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

$$\begin{pmatrix} a_1 & -1 \end{pmatrix} \mathbf{Z} = a_0 \tag{0.0.1}$$

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

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \ldots.., \begin{pmatrix} x_n \\ y_n \end{pmatrix} \tag{0.0.2}$$

Let us define few matrices as follows

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ y_{n-1} \\ y_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ . & \\ . & \\ 1 & x_{n-1} \\ 1 & x_n \end{pmatrix} \tag{0.0.3}$$

$$\mathbf{A} = \begin{pmatrix} a_1 \\ a_0 \end{pmatrix}, \mathbf{E} = \begin{pmatrix} e_1 \\ e_2 \\ . \\ . \\ e_{n-1} \\ e_n \end{pmatrix} \tag{0.0.4}$$

$$\mathbf{E} = \mathbf{Y} - \mathbf{X}\mathbf{A} \tag{0.0.5}$$

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.

$$SSE = \|\mathbf{E}^\top \mathbf{E}\| \tag{0.0.6}$$

$$= (\mathbf{Y} - \mathbf{X}\mathbf{A})^\top (\mathbf{Y} - \mathbf{X}\mathbf{A}) \tag{0.0.7}$$

$$= (\mathbf{Y}^\top - \mathbf{A}^\top \mathbf{X}^\top)(\mathbf{Y} - \mathbf{X}\mathbf{A}) \tag{0.0.8}$$

$$= (\mathbf{Y}^\top \mathbf{Y} - \mathbf{Y}^\top \mathbf{X}\mathbf{A} - \mathbf{A}^\top \mathbf{X}^\top \mathbf{Y} + \mathbf{A}^\top \mathbf{X}^\top \mathbf{X}\mathbf{A}) \tag{0.0.9}$$

$\mathbf{Y}^\top \mathbf{X}\mathbf{A}$ is of order (1×1). Therefore it is equal to its transpose.

$$\mathbf{Y}^\top \mathbf{X}\mathbf{A} = (\mathbf{Y}^\top \mathbf{X}\mathbf{A})^\top \tag{0.0.10}$$

$$= \mathbf{A}^\top \mathbf{X}^\top \mathbf{Y} \tag{0.0.11}$$

This will give,

$$SSE = (\mathbf{Y}^\top \mathbf{Y} - 2\mathbf{A}^\top \mathbf{X}^\top \mathbf{Y} + \mathbf{A}^\top \mathbf{X}^\top \mathbf{X}\mathbf{A}) \tag{0.0.12}$$

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

$$\nabla SSE = (\nabla \mathbf{Y}^\top \mathbf{Y} - 2\nabla \mathbf{A}^\top \mathbf{X}^\top \mathbf{Y} + \nabla \mathbf{A}^\top \mathbf{X}^\top \mathbf{X}\mathbf{A}) \tag{0.0.13}$$

$$= 2(\mathbf{X}^\top \mathbf{X}\mathbf{A} - \mathbf{X}^\top \mathbf{Y}) \tag{0.0.14}$$

Equate it to zero to get the minimum condition.

$$(\mathbf{X}^\top \mathbf{X}\mathbf{A} - \mathbf{X}^\top \mathbf{Y}) = 0 \tag{0.0.15}$$

$$\mathbf{A} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} \tag{0.0.16}$$

Using data given in question

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

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

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

$$A = \begin{pmatrix} \frac{10}{3} \\ \frac{4}{3} \end{pmatrix} \tag{0.0.19}$$

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

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

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

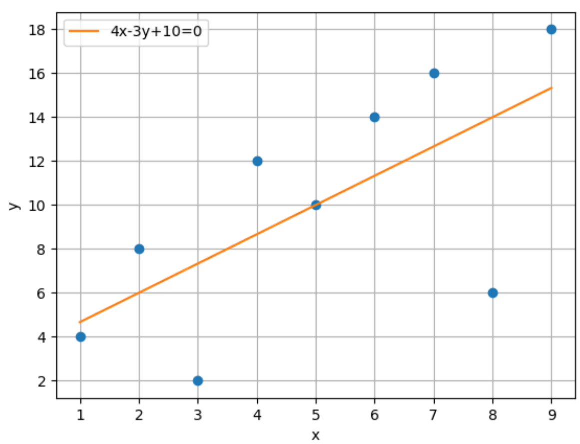

Fig. 0: Graph showing the Regression Line