# MULTIPLE REGRESSION

11 October 2021 08:02

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4.$$

$Y \rightarrow$ economic growth rate of a country.

$x_1 \rightarrow$ time period.

$x_2 \rightarrow$ size of population.

$x_3 \rightarrow$ level of employment

$x_4 \rightarrow$ literacy.

$b_0 \rightarrow$ intercept

$b_1, b_2, b_3, b_4 \rightarrow$ constant

$Y \rightarrow$ dependent variable.

Reg line

$y$ on $x_1$ and $x_2$

Suppose the number of independent variables is two, then Normal equations are

$n \rightarrow$ no. of observations.

Problem 1: The annual sales revenue(in crores of rupees ) of a product as a ffnction of sales force(number of salesmen) and annual advertising expenditure(in lakhs of rupees) for the past 10 year are summarized in the following table. Annual sales 20 revenue Y Sales force S 23 13 23 25 s 38 21 23 20 29 16 28 22 10 23 24 12 30 27 14 26 35 20 32 Annual advertising expenditures 28 16

$\left.\begin{matrix} x_1 = 25 \\ x_2 = 40 \end{matrix}\right\}$ ?

$$Y = b_0 + b_1 X_1 + b_2 X_2$$

| $Y$ | $X_1$ | $X_2$ | $X_1^2$ | $X_2^2$ | $X_1 X_2$ | $YX_1$ | $YX_2$ |
|---|---|---|---|---|---|---|---|
| 20 | | | | | | | |
| 23 | | | | | | | |
| 25 | | | | | | | |
| 27 | | | | | | | |
| 21 | | | | | | | |
| 29 | | | | | | | |
| 22 | | | | | | | |
| 24 | | | | | | | |
| 27 | | | | | | | |
| 35 | | | | | | | |

$\Sigma X_1 = 142$

$\Sigma Y = 253$

$\Sigma X_2 = 264$

$\Sigma X_1^2 = \cancel{20164}\ 2246$

$\Sigma X_2^2 = \cancel{7678}\ 7326$

$\Sigma X_1 X_2 = 3617$

$\Sigma YX_1 = 3678$

$\Sigma YX_2 = 6751$

~~se the number of independent variables is two, then Normal equations are~~

$253 = 10b_0 + 142 b_1 + 264 b_2$ — (1)

$3678 = 142 b_0 + 2246 b_1 + 3617 b_2$ — (2)

$6751 = 264 b_0 + 3617 b_1 + 7326 b_2$ — (3)

— (1)

— (2)

— (3)

Solve;

$b_0 = 5.1483$

$b_1 = 0.6190$

$b_2 = 0.4303$

regression model

$$Y = 5.1483 + 0.619 X_1 + 0.4304 X_2$$

when $X_1 = 25$, $X_2 = 40$

$Y = ?$ 37.835

For a multivariate data, the regression equation of $X$ on $Y$ and $Z$ is

$$(X-\bar{X})\frac{\omega_{11}}{\sigma_1}+(Y-\bar{Y})\frac{\omega_{12}}{\sigma_2}+(Z-\bar{Z})\frac{\omega_{13}}{\sigma_3}=0$$

where

$$\omega = det\begin{bmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{bmatrix}$$

$$\omega_{11} = det\begin{bmatrix} 1 & r_{23} \\ r_{23} & 1 \end{bmatrix}$$

$$\omega_{12} = -det\begin{bmatrix} r_{12} & r_{23} \\ r_{13} & 1 \end{bmatrix}$$

$$\omega_{13} = det\begin{bmatrix} r_{12} & 1 \\ r_{13} & r_{23} \end{bmatrix}$$

correl coeffs: $r_{12}, r_{13}, r_{23}$

SD: $\sigma_1, \sigma_2, \sigma_3$

mean: $\bar{X}_1, \bar{X}_1, \bar{Z}_1$

2) Find the reg. eqn of X on Y and Z given the fol.

| Variables | mean | SD | $r_{12}$ | $r_{23}$ | $r_{31}$ |
|---|---|---|---|---|---|
| X | $\bar{X}$ 35.8 | $\sigma_1$ 4.2 | 0.6 | – | – |
| Y | $\bar{Y}$ 52.4 | $\sigma_2$ 5.3 | – | 0.7 | 0.8 |
| Z | $\bar{Z}$ 48.8 | $\sigma_3$ 6.1 | – | – | |

$$\omega_{11} = \begin{vmatrix} 1 & r_{23} \\ r_{23} & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0.7 \\ 0.7 & 1 \end{vmatrix} = 1-(0.7)^2$$

$$= 0.51$$

$\omega_{12} =$

$\omega_{13} =$

$X =$

More problems on linear regression:-

1) The lines of regression of a bivariate population are:

$$8x - 10y + 66 = 0$$

$$40x - 18y = 214$$

The Variance of x is 9. Find

(i) mean ~~of~~ of X and Y

(ii) Correlation Coeff. bet' x and y

(iii) S D of Y

Solution

Solution: The regression lines given are SX-10Y+66=O 40-r-1SY -214 Since both the lines of regression pass through the mean values, the point (T , f) will satisfy both the equations. Hence these equations can be written as s — 10 + 66 o 40 r -IS? -214=0 Solving these two equations for and , we obtain 13 17

(ii) For correlation coefficient between X and Y, we have to calculate the values of by: and Rewriting the equations Similarly, + 66 b —+8/10=+4/5 401'= 1SY+214 b. = ISAO = 9/20 By these values, we can now work out the correlation coeffcient. r2=bF. b. r = + 9 / 25 = + 0.6 = 9/25 Both the values of the regression coefficients being positive, we have to consider only the positive value of the correlation coefficient. Hence r = 0.6

- If suppose we take the equations the otherway round
  8x=10y-66
    $b_{xy}$= 10/8
  18y=40x-214
    $b_{yx}$=40/18
  Hence r2= $b_{xy} . b_{yx}$ = 10/8 * 40/18 = 2.7
   r= 1.64 >1 is this possible ?
  No

(iii) We have been given variance of X i.e sx2=9

We consider 3 as SD is always positive Since Substituting the values of b,x, r and Sr we obtain, S = 4/5 x 3/0.6

Adjusted $R^2$

- $R^2$ shows how well terms (data points) fit a curve or line.
- Adjusted $R^2$ also indicates how well terms fit a curve or line, but adjusts for the number of terms in a model.
  - If you add more and more **useless variables** to a model, adjusted r-squared will

decrease. If you add more **useful** variables, adjusted r-squared will increase.
Adjusted $R^2$ will always be less than or equal to $R^2$.

The formula is:,

- where:
- n is the number of points in your data sample.
- k is the number of independent regressors, i.e. the number of variables in your model, excluding the constant.
- The range of $R^2$ is
  OSR (1
-

Problems :-

- For a sample of eight observations and two independent variables (years of experience and years of graduate education), R Square is 0.944346527. Find the adjusted R squared

  Given that n is 8, k is 2 with R Square is 0.944346527.

n-1 -2-1-(1-R2 8-1 8-(2+1) 0.922085138