Statistics and probability

Statistical methods:-

Mean:- If $x, x_2, x_3 \ldots\ldots x_n$ are the set of n values of a variant x, the mean is denoted by $\bar{x}$ and is defined as

$$\bar{x} = \Sigma x_i = \frac{\Sigma x}{n}$$

variance If a variant x take values $x, x_2, x_3 \ldots x_n$ the variance is defined as

$$V = \frac{\Sigma (x - \bar{x})^2}{n}$$

Standard deviation (SD):-

$$\sigma = \sqrt{V}$$

(or)

$$\sigma^2 = V$$

curve Fitting (least square method)

The method of finding a specific a relation $y = f(x)$ for the data to satisfy accurately as possible and such an equation is called as the method of least squares

1) Fitting of a straight line $y = a + bx$

The normal equation for fitting the straight line $y = a + bx$ in the least square seence (or) method are given by.

$$na + b\Sigma x = \Sigma y$$

$$a\Sigma x + b\Sigma x^2 = \Sigma xy$$

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

2) fitting of a second degree parabola

$$y = a + bx + cx^2$$

The normal equation for fitting the second degree parabola are given by

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

(or)

$$y = ax^2 + bx + c$$
$$a\Sigma x^2 + b\Sigma x + nc = \Sigma y$$
$$a\Sigma x^3 + b\Sigma x^2 + c\Sigma x = \Sigma xy$$
$$a\Sigma x^4 + b\Sigma x^3 + c\Sigma x^2 = \Sigma x^2 y$$

3) fitting of the curve $y = ab^x$

consider $y = ab^x$

Taking log on Both sides we get

$$\log_e y = \log_e (ab^x)$$
$$= \log_e a + \log_e b^x$$
$$\log_e y = \log_e a + x \log_e b$$
$$Y = A + BX \quad -①$$

where $Y = \log_e y$, $A = \log_e a$, $B = \log_e b$ and $X = x$

The normal equation for equation ① are given by

$$nA + B\Sigma X = \Sigma Y$$
$$A\Sigma X + B\Sigma X^2 = \Sigma XY$$

Solving these equation we get A and B hence a = ...
$b = e^B$.

Problem :-

1) find the values of a and b so that $y = a + bx$ fitting the data given in the data given in the table

x	0	1	2	3	4
y	1	2.9	4.8	6.7	8.6

$\Rightarrow$ consider $y = a + bx$

The normal equation are given by

$na + b\Sigma x = \Sigma y \rightarrow (1)$

$a\Sigma x + b\Sigma x^2 = \Sigma xy \rightarrow (2)$.

$n = 5$

x	y	xy	x^2
0	1	0	0
1	2.9	2.9	1
2	4.8	9.6	4
3	6.7	20.1	9
4	8.6	34.4	16
$\Sigma x = 10$	$\Sigma y = 24$	$\Sigma xy = 67$	$\Sigma x^2 = 30$

$(1) \rightarrow 5a + 10b = 24$

$(2) \rightarrow 10a + 30b = 67$

$a = 1$ $b = 1.9$

$\therefore y = 1 + (1.9)x$

(2) By the method of least squares find the straight line that fitts the following data

x	1	2	3	4	5
y	14	27	40	55	68

consider $y = a + bx$

The normal equation are given by $na + b\Sigma x$ - (1)

~~$15a + 55b = 748$~~

$a\Sigma x + b\Sigma x^2 = \Sigma xy$ - (2)

$\boxed{n=5}$

~~$a = 0$~~ , ~~$y = 13.6$~~

$\therefore$

x	y	xy	x^2
1	14	14	1
2	27	54	4
3	40	120	9
4	55	220	16
5	68	340	25
			$\Sigma x^2 = 55$
$\Sigma x = 15$	$\Sigma y = 204$	$\Sigma x = 748$	

(1) $\rightarrow 5a + 15b = 204$

(2) $\rightarrow 15a + 55b = 748$

$a = 0$, $y = 13.6$

$\therefore y = 0 + (13.6)x$

(3) Fit a straight line $y = a + bx$ in the least square sense for the data

x	1	3	4	6	8	9	11	14
y	1	2	4	4	5	7	8	9

$\Rightarrow$ consider $y = a + bx$

The normal equation are given by

$na + b\Sigma x = \Sigma y$ - (1)

$a\Sigma x + b\Sigma x^2 = \Sigma xy$ - (2)

$\boxed{n=8}$

x	y	xy	x^2
1	1	1	1
3	2	6	9
4	4	16	16
6	4	24	36
8	5	40	64
9	7	63	81
11	8	88	121
14	9	126	196
$\Sigma x = 56$	$\Sigma y = 40$	$\Sigma xy = 364$	$\Sigma x^2 = 524$

$(1) \rightarrow 5a + 18b = 15$

$(2) \rightarrow 18a + 110b = 71$

$a = 1.64$, $b = 0.376$

$y = 1.64 + (0.376)x$

(5) By the method least squares fit a straight lines to the following data.

x	0	1	2	3	4	5
y	9	8	24	28	26	20

consider, $y = a + bx$

The normal equation are given by.

$na + b\Sigma x = \Sigma y$ - ①

$a\Sigma x + b\Sigma x^2 = \Sigma xy$ - ② $\boxed{n=6}$

x	y	xy	x^2
0	9	0	0
1	8	8	1
2	24	48	4
3	28	84	9
4	26	104	16

5 20 100 25

$\Sigma x = 15$ $\Sigma y = 115$ $\Sigma xy = 344$ $\Sigma x^2 = 55$

(1) $\rightarrow$ $6a + 15b = 115$

(2) $\rightarrow$ $15a + 55b = 344$

$a = 11.09$ $b = 3.22$

$y = 11.09 + (3.22)x$

(6) By the method of least squares fit a straight line $y = ax + b$ for the following data

x	0	1	2	3	4
y	1.0	1.8	3.3	4.5	6.3

$\Rightarrow$ The normal equation are given by

$y = ax + b$:

$a\Sigma x + nb = \Sigma y$ — (1)

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

$\boxed{n = 5}$

x	y	xy	x^2
0	1.0	0	0
1	1.8	1.8	1
2	3.3	6.6	4
3	4.5	13.5	9
4	6.3	25.2	16

$\Sigma x = 10$ $\Sigma y = 16.9$ $\Sigma xy = 47.1$ $\Sigma x^2 = 30$.

(1) $\rightarrow$ $10a + 5b = 16.9$

(2) $\rightarrow$ $30a + 10b = 47.1$

$a = 1.33$, $b = 0.72$

$y = (1.33)x + 0.72$

7) fit a parabola of the form $y = a + bx + cx^2$ for the data.

x	0	1	2	3	4	5
y	1	3	7	13	21	31

=> The normal equation are given by.

$$na + b\Sigma x + c\Sigma x^2 = \Sigma y \quad \text{-} (1)$$

$$a\Sigma x + b\Sigma x^2 + c\Sigma x^3 = \Sigma xy \quad \text{-} (2)$$

$$a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 = \Sigma x^2 y \quad \text{-} (3)$$

$\boxed{n = 6}$

x	y	x^2	x^3	xy	x^4	x^2y
0	0	0	0	0	0	0
1	3	1	1	3	1	3
2	7	4	8	14	16	28
3	13	9	27	39	81	117
4	21	16	64	84	256	336
5	31	25	125	115	625	775
$\Sigma x = 15$	$\Sigma y = 76$	$\Sigma x^2 = 55$	$\Sigma x^3 = 225$	$\Sigma xy = 295$	$x^4 = 979$	$x^2y = 1259$

$(1) \rightarrow 6a + 15b + 55c = 76$

$(2) \rightarrow 15a + 55b + 225c = 295$

$(3) \rightarrow 55a + 225b + 979c = 1259$

$a = 1, \quad b = 1, \quad c = 1$

$$y = 1 + 1x + 1x^2$$

(8) fit a parabala of second degree $y = a + bx + cx^2$ for the data.

x	0	1	2	3	4
y	1	1.8	1.3	2.1	2.3

$\Rightarrow$ The normal equation are given by

$$na + b\Sigma x + c\Sigma x^2 = \Sigma y \quad -(1)$$
$$a\Sigma x + b\Sigma x^2 + c\Sigma x^3 = \Sigma xy \quad -(2)$$
$$a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 = \Sigma x^2 y \quad -(3)$$

$\boxed{n = 5}$

x	y	x^2	x^3	x^4	xy	x^2y
0	1	0	0	0	0	0
1	1.8	1	1	1	1.8	1.8
2	1.3	4	8	16	2.6	5.2
3	2.1	9	27	81	6.3	18.9
4	2.3	16	64	256	9.2	36.8
$\Sigma x = 10$	$\Sigma y = 8.5$	$\Sigma x^2 = 30$	$\Sigma x^3 = 100$	$\Sigma x^4 = 354$	$\Sigma xy = 19.9$	$\Sigma x^2y = 627$

(1) $\rightarrow 5a + 10b + 30c = 8.5$

(2) $\rightarrow 10a + 30b + 100c = 19.9$

(3) $\rightarrow 30a + 100b + 354c = 62.7$

$a = 1.134$, $b = 0.261$ $c = 0.007$

$$y = 1.134 + 0.261x + 0.007x^2$$

(9) Fit a parabola $y = a + bx + cx^2$ to the following data.

x	-3	-2	-1	0	1	2	3
y	4.63	2.11	0.67	0.09	0.63	2.15	4.58

⇒ The normal equation are given by.

$$na + b\Sigma x + c\Sigma x^2 = \Sigma y \quad -(1)$$
$$a\Sigma x + b\Sigma x^2 + c\Sigma x^3 = \Sigma xy \quad -(2)$$
$$a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 = \Sigma x^2 y \quad -(3)$$

$(n = 7)$

x	y	x^2	x^3	x^4	xy	x^2y
-3	4.63	9	~~9~~ -27	81 ~~27~~	~~81~~ -13.8	41.67
-2	2.11	4	-8	16	-4.22	8.44
-1	0.67	1	-1	1	-0.67	0.67
0	0.09	0	0	0	0	0
1	0.63	1	1	1	0.63	0.63
2	2.15	4	8	16	4.3	8.6
3	4.58	9	27	81	13.74	41.22
$\Sigma x = 0$	$\Sigma y = 14.86$	$\Sigma x^2 = 28$	$\Sigma x^3 = 0$	$\Sigma x^4 = 196$	$\Sigma xy = -0.11$	$\Sigma x^2y = 101.23$

(1) → $7a + 0b + 28c = 14.86$

(2) → $0a + 28b + 0c = -0.11$

(3) → $28a + 0b + 196c = 101.23$

~~a = 1.042 b = 0.24 c = 2.21~~

~~$y = 1.042 + 0.24x + 2.21x^2$~~

$a = 0.132, \quad b = -0.0039. \quad c = 0.4975$

$$y = 0.132 - 0.039x + 0.4975x^2$$

(10) fit parabola to each of the following data

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

⇒ The normal equation are given by

$y = a + bx + cx^2$

$na + b\Sigma x + c\Sigma x^2 = \Sigma y$ — (1).

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

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

$\boxed{n=5}$

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
$\Sigma x = 10$	$\Sigma y = 76$	$\Sigma x^2 = 30$	$\Sigma x^3 = 100$	$\Sigma x^4 = 354$	$\Sigma xy = 243$	$\Sigma x^2y = 851$

(1) → $5a + 10b + 30c = 76$

(2) → $10a + 30b + 100c = 243$

(3) → $30a + 100b + 354c = 851$

$a = 1.42$ $b = 0.24$ $c = 2.21$

$y = 1.42 + 0.24 + 2.21x^2$

(17) Fit a two curve of the form $y = ab^x$ by using the least square method to the following data.

x	1	2	3	4	5	6	7
y	87	97	113	129	202	195	193

consider $y = ab^x$

Taking log on BS

$\log y = \log(ab^x)$

$= \log a + \log b^x$

$\log_e y = \log a + x \log b$

$$\boxed{Y = A + BX} \rightarrow *$$

$nA + B\Sigma X = \Sigma Y$ — (1)

$A\Sigma X + B\Sigma X^2 = \Sigma XY$ — (2)

$\boxed{n = 7}$

$X = x$	y	x^2	$Y = \log_e y$	XY
1	87	1	4.46	4.46
2	97	4	4.59	9.14
3	113	9	4.92	14.16
4	129	16	4.85	19.4
5	202	25	5.30	20.6
6	195	36	5.27	31.62
7	193	49	5.26	36.82
$\Sigma x = 28$		$\Sigma x^2 = 140$	$\Sigma y = 34.43$	$\Sigma xy = 136.1$

(1) $\rightarrow 7A + 28B = 34.43$

(2) $\rightarrow 28A + 140B = 136.1$

The solving equation we get

$A = 5.15$ $\quad B = -0.05$

$a = e^A$

$b = e^B$

$a = e^{(5.15)}$

$b = e^{(-0.05)}$

$a = 172.43$

$b = 0.951$

$\therefore y = ab^x$

$y = (172.43)(0.95)^x$

(13) The growth of bateria (y) in a community after Hours(x) is given the following table.

Hours(x)	0	1	2	3	4	5	6
No of bacteria(y)	32	47	65	92	132	190	275

in the formula $y = ab^x$ to fit this data and and estimate y for $x - 7$.

$\Rightarrow$ consider $y = ab^x$

Taking log on BS

$\log_e y = \log_e (ab^x)$

$\log_e y = \log_e a + \log_e b^x$

$\log_e y = \log_e a + x \log_e b$

$y = A + BX$ — (*)

$nA + B\Sigma X = \Sigma y$ — (1)

$A\Sigma x + B\Sigma x^2 = \Sigma xy$ — (2)

$x = X$	y	$Y = \log_e y$	x^2	xY
0	32	3.48	0	0
1	47	3.85	1	3.85
2	65	4.17	4	8.34
3	92	4.52	9	13.56
4	132	4.88	16	19.52
5	190	5.24	25	26.2
6	275	5.61	36	33.66
		$\Sigma Y = 31.73$	$\Sigma x^2 = 91$	$\Sigma xY = 105.13$

$$\boxed{n = 7}$$

① → $7A + 21B = 31.73$

② → $21A + 91B = 105.13$

$A = 3.45$ $\quad\quad$ $B = 0.35.$

$a = 13.81$ $\quad\quad$ $b = 1.41$

$$y = (31.81)\,(1.41)^x$$

when $x = 7$; $y = (31.81)\,(1.41)^x$

when $x = 7$, $y = (31.81)\,(1.41)^7$

$$= 352.44.$$

(14) fit a curve of the form $y = ab^x$ in the least square sense for the following data

x	0	2	4	5	7	10
y	100	120	256	390	710	1600

→ consider $y = ab^x$

Taking log on BS.

$$\log_e y = \log_e (ab^x)$$

$$= \log a + \log b^x$$

$$\log_e y = \log a + x \log b$$

$$y = A + BX \quad - (*)$$

The normal equation,

$$nA + B\Sigma X = \Sigma y \quad - ①$$

$$A\Sigma x + B\Sigma X^2 = \Sigma xy \quad - ②$$

$X = x$	y	$y = \log_e y$	x^2	xy
0	100	4.60	0	0
2	120	4.78	4	9.56
4	256	5.54	16	22.16
5	390	5.96	25	29.8
7	710	6.56	49	45.92
10	1600	7.37	100	73.7
$\Sigma x = 28$		$\Sigma y = 34.81$		$\Sigma xy = 181.14$

$\boxed{n = 6}$

$$① \to 6A + 28B = 34.81$$

$$② \to 28A + 194B = 181.14$$

$A = 4.42$ $\quad$ $B = 0.029$

$a = e^A = e^{4.42}$ $\quad$ $b = e^B = e^{0.029}$

$= 83.09$ $\quad$ $= 1.033$

$$y = (83.09)(1.033)^x$$

15) fit a curve of the following data.

x	2	3	4	5	6
y	8.3	15.3	33.1	65.2	127.4

consider $y = ab^x$

Taking log on BS

$\log y = \log(ab^x)$

$= \log a + \log b^x$

$\log y = \log a + x \log b$

$Y = A + BX$ — (*)

The normal equation

$nA + B\Sigma X = \Sigma y$ — (1)

$A\Sigma x + B\Sigma x^2 = \Sigma xy$ — (2)

$X = x$	y	$Y = \log_e y$	x^2	xy
2	8.3	2.11	4	4.22
3	15.3	2.72	9	8.16
4	33.1	3.49	16	13.96
5	65.2	4.17	25	20.85
6	127.4	4.84	36	29.04
$\Sigma x = 20$		$\Sigma y = 17.33$	$\Sigma x^2 = 90$	$\Sigma xy = 76.23$

$n = 5$

(1) → $5A + 20B = 17.33$

(2) → $20A + 90B = 76.23$

$A = 0.702$ $\quad$ $B = 0.691$

$a = e^A = e^{0.702}$ $\quad$ $b = e^B = e^{0.091}$

$= 2.017$ $\quad$ $= 1.995$

$y = (2.017)(1.995)^x$

(16)

x	1	2	3	4	5	6	7	8
y	1.0	1.2	1.8	2.5	3.6	4.7	6.6	9.1

consider $\boxed{y = ab^x}$

Taking log on BS

$$\log_e y = \log_e (ab^x)$$

$$= \log_e a + \log_e b^x$$

$$Y = A + BX \quad - *$$

The normal equation.

$$nA + B\Sigma x = \Sigma y \quad - (1)$$

$$A\Sigma x + B\Sigma x^2 = \Sigma xy \quad - (2)$$

$X = x$	y	$Y = \log_e y$	x^2	xy
1	1.0	0	1	0
2	1.2	0.18	4	0.36
3	1.8	0.58	9	1.74
4	2.5	0.91	16	3.64
5	3.6	1.28	25	6.4
6	4.7	1.54	36	9.24
7	6.6	1.88	49	13.16
8	9.1	2.20	64	17.6
$\Sigma X = 36$		$\Sigma Y = 8.57$	$\Sigma x^2 = 204$	$\Sigma xy = 52.14$

$\boxed{n = 8}$

(1) $\rightarrow 8A + 36B = 8.57$

(2) $\rightarrow 36A + 204B = 52.14$

$A = -0.38 \qquad B = 0.32$

$a = e^A = e^{-0.38} \qquad b = e^B = e^{0.32}$

$= 0.683 \qquad = 1.377$

$y = (0.683)(1.377)$

correlation :-

co variance of two independent magnitudes is known as correlation.

The Numerical measure of correlation b/w 2 variables x and y is known as pearson's coefficient of correlation denoted by 'r' and is defined as

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n\,\bar{x}}$$

(or)

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}}$$

where $x = x - \bar{x}$
$y = y - \bar{y}$

NOTE :-
The coefficient of correlation value always lies b/w -1 and $+1$
i.e, $-1 \leq r \leq 1$

Regression :-

Regression is an estimation on one independent variable interms on the other.

(1) Regression line on y on x is given by

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

(or)

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

where $y = y - \bar{y}$
$x = x - \bar{x}$

① find the coefficient of correlation and regression lines for the following data

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

$\Rightarrow \bar{x} = \frac{1+2+3+4+5}{5} = 3$

$\Rightarrow \bar{y} = \frac{2+5+3+8+7}{5} = 5$

coefficient of corellation is given by.

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}} \quad -①$$

$X = x - \bar{x}$ $\quad$ $y = y - \bar{y}$.

$X = x - 3$ $\quad$ $y = y - 5$.

x	y	$x = x-3$	$y = y-5$	xy	x^2	y^2
1	2	-2	-3	6	4	9
2	5	-1	0	0	1	0
3	3	0	-2	0	0	4
4	8	1	3	3	1	9
5	7	2	2	4	4	4
				$\Sigma xy = 13$	$\Sigma x^2 = 10$	$\Sigma y^2 = 26$

$\therefore ① \rightarrow r = \frac{13}{\sqrt{10}\sqrt{26}}$

$\boxed{r = 0.806}$

y on x.

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

$y - \bar{y} = \frac{13}{10}(x - \bar{x})$

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

$y - 5 = (1.3)\,x - 3.9$

$y = (1.3)\,x - 3.9 + 5$

$y = (1.3)x + 1.1$

Regression line of ~~x~~ on y is given by

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

$y - \bar{y} = \frac{13}{10}(x - \bar{x})$

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

$y - 5 = (1.3)\,x - 3.9$

$y = (1.3)\,x - 3.9 + 5.$

Regression line of x on y is given by

$$x = \frac{\Sigma xy}{\Sigma y^2}(y).$$

$$x - \bar{x} = \frac{13}{26}(y - \bar{y})$$

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

$$= (0.5)\,y - 2.5$$

$$x = (0.5)\,y - 2.5 + 3$$

$$\boxed{x = (0.5)\,y + 0.5}$$

(2) Find the co-efficient of correlation and regression lines for the following data?

production (x)	55	56	58	59	60	61	62
Export (y)	35	38	38	39	44	43	45

respectively.

let x = production and y = export

$\bar{x} = \frac{411}{7} = 58.71$

$\bar{y} = \frac{282}{7} = 40.28$

$X = x - \bar{x} = x - 58.71$

$Y = y - \bar{y} = y - 40.28$

x	y	y = x − 58.71	y = y − 40.28	xy	x^2	y^2
55	35	−3.71	−5.28	19.58	13.76	27.83
56	38	−2.71	−2.28	6.17	7.34	5.19
58	38	−0.71	−2.28	1.61	0.50	5.19
59	39	0.29	−1.28	−0.37	0.08	1.63
60	44	1.29	3.72	4.79	1.66	13.83
61	43	2.29	2.72	6.22	5.24	7.39
62	45	3.29	4.72	15.52	10.82	22.27
				$\Sigma xy = 53.52$	$\Sigma x^2 = 39.4$	$\Sigma y^2 = 83.3$

coefficient of corelation is given by

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}} \quad -(1)$$

$$\therefore (1) \rightarrow r = \frac{53.52}{\sqrt{39.4}\sqrt{83.37}}$$

$$\boxed{r = 0.93}$$

Regression lies on y on x

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

$$y - 40.28 = \left(\frac{53.52}{39.4}\right)(x - 58.71)$$

$$y - 40.28 = (1.35)(x - 58.71)$$

$$y = (1.35)x - 58.71 + 40.28$$

$$y = (1.35)x - 38.96.$$

(3) In a study b/w the amount of rain fall quantity of rain pollution removed, the following data were collected find the regression lines.

Daily rain fall is (x) 4.3 4.5 5.9 6.1 5.2 3.8 2.1
0.01cm

Pollution removed 12.6 12.1 11.6 11.4 11.8 13.2 14.1
(mg/(y)

$\bar{x} = \frac{31.9}{7} \doteq 4.55$

$\bar{y} = \frac{86.8}{7} = 12.4$

$X = x - \bar{x} = x - 4.55$

$y = y - \bar{y} = y - 12.4$

x	y	$x = X - 4.55$	$y = y - 12.4$	xy	x^2	y^2
4.3	12.6	-0.25	0.2	-0.05	0.06	0.04
4.5	12.1	-0.05	-0.3	0.015	0.002	0.09
5.9	11.6	1.35	-0.8	-1.08	1.82	0.64
6.1	11.4	1.55	-1	-0.39	2.40	0.64
5.2	11.8	0.65	-0.6	-0.39	0.42	1
3.8	13.2	-0.75	0.8	-0.6	0.56	0.36
2.1	14.1	-2.45	1.7	-4.16	6.00	0.64
						2.89

$\Sigma xy = -7.8$ $\Sigma x^2 = 11.26$
$\Sigma y^2 = 5.66$

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

$y - 12.4 = \frac{-7.8}{11.26}(x - 4.55)$

$y - 12.4 = (-0.69)(x - 4.55)$

$y = (-0.69)x + 3.13 + 12.4$

$y = (-0.69)x + 15.53.$

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

$x - 4.55 = \left(\frac{-7.8}{5.66}\right)$

$y - 12.4)$

$x = (-1.37)y$

$+ 21.5$

(Q1)

x	1	2	3	4	5	6	7	8	9	10
y	10	12	16	28	25	36	41	49	40	50.

$\bar{x} = \frac{55}{10} = 5.5$

$\bar{y} = \frac{282}{10} = 28.2$

$x - \bar{x} = x - 5.5$

$y - \bar{y} = y - 28.2$

x	y	$X = x - 5.5$	$Y = y - 28.2$	XY	X^2
1	10	-4.5	-18.2	81.9	20.25
2	12	-3.5	-16.2	56.7	12.25
3	16	-2.5	-12.2	30.5	6.25
4	28	-1.5	-0.2	0.3	2.25
5	25	0.5	-3.2	1.6	0.25
6	36	1.5	7.8	3.9	0.25
7	41	2.5	12.8	19.2	2.25
8	49	3.5	20.8	52	6.25
9	40	4.5	11.8	41.3	12.25
10	50		21.8	98.1	20.25
				$\Sigma xy = 385.5$	$\Sigma x^2 = 82.5$

y^2

331.24
262.44
148.84
0.04
10.24
60.84
163.84
432.64
139.24
475.24

$\Sigma y^2 = 2024.6$

(1) $r = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}}$

$\therefore = \frac{385.5}{\sqrt{82.5}\sqrt{2024.6}}$

$= 0.94$

Regression eq y on x

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

$(y - \bar{y}) = \frac{385.5}{82.5}(x - \bar{x})$

$(y - 28.2) = 34.67(x - 5.5)$

$$y = (4.67)x - 25.68 + 28.2$$

$$y = (4.67)x + 53.88$$

Regression line x on y.

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

$$(x - \bar{x}) = \frac{385.5}{2024.6}(y - \bar{y})$$

$$(x - 5.5) = \frac{385.5}{2024.6}(y - 28.2)$$

$$(x - 5.5) = 0.190\,(y - 28.2)$$

$$x - 5.5 = (0.190)y - 5.35$$

$$x = (0.190)y - 5.35$$

$$x = (0.190)y - 5.35 + 5.5$$

$$x = 0.190y + 0.15$$

⑤

x	1	2	3	4	5	6	7
y	4	6	9	10	12	14	15

$$\bar{x} = \frac{20}{7} = 4$$

$$\bar{y} = \frac{70}{7} = 10$$

$$x - \bar{x} = x - 4$$

$$y - \bar{y} = y - 10$$

x	y	X = x−4	Y = y−10	XY	X^2	y^2
1	4	−3	0	0	9	0
2	6	−2	2	−4	4	4
3	9	−1	5	−5	1	25
4	10	0	6	0	0	36
5	12	1	8	8	1	64
6	14	2	10	20	4	100
7	15	3	11	33	9	121
				$\Sigma xy = 52$	$\Sigma x^2 = 28$	$\Sigma y^2 = 350$

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}} \quad (1)$$

$$\therefore (1) \rightarrow \frac{52}{\sqrt{28}\sqrt{350}}$$

$$\boxed{r = 0.52}$$

Regression lies. y on x

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

$$y - \bar{y} = \frac{52}{28}(x - \bar{x})$$

$$y - 10 = \frac{52}{28}(x - 4)$$

$$y - 10 = 1.85(x - 4)$$

$$y - 10 = (1.85)x - 7.4$$

Regression lies x on y.

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

$$(x - \bar{x}) = \frac{52}{350}(y - \bar{y})$$

$$(x - 4) = 0.14(y - 10)$$

$$x - 4 = (0.14)y - 1.4$$

$$x = (0.14)y - 1.4$$

$$x = (0.14)y - 1.4 + 4$$

$$x = (0.14)y + 2.6$$

(6)

x	5	7	8	10	11	13	16
y	~~30~~ 33	30	28	20	18	16	9

$$\bar{x} = \frac{70}{7} = 10$$

$$\bar{y} = \frac{154}{7} = 22$$

$$X = x - \bar{x} = x - 10$$

$$Y = y - \bar{y} = y - 22$$

x	y	$x = x-10$	$y = y-22$	xy	x^2	y^2
5	33	-5	11	-55	25	121
7	30	-3	8	-12	9	64
8	28	-2	6	0	4	36
10	20	0	-2	-4	0	4
11	18	1	-4	-18	1	16
13	16	3	-6	-78	9	36
16	9	6	-13		36	169
				$\Sigma xy = -191$	$\Sigma x^2 = 84$	$\Sigma y^2 = 446$

co-efficient of corelation is given by

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}} \quad -(1)$$

$$\therefore (1) \rightarrow \frac{-191}{\sqrt{84}\sqrt{446}}$$

$$\boxed{r = -0.987}$$

The regression lies you x

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

$$y - \bar{y} = \frac{181}{84}(x - \bar{x})$$

$$y - 22 = \frac{-191}{84}(x - 10)$$

$y - 22 = -2.027 (x - 10)$

$y - 22 = (-2.027) \cdot x + 22.7$

$y = -2.27 x + 22.7 + 22$

$y = (-2.027) x + 44.7$

Regression lies x on y

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

$$x - \bar{x} = \frac{-191}{446} (y - \bar{y})$$

$$x - 10 = \frac{-191}{446} (y - 22)$$

$x - 10 = -0.42 (y - 22)$

$x - 10 = (-0.42) x + 9.24$

$x = (-0.42) x + 9.24 + 10$

$x = (-0.42) x + 19.24$

Imp

The following calculation of prices of commodities have been made in bombay (y) and calcutta (x). coefficient of correlation b/w the prices of commodities in the 2 cities is 0.8 find the most price in bombay correspon-ding to the price of rupees 70 at calcutta

	x	y
mean	65	67
SD	2.5	3.5

$\Rightarrow$ Given data

$\bar{x} = 65 \qquad \bar{y} = 67$

$\sigma_x = 2.5 \qquad \sigma_y = 3.5$

$\boxed{r = 0.8}$

we have to find y when $x = 70$

Regression line of y on x is given by.

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

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

when $x = 70$ $\quad y - 67 = (1.12)(70 - 65)$

$$y - 67 = 5.6.$$

$$y = 5.6 + 67$$

$$y = 72.6 \approx 73$$

$\therefore$ The most price in bombay is rupees 73 corresponding to the price of ₹ 70 in calcutta.

(*) The following result were obtained from line ups in discrete mathametics (x) and data structure (y) in an examination and data structures (y) in an examination of lines of regression and hence find the value of y when $x = 30$

	x	y
mean	47.5	39.5
SD	16.8	10.8

=) Given data

$\bar{x} = 47.5$ $\quad \bar{y} = 39.5$

$\hat{x} = 16.8$ $\quad \hat{y} = 18.8$

The regression of y on x is given by

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

$$y - 39.5 = \frac{(0.95)(10.8)}{16.8} (x - 47.5$$

$$y - 39.5 = (0.61)(x - 47.5)$$

$$y = (0.61)x - 28.97 + 39.5$$

$$y = (0.61)x + 10.53$$

The regression of x on y given by

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

$$x - 47.5 = \frac{(0.95)(16.8)}{10.8}(y - 39.5)$$

$$= (1.47)(y - 39.5)$$

$$x = (1.47)y - 58.06 + 49.5$$

$$x = (1.4)y - 10.56$$

The regression lines y on x

$y = ?$, $x = 30$

$$y = (0.61)(30) + 10.53$$

$$y = 28.83 \approx 29$$

⑧ If $8x - 10y + 66 = 0$ and $40x - 18y = 214$ are the regression lines find the mean of x's and y's and the correlation coefficient also find $\bar{y}$ if $\bar{x} = 3$.

⇒ The given line passes through point $\bar{x}, \bar{y}$

we have $8\bar{x} - 10\bar{y} = +66$

$$40\bar{x} = 18\bar{y} = 214$$

on solving we get

$\bar{x} = 13$ $\qquad \bar{y} = 17$

consider $\quad 8x - 10y + 66 = 0$

$$10y = 8x + 66$$

$$y = \left(\frac{18}{10}\right)x + \frac{66}{10}$$

$y = 0.8x + 6.6$

is the regression line of y on x

consider $\quad 40x - 18y = 214$

$$40x = 18y + 214$$

$$x = \left(\frac{18}{40}\right)y + \left(\frac{214}{40}\right)$$

$x = 0.4y + 5.35$ is the regression

line of x on y

we have $r = \pm\sqrt{(\text{coeff of } x)\ (\text{coeff of } y)}$

$$= \pm\sqrt{(0.8)(0.45)}$$

$$\boxed{r = 0.6}$$

we have the regression lines of y on x

$$y - \bar{y} = \frac{r\,\tilde{y}}{\tilde{x}}(x - \bar{x})$$

and

$y = (0.8)x + 6.6$

on comparing above equations we get

$$r \cdot \frac{\tilde{y}}{\tilde{x}} = 0.8$$

$$(0.6)\left(\frac{\tilde{y}}{3}\right) = 0.8$$

$$\tilde{y} = \frac{0.8 \times 3}{0.6}$$

$$\boxed{\tilde{y} = 4}$$

* Given the lines of regression $x = 19.13 - 0.87y$ and $y = 11.64 - 0.5x$. compute the mean of x's and y's and the correlation coefficient also find $\tilde{y}$ if $\tilde{x} = 3$

$x + 0.87y = 19.3$ and $0.5x + y = 11.64$

The given lines passes through the point $\bar{x}, \bar{y}$

$\bar{x} + 0.87\bar{y} = 19.3$ $0.5\bar{x} + \bar{y} = 11.64$

on solving we get

$\bar{x} = 15.93$ $\bar{y} = 3.67$

from eqn (1) the regression line of x on y is given by

$x = (-0.87)y + 19.3$

$y = (-0.5)x + 11.64.$

The regression line of y on x is given by

$r = \pm\sqrt{(\text{coeff of } x)(\text{coeff of } y)}$

$= \sqrt{(0.5)(0.87)}$

$\boxed{r = -0.65}$

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

$y = (-0.5)x + 11.64$

equating coefficient of x the above eqn we get

$r\frac{\sigma_y}{\sigma_x} = -0.5$

$(-0.65)\left(\frac{\sigma_y}{3}\right) = -0.5$

$\sigma_y = \frac{-0.5 \times 3}{-0.65}$

$\sigma_y = 2.30$

* The given lines of regression are $-8x+5y+17=0$ and $2y-5x+14=0$ and $\bar{y}^2=16$
find
(1) means of x and y
(2) coefficient of correlation
(3) ex of σ_x^2

⇒ The given lines passing through point

$\bar{x}$ $\bar{y}$

$-8x+5y=-17$ $\quad$ $2y-5x=-14$

$-8\bar{x}+5\bar{y}=-17$ $\quad$ $2\bar{y}-5\bar{x}=-14$

on solving we get

$\bar{x}=4$ $\quad$ $\bar{y}=3$

$-8x+5y=-17$ — (1)

from eqn (1)

The regression line of y on x $5y=8x-17$

$$y=\left(\frac{8}{5}\right)x-\frac{17}{5}$$

$2y-5x=-14$ — (2)

The regression lines of x on y

$$5x=2y+14$$

$$x=\left(\frac{2}{5}\right)y+\left(\frac{14}{5}\right)$$

$$r=\pm\sqrt{(\text{coeff of }x)(\text{coeff of }y)}$$

$$=\sqrt{(8/5)(2/5)}$$

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

$$r\frac{\sigma_y}{\sigma_x}=8/5$$

$$(0.8)\left(\frac{(4)}{\sigma_x}\right) = 8/5$$

$$\sigma_y^2 = 16$$

$$\sigma_y = \sqrt{16} = 4$$

$$\sigma_x = \frac{5 \times 4 \times 0.8}{8}$$

$$\sigma_x = 2$$

$$\sigma_x^2 = 4$$

---

	x	y	
Given mean	18	100	and $r = 0.8$
S.D	14	20	

write down the equation of lines of regression and hence find most prob value y when $x = 70$

$\Rightarrow$ $\bar{x} = 18$ $\quad \bar{y} = 100$

$\sigma_x = 14$ $\quad \sigma_y = 20$

$r = 0.8$

we have to find y when $x = 70$

Regression line y on x is given by

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

$$y - 100 = (0.8)\frac{20}{14}(x - 18)$$

$$y - 100 = (1.14)(x - 18)$$

$$y - 100 = 1.14x - 20.52$$

$$y = 1.14x - 20.52 + 100$$

$$y = 1.14x + 79.48$$

The regression line x on y

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

$$x - 18 = (0.8)\frac{18}{20}(y - 100)$$

$x - 18 = 0.56(y - 100)$

$x - 18 = 0.56y - 56$

$x = 0.56y - 56 + 18$

$x = 0.56y - 56 + 18$

$x = 0.56y - 38.$

$x = 70$

$y = (1.11)(70) + 79.48$

$y = 159.28.$

---

* A partially destroyed laboratary of an analysis of carrilation data the following result only available $\bar{x}^2 = 9$

$8x - 10y + 66 = 0$ $\qquad$ $40x - 18y = 214$

1) mean of x and y
2) SD of by
3) coefficient of correlation

The given lines passing thraung teu point $\bar{x}, \bar{y}$

$8\bar{x} - 10\bar{y} = -66$ $\qquad$ $40\bar{x} - 18\bar{y} = 214$

on solving we get

$\bar{x} = 13$ $\qquad$ $\bar{y} = 17$

from ean by (1)

$8x - 10y = -66$

The regression lines y on x

$-10y = -8x - 66$

$y = \left(\frac{+8}{+10}\right)x + \left(\frac{66}{10}\right)$

from ean by (2)

$40x - 18y = 214$

$18y = 40x - 214$

$$x = \left(\frac{18}{40}\right) \cdot y + \frac{214}{40}$$

$$\boxed{r = 0.6}$$

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

$$r \cdot \frac{\sigma_x}{\sigma_y} = \frac{8}{10}$$

$$(0.6) \frac{3}{\sigma_y} = \frac{8}{10}$$

$$\sigma_y = \frac{10 \times 3 \times 0.8}{8}$$

$$\sigma_y = 4$$

$$\sigma_y^2 = 16$$

Obtain the lines of regression and hence find the coefficient of correlation for the following data.

x	1	3	4	2	5	8	9	10	13	15
y	8	6	10	8	12	16	16	10	32	32

$$x = \frac{70}{10} = 7$$

$$\bar{y} = \frac{150}{10} = 15$$

$X = x - \bar{x} \cdot \quad X = x - 7$

$Y = y - \bar{y} \quad Y = y - 15$

x	y	$X = x-7$	$Y = y-15$	XY	X^2	Y^2
1	8	-6	-7	42	36	49
3	6	-4	-9	36	16	81
4	10	-3	-5	15	9	25
2	8	-5	-7	35	25	49
5	12	-2	-3	6	4	9
8	16	1	1	1	1	1
9	16	2	1	2	4	1
10	10	3	-5	-15	9	25
13	32	6	17	102	36	289
15	32	8	17	136	64	289
		0	0	ΣXY 360	ΣX^2 204	$\Sigma Y^2 =$ 818

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}}$$

$$r = \frac{360}{\sqrt{204}\sqrt{818}} = 0.88$$

Regression y on x

$$y = \frac{\Sigma xy}{\Sigma x^2} X$$

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

$$y - 15 = \frac{360}{204}(x - 7)$$

$$y = 1.76x + 2.68$$

Regression x on y

$$X = \frac{\Sigma xy}{\Sigma y^2} Y$$

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

$$(x - 7) = \frac{360}{818}(y - 15)$$

$$x = 0.44y - 0.4$$ //