

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

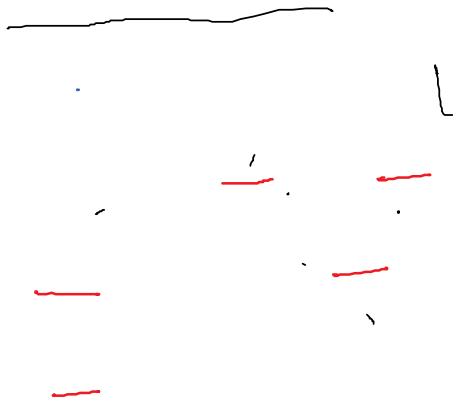
$b_1, b_2, b_3, b_4 \rightarrow$ constant

$Y \rightarrow$ 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 function of sales force (number of salesmen) and annual advertising expenditure (in lakhs of rupees) for the past 10 years are summarized in the following table. Annual sales revenue Y Sales force S 23 13 23 25 38 21 23 20 29 16 28 22 10 23 24 12 30 27 14 26 35 20 32 Annual advertising expenditures 28 16

$$\left. \begin{array}{l} X_1 = 25 \\ X_2 = 240 \end{array} \right\}$$

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

\bar{Y}	X_1	X_2	X_1^2	X_2^2	$X_1 X_2$	$Y X_1$	$Y X_2$
-----------	-------	-------	---------	---------	-----------	---------	---------

20

23

25

27

21

29

22

24

27

35

$$\sum X_1 = 142$$

$$\sum Y = 253$$

$$\sum X_2 = 264$$

$$\sum X_1^2 = 20164$$

$$2246$$

$$\sum X_2^2 = 7326$$

$$\sum X_1 X_2 = 3617$$

$$\sum Y X_1 = 3678$$

$$\sum Y X_2 = 6751$$

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

$$253 = 10b_0 + 142b_1 + 264b_2 \quad \text{--- (1)}$$

$$\text{--- (1)} \quad 3678 = 142b_0 + 2246b_1 + 3617b_2 \quad \text{--- (2)}$$

$$\text{--- (2)} \quad 6751 = 264b_0 + 3617b_1 + 7326b_2 \quad \text{--- (3)}$$

--- (3) solve;

$$b_0 = 5.1483 \approx 5.1$$

$$b_1 = 0.6190$$

$$b_2 = 0.4303$$

regression model

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

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

$$Y = ? \quad 37.835$$

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 coeff
SD → mean
 r_{12}
 r_{13}
 r_{23}
 σ_1
 σ_2
 σ_3
 \bar{X}_1
 \bar{Y}_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	σ_1 4.2	0.6	-	-
Y	\bar{Y} 52.4	σ_2 5.3	-	0.7	-
Z	\bar{Z} 48.8	σ_3 6.1	-	-	0.8

$$\omega_{11} = \begin{vmatrix} 1 & r_{23} \\ r_{23} & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0.7 \\ 0.7 & 1 \end{vmatrix} = 1 - (0.7)^2 = 0.51$$

$$\omega_{12} =$$

$$\omega_{13} =$$

$$X =$$

More problems on linear regression:-

i) 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 x and y
- (ii) Correlation Coeff. bet' x and y
- (iii) S.D of y

~~Soln~~

 Solution: The regression lines given are $SX - 10Y + 66 = 0$ and $40 - r - 15Y - 214 = 0$. Since both the lines of regression pass through the mean values, the point (\bar{x}, \bar{y}) will satisfy both the equations. Hence these equations can be written as $\bar{x} - 10 + 66 = 0$ and $40 - r - 15\bar{y} - 214 = 0$. Solving these two equations for \bar{x} and \bar{y} , we obtain $\bar{x} = 13$ and $\bar{y} = 7$.

 (ii) For correlation coefficient between X and Y , we have to calculate the values of b_1 and b_2 . Rewriting the equations similarly, $66 - 10\bar{y} = -\bar{x}$ and $40 - 15\bar{y} = r + 214$. $b_1 = \frac{66 - 10\bar{y}}{-\bar{x}} = \frac{66 - 10(7)}{-13} = \frac{66 - 70}{-13} = \frac{-4}{-13} = \frac{4}{13}$. $b_2 = \frac{40 - 15\bar{y}}{r + 214} = \frac{40 - 15(7)}{r + 214} = \frac{40 - 105}{r + 214} = \frac{-65}{r + 214}$. By these values, we can now work out the correlation coefficient. $r^2 = b_1 b_2$. $r = \frac{4}{13} \cdot \frac{-65}{r + 214} = -\frac{260}{13(r + 214)} = -\frac{20}{r + 214}$. 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

$$8x = 10y - 66$$

$$b_{xy} = 10/8$$

$$18y = 40x - 214$$


$$b_{yx} = 40/18$$

$$\text{Hence } r^2 = b_{xy} \cdot b_{yx} = 10/8 * 40/18 = 2.7$$

$$r = 1.64 > 1 \text{ is this possible?}$$

No

 (iii) We have been given variance of X i.e. $s_x^2 = 9$

 We consider 3 as SD is always positive Since Substituting the values of b_x , r and S_r we obtain, $S = 4/5 \times 3/0.6$

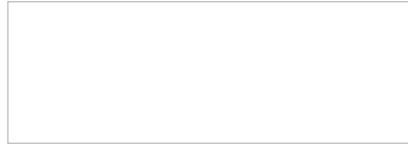
Adjusted R^2

- 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.

$n-1 - (2-1) \cdot (1-0.944346527) = 6 - 1 \cdot 0.055653473 = 5$