Name : Ayush Jain
SAP ID : 60004200132
Div : B
Computer Engineering

Maths - Tutorial 4

1) Find the equation of two regression lines for the following data.

x	1	2	3	4	5
y	2	5	3	8	7

Also find the most ~~proable~~ probable value of (i) y when $x = 10$
(ii) x when $y = 12$

Ans

Mean of x, $\bar{x} = \frac{1+2+3+4+5}{5}$ $\therefore \bar{x} = 3$

Mean of y, $\bar{y} = \frac{2+5+3+8+7}{5}$ $\therefore \bar{y} = 5$

x	y	x^2	y^2	xy
1	2	1	4	2
2	5	4	25	10
3	3	9	9	9
4	8	16	64	32
5	7	25	49	35
		$\Sigma x^2 = 55$	$\Sigma y^2 = 151$	$\Sigma xy = 88$

The regression coefficient for line of regression of y on x is $b_{yx} = \frac{\frac{1}{n}\Sigma xy - \bar{x}\bar{y}}{\frac{1}{n}\Sigma x^2 - \bar{x}^2}$

The regression coefficient for the line of regression of x on y is $b_{xy} = \frac{\frac{1}{n}\Sigma xy - \bar{x}\bar{y}}{\frac{1}{n}\Sigma y^2 - \bar{y}^2}$

Here, $n = 5$

$$b_{yx} = \frac{\frac{1}{5} \times 88 - 3 \times 5}{\frac{1}{5} \times 55 - (3)^2} = \frac{17.6 - 15}{11 - 9}$$

$$b_{yx} = 1.3$$

$$b_{xy} = \frac{\frac{1}{5} \times 88 - 3 \times 5}{\frac{1}{5} \times 151 - (5)^2} = \frac{17.6 - 15}{30.2 - 25}$$

$$b_{xy} = 0.5$$

The equation of line of regression of y on x is

$$(y - \bar{y}) = b_{yx} (x - \bar{x})$$

$$(y - 5) = 1.3 (x - 3)$$

$$10y - 50 = 13x - 39$$

$$13x - 10y + 11 = 0 \quad \text{—— (i)}$$

The equation of line of regression of x on y is

$$(x - \bar{x}) = b_{xy} (y - \bar{y})$$

$$(x - 3) = 0.5 (y - 5)$$

$$2x - 6 = y - 5$$

$$2x - y - 1 = 0 \quad \text{—— (2)}$$

(i) Value of y when $x = 10$,

Substitute $x = 10$ in eq (i)

$$13(10) - 10y + 11 = 0$$

$$10y = 130 + 11$$

$$10y = 141$$

$$y = 14.1$$

(ii) Value of x when $y = 12$

Substitute $y = 12$ in eq (2)

$$2x - 12 - 1 = 0$$

$$2x = 13$$

$$x = 6.5$$

Q.2) If the two regression equations are $4x - 5y - 33 = 0$ and $20x - 9y - 107 = 0$. Find the correlation coefficient between them (x and y). Also find the value of x when $y = 15$. Find standard deviation of x if varience of $y = 16$. Find the mean of x and mean of y. Also, find the angle between the two lines.

→ Given: $4x - 5y - 33 = 0$ — (i)

$20x - 9y - 107 = 0$ — (2)

Equation (i) is equation of line of regression of y on x

$$\therefore 4x - 5y - 33 = 0$$

$$5y = 4x - 33$$

$$y = \frac{4}{5}x - \frac{33}{5}$$

The regression coefficient, $b_{yx} = \frac{4}{5}$

Equation (2) is equation of line of regression of x on y

$$\therefore 20x - 9y - 107 = 0$$

$$\therefore x = \frac{9}{20}y + \frac{107}{20}$$

The regression coefficient $b_{xy} = 9/20$

The correlation coefficient is given by

$$r = \sqrt{b_{yx} \times b_{xy}}$$

$$= \sqrt{\frac{4}{5} \times \frac{1}{20}} = \frac{3}{5}$$

$$r = 0.6$$

For value of x on $y = 15$, use eq (2)

$$20x - 9y - 107 = 0$$

$$20x = 9y + 107$$

$$x = \frac{9 \times 15}{20} + \frac{107}{20}$$

$$x = 12.1$$

Now, varience of y, $\sigma_y^2 = 16$

$\therefore$ S.D of y : $\sigma_y = \sqrt{\sigma_y^2}$

$$= \sqrt{16}$$

$$\therefore \sigma_y = 4$$

Now, $b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x}$

$$\sigma_x = r \cdot \frac{\sigma_y}{b_{yx}}$$

$$= 0.6 \times 5$$

$$\sigma_x = 3$$

Now, the intersection point of eq (i) and (2) gives $\bar{x}$ and $\bar{y}$

$$4x - 5y - 33 = 0$$

$$20x - 9y - 107 = 0$$

$\therefore y = -\frac{58}{16} = -3.625$, $x = 3.71875$

Sundaram

FOR EDUCATIONAL USE

Scanned with CamScanner

Angle between 2 regression lines is given by

$$\tan\theta = \frac{(1-r^2)}{r}\left[\frac{\sigma_x \sigma_y}{\sigma_x^2+\sigma_y^2}\right]$$

$$= \frac{[1-(0.6)^2]}{(0.6)}\left[\frac{3\times 4}{3^2+4^2}\right]$$

$$= \frac{0.64}{0.6}\left[\frac{12}{9+16}\right]$$

$$= \frac{0.64}{0.6}\times\frac{12}{25}$$

$$\tan\theta = \frac{64}{125}$$

$$\theta = \tan^{-1}\left[\frac{64}{125}\right]$$

$$\theta = 27.1125^\circ$$

Q 3) Find the second order polynomial curve to the data given below.

x	2	4	6	8	10	12	14
y	175	600	1600	2250	3500	5000	6780

→ Let us consider the second order polynomial curve to be

$y = a + bx + cx^2$ — (i)

x	y	x^2	x^3	x^4	xy	x^2y
2	175	4	8	16	350	700
4	600	16	64	256	2400	9600
6	1600	36	216	1296	9600	57600
8	2250	64	512	4096	18000	144000
10	3500	100	1000	10000	35000	350000
12	5000	144	1728	20736	60000	720000
14	6780	196	2744	38416	94920	1328880
$\Sigma x = 56$	$\Sigma y = 19905$	$\Sigma x^2 = 560$	$\Sigma x^3 = 6272$	$\Sigma x^4 = 74816$	$\Sigma xy = 220270$	$\Sigma x^2y = 2610780$

Here, $n = 7$

Now put summation on LHS and RHS of eq (i)

$\therefore \Sigma y = a\Sigma 1 + b\Sigma x + cx^2$

$\Sigma y = an + b\Sigma x + c\Sigma x^2$ — (2)

Multiply eq (i) with x and put summation throughout

$xy = ax + bx^2 + cx^3$ ~~(3)~~

$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3$ — (3)

FOR EDUCATIONAL USE

Now, multiply eq (1) with x^2 and put summation throughout.

$$x^2y = ax^2 + bx^3 + cx^4$$

$$\Sigma x^2y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 \quad — (4)$$

Substituting values in eq (2), (3), (4) we get

$$19905 = 7a + 56b + 560c \quad — (5)$$

$$220270 = 56a + 560b + 6272c \quad — (6)$$

$$2610780 = 560a + 6272b + 74816c \quad — (7)$$

Solving eq (5), (6) and (7) we get

$a = -19.2857$

$b = 46.1012$

$c = 31.1756$

Thus, the second order polynomial curve for given data is

$$y = 31.1756x^2 + 46.1012x - 19.2857$$

Q. 4) A sample of 10 homes sold in goregoan each selected and the following data was gathered.

Home size (in sq ft)	1400	1300	1200	950	900	1000	1300	850	1100	800
Selling price (₹ Lakh)	70	62	65	45	40	53	68	40	55	38

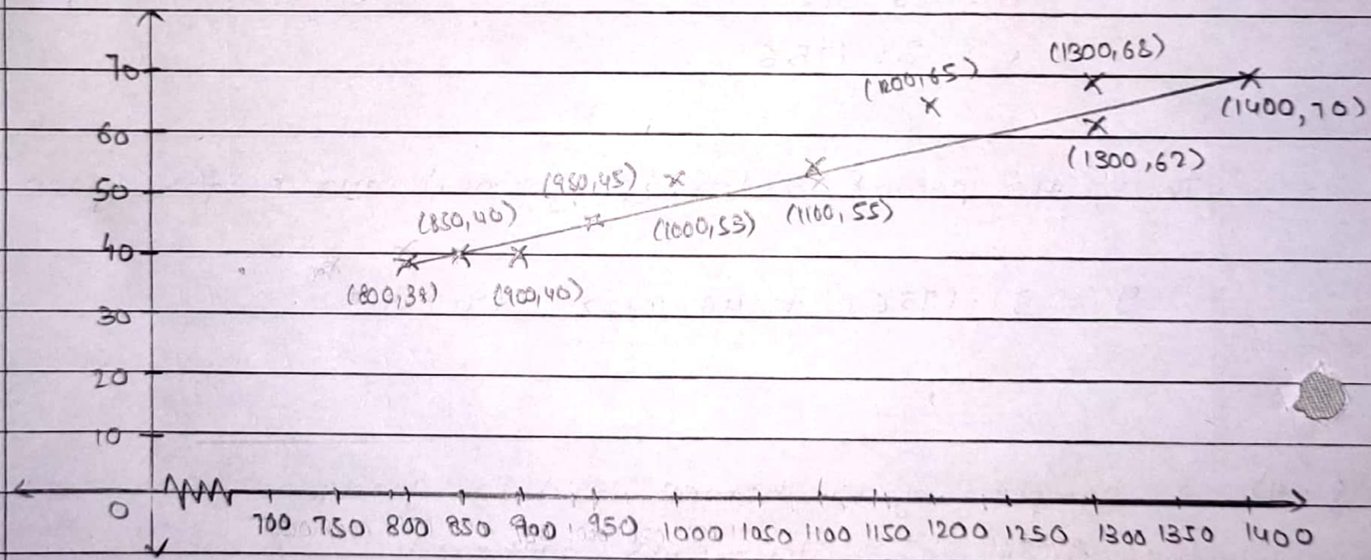
(1) Which variable is dependent variable.

(2) Comment on correlation between home size and selling price, by plotting scatter diagram.

(iii) Develop the regression eq of selling price on home size.
(iv) Comment on goodness of fit.
(v) Estimate the price of the 1000 sq.ft home using regression equation.

→ (i) Selling price is the dependent variable.
The value of selling price changes with change in value of home size.

(ii) Let home size be on x-axis
Selling price be on y-axis



COMMENT: The scatter points are all close to line of regression of y on x. Thus from the diagram we can say that the correlation is a positive correlation and that the value of variable can be determined accurately from another variable value.

(iii)

Home Size (x)	Selling Price (y)	x^2	xy
1400	70	1960000	98000
1300	62	1690000	80600
1200	65	1440000	78000
950	45	902500	42750
900	40	810000	36000
1000	53	1000000	53000
1300	68	1690000	88400
850	40	722500	34000
1100	55	1210000	60500
800	38	640000	30400
$\Sigma x = 10800$	$\Sigma y = 536$	$\Sigma x^2 = 12065000$	$\Sigma xy = 601650$

Here, $n = 10$

$$\bar{x} = \frac{\Sigma x}{10} = \frac{10800}{10} = 1080$$

$$\bar{y} = \frac{\Sigma y}{10} = \frac{536}{10} = 53.6$$

The regression coefficient is given by

$$byx = \frac{\frac{1}{n}\Sigma xy - \bar{x}\bar{y}}{\frac{1}{n}\Sigma x^2 - \bar{x}^2}$$

$$= \frac{\frac{1}{10} \times 601650 - 1080 \times 53.6}{\frac{1}{10} \times 12065000 - (1080)^2}$$

$$= \frac{2277}{40100}$$

$byx = 0.0568$

The regression evaluation of selling price on home size is given by

$$(y - \bar{y}) = b_{yx}(x - \bar{x})$$

$$(y - 53.6) = 0.0568(x - 1080)$$

$$0.0568x - y + 53.6 - 61.344 = 0$$

$$0.0568x - y - 7.744 = 0$$

$$x - 17.6056y - 136.338 = 0$$

(iv) COMMENT: From the curve equation and scatter diagram, we can say that the value of selling price can be accurately predicted from the equation of selling price on home size.

(v) Given, home size $x = 1000$

Let us consider the regression equation,

$$0.0568x - y - 7.744 = 0$$

$$y = 0.0568(1000) - 7.744$$

$$y = 49.0.56$$

$$y = 49.056$$

∴ The selling price of home of size 1000 sq.ft is ₹ 49.056 lakhs.