Multiple Linear Regression

Multiple Linear Regression (MLR) also known as simple multiple regression. It is a statistical technique that uses several explanatory variables to predict the outcomes of a response variable.

The aim of multiple linear regression is to model the linear relationship between the independent variables and dependent variables.

The multiple linear regression is the extension of Ordinary Least-Squares (OLS) regression because it involves more than one explanatory variable. MLR is used extensively in econometrics and financial inference.

Multiple Linear Regression Model

It is observed in agriculture that, the crop yield (Y) not only depends on the amount of rainfall (X_1) but also on the amount of fertilizer (X_2) applied, pesticides (X_3) used, quality of seeds (X_4), quality of soil (X_5) etc.

Thus in multiple regression, the dependent variable Y is a function of more than one independent variables, i.e.

$$Y = f(X_1, X_2, \ldots, X_n)$$

In multiple ~~nonlinear~~ linear regression, ~~f is non-linear~~ f is linear

$$\boxed{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_k X_k}$$

Suppose Y depends on two independent variables X_1 and X_2.

$$\boxed{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2}$$

To estimate the coefficients $\beta_0, \beta_1, \beta_2$ we apply the least square method to minimise

$$\boxed{\sum_{i=1}^{N} \{Y_i - (b_0 + b_1 X_{1i} + b_2 X_{2i})\}^2}$$

This results in three normal equations given by

$$\sum_{i=1}^{N} y_i = N b_0 + b_1 \sum_{i=1}^{N} x_{1i} + b_2 \sum_{i=1}^{N} x_{2i}$$

$$\sum_{i=1}^{N} x_{1i} y_i = b_0 \sum x_{1i} + b_1 \sum_{i=1}^{N} x_{1i}^2 + b_2 \sum x_{1i} \cdot x_{2i}$$

$$\sum x_{2i} y_i = b_0 \sum x_{2i} + b_1 \sum x_{1i} \cdot x_{2i} + b_2 \sum x_{2i}^2$$

Here b_0, b_1, b_2 are the least squares estimates of $\cancel{\beta_1}, \cancel{\beta_2}$ $\beta_0, \beta_1, \beta_2$

~~Linear~~ Multiple linear Regression in k-independent variables

The above analysis can be generalised to fit $N(k+1)$ tuples $(x_{1i}, x_{2i}, \ldots, x_{ki})$ $(i = 1 \text{ to } N)$, to the ~~tuples~~ equation.

$$y = \beta_0 + b_1 x_1 + \beta_1 x_2 + \ldots + \beta_K x_K$$

The (k+1) normal equations are:

$$\sum_{i=1}^{N} y_i = Nb_0 + b_1 \sum_{i=1}^{N} x_{1i} + b_2 \sum_{i=1}^{N} x_{2i} + \ldots + b_k \sum_{i=1}^{N} x_{ki}$$

PARSHWANATH CHARITABLE TRUST'S
A.P. SHAH INSTITUTE OF TECHNOLOGY
Department of Computer Science and Engineering
Data Science

Semester : VI   Subject :CSC601 Data Analytics and Visualization   Academic Year: 2023- 2024

Example 1 (multiple linear regression) [Dec 23] [10M]

Fit a regression equation to estimate $\beta_0, \beta_1, \beta_2$ to the following data of a transport company on the weights of 6 shipments, the distances they were moved and the damage of the goods that was incurred. Estimate the damage when a shipment of 3700 kg is moved to a distance of 260km.

Weight X_1 (1000 kg)	4.0	3.0	1.6	1.2	3.4	4.8
Distance X_2 (100 km)	1.5	2.2	1.0	2.0	0.8	1.6
Damage Y (Rs)	160	112	69	90	123	186

Solution :-

Let weight X_1 and distance X_2 be independent variables and the damage y be the dependent variable.

Let the equation of regression be,

$$y = b_0 + b_1 x_1 + b_2 x_2$$

where b_0, b_1, b_2 are estimates of $\beta_0, \beta_1, \beta_2$

The three normal equations become.

$$\sum_{i=1}^{6} y_i = n b_0 + b_1 \sum_{i=1}^{6} x_{1i} + b_2 \sum_{i=1}^{6} x_{2i}$$

$$\sum x_{1i} y_i = b_0 \sum x_{1i} + b_1 \sum x_{1i}^2 + b_2 \sum x_{1i} x_{2i}$$

$$\sum x_{2i} y_i = b_0 \sum x_{2i} + b_1 \sum x_{1i} \cdot x_{2i} + b_2 \sum x_{2i}^2$$

X_1 Weight (1000 kg)	X_2 Distance (100 km)	y Damage in Rs	X_1^2	X_2^2	$X_1 \cdot X_2$	$X_1 \cdot y$	$X_2 \cdot$
4.0	1.5	160	16	2.25	6.0	640	240
3.0	2.2	112	09	4.84	6.6	336	246.4
1.6	1.0	69	2.56	1.0	1.6	110.4	69
1.2	2.0	90	1.44	4.0	2.4	108	180
3.4	0.8	123	11.56	0.64	2.72	418.2	98.4
4.8	1.6	186	23.04	2.56	7.68	892.8	297.6
18	9.1	740	63.6	15.29	27	250.54	1131.4
ΣX_{1i}	ΣX_{2i}	Σy_i	ΣX_{1i}^2	ΣX_{2i}^2	$\Sigma X_{1i} \cdot X_{2i}$	$\Sigma X_{1i} \cdot y_i$	$\Sigma X_{2i} \cdot y$

Now, $n = 6$, $\Sigma X_{1i} = 18$, $\Sigma X_{2i} = 9.1$, $\cancel{\Sigma Y_i}$

$\Sigma Y_i = 740$

$\Sigma X_{1i}^2 = 63.6$

$\Sigma X_{2i}^2 = 15.29$

$\Sigma X_{1i} \cdot X_{2i} = 27$

$\Sigma X_{1i} \cdot Y_{i} = 250.54$

$\Sigma X_{2i} \cdot Y_i = 1131.4$

Normal equations become

$$740 = 6b_0 + 18b_1 + 9.1b_2$$

$$250.54 = 18b_0 + 63.6b_1 + 27b_2$$

$$1131.4 = 9.1b_0 + 27b_1 + 15.29b_2$$

Solving these equations we get

$b_0 = 14.56$
$b_1 = 30.109$
$b_2 = 12.16$

Thus the required regression equation is

$$\boxed{y = 14.56 + 30.109(x_1) + 12.16(x_2)}$$

For a weight of 3700 kg. $(x_1 = 3.7)$ and for a distance of 260 km $(x_2 = 2.6)$ the damage incurred in rupees is

$$y(x_1 = 3.7, x_2 = 2.6) = 14.56 + 30.109(3.7) + 12.16(2.6)$$
$$= 714.58$$
$$\boxed{= 715 \text{ Rs.}}$$