U3A Maths

16-Dec-2022 Term4/Week10

Linear Regression (Cont'd)

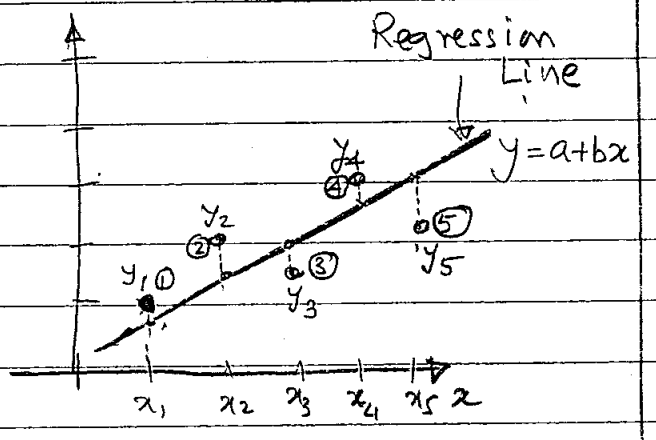
- Last week we developed simple Linear Regression equation based on 2 data values - (x_1, y_1) & (x_2, y_2)
- However, in practice we have tens or hundreds of data values.
- In such cases, Linear Regression equation is obtained based on Least Square (of Errors) method developed by Gauss!

- Linear Regression equation has the form as given below:

$$y = a + bx$$

We need to find the best estimates for 'a' & 'b' based on the given data values.

- The 'best estimate' provides minimum of sum of squares of the errors at each data point.



- Error at point ① $= (y - y_1)$ where y is the value obtained from $y = a + bx$
- Least Square Method Minimises $\sum_{i=1}^{N} (y - y_i)^2$
- In other words, 'a' & 'b' value provides lowest sum of square of errors.

- Details of the Least Square method is beyond the scope of this class. In fact, beyond the scope of statistics books!
- The statistics books only provide the final equations and not the proof!
- Regression Equation is $y = a + bx$.

where

$$b = \frac{N\Sigma(xy) - (\Sigma x)(\Sigma y)}{N\Sigma(x^2) - (\Sigma x)^2}$$

$$a = \frac{\Sigma y - b\Sigma x}{N}$$

Ex.1 The relationship between hours of sunshine and the number of ice creams sold in a shop is as given below:

No. of Hours (x)	No. of Ice Creams (y)
2	4
3	5
5	7
7	10
9	15

Find 'Linear Regression Equation and ice creams sold when hours of sunshine is 8.

∴ Linear Regression Eqn is

$$\boxed{y = 0.3049 + 1.5183x}$$

For $x = 8$ hrs

$$y = 0.3049 + 1.5183 \times 8$$
$$= 12.45$$

Say 13 ice creams

Note: It is always better to err on the safe side while estimating

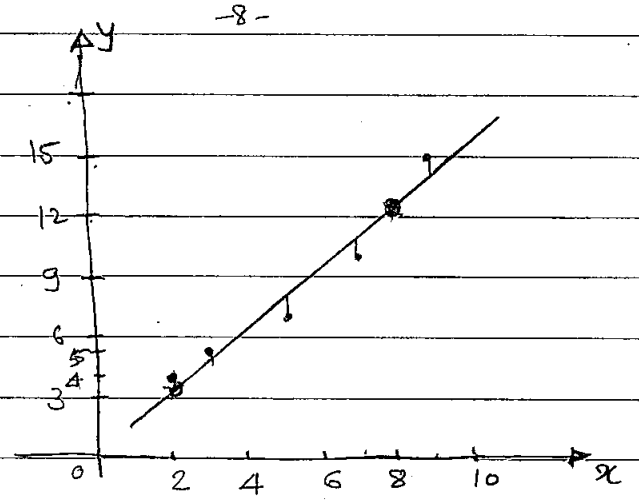
x	y	xy	x^2
2	4	8	4
3	5	15	9
5	7	35	25
7	10	70	49
9	15	135	81
26 (Σx)	41 (Σy)	263 (Σxy)	168 (Σx^2)

$$b = \frac{N\Sigma(xy) - (\Sigma x)(\Sigma y)}{N\Sigma x^2 - (\Sigma x)^2} = \frac{5 \times 263 - 26 \times 41}{5 \times 168 - (26)^2}$$

$$= 1.5183$$

$$a = \frac{\Sigma y - b\Sigma x}{N} = \frac{41 - 1.5183 \times 26}{5}$$

$$= 0.3049$$



For Regression Line

for $x = 2$, $\hat{y}_2 = 0.3049 + 1.5183 \times 2$
$= 3.0415$

for $x = 8$, $\hat{y}_8 = 12.45$

Ex.2

Calculate the error when $x = 2$

Error $= 4 - 3.0415$
$= 0.585$ ≈ 14.6% Error!