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 by

$$y = A + Bx \tag{0.0.1}$$

Here A is the Y-intercept and B is the slope of the line.
Using vectors, it can also be represented as:

$$\begin{pmatrix} B & -1 \end{pmatrix} \mathbf{X} = -A \tag{0.0.2}$$

Consider $\mathbf{x}, \mathbf{y}$ to be vectors.
From the definition of line regression, the value of B can be found from a set of data using following relation:

$$B = \frac{\Sigma\mathbf{xy} - \frac{\Sigma\mathbf{x}\Sigma\mathbf{y}}{n}}{\Sigma\mathbf{x^2} - \frac{(\Sigma\mathbf{x})^2}{n}} \tag{0.0.3}$$

Here,

$$\Sigma\mathbf{xy} = \mathbf{x}^\top\mathbf{y} \tag{0.0.4}$$
$$\Sigma\mathbf{x} = \mathbf{1}^\top\mathbf{x} \tag{0.0.5}$$
$$\Sigma\mathbf{y} = \mathbf{1}^\top\mathbf{y} \tag{0.0.6}$$
$$\Sigma\mathbf{x^2} = \mathbf{x}^\top\mathbf{x} \tag{0.0.7}$$

So, equation (0.0.2) can be written as:

$$\begin{pmatrix} \frac{\Sigma\mathbf{xy} - \frac{\Sigma\mathbf{x}\Sigma\mathbf{y}}{n}}{\Sigma\mathbf{x^2} - \frac{(\Sigma\mathbf{x})^2}{n}} & -1 \end{pmatrix} \mathbf{X} = -A \tag{0.0.8}$$

To find the value of A, we can substitute the point $\begin{pmatrix} \frac{\Sigma\mathbf{x}}{n} \\ \frac{\Sigma\mathbf{y}}{n} \end{pmatrix}$ in the equation (0.0.8) and solve for it.

Solving for A, we can simplify it as:

$$\begin{pmatrix} \frac{\Sigma\mathbf{xy} - \frac{\Sigma\mathbf{x}\Sigma\mathbf{y}}{n}}{\Sigma\mathbf{x^2} - \frac{(\Sigma\mathbf{x})^2}{n}} & -1 \end{pmatrix} \begin{pmatrix} \frac{\Sigma\mathbf{x}}{n} \\ \frac{\Sigma\mathbf{y}}{n} \end{pmatrix} = -A \tag{0.0.9}$$

Substituting this value of -A back in equation (0.0.9), we get the regression line of y on x as

$$\begin{pmatrix} \frac{\Sigma\mathbf{xy} - \frac{\Sigma\mathbf{x}\Sigma\mathbf{y}}{n}}{\Sigma\mathbf{x^2} - \frac{(\Sigma\mathbf{x})^2}{n}} & -1 \end{pmatrix} \mathbf{X} = \begin{pmatrix} \frac{\Sigma\mathbf{xy} - \frac{\Sigma\mathbf{x}\Sigma\mathbf{y}}{n}}{\Sigma\mathbf{x^2} - \frac{(\Sigma\mathbf{x})^2}{n}} & -1 \end{pmatrix} \begin{pmatrix} \frac{\Sigma\mathbf{x}}{n} \\ \frac{\Sigma\mathbf{y}}{n} \end{pmatrix} \tag{0.0.10}$$

The Regression line for the data given is shown in figure given below

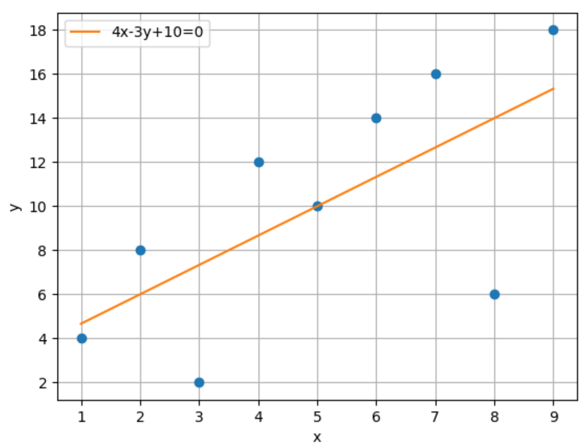

Fig. 0: Graph showing the Regression Line

From the data given in question,

$$\mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{pmatrix}, \mathbf{y} = \begin{pmatrix} 4 \\ 8 \\ 2 \\ 12 \\ 10 \\ 14 \\ 16 \\ 6 \\ 18 \end{pmatrix} \tag{0.0.11}$$

$$\Sigma\mathbf{x} = \begin{pmatrix}1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1\end{pmatrix}\begin{pmatrix}1\\2\\3\\4\\5\\6\\7\\8\\9\end{pmatrix} \quad (0.0.12)$$

$$\Sigma\mathbf{y} = \begin{pmatrix}1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1\end{pmatrix}\begin{pmatrix}4\\8\\2\\12\\10\\14\\16\\6\\18\end{pmatrix} \quad (0.0.13)$$

$$\Sigma\mathbf{xy} = \begin{pmatrix}1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9\end{pmatrix}\begin{pmatrix}4\\8\\2\\12\\10\\14\\16\\6\\18\end{pmatrix} \quad (0.0.14)$$

$$\Sigma\mathbf{x^2} = \begin{pmatrix}1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9\end{pmatrix}\begin{pmatrix}1\\2\\3\\4\\5\\6\\7\\8\\9\end{pmatrix} \quad (0.0.15)$$

These equations can be simplified to give,

$$\Sigma\mathbf{x} = 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 \quad (0.0.16)$$

$$\Sigma\mathbf{y} = 4 + 8 + 2 + 12 + 10 + 14 + 16 + 6 + 18 \quad (0.0.17)$$

$$\Sigma\mathbf{xy} = 4 + 16 + 6 + 48 + 50 + 84 + 112 + 48 + 162 \quad (0.0.18)$$

$$\Sigma\mathbf{x^2} = 1 + 4 + 9 + 16 + 25 + 36 + 49 + 64 + 81 \quad (0.0.19)$$

From these above equations we get,

$$\Sigma\mathbf{x} = 45 \quad (0.0.20)$$

$$\Sigma\mathbf{y} = 90 \quad (0.0.21)$$

$$\Sigma\mathbf{xy} = 530 \quad (0.0.22)$$

$$\Sigma\mathbf{x^2} = 285 \quad (0.0.23)$$

$$n = 9 \quad (0.0.24)$$

Substituting these values in equation (0.0.10) will give

$$\begin{pmatrix}\frac{530-\frac{45\times90}{9}}{285-\frac{(45)^2}{9}} & -1\end{pmatrix}\mathbf{X} = \begin{pmatrix}\frac{530-\frac{45\times90}{9}}{285-\frac{(45)^2}{9}} & -1\end{pmatrix}\begin{pmatrix}\frac{45}{9}\\\frac{90}{9}\end{pmatrix} \quad (0.0.25)$$

Simplifying we get

$$\begin{pmatrix}\frac{80}{60} & -1\end{pmatrix}\mathbf{X} = \begin{pmatrix}\frac{80}{60} & -1\end{pmatrix}\begin{pmatrix}5\\10\end{pmatrix} \quad (0.0.26)$$

$$\implies \begin{pmatrix}\frac{4}{3} & -1\end{pmatrix}\mathbf{X} = \frac{-10}{3} \quad (0.0.27)$$

Equation (0.0.27) gives us the linear equation

$$\frac{4}{3}x - y = \frac{-10}{3} \quad (0.0.28)$$

Simplifying we get,

$$4x - 3y + 10 = 0 \quad (0.0.29)$$

Equation (0.0.29) is the required equation of the regression line of y on x for the data given.

The second part of the question asks us to find the estimated value of y when x = 14.
Using equation (0.0.27), we can find y as follows:

$$\begin{pmatrix}\frac{4}{3} & -1\end{pmatrix}\begin{pmatrix}14\\y\end{pmatrix} = \frac{-10}{3} \quad (0.0.30)$$

$$\Rightarrow \frac{56}{3} - y = \frac{-10}{3} \quad (0.0.31)$$

$$\Rightarrow y = \frac{66}{3} \quad (0.0.32)$$

$$\Rightarrow y = 22 \quad (0.0.33)$$

Therefore, the value of y is 22 when x is 14.