SOLVED QUESTION BANK

[Sequence given as per syllabus]

INTRODUCTION

Statistics is the branch of mathematics which deals with the study of collection, organization, analysis, interpretation and presentation of data. Statistical method is the method of obtaining useful information from the statistics. This is useful in research when communicating the results of experiments. Statistics is closely related to probability theory with which it is often grouped.

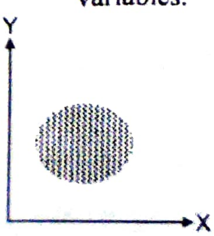
CURVE FITTING

Very often in practice, a relationship is found to exist between two (or more) variables and one wishes to express this relationship in mathematical form by determining an equation connecting the variables.

(1) The first step is the collection of data showing corresponding values of the variables. For example, suppose x and y denotes respectively the height and weight of an adult male. Then, a sample of n individuals would reveal the height $x_1, x_2, \ldots\ldots, x_n$ and the corresponding weights $y_1, y_2, \ldots\ldots, y_n$.

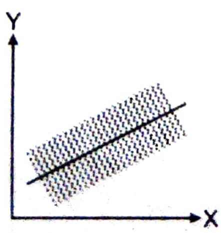
(2) The next step is to plot the points (x_1, y_1), $(x_2, y_2), \ldots\ldots$, (x_n, y_n) on a rectangular co-ordinate system. The resulting set of points is sometimes called **a scatter diagram or dot diagram** as shown in figure (a).

(3) From the **scatter diagram**, it is often possible to visualize a smooth curve approximating the data. Such a curve is called an approximating curve. In figure for example, the data appear to be approximated well by a straight line and we say that a linear relationship (fig. (b)) exists between the variables. However, although a relationship exists between the variables which is not a linear relationship and so it is a non linear relationship (fig.(c)). In figure, there appears to be no relationship between the variables.



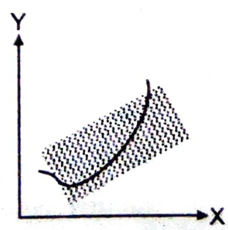
Scatter Diagram

(a)



Linear Relationship

(b)



Nonlinear Relationship

(c)

(4) The general problem of finding the equations of approximating curves that fits given sets of data is called **curve fitting.**

BEST FITTING CURVE :

(1) Let the given data be presented by a set of n points (x_i, y_i), $i = 1, 2, 3, \ldots.$ n. Let y = f(x) be an approximate curve which fits the given set of data.

(2) Let $Y_i = f(x_i)$, then Y_i is called the expected value of y corresponding to $x = x_i$. The value y_i is called the observed value of y corresponding to $x = x_i$.

(3) In general $Y_i \neq y_i$, as the point $P_i(x_i, y_i)$ does not necessarily lie on the curve y = f(x). $E_i = y_i - Y_i$ is called the error of estimate or the residual of y_i. Clearly some of the errors $E_1, E_2, \ldots\ldots E_n$ will be positive and others negative. Thus to give equal weightage to each others, we square each of these and form their sum.

(4) Of all curves approximating a given set of points, the curve for which $E = E_1^2 + E_2^2 + \ldots + E_n^2 = \sum_{i=1}^{n} E_i^2$ is minimum, is called the **best fitting** curve or the **least-square curve.** This is known as principle of least squares.

METHOD OF LEAST SQUARES

Let $(x_i, y_i), i = 1, 2, \ldots.. n$ be a set of n points and

$$y = a_0 + a_1 x + a_2 x^2, \qquad \ldots\ldots(1)$$

be the parabola of best fit to the set of n given points.

Expected value of y corresponding to $x = x_i$ is given by

$$Y_i = a_0 + a_1 x_i + a_2 x_i^2$$

Error of estimate for y_i is given by

$$E_i = y_i - Y_i = y_i - a_0 - a_1 x_i - a_2 x_i^2$$

Since (1) is the polynomial of best fit,

$$\therefore\ E = E_1^2 + E_2^2 + \ldots. + E_n^2 = \sum_{i=1}^{n} E_i^2 \text{ is minimum}$$

$$= \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i - a_2 x_i^2)^2 \text{ is minimum}$$

E is minimum, if $\frac{\partial E}{\partial a_0} = \frac{\partial E}{\partial a_1} = \frac{\partial E}{\partial a_2} = 0$

Now,

$$\frac{\partial E}{\partial a_0} = \Sigma 2(y_i - a_0 - a_1 x_i - a_2 x_i^2)(-1) = 0$$

$$\frac{\partial E}{\partial a_1} = \Sigma 2(y_i - a_0 - a_i x_i - a_2 x_i^2)(-x_i) = 0$$

$$\frac{\partial E}{\partial a_2} = \Sigma 2(y_i - a_0 - a_i x_i - a_2 x_i^2)(-x_i^2) = 0$$

which gives $\Sigma y_i = na_0 + a_1 \Sigma x_i + a_2 \Sigma x_i^2$

$\Sigma x_i y_i = a_0 \Sigma x_i + a_1 \Sigma x_i^2 + a_2 \Sigma x_i^3$

$\Sigma x_i^2 y_i = a_0 \Sigma x_i^2 + a_1 \Sigma x_i^3 + a_2 \Sigma x_i^4$

The above equations are called **normal equations** and can be solved simultaneously in a_0, a_1 and a_2. The values of these constants when substituted in (1) give the desired curve of best fit.

(1) Fitting of a Straight Line :

When the line to be fitted is $y = a_0 + a_1 x$,

the normal equations are $\Sigma y_i = na_0 + a_1 \Sigma x_i$

$\Sigma x_i y_i = a_0 \Sigma x_i + a_1 \Sigma x_i^2$

(2) Fitting of power curve $y = ax^b$

Taking logarithm on both sides, we get

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

$\Rightarrow Y = A + bX$ where

$Y = \log y$, $A = \log a$ and $X = \log x$

Normal equations for estimating A and b are :

$\Sigma Y = nA + b\Sigma X$

$\Sigma XY = A\Sigma X + b\Sigma X^2$

(3) Fitting of Exponential curves :

(i) $y = ab^x$

(ii) $y = ae^{bx}$

(i) Fitting of curve $y = ab^x$

Taking log on both sides, we get

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

$\Rightarrow Y = A + Bx$

where $Y = \log y$, $A = \log a$ and $B = \log b$.

Normal equations for estimating A & B are :

$\Sigma Y = nA + B\Sigma x$

$\Sigma xY = A\Sigma x + B\Sigma x^2$

(ii) Fitting of curve $y = ae^{bx}$

Taking log on both sides.

$\log y = \log a + bx \log e$

$\Rightarrow \log y = \log a + (b \log e)x$

$\Rightarrow Y = A + Bx$

where $Y = \log y$, $A = \log a$ and $B = b \log e$

Normal equations for estimating A & B are:

$\Sigma Y = nA + B\Sigma x$

$\Sigma xY = A\Sigma x + B\Sigma x^2$.

IMPORTANT FORMULAE :

(1) The normal equations to fit the straight line $y = a + bx$ are $\sum y = na + b\sum x$ and

$\sum xy = a\sum x + b\sum x^2$.

(2) The normal equations to fit the parabola $y = a + bx + cx^2$ are

$\sum y = na + b\sum x + c\sum x^2$,

$\sum xy = a\sum x + b\sum x^2 + c\sum x^3$ and

$\sum x^2 y = a\sum x^2 + b\sum x^3 + c\sum x^4$.

(3) The normal equations to fit the parabola $y = ax + bx^2$ are $\sum xy = a\sum x^2 + b\sum x^3$ and

$\sum x^2 y = a\sum x^3 + b\sum x^4$.

(4) The normal equations to fit the curve $y = ax^2 + \frac{b}{x}$ are

$\sum x^2 y = a\sum x^4 + b\sum x$ and

$\sum \frac{y}{x} = a\sum x + b\sum \frac{1}{x^2}$.

(5) To fit the curve $y = ax^b$, first take logarithm on both the sides which gives $\log y = \log a + b \log x$, put $Y = \log y$, $A = \log a$ and $X = \log x$, then we get a straight line equation $Y = A + bX$ and normal equations are $\sum Y = nA + b\sum X$ and

$\sum XY = A\sum X + b\sum X^2$.

(6) To fit the curve $y = ab^x$, take logarithm on both sides, $\log Y = \log a + x \log b$. Putting $Y = \log y$, $A = \log a$ and $B = \log b$, we get $Y = A + Bx$ and the normal equations are $\sum Y = nA + B\sum x$ and

$\sum xY = A\sum x + B\sum x^2$.

(7) To fit the curve $y = ae^{bx}$, take logarithm on both sides, we get $\log y = \log a + bx \log e$, put $Y = \log y$, $A = \log a$ and $B = b \log e$, then the equation becomes $Y = A + Bx$. Normal equations are :

$$\sum Y = nA + B\sum x \text{ and}$$

$$\sum xY = A\sum x + B\sum x^2 \text{ where } \log e = 0.4343.$$

SOLVED EXAMPLES :

Ex.1. Fit a straight line to the following data:

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

Soln. Let the straight line of fit be

$y = a + bx$ .....(1)

The normal equations are:

$\Sigma y = na + b\Sigma x$ .....(2)

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

Here, n = 5

x	y	xy	x^2
0	1	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
$\Sigma x = 10$	$\Sigma y = 16.9$	$\Sigma xy = 47.1$	$\Sigma x^2 = 30$

$\therefore$ (2) $\Rightarrow$ $5a + 10b = 16.9$ ......(4)

$\therefore$ (3) $\Rightarrow$ $10a + 30b = 47.1$ ......(5)

On solving (4) and (5), we get

$a = 0.72$

and $b = 1.33$

$\therefore$ The required curve is

$\boxed{y = 0.72 + 1.33x}$ **Ans.**

Ex.2. Fit a straight line $y = a + bx$ to the following data by the method of least squares:

x	0	1	3	6	8
y	1	3	2	5	4

Soln. Consider, $S = \Sigma(y_i - a - bx_i)^2$, then normal equations are given by $\frac{\partial S}{\partial a} = 0$ and $\frac{\partial S}{\partial b} = 0$ which gives

$\Sigma y_i = na + b\Sigma x_i$ .....(1)

$\Sigma x_i y_i = a\Sigma x_i + b\Sigma x_i^2$ ....(2)

Here, n = 5

x	y	xy	x^2
0	1	0	0
1	3	3	1
3	2	6	9
6	5	30	36
8	4	32	64
18	15	71	110

From eqn. (1) $\Rightarrow$ $15 = 5a + 18b$

From eqn. (2) $\Rightarrow$ $71 = 18a + 110b$

$$\therefore a = \frac{\begin{vmatrix} 15 & 18 \\ 71 & 110 \end{vmatrix}}{\begin{vmatrix} 5 & 18 \\ 18 & 110 \end{vmatrix}} = \frac{1650 - 1278}{550 - 324} = \frac{372}{226} = 1.65$$

$$\text{and } b = \frac{\begin{vmatrix} 5 & 15 \\ 18 & 71 \end{vmatrix}}{\begin{vmatrix} 5 & 18 \\ 18 & 110 \end{vmatrix}} = \frac{355 - 270}{550 - 324} = \frac{85}{226} = 0.38$$

$\therefore$ The required straight line is

$\boxed{y = 1.65 + 0.38x}$ **Ans.**

Ex.3. Find the straight line that best fit the following data :

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

Soln. Let

$y = a + bx$ ......(1)

The normal equations are:

$\Sigma y = na + b\Sigma x$ ......(2)

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

Here, n = 5

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
$\Sigma x = 15$	$\Sigma y = 204$	$\Sigma xy = 748$	$\Sigma x^2 = 55$

$\therefore$ Eqn. (2) $\Rightarrow 5a + 15b = 204$ .....(4)

$\therefore$ From eqn. (3) $\Rightarrow 15a + 55b = 748$ .....(5)

On solving eq.(4) & (5), we get

$a = 0$ and $b = 13.6$

$\therefore$ The required straight line is

$\boxed{y = 13.6x}$ **Ans.** *[$\because$ From eqn. (1)]*

Ex.4. Fit a straight line y = a + bx to the following data:

x	1	2	3	4	6	8
y	2.4	3	3.6	4	5	6

S-03

Soln. We have $y = a + bx$ .....(1)

The normal equations are

$\Sigma y = na + b\Sigma x$ .....(2)

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

Here, $n = 6$

x	y	xy	x^2
1	2.4	2.4	1
2	3	6	4
3	3.6	10.8	9
4	4	16	16
6	5	30	36
8	6	48	64
$\Sigma x = 24$	$\Sigma y = 24$	$\Sigma xy = 113.2$	$\Sigma x^2 = 130$

$\therefore (2) \Rightarrow 6a + 24b = 24$

or $a + 4b = 4$ ......(4)

and $(3) \Rightarrow 24a + 130b = 113.2$ ....(5)

Solving (4) and (5), we get

$a = 1.97648$ and $b = 0.50589$

$\therefore$ $\boxed{y = 1.97648 + 0.50589x}$ **Ans.**

Ex.5. Find the relation of type R = av + b from the following table:

V	60	65	70	75	80	85	90
R	109	114	118	123	127	130	133

Soln. Here, the figures are big in size. So we will use a short cut method by converting the independent variable v by V.

where, $V = \dfrac{v - \text{assumed mean}}{\text{interval}}$ where assumed mean is the middle value. If there are two middle values then take any one of them as the assumed means. If the interval is not unique then use v- assumed mean. Here, the interval of v is unique i.e. 5 and the middle value is 75.

So, we take $V = \dfrac{v - 75}{5}$

So, we consider $R = a'V + b'$

Now, let $S = \Sigma(R_i - a'V_i - b')^2$ be the sum of squares of deviations. By the method of least square 'S' should be minimum. But from the conditions of minima, we have

$$\frac{\partial S}{\partial a'} = 0 \text{ and } \frac{\partial S}{\partial b'} = 0$$

$$\frac{\partial S}{\partial a'} = 0 \Rightarrow \Sigma 2(R_i - a'V_i - b')(-V_i) = 0$$

$$\Rightarrow \Sigma(R_i - a'V_i - b')V_i = 0$$

$$\Rightarrow \Sigma R_i V_i = a'\Sigma V_i^2 + b'\Sigma V_i \quad ....(1)$$

$$\frac{\partial S}{\partial b'} = 0 \Rightarrow \Sigma 2(R_i - a'V_i - b')(-1) = 0 \quad ....(2)$$

$$\Rightarrow \Sigma R_i = a'\Sigma V_i + nb'$$

So we construct a table of the following form :

V	R	$V = \dfrac{v-75}{5}$	V^2	RV
60	109	−3	9	−327
65	114	−2	4	−228
70	118	−1	1	−118
75	123	0	0	0
80	127	1	1	127
85	130	2	4	260
90	133	3	9	399
	854	0	28	113

From (1), $113 = 28a' + b'.0$

$\therefore a' = \dfrac{113}{28} = 4.0357$

From (2), $854 = a'.0 + 7b'$

$\therefore b' = \dfrac{854}{7} = 122$

$\therefore R = 4.0357V + 122$

Now, replace V by $\dfrac{v - 75}{5}$

$$\therefore R = 4.0357\left(\frac{v - 75}{5}\right) + 122$$

$$= 0.80714v + 61.4645.$$

$\therefore$ $\boxed{R = 0.80714v + 61.4645}$ **Ans.**

Ex.6. Fit a curve $y = a + bx^2$ for the following data :

x	−1	0	1	2
y	2	5	3	0

Soln. Given parabola is

$y = a + bx^2$ ......(1)

$\therefore$ Normal equations are

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

$\Sigma x^2 y = a\,\Sigma x^2 + b\Sigma x^4$ ......(3)

Here, n = 4

x	y	x^2	x^4	x^2y
−1	2	1	1	2
0	5	0	0	0
1	3	1	1	3
2	0	4	16	0
Total	10	6	18	5

$\therefore$ (2) $\Rightarrow$ $4a + 6b = 10$ and ......(4)

(3) $\Rightarrow$ $6a + 18b = 5$ ......(5)

On solving (4) and (5), we get

$a = 4.17$

and $b = -1.12$

$\therefore$ The required curve is

$\boxed{y = 4.17 - 1.12\,x^2}$ **Ans.** *[$\because$ By equation (1)]*

Ex.7. Fit a curve $y = ax^2 + b$ for the following data:-

x	12	16	20	22	24	26	30
y	6.44	7.5	6.9	10.76	10.76	11.76	14.0

S-04,10

Soln. Given parabola is

$y = ax^2 + b$

$\therefore$ Normal equations are

$\sum y = nb + a\sum x^2$

$\sum x^2 y = b\sum x^2 + a\sum x^4$

Here, n = 7

The various calculations are shown in the table :

x	y	x^2	x^4	x^2y
12	6.44	144	20736	927.36
16	7.5	256	65536	1920
20	6.9	400	160000	2760
22	10.76	484	234256	5207.84
24	10.76	576	331776	6197.76
26	11.76	676	456976	7949.76
30	14.0	900	810000	12600
$\sum$	68.12	3436	2079280	37562.72

$\therefore$ We have,

$68.12 = 7b + 3436a$ ......(1)

$37562.72 = 3436b + 2079280a$ .....(2)

On solving (1) and (2), we get

$a = 0.0105$ and $b = 4.5746$

$\therefore$ $\boxed{\text{The required parabola is } y = 0.0105\,x^2 + 4.5746}$

Ans.

Ex.8. Fit a curve $y = ax + bx^2$ for the following data:-

x	1	2	3	4	5	6
y	2.51	5.82	9.93	14.84	20.55	27.06

S-09, W-06

Soln. We have to fit the curve

$y = ax + bx^2$ .....(1)

The normal equations are

$\Sigma xy = a\Sigma x^2 + b\Sigma x^3$ and

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

The various calculations are shown in the table :

x	y	x^2	x^3	x^4	xy	x^2y
1	2.51	1	1	1	2.51	2.51
2	5.82	4	8	16	11.64	23.28
3	9.93	9	27	81	29.79	89.37
4	14.84	16	64	256	59.36	237.44
5	20.55	25	125	625	102.75	513.75
6	27.06	36	216	1296	162.36	974.16
		91	441	2275	368.41	1840.51

$\therefore$ From (2), we get

$368.41 = 91a + 441b$ ......(3)

and $1840.51 = 441a + 2275b$ ......(4)

Solving eqn. (3) and eqn. (4) for a and b, we get

$a = 2.11$ and $b = 0.40$

$\therefore$ From (1),

$\therefore$ $\boxed{y = 2.11x + 0.4x^2}$ **Ans.**

Ex.9. Fit a parabola $y = a + bx^2$ for the following data by least square method:

x	1	2	3	4	5
y	1.8	5.1	8.9	14.1	19.8

S-07, W-11

Soln. We have to fit the parabola

$y = a + bx^2$ .....(1)

Here, n = 5

The normal equations are

$\Sigma y = na + b\Sigma x^2$ and $\Sigma x^2 y = a\Sigma x^2 + b\Sigma x^4$ .....(2)

x	y	x^2	x^4	x^2y
1	1.8	1	1	1.8
2	5.1	4	16	20.4
3	8.9	9	81	80.1
4	14.1	16	256	225.6
5	19.8	25	625	495
Σ	49.7	55	979	822.9

∴ From eq. (2), we get

$49.7 = 5a + 55b$ ......(3)

and $822.9 = 55a + 979b$ ......(4)

Solving eq. (3) and eq. (4) for a, b, we get

$a = 1.8165$ and $b = 0.7386$

From eq. (1), we get

$\boxed{y = 1.8165 + 0.7386\,x^2}$ **Ans.**

Ex.10. Fit a curve $y = ax + bx^2$ to the following data :

x	–2	–1	0	1	2
y	–72	–46	–12	35	93

Soln. Given parabola is $y = ax + bx^2$. ......(1)

The normal equations are

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

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

Here, n = 5

x	y	xy	x^2	x^2y	x^3	x^4
–2	–72	144	4	–288	–8	16
–1	–46	46	1	–46	–1	1
0	–12	0	0	0	0	0
1	35	35	1	35	1	1
2	93	186	4	372	8	16
	Total	411	10	73	0	34

∴ (2) ⇒ $10a + 0b = 411$

∴ $a = 41.1$

(3) ⇒ $0.a + 34b = 73$

∴ $b = 2.1471$

∴ The required curve is

$\boxed{y = 41.1x + 2.1471\,x^2}$ **Ans.** *[∵ By (1)]*

Ex.11. Fit a parabola $y = a + bx^2$ for the following data by least square method :

x	0	1	2	3
y	2	4	10	15

W-09

Soln. Given parabola is $y = a + bx^2$. ......(1)

∴ Normal equations are

$\Sigma y = na + b\,\Sigma x^2$ .....(2)

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

Here, n = 4

The various calculation are as shown in table :

n	x	y	x^2	x^4	x^2y
1	0	2	0	0	0
2	1	4	1	1	4
3	2	10	4	16	40
4	3	15	9	81	135
	Total	31	14	98	179

∴ (2) ⇒

$31 = 4a + b(14)$

∴ $4a + 14b = 31$ ......(4)

& (3) ⇒

$179 = 14a + 98b$

∴ $14a + 98b = 179$ ......(5)

On solving, we get

$a = 2.7143$ and $b = 1.4388$

∴ The required curve is

$\boxed{y = 2.7143 + 1.4388\,x^2}$ **Ans.** *[∵ By (1)]*

Ex.12. Fit a parabola of the type $y = a + bx + cx^2$ for

x	20	30	40	50	60
y	54	90	138	206	292

Soln. Consider, $X = \dfrac{x - 40}{10}$

∴ $a + bX + cX^2 = y$

Normal equations are $\sum y = na + b\sum X + c\sum X^2$ ....(1)

$\sum Xy = a\sum X + b\sum X^2 + c\sum X^3$ ....(2)

$\sum X^2y = a\sum X^2 + b\sum X^3 + c\sum X^4$ ....(3)

x	y	$X = \frac{x-40}{10}$	X^2	X^3	X^4	Xy	X^2y
20	54	−2	4	−8	16	−108	216
30	90	−1	1	−1	1	−90	90
40	138	0	0	0	0	0	0
50	206	1	1	1	1	206	206
60	292	2	4	8	16	584	1168
Total	780	0	10	0	34	592	1680

Equations (1), (2) and (3) gives

$780 = 5a + 10c$ .....(4)

$592 = 10b$

$\therefore b = 59.2$

$1680 = 10a + 34c$ .....(5)

On solving (4) and (5), we get

$a = \frac{972}{7}$ and $c = \frac{120}{14}$

$\therefore y = \frac{972}{7} + 59.2X + \frac{120}{14}X^2$

$= \frac{972}{7} + 59.2\left(\frac{x-40}{10}\right) + \frac{120}{14}\left(\frac{x-40}{10}\right)^2$

$\left[\because X = \frac{x-40}{10}\right]$

$= \frac{972}{7} + 5.92x - (5.92)(40) + \frac{120}{14}.$

$\times \frac{1}{100}(x^2 - 80x + 1600)$

$\therefore \boxed{y = 39.2 - 0.94x + 0.085x^2}$ **Ans.**

Ex.13. Fit a second degree parabola to the following data:

x	1	2	3	4	5	6	7	8	9
y	2	6	7	8	10	11	11	10	9

W-02

Soln. Consider, $X = \frac{x-5}{1} = x - 5$ and $n = 9$

$\therefore y = a + bX + cX^2$

Normal equations are

$\sum y = na + b\sum X + c\sum X^2$ .....(1)

$\sum Xy = a\sum X + b\sum X^2 + c\sum X^3$ .....(2)

$\sum X^2y = a\sum X^2 + b\sum X^3 + c\sum X^4$ .....(3)

x	y	$X = x - 5$	X^2	Xy	X^3	X^2y	X^4
1	2	−4	16	−8	−64	32	256
2	6	−3	9	−18	−27	54	81
3	7	−2	4	−14	−8	28	16
4	8	−1	1	−8	−1	8	1
5	10	0	0	0	0	0	0
6	11	1	1	11	1	11	1
7	11	2	4	22	8	44	16
8	10	3	9	30	27	90	81
9	9	4	16	36	64	144	256
Total	74	0	60	51	0	411	708

Equations (1), (2), (3) gives

$74 = 9a + 60c$ .....(4)

$51 = 60b$

$\therefore b = 0.85$

$411 = 60a + 708c$ .....(5)

Solving (4) and (5), we get

$a = 10.0041$ and $c = -0.2673$

$\therefore y = 10.0041 + 0.85X - 0.2673X^2$

$= 10.0041 + 0.85(x-5) - 0.2673(x-5)^2$

$\therefore \boxed{y = -0.9284 + 3.523x - 0.2673x^2}$ **Ans.**

Ex.14. Fit a second degree parabola $y = a + bx + cx^2$ to the data :-

x	20	40	60	80	100	120
y	5.5	9.1	14.9	22.8	33.3	46.0

S-06

Soln. Here, $\bar{x} = 70$, $h = 20$, $n = 6$

$\therefore$ Take $X = \frac{x - \bar{x}}{h}$

$\therefore X = \frac{x-70}{20}$

$y = a + bX + cX^2$ ......(1)

The normal equations are

$$\left.\begin{aligned} \sum y &= na + b\sum X + c\sum X^2 \\ \sum Xy &= a\sum X + b\sum X^2 + c\sum X^3 \\ \sum X^2y &= a\sum X^2 + b\sum X^3 + c\sum X^4 \end{aligned}\right\} \quad ......(2)$$

x	y	X	X^2	X^3	X^4	Xy	X^2y
20	5.5	– 2.5	6.25	– 15.625	39.0625	–13.75	34.375
40	9.1	– 1.5	2.25	– 3.375	5.0625	–13.65	20.475
60	14.9	– 0.5	0.25	– 0.125	0.0625	– 7.45	3.725
80	22.8	0.5	0.25	0.125	0.0625	11.4	5.7
100	33.3	1.5	2.25	3.375	5.0625	49.95	74.925
120	46.0	2.5	6.25	15.625	39.0625	115	287.5
Total	131.6	0	17.5	0	88.375	141.5	426.7

$\therefore$ From (2) $\Rightarrow 131.6 = 6a + 17.5c$ ......(3)

$141.5 = 17.5b \Rightarrow b = 8.085$

$426.7 = 17.5a + 88.375c$ ......(4)

On solving eq. (3) & (4), we get

$a = 18.5843$

$c = 1.1482$

$$\therefore\ y = 18.5843 + 8.085X + 1.1482X^2$$

$$\therefore\ y = 18.5843 + 8.085\left(\frac{x-70}{20}\right) + 1.1482\left(\frac{x-70}{20}\right)^2$$

$$\left[\because\ X = \frac{x-70}{20}\right]$$

$$\therefore\ \boxed{y = 4.35 + 0.00238x + 0.002871x^2}$$ **Ans.**

Ex.15. Fit a curve $y = ax^2 + \frac{b}{x}$ for the following data :

x	1	2	3	4
y	– 1.51	0.99	3.88	7.66

Soln. Given curve is

$$y = ax^2 + \frac{b}{x} \quad .....(1)$$

$\therefore$ Normal Equation are

$$\Sigma x^2 y = a\Sigma x^4 + b\Sigma x \quad ......(2)$$

$$\Sigma \frac{y}{x} = a\Sigma x + b\Sigma \frac{1}{x^2} \quad ......(3)$$

We have the table of the following type :

x	y	x^2	x^2y	x^4	$\frac{1}{x}$	$\frac{1}{x^2}$	$\frac{y}{x}$
1	– 1.51	1	–1.51	1	1	1	– 1.51
2	0.99	4	3.96	16	0.5	0.25	0.495
3	3.88	9	34.92	81	0.3333	0.1111	1.2933
4	7.66	16	122.56	256	0.25	0.0625	1.9150
10			159.93	354		1.4236	2.1933

$\therefore$ From (2) & (3), we have

$159.93 = 354a + 10b$ .....(4)

& $2.1933 = 10a + 1.4236b$ .....(5)

On solving (4) and (5), we get

$\therefore\ a = 0.509$ and $b = -2.037$

$$\therefore\ \boxed{y = 0.509x^2 - \frac{2.037}{x}}$$ is the required curve. **Ans.**

Ex.16. Fit the curve $y = ax + \frac{b}{x}$ to the following data:

x	1	2	3	4	5	6	7	8
y	5.43	6.28	8.23	10.32	12.63	14.86	17.27	19.51

W-04

Soln. The equation of the curve is

$$y = ax + \frac{b}{x} \quad .....(1)$$

The normal equations are :

$$\left.\begin{aligned} \Sigma xy &= a\Sigma x^2 + nb \\ \Sigma \frac{y}{x} &= na + b\Sigma \frac{1}{x^2} \end{aligned}\right\} \quad .....(2)$$

Here n = 8.

x	y	x^2	$\frac{1}{x^2}$	xy	$\frac{y}{x}$
1	5.43	1	1	5.43	5.43
2	6.28	4	0.25	12.56	3.14
3	8.23	9	0.1111	24.69	2.7433
4	10.32	16	0.0625	41.28	2.58
5	12.63	25	0.04	63.15	2.526
6	14.86	36	0.0277	89.16	2.4766
7	17.27	49	0.0204	120.89	2.4671
8	19.51	64	0.0156	156.08	2.4387
		204	1.5273	513.24	23.8017

From eqn. (2) $\Rightarrow 513.24 = 204a + 8b$

$23.8017 = 8a + 1.5273b$

On solving, we get

$a = 2.3971$ and $b = 3.0279$

$\therefore$ The required equation is

$$\boxed{y = 2.3971x + \frac{3.0279}{x}}$$ **Ans.**

Ex.17. Fit a least square geometric curve $y = ax^b$ to the following data

x	1	2	3	4	5
y	0.5	2	4.5	8	12.5

S-12

Soln. The curve to be fitted is $y = ax^b$ .....(i)

Taking log on both sides

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

$\Rightarrow y = A + bx$

where Y = log y, A = log a, X = log x

$\therefore$ The normal equations are

$\Sigma Y = nA + b\Sigma X$ .....(2)

$\Sigma XY = A\,\Sigma X + b\Sigma X^2$ .....(3)

Here, n = 5

x	X = log x	y	Y = log y	XY	X^2
1	0	0.5	– 0.3010	0	0
2	0.3010	2	0.3010	0.0906	0.0906
3	0.4771	4.5	0.6532	0.3116	0.2276
4	0.6020	8	0.9030	0.5436	0.3624
5	0.6989	12.5	1.0969	0.7666	0.4884
Total	2.079		2.6531	1.7124	1.169

$\therefore$ (2) $\Rightarrow 5A + 2.079b = 2.6531$ .....(4)

and (3) $\Rightarrow 2.079A + 1.169b = 1.7124$ .....(5)

On solving (4) and (5), we get

$A = -0.3011$ and $b = 2.0004$

But A = log a

$\therefore$ a = Antilog A

$\therefore$ a = Antilog (– 0.3011) = 0.4999

$\therefore$ From equation (1), the required curve is

$\boxed{y = 0.4999\, x^{2.0004}}$ **Ans.**

Ex.18. Fit a curve $y = ax^b$ to the following data :

x	1	2	3	4	5	6
y	2.98	4.26	5.21	6.10	6.80	7.5

W-10

Soln. The curve to be fitted is $y = ax^b$ .....(1)

Taking log on both sides, we get

log y = log a + b log x

$\Rightarrow$ Y = A + bX

where Y = log y, A = log a and X = log x

The normal equations for estimating A and b are

$\Sigma Y = nA + b\Sigma X$ .....(2)

$\Sigma XY = A\Sigma X + b\Sigma X^2$ .....(3)

Here, n = 6

x	$X = \log_{10} x$	y	$Y = \log_{10} y$	XY	X^2
1	0	2.98	0.4742	0	0
2	0.3010	4.26	0.6294	0.1894	0.0906
3	0.4771	5.21	0.7168	0.3420	0.2276
4	0.6021	6.10	0.7853	0.4728	0.3625
5	0.6990	6.80	0.8325	0.5819	0.4886
6	0.7782	7.50	0.8751	0.6810	0.6056
Total	2.8574		4.3133	2.2671	1.7749

$\therefore$ Eqn.(2) $\Rightarrow 6A + 2.8574\,b = 4.3133$ ......(4)

and eqn. (3) $\Rightarrow 2.8574\,A + 1.7749\,b = 2.2671$ ......(5)

On solving eq. (4) and (5), we get

A = 0.4740 and b = 0.5143

But A = log a

$\therefore$ a = Antilog A = Antilog (0.4740) = 2.9785

$\therefore$ From equation (1), the required curve is

$\boxed{y = 2.9785\, x^{0.5143}}$ **Ans.**

Ex.19. For the data given below fit the curve $y = ab^x$.

x	1	2	3	4	5	6
y	1.6	4.5	13.8	40.2	125	300

Soln. Given curve to be fitted is

$y = ab^x$

Taking logarithm on both sides, we get

log y = log a + x log b

$\Rightarrow$ Y = A + Bx

where Y = log y,

A = log a,

B = log b

The normal equations are

$\Sigma Y = nA + B\Sigma x$ ......(1)

$\Sigma xY = A\Sigma x + B\Sigma x^2$ ......(2)

x	y	Y = log y	xY	x^2
1	1.6	0.2041	0.2041	1
2	4.5	0.6532	1.3064	4
3	13.8	1.1398	3.4194	9
4	40.2	1.6042	6.4168	16
5	125	2.0969	10.4845	25
6	300	2.4771	14.8626	36
$\Sigma x = 21$		$\Sigma Y = 8.1753$	$\Sigma xY = 36.6938$	$\Sigma x^2 = 91$

$\therefore$ Eqn. (1) $\Rightarrow 8.1753 = 6A + 21B$ .....(3)

Eqn. (2) $\Rightarrow 36.6938 = 21A + 91B$ .....(4)

On solving (3) and (4), we get

$$\begin{array}{rl} 171.6813 &= 126A + 441B \\ 220.1628 &= 126A + 546B \\ \hline -48.4815 &= -105B \end{array}$$

$\therefore B = 0.4617286 = \log b$

$\therefore b = 2.8955$

$A = -0.2535 = \log a$

$\therefore a = 0.55783$

$\therefore \boxed{y = (0.55783)(2.8955)^x}$ **Ans.**

Ex.20. Fit an equation of the form $y = ab^x$ to the following data:

x	2	3	4	5	6
y	144	172.8	207.4	248.8	298.5

S-01, 08

Soln. Given curve is $y = ab^x$, taking logarithm on both sides, we get

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

$\Rightarrow Y = A + Bx$

where $Y = \log y$, $A = \log a$, $B = \log b$

The normal equations are

$\Sigma Y = nA + B\Sigma x$

$\Sigma xY = A\Sigma x + B\Sigma x^2$

So we have a table of the following form :

x	y	Y = log y	xY	x^2
2	144	2.1584	4.3168	4
3	172.8	2.2375	6.7125	9
4	207.4	2.3168	9.2672	16
5	248.8	2.3959	11.9795	25
6	298.5	2.4750	14.85	36
20		11.5836	47.126	90

$\therefore$ We have, $11.5836 = 5A + 20B$ ......(i)

$47.126 = 20A + 90B$ .....(ii)

(ii) – (i) × 4 $\Rightarrow 0.7916 = 10B$

$\therefore B = 0.07916$

$$A = \frac{11.5836 - 20\,(0.07916)}{5} = \frac{10.0008}{5} = 2 = \log a$$

$\therefore a = \text{Antilog } A = \text{Antilog } 2 = 10^2 = 100$

$b = \text{Antilog } B = \text{Antilog } (0.07916) = 1.199$

$\therefore$ The required curve is

$\boxed{y = 100\,(1.199)^x}$ **Ans.**

Ex.21. Fit a curve of the form $y = ab^x$ to the following data

x	2	3	4	5	6
y	144	172.3	207.4	248.8	298.5

S-11

Soln. Given curve is $y = ab^x$ .....(1)

Taking logarithm on both sides, we get

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

$\Rightarrow Y = A + Bx$

where $Y = \log y$, $A = \log a$, $B = \log b$

So, we have a table of the following form.

x	y	Y = log y	xY	x^2
2	144	2.1584	4.3167	4
3	172.3	2.2363	6.7089	9
4	207.4	2.3168	9.2672	16
5	248.8	2.3959	11.9795	25
6	298.5	2.4749	14.8394	36
$\Sigma x = 20$		$\Sigma Y = 11.5823$	$\Sigma xY = 47.1217$	$\Sigma x^2 = 90$

Normal equations are

$\Sigma Y = nA + B\Sigma x$ .....(2)

$\Sigma xY = A\Sigma x + B\Sigma x^2$ .....(3)

Putting the values of ΣY, n and Σx in (2), we get

$11.5823 = 5A + 20B$ .....(4)

Putting the values of Σx, Y, Σx and Σx^2 in (3), we get

$47.1217 = 20A + 90B$ ....(5)

Multiplying (4) by 4, we get

$46.3292 = 20A + 80B$ ....(6)

Subtracting (6) from (5), we get

$0.7925 = 10B \Rightarrow B = 0.07925$

By putting the value of B in (4), we get

$11.5823 = 5A + 20\,(0.07925)$

$$\Rightarrow 5A = 11.5823 - 1.583 \Rightarrow A = \frac{9.9973}{5} = 1.99946$$

$\Rightarrow A = 2$

Now, $A = \log a \Rightarrow 2 = \log a \Rightarrow a = \text{Antilog } 2$

$= 10^2 = 100$

$B = \log b \Rightarrow 0.07926 = \log b \Rightarrow b = \text{Antilog } 0.07926$

$= 10^{0.07926} = 1.2$

Putting the values of a and b in (1), we get

$\boxed{y = 100(1.2)^x}$ **Ans.**

Ex.22. Fit a curve $y = ae^{bx}$ to the following data :-

x	1	2	3	4	5	6
y	1.6	4.5	13.8	40.2	125	300

W-03, 08

Soln. The curve to be fitted is

$$y = ae^{bx} \quad .....(1)$$

Taking log on both sides, we get

$$\log y = \log a + bx \log_{10} e$$

$$\Rightarrow Y = A + Bx \quad .....(2)$$

where $A = \log a$, $B = b \log_{10} e$, $Y = \log y$

The normal equations for (2) are

$$\Sigma Y = nA + B\Sigma x$$

$$\Sigma xY = A\Sigma x + B\Sigma x^2 \quad .....(3)$$

Here n = 6

∴ The various calculations are shown below :

x	y	Y = log y	xY	x^2
1	1.6	0.2041	0.2041	1
2	4.5	0.6532	1.3064	4
3	13.8	1.1398	3.4194	9
4	40.2	1.6042	6.4168	16
5	125	2.0969	10.4845	25
6	300	2.4771	14.8626	36
$\Sigma x = 21$		$\Sigma Y = 8.1753$	$\Sigma xY = 36.6938$	$\Sigma x^2 = 91$

∴ Equation (3) ⇒

$$8.1753 = 6A + 21B \quad ......(4)$$

$$36.6938 = 21A + 91B \quad ....(5)$$

Solving equation (4) and (5), we get

$A = -0.2535$ and $B = 0.4617$

$B = 0.4617 = b\log_{10} e,$

$$\log a = -0.2535 \Rightarrow a = 10^{-0.2535} = 0.5579$$

$$\Rightarrow b = \frac{0.4617}{\log_{10} e} = \frac{0.4617}{0.4343} = 1.0631$$

∴ The required curve is

$$\boxed{y = (0.5579)\, e^{1.0631x}}$$ **Ans.**

Ex.23. Fit a curve $y = ae^{bx}$ to the following data :

x	1	2	3	4	5	6
y	2.98	4.61	7.93	18.54	51.83	128.92

W-07

Soln. The equation of the curve is

$$y = ae^{bx} \quad .....(1)$$

Taking logarithm on both sides, we get

$$\Rightarrow \log y = \log a + bx \log e$$

$$\Rightarrow \log y = \log a + (b \log e)\, x$$

$$\text{i.e.}\quad Y = A + Bx \quad ......(2)$$

where $Y = \log y$, $A = \log a$ and $B = b \log e$

The normal equations for the estimation of A & B are

$$\left.\begin{aligned}\Sigma Y &= nA + B\Sigma x\\ \Sigma xY &= A\Sigma x + B\Sigma x^2\end{aligned}\right\} \quad .....(3)$$

x	y	Y = log y	xY	x^2
1	2.98	0.4742	0.4742	1
2	4.61	0.6637	1.3274	4
3	7.93	0.8992	2.6976	9
4	18.54	1.2681	5.0724	16
5	51.83	1.7146	8.573	25
6	128.92	2.1103	12.6618	36
21		7.1301	30.8064	91

Here, n = 6

Eqn. (3) ⟹ $6A + 21B = 7.1301$ .....(4)

$21A + 91B = 30.8064$ .....(5)

Eqn. (4) divided by 2 gives

$$3A + \frac{21}{2}B = 3.56505 \quad .....(6)$$

Eqn. (5) divided by 7 gives

$$3A + 13B = 4.400914 \quad .....(7)$$

Eqn. (7) – Eqn. (6) gives

$$\frac{5}{2}B = 0.83586$$

$$\Rightarrow B = \frac{2}{5}(0.83586) = 0.33434$$

$$\therefore\ b = \frac{B}{\log e} = \frac{0.334344}{0.43429} = 0.7698$$

From eqn. (7),

$$A = \frac{4.400914 - 13\,(B)}{3}$$

$$= \frac{4.400914 - 13(0.33434)}{3}$$

$$= 0.018164$$

$$\therefore\quad a = 10^{0.018164} = 1.0427$$

∴ The required curve is

$$\boxed{y = (1.0427)\, e^{0.7698\,x}}$$ **Ans.**

Ex.24. Fit an exponential curve obeying the gas equation $PV^{\gamma} = K$ for the following data :

V	50	100	150	200
P	135	48	26	17

Soln. Given gas equation is $PV^{\gamma} = K$

Taking logarithm, we get

$\log P + \gamma \log V = \log K$

$\Rightarrow x + \gamma y = c$

where $x = \log P$, $y = \log V$, $c = \log K$

$\therefore x = c - \gamma y$

Normal equations are $\sum x = nc - \gamma \sum y$

$\sum xy = c\sum y - \gamma \sum y^2$

V	y = log V	P	x = log P	xy	y^2
50	1.6989	135	2.130	3.6186	2.8862
100	2	48	1.681	3.362	4
150	2.1760	26	1.414	3.0768	4.7349
200	2.3010	17	1.230	2.8302	5.2946
	8.1759		6.455	12.8876	16.9157

From the normal equations, we get

$6.455 = 4c - \gamma\,(8.1759)$ .....(1)

& $12.8876 = 8.1759\,c - 16.9157\,\gamma$ .....(2)

Multiplying eqn.(1) by 8.1759 and eqn.(2) by 4 and subtract eqn. (2) from eqn.(1), we get

$$\begin{array}{rcl} 52.7754 &=& 32.7036\,c - 66.8453\,\gamma \\ 51.5504 &=& 32.7036\,c - 67.6628\,\gamma \\ - & & -\quad\quad + \\ \hline 1.225 &=& 0.8175\gamma \end{array}$$

$\therefore \gamma = 1.4984$

Putting $\gamma = 1.4984$ in eq^n (1), we get

$c = 4.6764$

Now,

$K = \text{Antilog } c = \text{Antilog } (4.6764) = 47467.897$

$\therefore \boxed{PV^{1.4984} = 47467.897}$ **Ans.**

COEFFICIENT OF CORRELATION 'r'

(1) Two variables x and y are said to be correlated, if increase or decrease in one variable is accompanied by increase or decrease in the other variable. Coefficient of correlation measure the degree of relationship between two correlated variables x and y.

(2) **Coefficient of correlation** 'r' is defined as

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \, \Sigma Y^2}}$$

where X and Y are the standard deviation from their respective means

i.e. $X = x - \bar{x}$ and $Y = y - \bar{y}$

and $\bar{x} = \dfrac{\Sigma x}{n}$ and $\bar{y} = \dfrac{\Sigma y}{n}$

n is number of observations.

(3) The value of r always varies from −1 to 1. The sign of r determines the nature of correlation positive where r is positive and negative where r is negative.

LINE OF REGRESSION

(1) If two variables x and y are correlated i.e. there exists an association or relationship between them, then scatter will be more or less concentrated round a curve. This curve is known **as curve of regression**. When the curve is a straight line, it is known as a **line of regression** and regression is said to be **linear.**

(2) A line of regression is the straight line which gives the best fit in the least square sense to the given frequency.

(3) The equation of line of regression of y over x is

$$y - \bar{y} = r.\frac{\sigma_y}{\sigma_x}(x - \bar{x}) \quad \text{or}$$

$$y - \bar{y} = \frac{\Sigma XY}{\Sigma X^2}(x - \bar{x}) \quad .....(1)$$

and the equation of lines of regression of x over y is

$$x - \bar{x} = r.\frac{\sigma_x}{\sigma_y}(y - \bar{y}) \quad \text{or}$$

$$x - \bar{x} = \frac{\Sigma XY}{\Sigma Y^2}(y - \bar{y}) \quad ......(2)$$

where the coefficients $r.\dfrac{\sigma_y}{\sigma_x}$ and $r.\dfrac{\sigma_x}{\sigma_y}$ are called regression coefficients of y on x and x on y respectively, $\bar{x}, \bar{y}$ are the arithmetic means of x , y respectively and r is the coefficient of correlation.

(4) $a_1 = r\dfrac{\sigma_y}{\sigma_x}$ is called the regression coefficient of y on x.

(5) $b_1 = r\dfrac{\sigma_x}{\sigma_y}$ is called the regression coefficient of x on y.

(6) $a_1\, b_1 = r\dfrac{\sigma_y}{\sigma_x} \times r\dfrac{\sigma_x}{\sigma_y} = r^2$ Or $r = \sqrt{a_1\, b_1}$.

PROPERTIES OF CORRELATION AND REGRESSION COEFFICIENT

(1) $-1 \le r \le 1$

(2) The correlation coefficient and the two regression coefficient have same sign.

(3) If one of the regression coefficient is greater than unity, then other must be less than unity.

(4) Regression coefficients are independent of the origin but not of scale.

(5) Arithmetic mean of regression coefficient is greater than the correlation coefficient.

IMPORTANT FORMULAE :

(1) The correlation coefficient, $r = \dfrac{\sum XY}{\sqrt{\sum X^2 . \sum Y^2}}$

where $\sigma_x^2 = \dfrac{\sum X^2}{n}$, $\sigma_y^2 = \dfrac{\sum Y^2}{n}$

(2) The equation of the regression line of y on x is

$$y - \bar{y} = \frac{\sum XY}{\sum X^2}(x - \bar{x})$$

or $y - \bar{y} = r\left(\dfrac{\sigma_y}{\sigma_x}\right)(x - \bar{x})$

(3) The equation of the regression line of x on y is

$$x - \bar{x} = \frac{\sum XY}{\sum Y^2}(y - \bar{y})$$

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

(4) $a_1 = r\dfrac{\sigma_y}{\sigma_x}$ is called regression coefficient of y on x.

(5) $b_1 = r\dfrac{\sigma_x}{\sigma_y}$ is called regression coefficient of x on y.

(6) $r = \sqrt{a_1 b_1}$

(7) $\bar{x} = \dfrac{\Sigma x}{n}$ and $\bar{y} = \dfrac{\Sigma y}{n}$

(8) $a_1 = \dfrac{\Sigma XY}{\Sigma X^2}$, $b_1 = \dfrac{\Sigma XY}{\Sigma Y^2}$

SOLVED EXAMPLES :

Ex.25. Find the co-efficient of correlation and obtain the equation to the lines of regression for the data :

x	6	2	10	4	8
y	9	11	5	8	7

S-08

Soln. Here, n = 5

$\therefore \bar{x} = \dfrac{\Sigma x}{n} = \dfrac{30}{5} = 6$

and $\bar{y} = \dfrac{\Sigma y}{n} = \dfrac{40}{5} = 8$

x	y	X = x − 6	Y = y − 8	X^2	Y^2	XY
6	9	0	1	0	1	0
2	11	− 4	3	16	9	− 12
10	5	4	−3	16	9	− 12
4	8	− 2	0	4	0	0
8	7	2	−1	4	1	− 2
Σx = 30	Σy = 40			40	20	− 26

Coefficient of correlation is given by

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 . \Sigma Y^2}} = \frac{-26}{\sqrt{40 \times 20}} = \frac{-26}{28.2842} = -0.919$$

$\therefore$ $\boxed{r = -0.919}$ **Ans.**

The regression coefficient of y on x is

$$a_1 = \frac{\Sigma XY}{\Sigma X^2} = -\frac{26}{40} = -0.65$$

The equation of line of regression of y on x is

$y - \bar{y} = a_1 (x - \bar{x})$

$\Rightarrow y - 8 = -0.65 (x - 6)$

$\therefore$ $\boxed{y = -0.65x + 11.9}$ **Ans.**

The regression coefficient of x on y is

$$b_1 = \frac{\Sigma XY}{\Sigma Y^2} = -\frac{26}{20} = -1.3$$

The equation of line of regression of x on y is

$x - \bar{x} = b_1 (y - \bar{y})$

$\Rightarrow x - 6 = -1.3 (y - 8)$

$\therefore$ $\boxed{x = -1.3y + 16.4}$ **Ans.**

Ex.26. Find the equations of lines of regression and the coefficient of correlation for the following data :-

x	2	4	5	6	8	11
y	18	12	10	8	7	5

W-05

Soln. Here, n = 6

Mean of $x = \bar{x} = \dfrac{\Sigma x}{6} = \dfrac{36}{6} = 6$

Mean of $y = \bar{y} = \frac{\Sigma y}{6} = \frac{60}{6} = 10$

x	y	X = x − 6	Y = y − 10	X²	Y²	XY
2	18	−4	8	16	64	−32
4	12	−2	2	4	4	−4
5	10	−1	0	1	0	0
6	8	0	−2	0	4	0
8	7	2	−3	4	9	−6
11	5	5	−5	25	25	−25
Σx = 36	Σy = 60		Total	50	106	−67

Coefficient of correlation,

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}}$$

$$= \frac{-67}{\sqrt{50 \times 106}}$$

$$= -0.92$$

$\therefore$ $\boxed{r = -0.92}$ **Ans.**

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

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

$$\Rightarrow y - 10 = \frac{-67}{50}(x - 6)$$

$$\Rightarrow y - 10 = -1.34x + 8.04$$

$\therefore$ $\boxed{y = -1.34x + 18.04}$ **Ans.**

(ii) The equation of regression line of x on y is

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

$$\Rightarrow x - 6 = \frac{-67}{106}(y - 10)$$

$$\Rightarrow x - 6 = -0.6321y + 6.321$$

$\therefore$ $\boxed{x = -0.6321\,y + 12.321}$ **Ans.**

Ex.27. Find the correlation coefficient and the equations of regression lines from the following data :

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

Soln. Here, n = 5

$$\bar{x} = \frac{1}{n}\Sigma x = \frac{1}{5}15 = 3$$

$$\& \ \bar{y} = \frac{1}{n}\Sigma y = \frac{1}{5}25 = 5$$

x	y	X = x − 3	Y = y − 5	X²	Y²	XY
1	2	−2	−3	4	9	6
2	5	−1	0	1	0	0
3	3	0	−2	0	4	0
4	8	1	3	1	9	3
5	7	2	2	4	4	4
Σx = 15	Σy = 25		Total	10	26	13

Correlation Coefficient, 'r' $= \frac{\Sigma XY}{\sqrt{\Sigma X^2}\sqrt{\Sigma Y^2}}$

$$= \frac{13}{\sqrt{10 \times 26}}$$

$$= 0.8062$$

$\therefore$ $\boxed{r = 0.8062}$ **Ans.**

The regression coefficient of y on x is

$$a_1 = \frac{\Sigma XY}{\Sigma X^2} = \frac{13}{10} = 1.3$$

The regression coefficient of x on y is

$$b_1 = \frac{\Sigma XY}{\Sigma Y^2} = \frac{13}{26} = 0.5$$

Equation of regression line of y on x is $Y = a_1 X$

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

$\therefore$ $\boxed{y = 1.3x + 1.1}$ **Ans.**

Equation of regression line of x on y is $X = b_1 Y$,

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

$\therefore$ $\boxed{x = 0.5y + 0.5}$ **Ans.**

Ex.28. Calculate the coefficient of correlation and obtain the two regression lines for the following data:

x	1	2	3	4	5	6	7	8	9
y	9	8	10	12	11	13	14	16	15

Also obtain an estimate of y when x = 6.2.

W-06, 08

Soln. Here, n = 9

Mean of $x = \bar{x} = \frac{\Sigma x}{n} = \frac{45}{9} = 5$

Mean of $y = \bar{y} = \frac{\Sigma y}{n} = \frac{108}{9} = 12$

x	y	X = x – 5	Y = y – 12	X^2	Y^2	XY
1	9	– 4	–3	16	9	12
2	8	– 3	– 4	9	16	12
3	10	– 2	– 2	4	4	4
4	12	– 1	0	1	0	0
5	11	0	–1	0	1	0
6	13	1	1	1	1	1
7	14	2	2	4	4	4
8	16	3	4	9	16	12
9	15	4	3	16	9	12
$\Sigma x = 45$	$\Sigma y = 108$		Total	60	60	57

Coefficient of correlation,

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 . \Sigma Y^2}}$$

$$= \frac{57}{\sqrt{60 \times 60}} = 0.95$$

$\therefore$ $\boxed{r = 0.95}$ **Ans.**

(1) The regression line of y on x is

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

$$\Rightarrow y - 12 = \frac{57}{60}(x - 5)$$

$$\Rightarrow y - 12 = 0.95\,x - 4.75$$

$$\Rightarrow y = 0.95\,x + 12 - 4.75$$

$\therefore$ $\boxed{y = 0.95\,x + 7.25}$ **Ans.**

The value of y when x = 6.2 is

$$y = 0.95\,(6.2) + 7.25$$

$\therefore$ $\boxed{y = 13.14 \text{ when } x = 6.2}$ **Ans.**

(2) The regression line of x on y is

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

$$\Rightarrow x - 5 = \frac{57}{60}(y - 12)$$

$$\Rightarrow x - 5 = 0.95\,y - 11.4$$

$$\Rightarrow x = 0.95\,y + 5 - 11.4$$

$\therefore$ $\boxed{x = 0.95\,y - 6.4}$ **Ans.**

Ex.29. Find the coefficient of correlation and obtain equation to the lines of regression to the following data :

x	10	15	25	20	35	40	50	45	30
y	7	8	3	5	9	7	19	15	17

W-10

Soln. Here n = 9, $\bar{x} = \frac{270}{9} = 30$

$$\bar{y} = \frac{90}{9} = 10$$

$\therefore$ $X = x - \bar{x} = x - 30$

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

The various calculations are shown in the following table

x	y	X = x – 30	Y = y –10	XY	X^2	Y^2
10	7	– 20	– 3	60	400	9
15	8	– 15	– 2	30	225	4
25	3	– 5	– 7	35	25	49
20	5	– 10	– 5	50	100	25
35	9	5	– 1	– 5	25	1
40	7	10	– 3	– 30	100	9
50	19	20	9	180	400	81
45	15	15	5	75	225	25
30	17	0	7	0	0	49
$\Sigma x = 270$	$\Sigma y = 90$			$\Sigma XY = 395$	$\Sigma X^2 = 1500$	$\Sigma Y^2 = 252$

Coefficient of correlation,

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}}$$

$$= \frac{395}{\sqrt{1500 \times 252}}$$

$\therefore$ $\boxed{r = 0.6425}$ **Ans.**

Now,

$$r\frac{\sigma_y}{\sigma_x} = r\sqrt{\frac{\Sigma Y^2}{\Sigma X^2}}$$

$$= 0.6425\sqrt{\frac{252}{1500}}$$

$$\therefore\ r\frac{\sigma_y}{\sigma_x} = 0.2634$$

$$\text{and } r\frac{\sigma_x}{\sigma_y} = 0.6425\sqrt{\frac{\Sigma X^2}{\Sigma Y^2}}$$

$$= 0.6425\sqrt{\frac{1500}{252}}$$

$$\therefore\ r\frac{\sigma_x}{\sigma_y} = 1.5676$$

$\therefore$ Equation of line of regression of y on x is

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

$\Rightarrow y - 10 = (0.2634)(x - 30)$

$\Rightarrow y - 10 = 0.2634\,x - 7.902$

$\therefore \boxed{y = 0.3624\,x + 2.098}$ **Ans.**

and equation of line of regression of x on y is

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

$\Rightarrow x - 30 = 1.5576\,(y - 10)$

$\Rightarrow x - 30 = 1.5676\,y - 15.676$

$\therefore \boxed{x = 1.5676\,y + 14.324}$ **Ans.**

Ex.30. Find the coefficient of correlation and regression lines to the following data :-

x	5	7	8	10	11	13	16
y	33	30	28	20	18	16	09

W-03

Soln. Here, $n = 7,\ \bar{x} = \dfrac{\Sigma x}{n} = \dfrac{70}{7} = 10$

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

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

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

The various calculations are shown in the following table :

x	y	X = x − 10	Y = y − 22	XY	X²	Y²
5	33	−5	11	−55	25	121
7	30	−3	8	−24	9	64
8	28	−2	6	−12	4	36
10	20	0	−2	0	0	4
11	18	1	−4	−4	1	16
13	16	3	−6	−18	9	36
16	9	6	−13	−78	36	169
Σx = 70	Σy = 154			ΣXY = −191	ΣX² = 84	ΣY² = 446

Coefficient of correlation,

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 . \Sigma Y^2}}$$

$$= \frac{-191}{\sqrt{84 \times 446}} = -0.9868$$

$\therefore \boxed{r = -0.9868}$ **Ans.**

Now,

$$r\frac{\sigma_y}{\sigma_x} = r\sqrt{\frac{\Sigma Y^2}{\Sigma X^2}}$$

$$= -0.9868\sqrt{\frac{446}{84}} = -2.2739$$

and $\quad r\dfrac{\sigma_x}{\sigma_y} = r\sqrt{\dfrac{\Sigma X^2}{\Sigma Y^2}}$

$$= -0.9868\sqrt{\frac{84}{446}}$$

$$= -0.4283$$

$\therefore$ Equation of line of regression of y on x is

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

$\Rightarrow y - 22 = -2.2739\,(x - 10)$

$\Rightarrow y - 22 = -2.2739x + 22.739$

$\therefore \boxed{y = -2.2739x + 44.739}$ **Ans.**

and equation of line of regression of x on y is

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

$\Rightarrow x - 10 = -0.4283\,(y - 22)$

$\Rightarrow x - 10 = -0.4283\,y + 9.4226$

$\therefore \boxed{x = -0.4283\,y + 19.4226}$ **Ans.**

Ex.31. Two lines of regression are given by $8x - 10y + 66 = 0$ and $40x - 18y = 214$.

If $\sigma_x^2 = 9$, find

(i) mean values of x and y,

(ii) the coefficient of correlation between x and y,

(iii) the standard deviation of y,

(iv) variance of y.

S-02, 04, W-07

Soln. (i) Since both the lines of regression pass through the point $(\bar{x}, \bar{y})$ therefore, we have

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

$40\bar{x} - 18\bar{y} - 214 = 0$ .....(2)

Multiplying equation (1) by 5

$\therefore\ 40\bar{x} - 50\bar{y} + 330 = 0$ .....(3)

Subtracting equation (3) from (2), we get

$32\bar{y} - 544 = 0$

$$\therefore\ \bar{y} = \frac{544}{32} = 17$$

$\therefore$ From (1), $8\bar{x} - 170 + 66 = 0$

$\Rightarrow 8\bar{x} = 104$

$\therefore\ \bar{x} = 13$

$\therefore \boxed{\begin{array}{l} \bar{x} = 13 \\ \bar{y} = 17 \end{array}}$ **Ans.**

(ii) Variance of $x = \sigma_x^2 = 9$ ......(given)

$\therefore \boxed{\sigma_x = 3}$

From (1) and (2), the equations of lines of regression can be written as

$y = 0.8x + 6.6 = a_0 + a_1 x$

and $x = 0.45y + 5.35 = b_0 + b_1 y$

$\therefore$ The regression co-efficient of y on x is

$$r\frac{\sigma_y}{\sigma_x} = 0.8 \quad .....(4)$$

The regression co-efficient of x on y is

$$r\frac{\sigma_x}{\sigma_y} = 0.45 \quad ....(5)$$

Multiplying (4) and (5), we get

$r^2 = 0.8 \times 0.45 = 0.36$

$\therefore$ $\boxed{\text{Coefficient of correlation, } r = 0.6}$ **Ans.**

(+ve sign with square root is taken because regression coefficients are +ve)

(iii) From (4), $\sigma_y = \frac{0.8\,\sigma_x}{r} = \frac{0.8 \times 3}{0.6} = 4$

$\therefore \boxed{\sigma_y = 4}$ **Ans.**

(iv) $\therefore$ Variance of $y = \sigma_y^2 = (4)^2 = 16$

$\therefore \boxed{\sigma_y^2 = 16}$ **Ans.**

Ex.32. Find the coefficient of correlation between the variables x and y and hence find the regression lines

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

S-09

Soln. Here n = 8,

$\bar{x} = \frac{\Sigma x}{n} = \frac{56}{8} = 7, \quad \bar{y} = \frac{\Sigma y}{n} = \frac{40}{8} = 5$

$\therefore X = x - \bar{x} = x - 7, \quad Y = y - \bar{y} = y - 5$

x	y	X = x − 7	Y = y − 5	XY	X^2	Y^2
1	1	−6	−4	24	36	16
3	2	−4	−3	12	16	9
4	4	−3	−1	3	9	1
6	4	−1	−1	1	1	1
8	5	1	0	0	1	0
9	7	2	2	4	4	4
11	8	4	3	12	16	9
14	9	7	4	28	49	16
Σx = 56	Σy = 40		Total	84	132	56

$\therefore$ Coefficient of correlation,

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \; \Sigma Y^2}} = \frac{84}{\sqrt{132 \times 56}} = 0.977$$

$\therefore \boxed{r = 0.977}$ **Ans.**

(1) The equation of the line of regression y on x is

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

Now,

$$r\frac{\sigma_y}{\sigma_x} = r\sqrt{\frac{\Sigma Y^2}{\Sigma X^2}} = 0.977 \times \sqrt{\frac{56}{132}} = 0.64$$

$\therefore\; y - 5 = 0.64\,(x - 7)$

$\therefore \boxed{y = 0.64x + 0.52}$ **Ans.**

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

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

Now

$$r\frac{\sigma_x}{\sigma_y} = r\sqrt{\frac{\Sigma X^2}{\Sigma Y^2}} = 0.977 \times \sqrt{\frac{132}{56}} = 1.5$$

$\therefore\; x - 7 = 1.5\,(y - 5)$

$\therefore \boxed{x = 1.5y - 0.5}$ **Ans.**

Ex.33. Find the equation of regression lines and the coefficient of correlation for the following data :

x	3	5	6	8	9	11
y	2	3	4	6	5	8

Soln. Here, n = 6

Mean of $x = \bar{x} = \frac{\Sigma x}{6} = \frac{42}{6} = 7$

Mean of $y = \bar{y} = \frac{\Sigma y}{6} = \frac{28}{6} = 4.7$

$\therefore dx = x - 7$ and $dy = y - 5$

x	y	dx = x − 7	dy = y − 5	dx^2	dy^2	dx dy
3	2	−4	−3	16	9	12
5	3	−2	−2	4	4	4
6	4	−1	−1	1	1	1
8	6	1	1	1	1	1
9	5	2	0	4	0	0
11	8	4	3	16	9	12
Σx = 42	Σy = 28	0	−2	42	24	30

$\therefore$ Coefficient of correlation,

$$r = \frac{n\sum dx\,dy - \sum dx.\sum dy}{\sqrt{n\sum dx^2 - (\sum dx)^2}\ \sqrt{n\sum dy^2 - (\sum dy)^2}}$$

$$= \frac{6\times 30 - 0\,(-2)}{\sqrt{[6\times 42 - 0]\,[6\times 24 - (-2)^2]}}$$

$$= \frac{6\times 30}{\sqrt{6\times 42\times 140}}$$

$$= 0.9583$$

$\therefore$ $\boxed{r = 0.9583}$ **Ans.**

(i) Equation of the lines of regression of y on x is

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

Now,

$$\sigma_x = \sqrt{\frac{1}{n}\sum dx^2 - \left(\frac{\sum dx}{n}\right)^2} = \sqrt{\frac{42}{6} - 0} = 2.6458$$

$$\sigma_y = \sqrt{\frac{1}{n}\sum dy^2 - \left(\frac{\sum dy}{n}\right)^2}$$

$$= \sqrt{\frac{24}{6} - \left(\frac{-2}{6}\right)^2} = 1.9720$$

$$\therefore\ y - 4.7 = 0.9583\times\frac{1.9720}{2.6458}(x-7)$$

$$\therefore\ y = 4.7 + 0.7143x - 5.0001$$

$\therefore$ $\boxed{y = 0.7143x - 0.3001}$ **Ans.**

Equation of the line of regression of x on y is

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

$$\Rightarrow\ x - 7 = 0.9583\times\frac{2.6458}{1.9720}(y - 4.7)$$

$$\therefore\quad x = 7 + 1.2857\,y - 6.0430$$

$\therefore$ $\boxed{x = 1.2857y + 0.9570}$ **Ans.**

Ex.34. Calculate the coefficient for the following data. Also find the equation to the lines of regression.

x	24	13	27	12	31	42	13	29	17	11
y	24	25	21	25	22	19	24	20	25	26

Soln. Here n = 10

$$\bar{x} = \frac{\sum x}{n} = \frac{219}{10} = 21.9$$

$$\bar{y} = \frac{\sum y}{n} = \frac{231}{10} = 23.1$$

$\therefore$ dx = x − 22 and dy = y − 23

x	y	dx = x − 22	dy = y − 23	dx^2	dy^2	dxdy
24	24	2	1	4	1	2
13	25	−9	2	81	4	−18
27	21	5	−2	25	4	−10
12	25	−10	2	100	4	−20
31	22	9	−1	81	1	−9
42	19	20	−4	400	16	−80
13	24	−9	1	81	1	−9
29	20	7	−3	49	9	−21
17	25	−5	2	25	4	−10
11	26	−11	3	121	9	−33
$\sum x$ = 219	$\sum y$ =231	$\sum dx$ = −1	$\sum dy$ = 1	$\sum dx^2$ = 967	$\sum dy^2$ = 53	$\sum dx\,dy$ = −208

$\therefore$ Coefficient of correlation,

$$r = \frac{n\sum dx\,dy - \sum dx.\sum dy}{\sqrt{n\sum dx^2 - (\sum dx)^2}\ .\sqrt{n\sum dy^2 - (\sum dy)^2}}$$

$$= \frac{(10)\,(-208) - (-1)(1)}{\sqrt{(10)\,(967) - (-1)^2}\ \sqrt{(10)\,(53) - (1)^2}}$$

$$= \frac{-2080 + 1}{\sqrt{9670 - 1}\ .\sqrt{530 - 1}}$$

$$= \frac{-2079}{\sqrt{9669}\ .\sqrt{529}}$$

$\therefore$ $\boxed{r = -0.9194}$ **Ans.**

Equation of lines of regression of y on x is

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

Now

$$\sigma_x = \sqrt{\frac{1}{n}\sum dx^2 - \left(\frac{\sum dx}{n}\right)^2}$$

$$= \sqrt{\frac{967}{10} - \left(\frac{-1}{10}\right)^2}$$

$$\therefore\ \sigma_x = \sqrt{96.7 - 0.01}$$

$$\therefore\ \sigma_x = 9.8332$$

$$\text{And } \sigma_y = \sqrt{\frac{1}{n}\sum dy^2 - \left(\frac{\sum dy}{n}\right)^2}$$

$$= \sqrt{\frac{53}{10} - \left(\frac{1}{10}\right)^2}$$

$$= \sqrt{5.3 - 0.01}$$

$$\therefore\ \sigma_y = 2.3$$

$$\therefore\ y - 23 = -0.9193\ \frac{(2.3)}{9.8332}(x - 22)$$

$\Rightarrow y - 23 = -0.215\,(x - 22)$

$\Rightarrow y - 23 = -0.215\,x + 4.73$

$\therefore$ $\boxed{y = -0.215\,x + 27.73}$ **Ans.**

Equation of the line of regression of x on y is

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

$$\Rightarrow (x - 22) = (-0.9193)\,\frac{9.8332}{(2.3)}\,(y - 23)$$

$$\Rightarrow (x - 22) = 3.93(y - 23)$$

$$\Rightarrow x = 3.93\,y - 90.39 + 22$$

$\therefore$ $\boxed{x = 3.93\,y - 68.39}$ **Ans.**

Ex.35. While calculating coefficients of correlation between two variables x and y from 25 pairs of observations, the following results were obtained:-

$n = 25$, $\sum x = 125$, $\sum x^2 = 840$, $\sum y = 98$, $\sum y^2 = 323$ and $\sum xy = 620$

It was however, later discovered at the time of checking that he had copied down two pairs as

x	y
5	12
9	8

with correct values

x	y
9	2
5	18

Find the correct value of coefficients of correlation.

S-10

Soln. **Corrected :-**

$$\sum x = 125 - 5 - 9 + 5 + 9 = 125$$

$$\sum y = 98 - 12 - 8 + 2 + 18 = 98$$

$$\sum x^2 = 840 - 5^2 - 9^2 + 5^2 + 9^2 = 840$$

$$\sum y^2 = 323 - 12^2 - 8^2 + 2^2 + 18^2 = 443$$

$$\sum xy = 620 - 60 - 72 + 18 + 90 = 596$$

$$\therefore\ r = \frac{n\sum xy - \sum x.\sum y}{\sqrt{[n\sum x^2 - (\sum x)^2]\times[n\sum y^2 - (\sum y)^2]}}$$

$$= \frac{(25)(596) - (125)(98)}{\sqrt{[(25)(840) - (125)^2]\times[(25)(443) - (98)^2]}}$$

$$= \frac{2650}{\sqrt{5375 \times 1471}}$$

$\therefore$ $\boxed{r = 0.9425}$ **Ans.**

Ex.36. Two lines of regression are given by $5y - 8x + 17 = 0$ and $2y - 5x + 14 = 0$.

If $\sigma_y^2 = 16$, find (i) the mean values of x and y, (ii) the coefficient of correlation between x and y, (iii) σ_x^2

S-12, W-09

Soln. i) We have, $5y - 8x + 17 = 0$ .....(1)

$2y - 5x + 14 = 0$

Since $(\bar{x}, \bar{y})$ is a common point of the two lines of regression, we have

$5\bar{y} - 8\bar{x} + 17 = 0$

$2\bar{y} - 5\bar{x} + 14 = 0$

$$\Rightarrow \frac{\bar{x}}{17\times2 - 14\times5} = \frac{\bar{y}}{-8\times14 - (-5\times17)}$$

$$= \frac{1}{5\times-5 - (-8)\times2}$$

$$\Rightarrow \frac{\bar{x}}{34 - 70} = \frac{\bar{y}}{-112 + 85} = \frac{1}{-25 + 16}$$

$$\Rightarrow \frac{\bar{x}}{-36} = \frac{\bar{y}}{-27} = \frac{-1}{9}$$

$$\therefore\ \bar{x} = \frac{36}{9} = 4 \text{ and } \bar{y} = \frac{27}{9} = 3$$

$\therefore$ $\boxed{\bar{x} = 4 \text{ and } \bar{y} = 3}$ **Ans.**

ii) The equations of line of regression can be written as

$$y = \frac{8}{5}x - \frac{17}{5} \quad\text{and}\quad x = \frac{2}{5}y + \frac{14}{5}$$

$$\therefore\ a_1 = \frac{8}{5} \text{ and } b_1 = \frac{2}{5}$$

Correlation coefficient, $r = \sqrt{a_1 b_1} = \sqrt{\frac{8}{5}\times\frac{2}{5}} = \frac{4}{5} = 0.8$

$\therefore$ $\boxed{r = 0.8}$ **Ans.**

iii) But $a_1 = r\frac{\sigma_y}{\sigma_x}$

and we are given $\sigma_y^2 = 16$

$\therefore\ \sigma_y = 4$

Hence we have $\frac{8}{5} = \frac{4}{5}\times\frac{4}{\sigma_x}$

$$\therefore\ 8 = \frac{16}{\sigma_x} \Rightarrow 8\sigma_x = 16$$

$$\therefore\ \sigma_x = \frac{16}{8} = 2$$

$\therefore$ $\boxed{\sigma_x^2 = 4}$ **Ans.**

Ex.37. Two lines of regression are given by

$x + 2y - 5 = 0$ **and** $2x + 3y - 8 = 0$

If $\sigma_x^2 = 12$, **find**

(i) the mean values of x and y,

(ii) the coefficient of correlation between x and y,

(iii) the standard deviation of y.

W-11

Soln. (i) Since both the lines of regression pass through the point $(\bar{x}, \bar{y})$, therefore, we have

$\bar{x} + 2\bar{y} - 5 = 0$

$2\bar{x} + 3\bar{y} - 8 = 0$

$$\Rightarrow \frac{\bar{x}}{-1} = -\frac{\bar{y}}{2} = \frac{1}{-1}$$

$\therefore \boxed{\bar{x} = 1,\ \bar{y} = 2}$ **Ans.**

are the required mean values.

(ii) Given : $\sigma_x^2 = 12 \Rightarrow \sigma_x = 3.464$

The equations of line of regression can be written as

$$y = \frac{-1}{2}x + \frac{5}{2} = -0.5x + 2.5$$

and $x = -\frac{3}{2}y + 4 = -1.5x + 4$

$$\therefore a_1 = r\frac{\sigma_y}{\sigma_x} = -0.5 \text{ and } b_1 = r\frac{\sigma_x}{\sigma_y} = -1.5$$

$$\therefore r = \sqrt{a_1 b_1} = \sqrt{(-0.5)(-1.5)} = 0.87$$

$\therefore$ $\boxed{\text{Coefficient of correlation, } r = 0.87}$ **Ans.**

(iii) Now

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

$$\Rightarrow 0.87 \times \frac{\sigma_y}{3.464} = -0.5$$

$\therefore \boxed{\sigma_y = 2}$ **Ans.**

[∵ + ve sign is taken]

Ex.38. If θ is the angle between the two regression lines, show that

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

where r, σ_x and σ_y have their usual meaning. Explain the significance when r = 0 and r = ± 1.

W-04

Soln. The equations to the lines of regression of y on x and x on y are

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

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

$\therefore$ Their slopes are $m_1 = r\frac{\sigma_y}{\sigma_x}$ and $m_2 = \frac{\sigma_y}{r\sigma_x}$.

$$\text{Thus, } \tan\theta = \frac{m_2 - m_1}{1 + m_1 m_2}$$

$$= \frac{\sigma_y / r\sigma_x - r\sigma_y/\sigma_x}{1 + \sigma_y^2/\sigma_x^2}$$

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

$\therefore$ $\boxed{\tan\theta = \frac{1-r^2}{r} \cdot \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}}$ **Ans.**

When $r = 0$, $\tan\theta \to \infty$ or $\theta = \pi/2$ i.e. when the variables are independent, the two lines of regression are perpendicular to each other.

When $r = \pm 1$, $\tan\theta = 0$ i.e. $\theta = 0$ or π. Thus, the lines of regression coincide i.e. there is a perfect correlation between the two variables.

CORRELATION BY RANK OR RANK CORRELATION

Some times it is more convenient to deal with the rank or order of the values of the two sets of variables than with their actual values. A method was developed by C. Spearman. Let us explain this by an illustration.

Suppose n students are examined in statistics and the top scorer of marks is assigned a number 1, the second number 2, the third number 3 and so on. Similarly, the same n students are graded according to the marks obtained by them in maths and assigned number as before. Our problem is to find the coefficient of correlation between their grades in maths and statistics.

Let the ranks given to statistics be x_i,

where $i = 1, 2, \ldots. n$.

Similarly, let the ranks given to mathematics be y_i,

where $i = 1, 2, \ldots.n$.

Let $d_i = x_i - y_i$, the difference between the ranks of x_i and y_i.

Then the rank correlation coefficient, $r = 1 - \dfrac{6\sum d_i^2}{n(n^2-1)}$

Note :

(1) If some item is repeated for p times in a given data then add a factor $\frac{1}{12}(p^3 - p)$ to every repeated item.

$$\therefore \quad r = 1 - \frac{6\left[\Sigma d^2 + \frac{1}{12}(p^3-p) + \frac{1}{12}(q^3-q) + \ldots\ldots\right]}{n(n^2-1)}$$

The value of r given above is known as Spearman's coefficient of rank correlation.

(2) If there are more than one item with the same rank. The rank to the equal items is assigned by average rank to each of these individuals.

For example : Suppose an item is repeated at the rank 5^{th} (i.e. the 5^{th} and 6^{th} item and having the same values than the common rank is assigned to 5^{th} and 6^{th} item is $\frac{5+6}{2} = 5.5$ which is the average of 5 and 6. The next rank assigned will be seven.

If an item is repeated thrice at rank 2, then the common rank assigned to each value will be $\frac{2+3+4}{3} = 3$ which is the arithmetic mean of 2, 3 and 4. Then next rank to be assigned would be 5.

IMPORTANT FORMULAE :

(1) The rank correlation coefficient for non-repeated rank is $r = 1 - \dfrac{6\sum d^2}{n(n^2-1)}$, d = difference in ranks,

n = number of item.

(2) The rank correlation coefficient for repeated ranks is

$$r = 1 - \frac{6\left[\sum d^2 + \frac{1}{12}(m^3-m) + \frac{1}{12}(l^3-l) + \frac{1}{12}(k^3-k) + \ldots\right]}{n(n^2-1)}$$

where one rank is repeated m times, another l times etc.

SOLVED EXAMPLES :

Ex.39. Calculate the coefficient of rank correlation for the following data :

x	2	4	5	6	8	11
y	18	12	10	8	7	5

Soln. Here, n = 6

The various calculations are shown in table :

x	y	R_1	R_2	$d = R_1 - R_2$	d^2
2	18	6	1	5	25
4	12	5	2	3	9
5	10	4	3	1	1
6	8	3	4	–1	1
8	7	2	5	–3	9
11	5	1	6	–5	25
					$\Sigma d^2 = 70$

∴ Coefficient of rank correlation is

$$r = 1 - \frac{6\,\Sigma d^2}{n(n^2-1)}$$

$$= 1 - \frac{6 \times 70}{6(36-1)}$$

$$= 1 - \frac{70}{35}$$

$$= 1 - 2$$

∴ $\boxed{\text{Rank correlation coefficient, } r = -1}$ **Ans.**

Ex.40. Obtain the rank correlation coeffici.nt for the following data :

x	68	64	75	50	64	80	75	40	55	64
y	62	58	68	45	81	60	68	48	50	70

S-02, 06, W-02

Soln. The various calculations are shown in table :

x	y	Rank in x x'	Rank in y y'	d = x' – y'	d^2
68	62	4	5	–1	1
64	58	6	7	–1	1
75	68	2.5	3.5	–1	1
50	45	9	10	–1	1
64	81	6	1	5	25
80	60	1	6	–5	25
75	68	2.5	3.5	–1	1
40	48	10	9	1	1
55	50	8	8	0	0
64	70	6	2	4	16
				Total	72

Rank correlation coefficient,

$$r = 1 - \frac{6\left[\Sigma d^2 + \frac{1}{12}(m^3-m) + \frac{1}{12}(p^3-p) + \frac{1}{12}(q^3-q)\right]}{n(n^2-1)}$$

$$\therefore r = 1 - \frac{6\left[72 + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(3^3 - 3)\right]}{10\,(100 - 1)}$$

$$= 1 - \frac{6\,(72 + 0.5 + 0.5 + 2)}{10\,(99)}$$

$$= 0.545.$$

$\therefore$ $\boxed{\text{Rank correlation coefficient, } r = 0.545}$ **Ans.**

Ex.41. Find rank correlation coefficient to the following data :

x	65	63	67	64	68	62	70	66	68	67	69	71
y	68	66	68	65	69	66	68	65	71	67	68	70

S-07

Soln. Here, we assign rank to the values of x and y and we have a table of the following form :

x	y	Rank in x x'	Rank in y y'	$d = x' - y'$	d^2
65	68	9	5.5	3.5	12.25
63	66	11	9.5	1.5	2.25
67	68	6.5	5.5	1	1
64	65	10	11.5	− 1.5	2.25
68	69	4.5	3	1.5	2.25
62	66	12	9.5	2.5	6.25
70	68	2	5.5	− 3.5	12.25
66	65	8	11.5	− 3.5	12.25
68	71	4.5	1	3.5	12.25
67	67	6.5	8	− 1.5	2.25
69	68	3	5.5	− 2.5	6.25
71	70	1	2	− 1	1
				Total	72.5

The rank correlation coefficient r is given by

$$r = \frac{1 - 6\left[\begin{array}{l}\Sigma d^2 + \frac{1}{12} m\,(m^2 - 1) + \frac{1}{12} l\,(l^2 - 1) \\ + \frac{1}{12} p\,(p^2 - 1) + \frac{1}{12} q\,(q^2 - 1) + \frac{1}{12} k\,(k^2 - 1)\end{array}\right]}{n(n^2 - 1)}$$

Here, n = 12, two x values are repeated twice. Two y values are repeated twice and one y values is repeated four times.

$$\therefore r = \frac{1 - 6\left[\begin{array}{l}72.5 + \frac{1}{12} 2(2^2 - 1) + \frac{1}{12} 2(2^2 - 1) \\ + \frac{1}{12} 2(2^2 - 1) + \frac{1}{12} 2(2^2 - 1) + \frac{1}{12} 4(4^2 - 1)\end{array}\right]}{12\,(144 - 1)}$$

$$= 1 - \frac{6(72.5 + 2 + 5)}{12 \times 143}$$

$$= 1 - 0.2779$$

$$= 0.722$$

$\therefore$ $\boxed{\text{Rank correlation coefficient, } r = 0.722}$ **Ans.**

Ex.42. Find rank correlation coefficient for following data :

x	74	75	78	72	78	77	79	81	79	76	72	71
y	47	44	40	48	49	45	46	42	42	39	46	40

S-05

Soln. Here, n = 12

The various calculation are shown in table :

x	y	Rank in x'	Rank in y'	Rank diff $d = x' - y'$	d^2
74	47	9	3	6	36
75	44	8	7	1	1
78	40	4.5	10.5	− 6	36
72	48	10.5	2	8.5	72.25
78	49	4.5	1	3.5	12.25
77	45	6	6	0	0
79	46	2.5	4.5	−2	4
81	42	1	8.5	− 7.5	56.25
79	42	2.5	8.5	− 6	36
76	39	7	12	− 5	25
72	46	10.5	4.5	6	36
71	40	12	10.5	1.5	2.25
					$\Sigma d^2 = 317$

Coefficient of Rank correlation,

$$r = 1 - \frac{6\left[\Sigma d^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \ldots\right]}{n(n^2 - 1)}$$

$$= 1 - \frac{6\left[\begin{array}{l}317 + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2) \\ + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2)\end{array}\right]}{12(12^2 - 1)}$$

$$= 1 - \frac{6[317 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5 + 0.5]}{12 \times 143}$$

$$= 1 - 6\frac{(317 + 3)}{12 \times 143}$$

$\therefore r = 1 - 1.1189 = -0.1189.$

$\therefore$ $\boxed{\text{Coefficient of Rank correlation, } r = -0.1189}$ **Ans.**

Ex.43. Marks of twelve students in mathematics and statistics are given below:

Mathematics	60	34	40	50	45	40	22	43	42	66	64	46
Statistics	75	32	33	40	45	33	12	30	34	72	41	57

Calculate rank correlation coefficient.

Soln. Let the marks of mathematics be denoted by x and the marks of statistics be denoted by y. Here n = 12. Also, here we observe that two students have secured equal marks in mathematics and too in statistics.

x	y	Rank in x'	Rank in y'	Rank diff $d = x' - y'$	d^2
60	75	3	1	2	4
34	32	11	10	1	1
40	33	9.5	8.5	1	1
50	40	4	6	−2	4
45	45	6	4	2	4
40	33	9.5	8.5	1	1
22	12	12	12	0	0
43	30	7	11	−4	16
42	34	8	7	1	1
66	72	1	2	−1	1
64	41	2	5	−3	9
46	57	5	3	2	4
				Total	46

$\therefore$ The rank correlation coefficient in this case is given by

$$r = 1 - 6\frac{\left[\Sigma d^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(p^3 - p) + \ldots\ldots\ldots\ldots\right]}{n(n^2 - 1)}$$

$$= 1 - 6\frac{\left[46 + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2)\right]}{12\,(12^2 - 1)}$$

[$\because$ m = 2 for one series and m = 2 for the other]

$$\therefore r = 1 - \frac{6[46 + 0.5 + 0.5]}{12 \times 143} = 0.84$$

$\therefore$ $\boxed{\text{Coefficient of Rank correlation, } r = 0.84}$ **Ans.**

COMPLETE LIST OF FORMULAE:

(1) The normal equations to fit the straight line $y = a + bx$ are $\sum y = na + b\sum x$ and $\sum xy = a\sum x + b\sum x^2$

(2) The normal equations to fit the parabola $y = a + bx + cx^2$ are

$\sum y = na + b\sum x + c\sum x^2$,

$\sum xy = a\sum x + b\sum x^2 + c\sum x^3$ and

$\sum x^2 y = a\sum x^2 + b\sum x^3 + c\sum x^4$

(3) The normal equations to fit the parabola $y = ax + bx^2$ are

$\sum xy = a\sum x^2 + b\sum x^3$

and $\sum x^2 y = a\sum x^3 + b\sum x^4$

(4) The normal equations to fit the curve $y = ax^2 + \frac{b}{x}$ are $\sum x^2 y = a\sum x^4 + b\sum x$ and

$\sum \frac{y}{x} = a\sum x + b\sum \frac{1}{x^2}$

(5) To fit the curve $y = ax^b$, first take logarithm on both the sides which gives $\log y = \log a + b \log x$, put $Y = \log y$, $A = \log a$ and $X = \log x$, then we get a straight line equation $Y = A + bX$ and normal equations are $\sum Y = nA + b\sum X$ and

$\sum XY = A\sum X + b\sum X^2$.

(6) To fit the curve $y = ab^x$, take logarithm on both sides which gives $\log Y = \log a + x \log b$ putting $Y = \log y$, $A = \log a$ and $B = \log b$, we get $Y = A + Bx$ and the normal equations are

$\sum Y = nA + B\sum x$ and

$\sum xY = A\sum x + B\sum x^2$

(7) To fit the curve $y = ae^{bx}$, take logarithm on both sides, we get $\log y = \log a + bx \log e$, put $Y = \log y$, $A = \log a$ and $B = b \log e$, then the equation becomes $Y = A + Bx$. Normal equations are $\sum Y = nA + B\sum x$ and $\sum xY = A\sum x + B\sum x^2$

(8) The correlation coefficient, $r = \frac{\sum XY}{\sqrt{\sum X^2 . \sum Y^2}}$

where $\sigma_x^2 = \frac{\sum X^2}{n}$, $\sigma_y^2 = \frac{\sum Y^2}{n}$

(9) The equation of the regression line of y on x is

$$y - \bar{y} = \frac{\sum XY}{\sum X^2}(x - \bar{x}) \text{ or}$$

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

(10) The equation of the regression line of x on y is

$$x - \bar{x} = \frac{\sum XY}{\sum Y^2}(y - \bar{y})$$

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

(11) $a_1 = r\frac{\sigma_y}{\sigma_x}$ is called regression coefficient of y on x.

(12) $b_1 = r\frac{\sigma_x}{\sigma_y}$ is called regression coefficient of x on y.

(13) $r = \sqrt{a_1 b_1}$

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

(15) $a_1 = \frac{\Sigma XY}{\Sigma X^2}$, $b_1 = \frac{\Sigma XY}{\Sigma Y^2}$

(16) The rank correlation coefficient for non-repeated rank

is $r = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$, d = difference in ranks,

n = number of data.

(17) The rank correlation coefficient for repeated ranks is

$$r = 1 - \frac{6\left[\sum d^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(l^3 - l) + \frac{1}{12}(k^3 - k) + ...\right]}{n(n^2 - 1)}$$

where one rank is repeated m times, another l times etc.

PROBLEMS FOR PRACTICE :

Ex.1. Find the coefficient of correlation and obtain the equations to the lines of regression for the following data:

x	1	2	3	4	5	6	7
y	7	8	9	11	10	13	12

Also find the value of y at x = 9.

Ans. $r = 0.93$, $y = 0.93\,x + 6.28$, $x = 0.93\,y - 5.3$, $y = 14.65$.

Ex.2. Calculate the rank correlation coefficient from the data given below :

x	78	89	97	69	50	79	68	57
y	125	137	156	112	107	136	123	108

Ans. . $r = 0.95$.

Ex.3. Fit a parabola $y = ax + bx^2$ to the following data :

x	1	2	3	4	5
y	1.8	5.1	8.9	14.1	19.8

Ans. $y = 1.521\,x + 0.49\,x^2$.

Ex.4. Fit a parabola $y = a + bx + cx^2$ for the following data :

x	20	30	40	50	60
y	54	90	138	206	292

Ans. $y = 38.2 - 0.88\,x + 0.085\,x^2$.

Ex.5. Fit a curve of the term $y = ae^{bx}$ to the following data :

x	0	2	4
y	5.012	10	31.62

Ans. $y = (4.642)e^{0.46x}$.

Ex.6. Fit a curve $PV^n = C$ for the following data :

V	2	4	6	8	10
P	91.5	35.9	19.9	14.1	10.4

Ans. $PV^{1.35} = 233$