(Lecture No: - 17) (30/10/23)

(Multiple Regression)

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + e_i$$

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e_i$$

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e_i$$

$\beta_0 \longrightarrow Y$ - Intercept.

$\beta_1, \beta_2, \ldots, \beta_K$

Partial regression Coefficient

or also called the parameters of regression line including β_0.

$$Y_i = \underline{b}_0 + \underline{b}_1 X_1 + \underline{b}_2 X_2 + \cdots + b_K X_K + e_i$$

Sample Regression Function (SRF)

The partial regression coefficients are obtained by minimizing the

Sum of squares of residuals : i.e

$$S = \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2$$

$y_i \longrightarrow$ original values of dependent variable

$\hat{y}_i \longrightarrow$ Estimated values of dependent variable.

(Coefficient of Determination)

It is the ratio between

the explained variation to the total variation.

It is denoted by R^2

$$R^2 = \frac{\text{Explained Variation}}{\text{Total Variation}}$$

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}-\bar{y})^2}{\sum_{i=1}^{n}(y-\bar{y})^2}$$

It lies b/w '0' and '1'

$$0 \leq R^2 \leq 1$$

It is used to find the strength of relationship among the variables.

If $R^2 = 0$, the strength of relationship among the variable does not exist.

(Correlation)

$r \longrightarrow$ Correlation Coefficient

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$$

$$-1 \leq r \leq +1.$$

$r = -1$, Perfect Negative Correlation

$r = 1$, Perfect Positive Corrdation

(Orion data output Interpretation)

$R^2 = 0.9361$

<u>93.61</u>% variation / change in price of a car is explained by the age of the car and miles driven.

Intercept $= \beta_0 = \$18303.50$

The price of a brand

new car is $ 18303.50

$\beta_1 = -950.427$

$Y_i = 18303.50 - 950.427 X_1 - 0.08215 X_2$

$X_1 \longrightarrow$ Age

$X_2 \longrightarrow$ miles

$\beta_1 = -950.427$

The Price of car decreases by $ 950.427

If the age of the car is increased by 1 year keeping the effect of miles driven as fixed

$\beta_2 = -0.08215$

The price of a car decreases by $\underline{\underline{0.08215}}$ if the car is driven $\underline{1}$ mile keeping the effect of age of the car as fixed.

We can also say that the price would decrease by \$82.15 after being driven 1,000 miles.

$H_0: \beta_1 = 0$ (The coefficient is insignificant)

$H_1: \beta_1 \neq 0$ (The coefficient is significant)

$$t = \frac{\beta_1 - \beta_{10}}{S.E(\beta_1)}$$

All the β - Coefficients $\beta_0, \beta_1, \beta_2$ are significant since the P-values of all the β - Coefficients is less than α which mean H_0 is rejected.

$$b \pm t_{\alpha/2(V)} \cdot S_b$$

C.I. for 'β' Coefficient

$$b_0 \pm t_{\alpha/2}(\gamma) \cdot S_{b_0}$$

C.I. for the Intercept.

ANOVA is used to find out the overall significance of the fitted regression model.

$H_0: \beta_1 = \beta_2 = \dots = \beta_K = 0$

$H_1: \beta_1 \neq \beta_2 \dots \neq \beta_K \neq 0$.

$H_0 \longrightarrow$ The overall regression model is not significant.

H_1: The overall regression model is significant.

(ANOVA Table)

S.O.V	d.f.	SS	MS	F-ratio
Regression	$K-1$ ✓	SSR	MSR	$F = \frac{MSR}{MSE}$
Error	$n-K$	SSE	MSE	–
Total	$n-1$ ✓	SST	–	–

$K \longrightarrow$ No. of 'β' coefficients or Parameter

$$MSR = \frac{SSR}{K-1}$$

$$MSE = \frac{SSE}{n-K}$$

The P-value $= 0.0000167$ is less than α.

Therefore H_0 is rejected meaning.

The overall regression is significant.

The regression line is mainly used to predict the dependent variable on the basis of Independent variables.

$\hat{Y}_i = 18303.5 - 950.427 X_1 - 0.08215 X_2.$

Predict the price of onion which is 4 years old already driven 50,000 miles

$X_1 = 4, \quad X_2 = 50,000.$

$\hat{y} = 18303.5 - 950.427(4) - 0.08215(50,000)$

$\hat{y} = \$10396.90$

't' distribution degrees of freedom

$$df = \underline{n - K}$$
$$= 11 - 3$$
$$= 8$$