B = \sum_{i=1}^{n} Y_i^2, \quad D = N(\overline{Y})^2 = \frac{A^2}{N}.$$

Step2.

Calculate the $(n \times 1)$ -vector $\hat{Y} = Xb$ and obtain

$$H = \sum_{i=1}^{n} [\hat{Y}_i]^2 = \hat{b}' X' X \hat{b} = b' X' Y.$$

ANOVA Table for Multiple Regression Model

Step3.

Obtain $SST = B - D$.

Step4.

Obtain $SSR = H - D$.

Step5.

Obtain $SSE = B - H$

(or use the ANOVA Identity $SSE = SST - SSR$).