Date __/__/__

(5) The angle between the two regression lines depends on the correlation coefficient r. when $r=0$, the two lines are perpendicular to each other; when $r=1$ or $r=-1$, the coincide, the angle between the regression lines dimishes from 90° to 0°.

(6) The two regression equation are usually diffrent. Howeever, when $r=\pm 1$ they become identical; and in this case there is an exact liniar relationship between the variables. when $r=0$, the regression equation reduce to $y=\bar{y}$ & $x=\bar{x}$ & neither y nor x can be estimated from linear regression equation.

Angle between Two regression lines

Que If θ is the acute angle between the two regression lines in case of two variables x & y shows that

$$\tan\theta = \frac{1-r^2}{r} \cdot \frac{\sigma_x \cdot \sigma_y}{\sigma_x^2 + \sigma_y^2}$$

where, r, σx, σy have their usual meaning explain the significance of the formula when $r=0$ and $r=\pm 1$

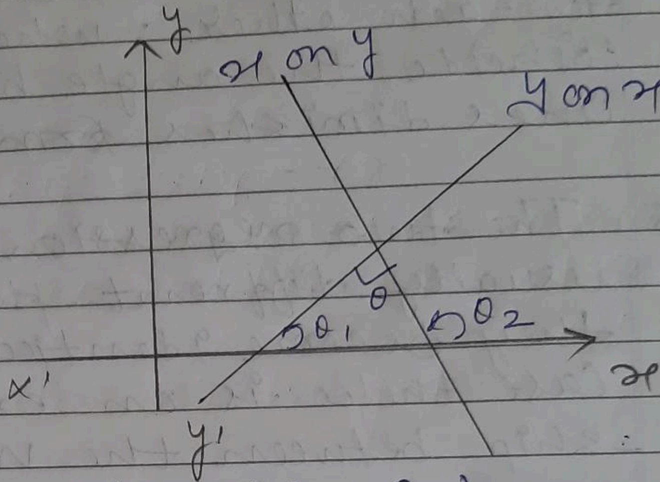
Soln If θ_1, θ_2 be the angle which the two regression lines make with x-axis then

$$\tan\theta_1 = b_{yx} = r\frac{\sigma_y}{\sigma_x}$$

Date ___ / ___ / ___

and

$$\tan\theta_2 = \frac{1}{b_{xy}} = \frac{\sigma_y}{r\sigma_x}$$



hence,

$$\tan\theta = \tan(\theta_2 - \theta_1)$$

$$= \frac{\tan\theta_2 - \tan\theta_1}{1 + \tan\theta_2 \tan\theta_1}$$

$$= \frac{\frac{\sigma_y}{r\sigma_x} + \frac{r\sigma_y}{\sigma_x}}{1 + \left(\frac{\sigma_y}{r\sigma_x}\right)\left(\frac{r\sigma_y}{\sigma_x}\right)}$$

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

when $r = \pm 1$, $\tan\theta = 0 \Rightarrow \theta = 0$ or π

Hence, the lines of regression coincide & there is a perfect correlation between the two variable x and y.

Date __/__/__

When $r=0$, $\theta = \frac{\pi}{2}$ i.e. the two lines of regression are perpendicular to each other. the estimated value of y is the same for all values of x or vice-versa.

Qus Prove that the coefficient of correlation is the geometrical mean of the coefficient of regression.

Soln The regression coefficient b_{yx} & b_{xy} may be expressed in terms of the correlation coefficient (r) and the standred deviation of x and y (veiz σ_x and σ_y) by the relation

$$b_{yx} = \frac{r\sigma_y}{\sigma_x}, \quad b_{xy} = r\frac{\sigma_x}{\sigma_y}$$

$$\therefore b_{yx} b_{xy} = \left(r\frac{\sigma_y}{\sigma_x}\right)\left(r\frac{\sigma_x}{\sigma_y}\right) = r^2$$

or

$$r^2 = b_{xy} \cdot b_{yx}$$

or

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

Thus the coefficient r is the geometric mean of the two regression coefficient b_{xy} and b_{yx}.

Qus You are given that the variance of x is 9. the regression equations are $8x-10y+66=0$ and $40x-18y=214$ find-

(1) Averge value of x and y.

Page No. ____

Date __/__/__

(II) Coefficient of correlation between the two variables.

(iii) Standred duiation of y.

Soln (i) let $\bar{x}$ & $\bar{y}$ be the average value of x & y

$\therefore$ $8\bar{x} - 10\bar{y} + 66 = 0$ —(1)

and $40\bar{x} - 18\bar{y} = 214$ —(2)

$\because$ both lines $8x - 10y + 66 = 0$ & $40x - 18y = 214$ are intersecting at $\bar{x}$ and

on solving ① & ② we get

$$\bar{x} = 13, \quad \bar{y} = 17$$

(ii) let us assume that $8x - 10y + 66 = 0$ represents the regression equation of x and y and $40x - 18y = 214$. the regression eqn of y on x. These equati can be re-written as

$$x = \frac{5y}{4} - \frac{33}{4} \text{ and } y = \frac{20x}{9} - \frac{107}{9}$$

representing. The regression coefficient should then be $b_{xy} = \frac{5}{4}$. and $b_{yx} = \frac{20}{9}$

$$r^2 = b_{xy} \cdot b_{yx}$$

$$= (5/4)(20/9)$$

$$= 25/4 > 1$$

Date __/__/__

$r^2 > 1$ (which is not possible)
so our supposition must be wrong.

$\therefore$ eq^n $8x - 10y + 66 = 0$ is regression equation of y on x and $40x - 18y = 214$ the regression eq^n of x on y. The two equation when expressed in the usual form given.

$$y = \left(\frac{4}{5}\right)x + \frac{33}{5} \quad \text{and} \quad x = \left(\frac{9}{20}\right)y + \left(\frac{107}{20}\right)$$

The correct value of regression coefficient are $b_{yx} = 4/5$ and $b_{xy} = 9/20$

i.e. $r = \sqrt{b_{xy} \cdot b_{yx}} = \sqrt{4/5 \times 9/20}$

$$\left[r = +0.6\right]$$

iii) given $\sigma_x^2 = 9$

i.e. $\sigma_x = 3$

$b_{yx} = 4/5$

$$r\frac{\sigma_y}{\sigma_x} = 4/5$$

$$(0.5)\frac{\sigma_y}{3} = \frac{4}{5}$$

This given $\boxed{\sigma_y = 4}$

Date __/__/__

Que find the most likely price in Mumbai corresponding to the price of Rs 70 at kolkata from the following.

	Kolkata	Mumbai
Average price	65	67
Standard deviation	2.5	3.5

Correlation coefficient between the prices of commodities in the two cities is 0.8.

Soln Let the prices [in Rupees] in kolkata and Mumbai be denoted by X & Y respectively.

Then we are given

$\bar{X} = 65, \ \bar{Y} = 67, \ \sigma_x = 2.5, \ \sigma_y = 3.5$ and

$r = 0.8$

We want y for $x = 70$

line of regression of y on x is

$$y - \bar{y} = r \frac{\sigma_y}{\sigma_x} (x - \bar{x})$$

$$y = 67 + 0.8 \times \frac{3.5}{2.5} (x - 65)$$

When $x = 70$

$$y = 67 + (0.8) \times \frac{(3.5)}{(2.5)} (5)$$

$$y = 72.6$$

Hence, the most likely price in mumbai corresponding to the price of Rs 70 at kolkata is Rs 72.60.

Date __/__/__

Que Obtain the equation of two lines of regression for the following data. Also obtain the estimate of X for $Y=70$

X	65	66	67	67	68	69	70	72
Y	67	68	65	68	72	72	69	71

Soln let $U = X-68$ and $V = Y-69$ then

X	Y	U = X-68	V = Y-69	U^2	V^2	UV
65	67	-3	-2	9	4	6
66	68	-2	-1	4	1	2
67	65	-1	-4	1	16	4
67	68	-1	-1	1	1	1
68	72	0	3	0	9	0
69	72	1	3	1	9	3
70	69	2	0	4	0	0
72	71	4	2	16	4	8
		0	0	36	44	24

$\bar{U}=0,\ \bar{V}=0,\ \sigma_U^2 = \frac{1}{n}\Sigma u^2 - (\bar{u})^2,\ \frac{1}{8}\times 36 = 4.5$

$\sigma_V^2 = \frac{1}{n}\Sigma V^2 - (\bar{V})^2 = \frac{1}{8}\times 44 = 5.5$

$r(u,v) = \frac{Cov(u,v)}{\sigma_u \sigma_v} = \frac{\frac{1}{n}\Sigma uv - \bar{u}\bar{v}}{\sigma_u \sigma_v} = \frac{3}{\sqrt{(4.5)(5.5)}}$

$= 0.603 = 0.6$

Date __/__/__

$u = x - 68 \Rightarrow \bar{u} = \bar{x} - 68 \Rightarrow \bar{x} = 68$

$v = y - 69 \Rightarrow \bar{v} = \bar{y} - 69 \Rightarrow \bar{y} = 69$

$\sigma_x = \sigma_u = \sqrt{4.5} = 2.12$

$\sigma_y = \sigma_v = \sqrt{5.5} = 2.35$

Equation of line of regression of y on x is

$$(y - \bar{y}) = r\frac{\sigma_y}{\sigma_x}(x - \bar{x})$$

$$y = 69 + (0.6)\left(\frac{2.35}{2.12}\right)(x - 68)$$

$$[y = 0.665x + 23.58]$$

eq^n of line of regression x on y is

$$(x - \bar{x}) = r\frac{\sigma_x}{\sigma_y}(y - \bar{y})$$

$$x = 68 + (0.6)\left(\frac{2.12}{2.35}\right)(x - 69)$$

$$x = 0.544 + 30.74$$

if $y = 70$, $x = (0.54)(70) + 30.74$

$x = 68.54$

estimate value of x is 68.54 when x is 70.

Date __/__/__

Curve fitting :-

Curve fitting means an expression of the relationship between two variables by algebraic equations on the basis of observed data.

In practical statistics we are required to find a functional ~~of the~~ relation between x and y where the dependent variable say y, is expressed as a function of the independent variable say x, which may involve power upto n where $n = 1, 2, 3 \ldots n$

The general problem to find equation of approximating curves which fit given sets of data is called curve fitting.

** The method of least square used in the curve fitting.

fitting Straight Line -

let $y = a + bx$ be the equation of straight line to be fitted to a given set of n parts of observations $(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)$

Applying the method of least square, the value of a and b are to determined as the minimise -

$$\sum_{i=1}^{n} (y_i - a - bx_i)^2$$

Date __/__/__

This leads to the normal equations

$$\Sigma y = an + b\Sigma x \quad \text{—— (1)}$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2 \quad \text{—— (2)}$$

Here the value of n, Σx, Σx^2, Σy, Σxy are substituted on the basis of given data. We have them two equations involving a and b, solving which the values of a and b are obtained.

Example - Determine the equation of straight line which best fix the following data.

x	10	12	13	16	17	20	25
y	19	22	24	27	29	33	39

Solⁿ first Method →

let $y = a + bx$, be the eqn of the best fitting straight line by method of least square. The constants a & b are obtained by solving the normal equations.

$$\Sigma y = ax + b\Sigma x, \quad \Sigma xy = a\Sigma x + b\Sigma x^2$$

where

x is the no. of pairs of observations

Calculating for fitting straight line

putting $\Sigma y = 191$, $\Sigma x = 113$

Page No.

Date __/__/__

~~putting~~ $\Sigma xy = 3276$, $\Sigma x^2 = 1983$ in the normal eqn we have

x	y	x^2	xy
10	19	100	190
12	22	144	264
13	24	169	312
16	27	256	432
17	29	289	493
20	33	400	660
25	37	625	925
total 113	191	1983	3276

$$191 = 7a + 113b \quad \text{---(1)}$$
$$3276 = 113a + 1983b \quad \text{---(2)}$$

multiplying (1) by 113 & (2) by 7 & subtracting

$$21,583 = 791a + 12,769\,b$$
$$22,932 = 791a + 13,881\,b$$
$$\underline{-\qquad\quad -\qquad\quad -}$$
$$-1349 = 0 + (-1,112)\,b$$

$$[b = 1.21]$$

putting the value of b in (1), we have

$$7a = 191 - 113 \times 1.21$$
$$= 54.21$$
$$\text{so } a = 7.75$$

the equation of fitted straight line is

$$y = 7.75 + 1.21x$$

Date __/__/__

2^{nd} Method -

Let $y = a + bx$ be the equation of best fitting straight line, where, $X = x - 16$ & $Y = y - 26$.

The normal eq^n^s are $\Sigma Y = an + b\Sigma X$,

$\Sigma XY = a\Sigma X + b\Sigma X^2$.

Calculation for fitting straight line

x	y	$X = x-16$	$Y = y-26$	X^2	XY
10	19	-6	-7	36	42
12	22	-4	-4	16	16
13	24	-3	-2	9	6
16	27	0	1	0	0
17	29	1	3	1	3
20	33	4	7	16	28
25	37	9	11	81	99
Total		1	9	159	194

Substituting the values in normal eq^n^, we get

$$9 = 7a + b \quad —(3)$$

$$194 = a + 159 \quad —(4)$$

multiplying (4) by (7) & substituting from (3)

$$9 = 7a + b$$

$$-\underline{1358 = 7a + 1113b}$$

$$-1349 = -1112b \qquad b = 1.21$$

putting value in (3), we get $a = 1.11$

$$\therefore Y = 1.11 \times 1.21 X$$

$$y - 26 = 1.11 + 1.21(x - 16)$$

or $$y = 7.75 + 1.21x$$

Date __/__/__

Fitting a Parabola -

let $y = a + bx + cx^2$ ——(1) be the equation of parabola to be fitted to a given set of n pair of observation $(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)$. using the method of least square. the constant a, b & c of the best fitting Parabola are obtained by solving the normal equations.

$$\Sigma y = an + b\Sigma x + c\Sigma x^2$$
$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3$$
$$\Sigma x^2y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4.$$

Eg. find a Second degree parabola to the following data

x	0	1	2	3	4
y	1	5	10	22	38

Sol^n

	x	y	x^2	x^3	x^4	xy	x^2y
	0	1	0	0	0	0	0
	1	5	1	1	1	5	5
	2	10	4	8	16	20	40
	3	22	9	27	81	66	198
	4	38	16	64	256	152	608
total	10	76	30	100	354	243	851

let the parabola to be fit be $y = a + bx + cx^2$

on substituting the values of $\Sigma x_i, \Sigma y_i \ldots$ etc.
the normal equation are

Date __/__/__

$$76 = 5a + 10b + 30c$$
$$243 = 10a + 30b + 100c$$
$$851 = 30a + 100b + 354c$$

On solving there simulataneous eqⁿ, we get
$a = 1.43, \quad b = 0.24, \quad c = 2.21$
Hence, the equation of the curve is
$$y = 1.43 + 2.24x + 2.21x^2$$

Eg. fit a second degree parabola to the given data, using method of least square:

length :	7.5	10.0	12.5	15.0	17.5	20.8	22.5
weight :	1.9	4.5	10.1	17.6	27.8	40.8	56.9

Ans let x and y represent the length & weight the equation at the parabola may be written as -
$$y = a + bu + cu^2$$
where $u = \frac{x - 15}{2.5}$

Using the method of least square, the normal equations for determining the constants, a, b, c, are

$$\Sigma y = an + b\Sigma u + c\Sigma u^2$$
$$\Sigma uy = a\Sigma u + b\Sigma u^2 + c\Sigma u^3$$
$$\Sigma u^2 y = a\Sigma u^2 + b\Sigma u^3 + c\Sigma u^4$$

Date __/__/____

Calculation for fitting parabola-

x	y	u	u^2	u^3	u^4	uy	u^2y
7.5	1.9	-3	9	-27	81	-5.7	17.1
10.0	4.5	-2	4	-8	16	-9.0	18.0
12.5	10.1	-1	1	-1	1	-10.1	10.1
15.0	17.6	0	0	0	0	0	0
17.5	27.8	1	1	1	1	27.8	27.8
20.0	40.8	2	4	8	16	81.6	163.2
22.5	56.9	3	9	27	81	170.7	512.1

Substituting the values of Σy, Σu, Σu^2... etc in the normal equation (here $n=7$), we get

$$159.6 = 7a + 28c$$
$$255.3 = 28b$$
and $748.3 = 28a + 196c$

Solving $a = 17.6$, $b = 9.1$ and $c = 1.3$

Putting these values in ①, the equation of the parabola is

$$y = 17.6 + 9.1u + 1.3u^2$$
where $u = \frac{(x-15)}{2.5} = 0.4x - 6$

Substituting for u we have

$$y = 17.6 + 9.1(0.4x-6) + 1.3(0.4x-6)^2$$
$$= 17.6 + 3.64x - 54.6 + 1.3(0.16x^2 - 48x + 36)$$

Date __/__/__

$= -0.37 + 3.64x - 54.6 + 1.3(0.16x^2 - 4.8x + 36)$

$= -0.37 + 3.64 + 0.208x^2 - 6.24x + 46.8$

$y = 9.8 - 2.60 + 0.208x^2$

This is the required equation of the parabola.