MODULE – 04

Statistical Methods

Correlation and Regression

This topic deals with data concerning independent observations.

Examples:

1. Marks of individuals in two subjects
2. Height and weight of individuals
3. Amount involved in advertising a product and sale of product etc.,
   We discuss the aspect of inter – relation between the independent variables.

- **Correlation and Correlation Coefficient**

The numerical measure of correlation between two variables x and y is known as

Pearson's coefficient of correlation usually denoted by r and is defined as follows.

$$r = \frac{\sum_1^n (x-\bar{x})(y-\bar{y})}{n\sigma_x\sigma_y} \qquad \dots (1)$$

This can be put in the alternative form as follows.

If $X = x - \bar{x}$, $Y = y - \bar{y}$ we can write

Thus (1) becomes

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

Property :

The coeeficient of correlation numerically does not exceed unity.

ie $-1 \leq r \leq +1$.

Note : If r = ± 1 we say that x and y are perfectly correlated and if r = 0 we say that x and y are non correlated.

- **<u>Alternative formula for the Correlation Coefficient r</u>**

$$r = \frac{\sigma_x^2+\sigma_y^2-\sigma_{x-y}^2}{2\sigma_x\sigma_y}$$

<u>Problems:</u>

1. Calculate the karl – pearson co – efficient for the following ages of husband and wife's.

Roll No.	1	2	3	4	5	6	7	8	9	10
Husband's age (x)	36	23	27	28	28	29	30	31	33	35
Wife's age (y)	29	18	20	22	27	21	29	27	29	28

Solu: $\bar{x} = \frac{\sum X}{10} = \frac{300}{10} = 30$ $\bar{y} = \frac{\sum Y}{10} = \frac{250}{10} = 25$ Here n = 10

x	y	X= x - $\bar{x}$	Y = y -$\bar{y}$	XY	X^2	Y^2
36	29	-7	-7	49	49	49
23	18	-3	-5	15	9	25
27	20	-2	-3	6	4	9
28	22	-2	2	-4	4	4
28	27	-1	-4	4	1	16
29	21	0	4	0	0	16
30	29	1	2	2	1	4
31	27	3	4	12	9	16
33	29	5	3	15	25	9
35	28	6	4	24	36	16
$\sum x = 300$	$\sum y = 250$			$\sum XY = 123$	$\sum X^2 = 138$	$\sum Y^2 = 164$

We have

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

$$r = \frac{123}{\sqrt{(138)(164)}} = 0.817$$

It is a positive correlation.

2. Obtain the correlation of the following data:

x	10	14	18	22	26	30
y	18	12	24	6	30	36

Solu. We have $r = \frac{\sum XY}{\sqrt{\sum X^2 \ \sum Y^2}}$ where X = x - $\bar{x}$, Y = y -$\bar{y}$

Here n = 6

$$\bar{x} = \frac{\sum X}{10} = \frac{120}{6} = 20 \qquad \bar{y} = \frac{\sum Y}{10} = \frac{126}{6} = 21$$

x	y	X= x -$\bar{x}$	Y = y-$\bar{y}$	XY	X^2	Y^2
10	18	-10	-3	30	100	9
14	18	-6	-9	53	36	81
18	18	-2	3	-6	4	9
22	18	2	-15	-30	4	225
26	18	6	9	54	36	81
30	18	10	15	150	100	225
$\sum x = 120$	$\sum y = 126$			$\sum XY = 252$	$\sum X^2 = 280$	$\sum Y^2 = 630$

We have $r = \frac{252}{\sqrt{(280)(630)}} = 0.6$

It is a positive correlation.

3. Calculate Karl pearson co-efficient of correlation b/w the marks obtained by 8 students in mathematics and statistics :

Statistics	8	10	15	17	20	23	24	25
Mathematics	25	30	32	35	37	40	42	45

Solu: Let statistics = x , Mathematics = y

We have $r = \frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}}$ where $X = x - \bar{x}$, $Y = y - \bar{y}$

Here n = 8

$\bar{x} = \frac{\sum X}{n} = \frac{142}{8} = 17.75$ $\bar{y} = \frac{\sum Y}{n} = \frac{286}{8} = 35.75$

x	y	$X = x - \bar{x}$	$Y = y - \bar{y}$	XY	X^2	Y^2
8	25	-9.75	-10.75	104.81	95.06	115.56
10	30	-7.75	-5.75	44.56	60.06	33.06
15	32	-2.75	-3.75	10.31	7.56	14.06
17	35	-0.75	-0.75	0.56	0.56	0.56
20	37	2.25	1.25	2.81	5.06	1.56
23	40	5.25	4.25	22.31	27.56	18.06
24	42	6.25	6.25	39.06	39.06	39.06
25	45	7.25	9.25	67.06	56.56	85.56
$\sum x = 17.75$	$\sum y = 35.75$			$\sum XY = 291.48$	$\sum X^2 = 291.480$	$\sum Y^2 = 307.48$

We have $r = \frac{291.48}{\sqrt{(291.48)(307.48)}} = 0.97$

It is a positive correlation.

Regression

Regression is an estimation of one independent variable in terms of the other. If x & y are correlated, the best fitting straight line in the least square sense gives reasonably a good relation between x & y.

The best fitting straight line of the form y = ax + b (x being the independent variable) is called the regression line of y on x & x = ay + b (y being the independent variable) is called the regression line of x on y.

Formulas for line of regression

Let y = ax + b be the equation of the regression line of y on x for a given set of n values (x,y).

Then y- $\bar{y} = r\,\frac{\sigma_y}{\sigma_x}$ (x-$\bar{x}$)…………(1)

This is the regression of y on x.

Similarly,

x- $\bar{x} = r\,\frac{\sigma_x}{\sigma_y}$ (y-$\bar{y}$)………..(2)

This is the regression line of x on y.

The coefficient of x in (1) & the coefficient of y in (2) respectively given by r $\frac{\sigma_y}{\sigma_x}$ and r $\frac{\sigma_x}{\sigma_y}$ are known as the regression coefficients.

Their product is equal to r^2.

Thus we can conclude that r is the geometric mean (GM) of the regression coefficients since the GM of two numbers a, b is $\sqrt{ab}$. That is

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

The sign of r will be positive or negative according as the regression coefficients are positive or negative.

Note: The lines of regression (1) & (2) are also of the form

$$Y=\frac{\sum XY}{\sum X^2}(X) \quad \text{and} \quad X=\frac{\sum XY}{\sum Y^2}(Y)$$

Where X= x- $\bar{x}$ & Y= y- $\bar{y}$.

This form will be useful to find out the coefficient of correction by first obtaining the lines of regression as we have deduced that

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

Problems:

1. Compute the coefficient of correlation & the equation of the lines of regression for the following data.

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

Solu: We have coefficient of correlation $r = \frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}}$

Where $Y = y - \bar{y}$ $X = x - \bar{x}$ Here n = 7

$\bar{x} = \frac{\sum x}{n}$ = 28/7 = 4 $\bar{y} = \frac{\sum y}{n}$ = 77/7 = 11

x	y	$X = x - \bar{x}$	$Y = y - \bar{y}$	XY	X^2	Y^2
1	9	-3	-2	6	9	4
2	8	-2	-3	6	4	9
3	10	-1	-1	1	1	1
4	12	0	1	0	0	1
5	11	1	0	0	1	0
6	13	2	2	4	4	4
7	14	3	3	9	9	9
$\sum x = 28$	$\sum y = 77$			$\sum XY = 26$	$\sum X^2$=28	$\sum Y^2$=28

Line regression y on x

$Y = \frac{\sum XY}{\sum X^2}.X$

$y - \bar{y} = \frac{26}{28}(x - \bar{x})$

y-11 = 0.928 (x-4)

y = 0.92x + 7.29

Line regression x on y

$X = \frac{\sum XY}{\sum Y^2}.Y$

x-4 = $\frac{26}{28}$ (y-11)

x-4 = 0.928y -10.208

x = 0.928y-6.208

2. Obtain the lines of regression and hence find the co-efficient 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

Solution:

$\bar{x} = \frac{\sum x}{n} = 70/10 = 7$ $\bar{y} = \frac{\sum y}{n} = 150/10 = 15$

x	y	X= x- $\bar{x}$	Y= y- $\bar{y}$	xy	X^2	Y^2
1	8	-6	-7	42	36	49
3	6	-4	-9	35	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
$\sum x = 10$	$\sum y = 150$			$\sum XY = 360$	$\sum X^2 = 204$	$\sum Y^2 = 818$

Line of regression y on x

$Y = \frac{\sum XY}{\sum X^2} . X$

y- $\bar{y}$ = 360/204(x- $\bar{x}$)

y-15 = 1.764 (x-7)

y-15 = 1.764x-12.348

y = 1.764x-12.348+15

y = 1.764x + 2.652

line of regression x on y

$X = \frac{\sum XY}{\sum Y^2} . Y$

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

x- 7 = 0.44y + 6.6

x = 0.44y + 0.4

We have

Co-efficient of correlation r = $\pm\sqrt{(coeff. of\ x)(coeff. of\ y)}$

= 0.88

3. Compute the coefficient of correlation & the equation of the lines of regression for the following data.

x	10	14	18	22	26	30
y	18	12	24	6	30	36

Solu: $\bar{x} = \frac{\sum x}{n} = \frac{120}{6} = 20$ $\quad \bar{y} = \frac{\sum y}{n} = \frac{126}{6} = 21$

x	y	X= x- $\bar{x}$	Y= y- $\bar{y}$	XY	X^2	Y^2
10	18	-10	-3	30	100	9
14	12	-6	-9	54	36	81
18	24	-2	3	-6	4	9
22	6	2	-15	-30	4	225
26	30	6	9	54	36	81
30	36	10	15	150	100	225
$\sum x = 120$	$\sum y = 126$			$\sum XY = 252$	$\sum X^2$=280	$\sum Y^2$=630

We have, Coefficient of correlation r = $\frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}}$ = 0.6

Line of regression y on x

$Y = \frac{\sum XY}{\sum X^2}. X$

$y- \bar{y} = \frac{252}{280} (x- \bar{x})$

y – 21 = 0.9 (x – 20)

y = 0.9 x – 18+21

y = 0.9 x + 3

line of regression x on y

$X = \frac{\sum XY}{\sum Y^2}. Y$

$(x- \bar{x}) = \frac{252}{630} (y- \bar{y})$

x – 20 = 0.4 (y – 21)

x = 0.4y -8.4 + 20

x = 0.4 y +11.6

4. If Θ is the acute angle between the lines of regression, then show that $\tan\Theta = \frac{\sigma_x \sigma_y}{\sigma_x{}^2+\sigma_y{}^2}\left(\frac{1-r^2}{r}\right)$.Explain the significance when r = 0 & r = ± 1.

Solu: W. K. T If is Θ acute, the angle between the lines $y = m_1 x + c_1$ and $y = m_2 x + c_2$ is given by

$$\tan\Theta = \frac{m_2 - m_1}{1+ m_1 m_2}$$

We have the lines of regression ,

$$y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x - \bar{x}) \quad \text{.... (1)}$$

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

We write the second of the equation as $y - \bar{y} = \frac{\sigma_y}{r\sigma_x}(x - \bar{x})$ ….(2)

Slope of (1) and (2) are respectively given by

$$m_1 = r\frac{\sigma_y}{\sigma_x} \quad \text{and} \quad m_2 = \frac{\sigma_y}{r\sigma_x}$$

Substituting these in the formula for tan Θ we have,

$$\tan\Theta = \frac{r\frac{\sigma_y}{\sigma_x} - \frac{\sigma_y}{r\sigma_x}}{1+ r\frac{\sigma_y}{\sigma_x}\frac{\sigma_y}{r\sigma_x}}$$

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

If r = ± 1 , tan Θ = 0 → Θ = 0,which implies that the two regression lines coincide and hence the variables are perfectely correlated. Also if r = 0 , tan Θ = ∞ or $\Theta = \frac{\pi}{2}$. This implies that the lines are perpendicular and hence the variables are uncorrelated.

5. In a partially destroyed record,only the lines of regression of y on x and x on y are available as $4x - 5y + 33 = 0$ and $20x - 9y = 107$ respectively. Calculate $\overline{x}$, $\overline{y}$ and the coefficient of correlation between x and y.

Solu. W .K .T regression lines passes through $\bar{x}$ and $\bar{y}$

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

By solving we get $\bar{x} = 13$ $\qquad \bar{y} = 17$

We shall now rewrite the equation of the regression lines to find the regression coefficients.

$5y = 4x + 33$ or $y = 0.8\,x + 6.6$ ….(1)

$20x = 9y + 107$ or $x = 0.45y + 4.35$ …(2)

$$r = \pm\sqrt{(cofficient\ of\ x\ )(coefficient\ of\ y)}$$

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

$r = 0.6$

It is a positive correlation.

Rank Correlation and an expression for the rank correlation coefficient.

- The coefficient of correlation in respect of the ranks of some two characteristics of an individual or an observation is called Rank Correlation Coefficient usually denoted by ρ

We now proceed to derive an expression for ρ in the following form.

$$\rho = 1 - \frac{6\sum(x-y)^2}{n\,(n^2-1)} \quad \text{or} \quad 1 - \frac{6\sum d^2}{n\,(n^2-1)}$$

Note:

(1) If the ranking of x ,y are entirely in the same order like for example, x :1 1,2,3,4,5 ; y : 1,2,3,4,5 then $\sum d^2 = \sum(x-y)^2 = 0$. This will give us $\rho = \pm 1$ and is called perfect direct correlation.
If the ranking of x and y are entirely in the opposite order like for example, x : 1,2,3,4,5 y : 5,4,3,2,1 then $\sum d^2 = 40$.This will give us ρ = -1 and is called perfect inverse correlation.

Problems:

1. Ten competitors in a beauty contest are ranked by two judges in the following order. Compute the coefficient of correlation

I	1	6	5	3	10	2	4	9	7	8
II	6	4	9	8	1	2	3	10	5	7

Solu : We have $\rho = 1 - \frac{6\sum d^2}{n\,(n^2-1)}$

For the given data, n = 10 and

$\sum d^2 = (1-6)^2 + (6-4)^2 + (5-9)^2 + (3-8)^2 + (10-1)^2 + (2-2)^2 + (4-3)^2 + (9-10)^2$

$+ (7-5)^2 + (8-7)^2$

= 25 + 4+ 16 + 25 + 81 + 0 +1 +1 + 4 + 1 =158

Hence $\rho = 1 - \frac{6\,(158)}{10(10^2-1)}$ = 0.042

2. Ten students got the following percentage of marks in two subjects x and y. Compute their rank correlation coefficient.

Marks in x	78	36	98	25	75	82	90	62	65	39
Marks in y	84	51	91	60	68	62	86	58	53	47

Solu : We prepare the table consisting of the given data along with the ranks assigned according to their order of the magnitude.In the subject x, 98 will be awarded rank 1, 90 as rank 2 and so on.

Marks in x	Rank(x)	Marks in y	Rank(y)	d = (x-y)	$d^2 = (x-y)^2$
78	4	84	3	1	1
36	9	51	9	0	0
98	1	91	1	0	0
25	10	60	6	4	16
75	5	68	4	1	1
82	3	62	5	-2	4
90	2	86	2	0	0
62	7	58	7	0	0
65	6	53	8	-2	4
39	8	47	10	-2	4
					$\sum d^2 = 30$

We have $\rho = 1 - \frac{6\sum d^2}{n\,(n^2-1)}$ and n = 10 for the given data.

$= 1 - \frac{6\,(30)}{10(10^2-1)}$

$= 0.82$

3. Ten competitors in music contest are ranked by 3 judges A, B,C in the following order. Use the rank correlation coefficient to decide which pair of judges have the nearest . approach to common taste of music

A	1	6	5	10	3	2	4	9	7	8
B	3	5	8	4	7	10	2	1	6	9
C	6	4	9	8	1	2	3	10	5	7

Solu : We shall compute ρ_{AB} , ρ_{BC} , ρ_{CA} with the help of the following table where d is the difference in ranks.

A	B	C	d_{AB}^2	d_{BC}^2	d_{CA}^2
1	3	6	4	9	25
6	5	4	1	1	4
5	8	9	9	1	16
10	4	8	36	16	4
3	7	1	16	36	4
2	10	2	64	64	0
4	2	3	4	1	1
9	1	10	64	81	1
7	6	5	1	1	4
8	9	7	1	4	1
			$\sum d_{AB}^2 = 200$	$\sum d_{BC}^2 = 214$	$\sum d_{CA}^2 = 60$

We have $\rho = 1 - \frac{6\sum d^2}{n\,(n^2 - 1)}$ and n = 10 for the given data.

Now, $\rho_{AB} = 1 - \frac{6\,(200)}{10(10^2 - 1)} = -\,0.21$

$\rho_{BC} = 1 - \frac{6\,(214)}{10(10^2 - 1)} = -\,0.297$

$\rho_{CA} = 1 - \frac{6\,(60)}{10(10^2 - 1)} = +\,0.636$

It may be observed that ρ_{AB} and ρ_{BC} are negative which means their tastes (A &B; B &C) are opposite. But ρ_{CA} is positive and is nearer to 1.(perfect correlation)

CURVE FITTING [BY THE METHOD OF LEAST SQUARE]:

We can plot 'n' points (x_i, y_i) where i=0,1,2,3,.........
At the XY plane. It is difficult to draw a graph y=f(x) which passes through all these points but we can draw a graph which passes through maximum number of point. This curve is called the curve of best fit. The method of finding the curve of best fit is called the curve fitting.

FITTING A STRAIGHT LINE Y = AX + B

We have straight line that sounds as best approximate to the actual curve y=f(x) passing through 'n' points (x_i, y_i), i=0,1.2...........n equation of a straight line is
$y = a + bx$ (1)

Then for 'n' points $Y_i = a + bx_i$ (2)

Where a and b are parameters to be determined; Y_i is called the estimated value. The given value Y_i corresponding to x_i.

$$Let \quad S = \sum \ y_i - Y_i \ ^2 \qquad (3)$$

$$= \sum \left[y_i - \ a + bx_i \ \right]^2$$

$$S = \sum \ y_i - a - bx_i \ ^2 \qquad (4)$$

We determined a and b so that S is minimum (least). Two necessary conditions for this $\frac{\partial S}{\partial a} = 0$ and $\frac{\partial S}{\partial b} = 0$

differentiate (4) w.r.t a and b partially

$$\frac{\partial S}{\partial a} = \sum \ y_i - a - bx_i$$

$$0 = \sum \ y_i - a - bx_i$$

$$0 = \sum y_i - \sum a - b\sum x_i$$

$$0 = \sum y_i - na - b\sum x_i$$

$$\boxed{\sum y_i = na + b\sum x_i}$$

$$\text{or } \boxed{\sum y = na + b\sum x}$$

$$\frac{\partial S}{\partial b} = 2\sum \ y_i - a - bx_i \quad -x_i$$

$$0 = -2\sum \ x_i y_i - ax_i - bx_i^2$$

$$0 = \sum x_i y_i - a\sum x_i - b\sum x_i^2$$

$$\sum x_i y_i = a\sum x_i + b\sum x_i^2$$

or $\boxed{\sum xy = a\sum x + b\sum x^2}$

where n = number of points or value.

FITTING A SECOND DEGREE PARABOLA Y=AX² + BX + C

Let us take equation of parabola called parabola of best fit in the form

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

Where a, b, c are parameters to be determined. Let be the value of corresponding to the v

$Y_i = a + bx_i + cx_i^2$ (2)

Also

S= $\sum y_i - Y_i \;^2$ (3)

$= \sum \left[y_i - a + bx_i + cx_i^2 \right]^2$

$\boxed{S = \sum y_i - a - bx_i - cx_i^2}$ (4)

We determine a, b, c so that S is least (minimum).
The necessary condition for this are

$\frac{\partial S}{\partial a} = 0, \frac{\partial S}{\partial b} = 0$ & $\frac{\partial S}{\partial c} = 0$

diff (4) w.r.t 'a' partially

$$\frac{\partial S}{\partial a} = -2\sum y_i - a - bx_i - cx_i^2$$

$$0 = -2\sum y_i - a - bx_i - cx_i^2$$

$$0 = \sum y_i - a - bx_i - cx_i^2$$

$$0 = \sum y_i - \sum a - b\sum x_i - c\sum x_i^2$$

$$\sum y_i = \sum a + b\sum x_i + c\sum x_i^2$$

$$\boxed{\sum y_i = na + b\sum x_i + c\sum x_i^2}$$

or $\boxed{\sum y = na + b\sum x + c\sum x^2}$

diff (4) w.r.t 'b' partially

$$\frac{\partial S}{\partial b} = 2\sum y_i - a - bx_i - cx_i^2 \quad -x_i$$

$$0 = -2\sum x_i y_i - ax_i - bx_i^2 - cx_i^3$$

$$0=\sum \left(x_i y_i - a x_i - b x_i^2 - c x_i^3\right)$$

$$0=\sum x_i y_i - a\sum x_i - b\sum x_i^2 - c\sum x_i^3$$

$$\boxed{\sum x_i y_i = a\sum x_i + b\sum x_i^2 + c\sum x_i^3}$$

$$or\ \boxed{\sum xy = a\sum x + b\sum x^2 + c\sum x^3}$$

diff (4) w.r.t 'c' partially

$$\frac{\partial S}{\partial c} = 2\sum \left(y_i - a - b x_i - c x_i^2\right)\left(-x_i^2\right)$$

$$0=-2\sum \left(x_i^2 y_i - a x_i^2 - b x_i^3 - c x_i^4\right)$$

$$0=\sum \left(x_i^2 y_i - a x_i^2 - b x_i^3 - c x_i^4\right)$$

$$0=\sum x_i^2 y_i - a\sum x_i^2 - b\sum x_i^3 - c\sum x_i^4$$

$$\boxed{\sum x_i^2 y_i = a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4}$$

$$or\ \boxed{\sum x^2 y = a\sum x^2 + b\sum x^3 + c\sum x^4}$$

Hence the normal equation for second degree parabola are

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

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

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

PROBLEMS:

1) **Fit a straight line** $y = a + bx$ **to the following data**

x:	5	10	15	20	25
y:	16	19	23	26	30

Solution: Let $y = a + bx$ (1)

Normal equation

$$\sum y = na + b\sum x \qquad (2)$$

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

x	y	x^2	xy
5	16	25	80
10	19	100	190
15	23	225	345
20	26	400	520
25	30	625	750
$\sum x = 75$	$\sum y = 114$	$\sum x^2 = 1375$	$\sum xy = 1885$

(1) & (2) $\Rightarrow$

$114 = 15a + b \times 75$

$1885 = 75a + b \times 1375$

$$\boxed{\begin{array}{l} a = 12.3 \\ b = 0.7 \end{array}}$$

(1) $\Rightarrow$ $\boxed{y = 12.3 + 0.7x}$

2) Fit equation of straight line of best fit to the following data

x:	1	2	3	4	5
y:	14	13	9	5	2

Solution:

Let $y = a + bx$ (1)

Normal equation

$$\sum y = na + b\sum x \qquad (2)$$

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

x	y	x^2	xy
1	14	1	14
2	13	4	26
3	9	9	27
4	5	16	20
5	2	25	10
$\sum x = 15$	$\sum y = 43$	$\sum x^2 = 55$	$\sum xy = 97$

(2) & (3) $\Rightarrow$

$$43 = 15a + 15b$$

$$97 = 15a + 56b$$

Solving above equations we get

$$\boxed{\begin{matrix} a = 18.2 \\ b = -3.2 \end{matrix}}$$

(1) $\Rightarrow$ $\boxed{y = 18.2 - 3.2x}$

3) The equation of straight line of best fit find the equation of best fit

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

Solution: let $y = a + bx$ (1)

Normal equation

$$\sum y = na + b\sum x \qquad (2)$$

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

x	y	x^2	xy
0	1	0	0
1	1.8	1	1.8
2	3.3	4	6.6
3	4.5	9	13.5
4	6.3	16	25.2
$\sum x = 10$	$\sum y = 16.9$	$\sum x^2 = 30$	$\sum xy = 47.1$

(2) & (3) $\Rightarrow$

$$16.9 = 5a + 10b$$

$$47.1 = 10a + 30b$$

$$\boxed{\begin{matrix} a = 0.72 \\ b = 1.33 \end{matrix}}$$

(1) $\Rightarrow$ $\boxed{y = 0.72 + 1.33x}$

4) If p is the pull required to lift a load by means of pulley block. Find a linear block of the form p=MW+C Connected p &w using following data

ω:	50	70	100	120
p:	12	15	21	25

Compute p when W=150.

Solution: Given p=y & W=x

$\therefore$ equation of straight line is

:

let $y = a + bx$ (1)

Normal equations

$\sum y = na + b\sum x$ (2)

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

$x = \omega$	$p = y$	x^2	xy
50	12	2500	600
70	15	4900	1050
100	21	10000	2100
120	25	144000	3000
$\sum x = 10$	$\sum y$=16.9	$\sum x^2$=30	$\sum xy$=47.1

(2) & (3) $\Rightarrow$

$73 = 4a + 340b$

$6750 = 340a + 318000b$

$\boxed{\begin{array}{l} a = 2.27 \\ b = 0.187 \end{array}}$

(1) $\Rightarrow$ $\boxed{y = 2.27 + 0.187x}$

put $\omega = 150$

$y = 30.32$

5) Fit a curve of the form $y = ab^x$

Solution: Consider

$y = ab^x$ (1)

Take log on both side

logy=log ab^x

=loga+logb^x

logy=loga+logb^x

$y = A + Bx$ (2) ; logy=Y $\Rightarrow$ y=e^Y

Corresponding normal equation loga=A $\Rightarrow$ a=e^A

$\sum Y = nA + B\sum x$ (3) logb=B $\Rightarrow b$=e^B

$\sum xY = A\sum x + B\sum x^2$ (4)

Solving the normal equation (3) & (4) for a & b . Substitute these values in (1) we get curve of best fit of the form $y = ab^x$

6) Fit a curve of the curve $y = ab^x$ for the data

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

Solution:

let $y = ab^x$ (1)

Normal equations are

$\sum Y = nA + B\sum x$ (2); A=loga

$\sum xY = A\sum x + B\sum x^2$ (3) Y=logy, B=logb

x	y	Y=logy	x^2	xY
1	1.0	0	1	0
2	1.2	0.182	4	0.364
3	1.8	0.587	9	1.761
4	2.5	0.916	16	3.664
5	3.6	1.280	25	6.4
6	4.7	1.547	36	9.282
7	6.6	1.887	49	13.209
8	9.1	2.208	64	17.664
$\sum x = 36$		$\sum Y = 8.607$	$\sum x^2 = 204$	$\sum xY = 52.34$

(1) & (2) $\Rightarrow$

$8.607 = 8A + 36B$

$52.34 = 36A + 203B$

$$\boxed{\begin{aligned} A &= -0.382 \\ B &= 0.324 \end{aligned}}$$

then $e^A = a = 0.682$

$e^B = b = 1.382$

(1) $\Rightarrow$ $\boxed{y = 0.682(1.382)^x}$

7) Fit a curve of the form $y = ax^b$ for the following data

x:	1	1.5	2	2.5
y:	2.5	5.61	10.0	15.6

Solution:: Consider

$let\ y = ax^b$ (1)

Normal equations are

$\sum Y = nA + b\sum x$ (2); Y=logy

$\sum XY = A\sum X + B\sum X^2$ (3) A=loga, X=logx

x	y	$X = \log x$	X^2	$Y = \log y$	XY
1	2.5	0	0	0.916	0
1.5	5.62	0.405	0.164	1.726	0.699
2	10.0	0.693	0.480	2.302	1.595
2.5	15.6	0.916	0.839	2.747	2.516
		$\sum$X=2.014	$\sum X^2$=1.483	$\sum$Y=7.691	$\sum$XY=4.81

(1) & (2) $\Rightarrow$

$7.691 = 4A + 2.014b$

$4.81 = 2.014A + 1.483b$

A=0.916, $e^A = a = 2.499 \approx 2.5$

$b = 1.999 \approx 2$

(1) $\Rightarrow$ $\boxed{y = 2.5x^2}$

8) Fit a parabola $y = a + bx + cx^2$ **for the following data**

x: 1 2 3 4

y: 1.7 1.8 2.3 3.2

Sol:

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

Normal equation

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

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

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

x	y	x^2	x^3	x^4	xy	x^2y
1	1.7	1	1	1	1.7	1.7
2	1.8	4	8	16	3.6	7.2
3	2.3	9	27	81	6.9	20.7
4	3.2	16	64	256	12.8	51.2
$\sum x$	$\sum y = 9$	$\sum x^2 = 30$	$\sum x^3 = 100$	$\sum x^4 = 354$	$\sum xy = 25$	$\sum x^2 y = 80.8$

(1),(3) & (4)

$9 = 4a + 10b + 30c$

$25 = 10a + 30b + 100c$

$80.8 = 30a + 100b + 354c$

$a = 2$

$b = -0.5$

$c = 0.2$

(1) $\Rightarrow$ $y = 2 - 0.5x + 0.2x^2$

9) Fit a curve of the form $y = ae^{bx}$ for the following

x: 0 2 4

y: 8.12 10 31.82

Sol:

$y = ae^{bx}$ (1)

Normal equation

$\sum$ Y=nA+b$\sum x$ (2); Y=logy

$\sum xy = A\sum x + b\sum x^2$ (3) A=loga

x	y	Y=logy	x^2	xY
0	8.12	2.094	0	0
2	10	2.302	4	4.604
4	31.82	3..46	16	13.84
$\sum x = 6$		$\sum$Y=7.86	$\sum x^2 = 20$	$\sum xY = 18.444$

(2) & (3) $\Rightarrow$

7.856=3A+6b

18.444=6A+20b

A=1.935, $a = e^A = 6.924$

b=0.341

(1) $\Rightarrow$ $y = 6.924e^{0.341x}$

10) Fit a II degree parabola ax^2+bx+c to the least square method & hence find y when x=6

x: 1 2 3 4 5

y: 10 12 13 16 19

Sol:

$y = ax^2 + bx + c \qquad (1)$

$y = c + bx + ax^2 \qquad (1)$

Normal equation

$\sum y = nc + b\sum x + a\sum x^2 \qquad (2)$

$\sum xy = c\sum x + b\sum x^2 + a\sum x^3 \qquad (3)$

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

x	y	x^2	x^3	x^4	xy	x^2y
1	10	1	1	1	10	10
2	12	4	8	16	24	14
3	13	9	27	81	39	117
4	16	16	64	256	64	256
5	19	25	125	625	95	475
=15	=70	=55	=225	=979	=232	=906

$70 = 5c + 15b + 55a$

$232 = 15c + 55b + 225a$

$906 = 55c + 225b + 979a$

$a = 0.285$

$b = 0.485$

$c = 9.4$

$(1) \Rightarrow y = 0.285x^2 + 0.485x + 9.4$

at $x = 6, \quad y = 22.6$