

# **LINEAR REGRESSION ANALYSIS**

## **MODULE – III**

### **Lecture - 8**

# **Multiple Linear Regression Analysis**

**Dr. Shalabh**

**Department of Mathematics and Statistics**

**Indian Institute of Technology Kanpur**

We consider now the problem of regression when study variable depends on more than one explanatory or independent variables, called as multiple linear regression model. This model generalizes the simple linear regression in two ways. It allows the mean function  $E(y)$  to depend on more than one explanatory variables and to have shapes other than straight lines, although it does not allow for arbitrary shapes.

### The multiple linear regression model

Let  $y$  denotes the dependent (or study) variable that is linearly related to  $k$  independent (or explanatory) variables  $X_1, X_2, \dots, X_k$  through the parameters  $\beta_1, \beta_2, \dots, \beta_k$  and we write

$$y = X_1\beta_1 + X_2\beta_2 + \dots + X_k\beta_k + \varepsilon.$$

This is called as the multiple linear regression model. The parameters  $\beta_1, \beta_2, \dots, \beta_k$  are the regression coefficients associated with  $X_1, X_2, \dots, X_k$  respectively and  $\varepsilon$  is the random error component reflecting the difference between the observed and fitted linear relationship. There can be various reasons for such difference, e.g., joint effect of those variables not included in the model, random factors which can not be accounted in the model etc.

The  $j^{\text{th}}$  regression coefficient  $\beta_j$  represents the expected change in  $y$  per unit change in  $j^{\text{th}}$  independent variable  $X_j$ . Assuming  $E(\varepsilon) = 0$ ,

$$\beta_j = \frac{\partial E(y)}{\partial X_j}.$$

## Linear model

A model is said to be linear when it is linear in parameters. In such a case  $\frac{\partial y}{\partial \beta_j}$  (or equivalently  $\frac{\partial E(y)}{\partial X_j}$ ) should not depend on any  $\beta$ 's. For example

i)  $y = \beta_1 + \beta_2 X$  is a linear model as it is linear in parameter.

ii)  $y = \beta_1 X^{\beta_2}$  can be written as

$$\log y = \log \beta_1 + \beta_2 \log X$$

$$y^* = \beta_1^* + \beta_2 x^*$$

which is linear in parameter  $\beta_1^*$  and  $\beta_2$ , but nonlinear in variables  $y^* = \log y, x^* = \log x$ . So it is a linear model.

iii)  $y = \beta_1 + \beta_2 X + \beta_3 X^2$

is linear in parameters  $\beta_1, \beta_2$  and  $\beta_3$  but it is nonlinear in variables  $X$ . So it is a linear model.

iv)  $y = \beta_1 + \frac{\beta_2}{X - \beta_3}$

is nonlinear in parameters and variables both. So it is a nonlinear model.

v)  $y = \beta_1 + \beta_2 X^{\beta_3}$

is nonlinear in parameters and variables both. So it is a nonlinear model.

vi)  $y = \beta_1 + \beta_2 X + \beta_3 X^2 + \beta_4 X^3$

is a cubic polynomial model which can be written as  $y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$

which is linear in parameters  $\beta_1, \beta_2, \beta_3, \beta_4$  and linear in variables  $X_2 = X, X_3 = X^2, X_4 = X^3$ . So it is a linear model.

### Example

The income and education of a person are related. It is expected that, on an average, higher level of education provides higher income. So a simple linear regression model can be expressed as

$$\text{income} = \beta_1 + \beta_2 \text{ education} + \varepsilon.$$

Not that  $\beta_2$  reflects the change in income with respect to per unit change in education and  $\beta_1$  reflects the income when education is zero as it is expected that even an illiterate person can also have some income.

Further this model neglects that most people have higher income when they are older than when they are young, regardless of education. So  $\beta_2$  will over-state the marginal impact of education. If age and education are positively correlated, then the regression model will associate all the observed increase in income with an increase in education. So better model is

$$\text{income} = \beta_1 + \beta_2 \text{ education} + \beta_3 \text{ age} + \varepsilon.$$

Usually it is observed that the income tends to rise less rapidly in the later earning years than in early years. To accommodate such possibility, we might extend the model to

$$\text{income} = \beta_1 + \beta_2 \text{ education} + \beta_3 \text{ age} + \beta_4 \text{ age}^2 + \varepsilon.$$

This is how we proceed for regression modeling in real life situation. One needs to consider the experimental condition and the phenomenon before taking the decision on how many, why and how to choose the dependent and independent variables.

## Model set up

Let an experiment be conducted  $n$  times and the data is obtained as follows:

Observation number	Response $y$	Explanatory variables $X_1 \quad X_2 \quad \cdots \quad X_k$
1	$y_1$	$x_{11} \quad x_{12} \quad \cdots \quad x_{1k}$
2	$y_2$	$x_{21} \quad x_{22} \quad \cdots \quad x_{2k}$
$\vdots$	$\vdots$	$\vdots \quad \vdots \quad \ddots \quad \vdots$
$n$	$y_n$	$x_{n1} \quad x_{n2} \quad \cdots \quad x_{nk}$

Assuming that the model is

$$y = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon,$$

the  $n$ -tuples of observations are also assumed to follow the same model. Thus they satisfy

$$y_1 = \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_k x_{1k} + \varepsilon_1$$

$$y_2 = \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_k x_{2k} + \varepsilon_2$$

$$\vdots \qquad \qquad \qquad \vdots$$

$$y_n = \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_k x_{nk} + \varepsilon_n.$$

These  $n$  equations can be written as

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

or

$$y = X\beta + \varepsilon$$

where  $y = (y_1, y_2, \dots, y_n)'$  is a  $n \times 1$  vector of  $n$  observation on study variable,

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}$$

is a  $n \times k$  matrix of  $n$  observations on each of the  $k$  explanatory variables,  $\beta = (\beta_1, \beta_2, \dots, \beta_k)'$  is a  $k \times 1$  vector of regression coefficients and  $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$  is a  $n \times 1$  vector of random error components or disturbance term.

If intercept term is present, take first column of  $X$  to be  $(1, 1, \dots, 1)'$ . So that

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}.$$

In this case, there are  $(k - 1)$  explanatory variables and one intercept term.

## Assumptions in multiple linear regression model

Some assumptions are needed in the model  $y = X\beta + \varepsilon$  for drawing the statistical inferences. The following assumptions are made:

- (i)  $E(\varepsilon) = 0$
- (ii)  $E(\varepsilon\varepsilon') = \sigma^2 I_n$
- (iii)  $\text{Rank}(X) = k$
- (iv)  $X$  is a non-stochastic matrix.
- (v)  $\varepsilon \sim N(0, \sigma^2 I_n)$

These assumptions are used to study the statistical properties of estimator of regression coefficients. The following assumption is required to study particularly the large sample properties of the estimators

- (vi)  $\lim_{n \rightarrow \infty} \left( \frac{X'X}{n} \right) = \Delta$  exists and is a non-stochastic and nonsingular matrix (with finite elements).

The explanatory variables can also be stochastic in some cases. We assume that  $X$  is non-stochastic unless stated separately.

We consider the problems of estimation and testing of hypothesis on regression coefficient vector under the stated assumption.

## Estimation of parameters

A general procedure for the estimation of regression coefficient vector is to minimize

$$\sum_{i=1}^n M(\varepsilon_i) = \sum_{i=1}^n M(y_i - x_{i1}\beta_1 - x_{i2}\beta_2 - \dots - x_{ik}\beta_k)$$

for a suitably chosen function  $M$ .

Some examples of choice of  $M$  are

$$M(x) = |x|$$

$$M(x) = x^2$$

$$M(x) = |x|^p, \text{ in general.}$$

We consider the principle of least square which is related to  $M(x)=x^2$  and method of maximum likelihood estimation for the estimation of parameters.



## Principle of ordinary least squares (OLS)

Let  $B$  be the set of all possible vectors  $\beta$ . If there is no further information, then  $B$  is  $k$ -dimensional real Euclidean space. The object is to find a vector  $b' = (b_1, b_2, \dots, b_k)$  from  $B$  that minimizes the sum of squared deviations of  $\varepsilon_i$ 's, i.e.,

$$S(\beta) = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon' \varepsilon = (y - X\beta)'(y - X\beta)$$

for given  $y$  and  $X$ . A minimum will always exist as  $S(\beta)$  is a real valued, convex and differentiable function. Write

$$S(\beta) = y'y + \beta'X'X\beta - 2\beta'X'y.$$

Differentiate  $S(\beta)$  with respect to  $\beta$

$$\frac{\partial S(\beta)}{\partial \beta} = 2X'X\beta - 2X'y$$

$$\frac{\partial^2 S(\beta)}{\partial \beta \partial \beta'} = 2X'X \quad (\text{atleast non-negative definite}).$$

The normal equation is

$$\frac{\partial S(\beta)}{\partial \beta} = 0$$

$$\Rightarrow X'Xb = X'y$$

where the following result is used:

**Result:** If  $f(z) = Z'AZ$  is a quadratic form,  $Z$  is a  $m \times 1$  vector and  $A$  is any  $m \times m$  symmetric matrix then  $\frac{\partial}{\partial z} F(z) = 2Az$ .

Since it is assumed that  $\text{rank}(X) = k$  (full rank), then  $X'X$  is positive definite and unique solution of normal equation is

$$b = (X'X)^{-1}X'y$$

which is termed as **ordinary least squares estimator** (OLSE) of  $\beta$ .

Since  $\frac{\partial^2 S(\beta)}{\partial \beta^2}$  is at least non-negative definite, so  $b$  minimizes  $S(\beta)$ .

In case,  $X$  is not of full rank, then

$$b = (X'X)^- X'y + [I - (X'X)^- X'X] \omega$$

where  $(X'X)^-$  is the generalized inverse of  $X'X$  and  $\omega$  is an arbitrary vector. The generalized inverse  $(X'X)^-$  of  $X'X$  satisfies

$$X'X(X'X)^- X'X = X'X$$

$$X(X'X)^- X'X = X$$

$$X'X(X'X)^- X' = X'.$$