Linear Regression

Data ÷

X	Y
2	3
4	7
6	5
8	10

We know,

$$b = \bar{y} - m\bar{x} \nearrow ①$$ and

$$m = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x_1})^2}$$

$$\frac{\sum_{i=1}^{n} x_i}{n} \Rightarrow \boxed{\bar{x} = 5}$$

or we can also write it as

$$m = \frac{\overline{xy} - b\bar{x}}{\overline{x^2}} \rightarrow ②$$

$$\frac{\sum_{i=1}^{n} y_i}{n} \Rightarrow \boxed{\bar{y} = 6.25}$$

where m is slope & b is intercept

Calculating all the parameters

$$\frac{\sum_{i=1}^{n}(x_i * y_i)}{n} \Rightarrow \boxed{\overline{xy} = 36}$$

$$\frac{\sum_{i=1}^{n}(x_i^2)}{n} \Rightarrow \boxed{\overline{x^2} = 30}$$

Using values of $\bar{x}$, $\bar{y}$, $\overline{xy}$ and $\overline{x^2}$ to find the values of slope and intercept.

- Slope $\Rightarrow m = 0.95$
- Intercept $\Rightarrow b = 1.5$

Equation of best fit line $\rightarrow$ $\boxed{Y = 0.95 X + 1.5}$

Linear Regression calculator

Data Set X

2,4,6,8

Data Set Y

3,7,5,10

Reset ↰   Calculate →

Step by step solution :

Data :

$X = 2, 4, 6, 8$

$Y = 3, 7, 5, 10$

Solution :

Now we have to find mean of both datasets:

Mean for Dataset X :

$$\overline{X} = \frac{20}{4}$$

$$\overline{X} = 5$$

Mean for Dataset Y :

$$\overline{Y} = \frac{25}{4}$$

$$\overline{Y} = 6.25$$

x_i	y_i	$x_i.y_i$	x^2	y^2
2	3	6	4	9
4	7	28	16	49
6	5	30	36	25
8	10	80	64	100
$\sum x_i = 20$	$\sum y_i = 25$	$\sum x_i.y_i = 144$	$\sum x^2 = 120$	$\sum y^2 = $

The slope m,

$$m = \frac{(n \times \sum x_i \times y_i) - (\sum x_i \times \sum y_i)}{n \times \sum x^2 - (\sum x)^2}$$

Now putting values in the above equation:

$$m = \frac{(4 \times 144) - (20 \times 25)}{4 \times 120 - 400}$$

$$m = \frac{76}{80}$$

$$m = 0.95$$

The intercept b,

$$b = \frac{(\sum y_i) - (m \times \sum x_i)}{n}$$

$$b = \frac{(25) - (0.95 \times 20)}{4}$$

$$b = 1.5$$

The equation of Linear Regression,

$$y = mx + b$$

$$y = 0.95x + 1.5$$