

# Assignment 2

Challa Akshay Santoshi-CS21BTECH11012

**ICSE 12 2019**

**Question: 20 (b)**

Find the equation of the regression line of y on x, if the observations (x, y) are as follows:

(1, 4), (2, 8), (3, 2), (4, 12), (5, 10), (6, 14), (7, 16), (8, 6), (9, 18)

Also find the estimated value of y when x = 14.

**Solution:**

Let the regression line of y on x be given as

$$(a_1 \quad -1)Z = a_0 \quad (0.0.1)$$

Here,  $Z = \begin{pmatrix} x \\ y \end{pmatrix}$ . The observations given can be written as follows

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}, \dots, \begin{pmatrix} x_n \\ y_n \end{pmatrix} \quad (0.0.2)$$

Let us define few matrices as follows

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_{n-1} \\ 1 & x_n \end{pmatrix} \quad (0.0.3)$$

$$A = \begin{pmatrix} a_1 \\ a_0 \end{pmatrix}, E = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_{n-1} \\ e_n \end{pmatrix} \quad (0.0.4)$$

$$E = Y - XA \quad (0.0.5)$$

**E** is the matrix of residual errors.

To get the regression line which is the best fit line equation for a given data, the sum of squares of residual errors has to be minimum.

Let SSE be the sum of square of errors.

$$SSE = \|E^T E\| \quad (0.0.6)$$

$$= (Y - XA)^T (Y - XA) \quad (0.0.7)$$

$$= (Y^T - A^T X^T) (Y - XA) \quad (0.0.8)$$

$$= (Y^T Y - Y^T XA - A^T X^T Y + A^T X^T XA) \quad (0.0.9)$$

$Y^T XA$  is of order  $(1 \times 1)$ . Therefore it is equal to its transpose.

$$Y^T XA = (Y^T XA)^T \quad (0.0.10)$$

$$= A^T X^T Y \quad (0.0.11)$$

This will give,

$$SSE = (Y^T Y - 2A^T X^T Y + A^T X^T XA) \quad (0.0.12)$$

For SSE to be minimum, take the gradient with respect to A.

$$\nabla SSE = (\nabla Y^T Y - 2\nabla A^T X^T Y + \nabla A^T X^T XA) \quad (0.0.13)$$

$$= 2(X^T XA - X^T Y) \quad (0.0.14)$$

Equate it to zero to get the minimum condition.

$$(X^T XA - X^T Y) = 0 \quad (0.0.15)$$

$$A = (X^T X)^{-1} X^T Y \quad (0.0.16)$$

Using data given in question

$$\begin{pmatrix} x \\ y \end{pmatrix} : \begin{pmatrix} 1 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ 8 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 \\ 12 \end{pmatrix}, \begin{pmatrix} 5 \\ 10 \end{pmatrix}, \begin{pmatrix} 6 \\ 14 \end{pmatrix}, \begin{pmatrix} 7 \\ 16 \end{pmatrix}, \begin{pmatrix} 8 \\ 6 \end{pmatrix}, \begin{pmatrix} 9 \\ 18 \end{pmatrix} \quad (0.0.17)$$

$$\mathbf{Y} = \begin{pmatrix} 4 \\ 8 \\ 2 \\ 12 \\ 10 \\ 14 \\ 16 \\ 6 \\ 18 \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \\ 1 & 7 \\ 1 & 8 \\ 1 & 9 \end{pmatrix} \quad (0.0.18)$$

Using equations (0.0.16) and (0.0.18), solving will give

$$A = \begin{pmatrix} \frac{10}{3} \\ \frac{4}{3} \\ \frac{4}{3} \end{pmatrix} \quad (0.0.19)$$

Using equation (0.0.1), regression line equation can be obtained.

$$\begin{pmatrix} \frac{4}{3} & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \frac{10}{3} \quad (0.0.20)$$

From this equation we get the value of y as 22 when x is 14.

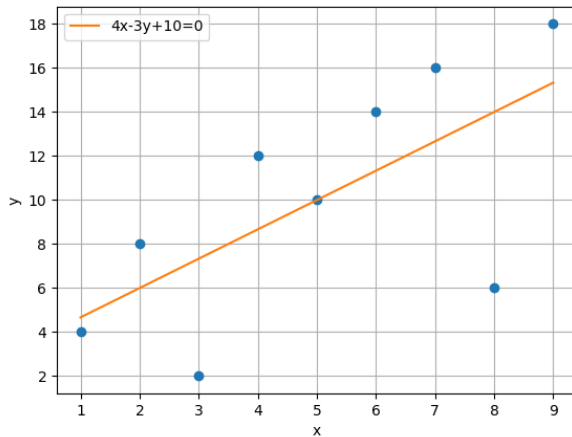


Fig. 0: Graph showing the Regression Line