MODULE 3 : STATISTICAL METHODS AND CURVE FITTING (CO_3)

* Curve fitting :-

i) Fitting of a straight line [$y = ax+b$] :-

In the above form 'a' and 'b' are the constants to be determined by the method of least squares.

The normal equations are : $y = ax+b$ —①

$\Sigma y = a\Sigma x + nb$ —②

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

solving these 2 equations, we, get, a & b then substituting in eqn ① which is the required fitting of straight line.

Problems :

Q:- Fit a straight line $y = ax+b$ for the following

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

Sol :- Let straight line fit be eqn ① where a and b are the constants to be determined by the method of least squares.

The normal equations of ① are :-

$y = ax+b$ —①

$\Sigma y = a\Sigma x + nb$ —②

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

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
56	40	364	524

substitute these values in eqn ①, ② & ③

$$40 = 56a + 8b$$

$$364 = 524a + 56b.$$

$$a = 0.64 \quad \& \quad b = 0.55$$

∴ The required fitting of straight line is $y = (0.64)x + 0.55$

Q: Find the equation of best fitting straight line for the following data and hence estimate the value of dependent variable corresponding to the value 30 of the independent variable.

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

Sol:-

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

$y = ax + b$ —— (1)

$\Sigma y = a\Sigma x + nb$ —— (2)

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

Substitute all the values in eqn (2) & (3).

$114 = 75a + 5b$

$1885 = 1375a + 75b$ $\quad a = 0.7, \quad b = 12.3$

$y = 0.7x + 12.3$

$y = (0.7)(30) + 12.3$

$y = 33.3$

Q: A simply supported beam carries a concentrated load P at its midpoint corresponding to various val

P	100	120	140	160	180	200
Y	0.45	0.55	0.60	0.70	0.80	0.85

find the law of the form $y = a + bP$ and hence estimate 'y' when 'P' is 150

Sol:- Let the straight line fit be $Y = a + bP$ — (1)

$\therefore \sum Y = na + b\sum P$ — (2)

$\sum PY = a\sum P + b\sum P^2$. — (3)

P	Y	PY	P^2
100	0.45	45	10000
120	0.55	66	14400
140	0.6	84	19600
160	0.7	112	25600
180	0.80	144	32400
200	0.85	170	40000
900	3.95	621	14200

$3.95 = 6a + 900b$

$621 = 900a + 14200b$

$a = 4.07 + 14 \times 10^{-3}$, $b = 0.0041$

$Y = 0.0476 + 0.0041P$; $y(P) = 0.0476 + 0.0041P \Rightarrow y(150) = 0.6626$

Q:- Fit a straight line in at least square sense for the following data:

x	50	70	100	120
y	12	15	21	25

Sol:- $y = ax + b$ —(1)

$\Sigma y = a\Sigma x + nb$ —(2)

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

X	Y	XY	X^2
50	12	600	2500
70	15	1050	4900
100	21	2100	10000
120	25	3000	14400
340	73	6750	31800

$73 = 340a + 4b$

$6750 = 31800a + 340b.$

$a = 0.19$; $b = 2.28$

$y = 0.19x + 2.28$

$y = 0.19\ (50) + 2.28$

$y = 11.78$

$y \approx 12.$

* Fitting of a second degree polynomial or a parabola of the form $y = ax^2 + bx + c$

Q:- Fit a second degree parabola $y = ax^2 + bx + c$ in the least square sense for the following data and estimate y at $x = 6$.

Sol:- Let parabola fit be $y = ax^2 + bx + c$ — (1) where a, b, c are constants to be determined by the method of least squares. Hence the normal eqn (1) are:

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

x	y	xy	x^2	x^2y	x^3	x^4
1	10	10	1	10	1	1
2	12	24	4	48	8	16
3	13	39	9	117	27	81
4	16	64	16	256	64	256
5	19	95	25	475	125	625
15	70	232	55	906	225	979.

$$70 = 55a + 15b + 5c$$
$$232 = 225a + 55b + 15c$$
$$906 = 979a + 225b + 55c$$
$$a = 0.2857, \quad b = 0.4857, \quad c = 9.4.$$

$$y = 0.2857x^2 + 0.4857x + 9.4$$
$$y = 10.44 + 2.94 + 9.4.$$
$$y = 22.599$$

Q:- Fit a parabola $y = a+bx+cx^2$ for the following data:

x	0	1	2	3	4
y	1	1.8	1.3	2.5	2.3

Sol:- Let the parabola fit be $y=a+bx+cx^2$ —① where a, b, c are constants to be determined by the method of least squares.

Normal equations are given by :-

$\Sigma y = na + b\Sigma x + c\Sigma x^2$ —②

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

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

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	2.5	9	27	81	7.5	22.5
4	2.3	16	64	256	9.2	36.8
10	8.9	30	100	354	21.1	66.3.

$n=5$.

~~$\Sigma y = na + b\Sigma x + c\Sigma x^2$~~ —②

$8.9 = 5a + 10b + 30c$

$21.1 = 10a + 30b + 100c$

$66.3 = 30a + 100b + 354c$.

$a = 1.0771$, $b = 0.4157$, $c = -0.0214$

~~eqn ① $\Rightarrow y = 1.0771 + (0.4157)x + (-0.0214)x^2$~~

Q:- Fit a parabola $y = a + bx + cx^2$ for the following data:

x	-2	-1	0	1	2
y	-3.150	-1.390	0.620	2.880	5.378

Sol:- Let the parabola fit be $a + bx + cx^2$ —① where a, b, c are constants to be determined by the method of least squares.

Normal equations are given by:

$$\Sigma y = na + b\Sigma x + c\Sigma x^2 \quad —②$$

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

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

x	y	x^2	x^3	x^4	xy	x^2y
-2	-3.150	4	-8	16	6.3	-12.6
-1	-1.390	1	-1	1	1.390	-1.390
0	0.620	0	0	00	0	0
1	2.880	1	1	1	2.88	2.88
2	5.378	4	8	16	10.756	21.512
0	4.338	10	0	34	21.326	10.402

$n = 5$.

$$4.338 = 5a + 0 + 10c$$

$$21.326 = 0 + 10b + 0c$$

$$10.402 = 10a + 0b + 34c.$$

$a = 0.621$, $b = 2.1326$, $c = 0.1233$

① $\rightarrow$ $y = 0.621 + 2.1326x + 0.1233x^2$

Fitting of a curve of the form $y = ax^b$

Consider $y = ax^b \rightarrow (1)$; Taking logarithms (to the base e) on both sides we get,

$$\log_e y = \log_e a + b \log_e x \iff Y = A + BX \rightarrow (2)$$

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

The normal equations are ;

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

But $\log_e a = A \iff a = e^A$ also $B = b$.

Solving these equations we get a & b then substitute in equation (1). which is the required fit.

Examples :-

1. Fit a least square geometric curve $y = ax^b$ for the following data ⇒

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

Sol Consider $y = ax^b \rightarrow (1)$; Taking $\log_e$ on both sides

$\log_e y = \log_e a + b \log_e x$ and let $Y = \log_e y$; $A = \log_e a$; $X = \log_e x$ and take $b = B$ then,

$Y = A + BX \rightarrow (2)$; The normal equations are,

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

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

x	y	$X = \log_e x$	$Y = \log_e y$	XY	X^2
1	0.5	0	−0.6931	0	0
2	2	0.6931	0.6931	0.4804	0.4804
3	4.5	1.0986	1.5041	1.6524	1.2069
4	8	1.3863	2.0794	2.8827	1.9218
5	12.5	1.6094	2.5257	4.0649	2.5902
		$\Sigma X = 4.7874$	$\Sigma Y = 6.1092$	$\Sigma XY = 9.0804$	$\Sigma X^2 = 6.1993$

Substituting in the normal equations,

$$5A + 4.7874B = 6.1092$$

$$4.7874A + 6.1993B = 9.0804$$

Solving these equations, $A = -0.69315$, $B = b = 2$

$\therefore \log_e a = A = -0.69315 \iff a = e^{-0.69315} = 0.5$

$\therefore$ The required fit is, $y = 0.5x^2$.

---

2. Fit a curve of the form $y = ax^b$ for the following data

x :	0	2	4
y :	8.12	10	31.82

Sol. $y = ax^b \to (1) \iff \log_e y = \log_e a + b \log_e x$

$y = A + BX \to (2)$; $Y = \log_e y$; $A = \log_e a$

The normal equations are, $B = b$ & $X = \log_e x$

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

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

x	y	$X = \log_e x$	$Y = \log_e y$	X^2	XY
0	8.12	0	2.0943	0	0
2	10	0.69315	2.3026	0.4805	1.5961
4	31.82	1.38629	3.4601	1.9278	4.7967
		2.0794	7.857	2.4023	6.3928

$$7.857 = 3A + B(2.0794)$$

$$6.3928 = 2.0794A + B(2.4023)$$

$A = 1.9361$, $B = 0.98528 = b$

Here $A = \log_e a = 1.9361 \Leftrightarrow a = e^A = e^{1.9361} = 6.9317$

The required fit is $y = (6.9317)x^{(0.98528)}$

3. Fit a curve of the form $y = ax^b$ by the method of least squares for the following data.

No of Petals :	5	6	7	8	9	10
No of flowers :	133	55	23	7	2	2

Sol Let $y = ax^b \longrightarrow (1)$

$$\log_e y = \log_e a + b \log_e x \Leftrightarrow Y = A + BX \longrightarrow (2)$$

where, $Y = \log_e y$; $A = \log_e a$; $b = B$

$$X = \log_e x$$

The normal equations are,

$$\Sigma y = nA + B\Sigma X \quad \& \quad \Sigma XY = A\Sigma X + B\Sigma X^2$$

x	y	$X = \log_e^x$	$Y = \log_e^y$	XY	x^2
5	133	1.6094	4.8903	7.8704	2.5902
6	55	1.7918	4.0073	7.1803	3.2105
7	23	1.9459	3.1355	6.1014	3.7865
8	7	2.07944	1.9459	4.0464	4.3241
9	2	2.1972	0.6931	1.5228	4.8277
10	2	2.3026	0.6931	1.5959	5.302
		11.9263	15.3652	28.3172	24.041

$$6A + 11.9263B = 15.3652$$

$$11.9263A + 24.041B = 28.3172$$

$$A = 15.7638 \quad , \quad B = -6.6423 = b$$

$$A = \log_e^a = 15.7638 \iff a = e^A = e^{15.7638}$$

$$\therefore a = 7016674.88 \quad ; \quad b = -6.6423$$

∴ The required fit is,

$$y = 7016674.88\, x^{-6.6423}$$

=

4. Fit a curve of the form $y = ax^b$ for the following data:

x	1	2	3	4	5	6
y	2.98	4.26	5.21	6.1	6.8	7.5

Sol Let the curve be $y = ax^b \rightarrow (1)$

$$\log_e y = \log_e a + b \log_e x \Leftrightarrow Y = A + BX \rightarrow (2)$$

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

x	y	$X = \log_e x$	$Y = \log_e y$	XY	X^2
1	2.98	0	1.0919	0	0
2	4.26	0.6931	1.4493	1.0045	0.4804
3	5.21	1.0986	1.6506	1.8133	1.2069
4	6.1	1.3863	1.8083	2.5068	1.9218
5	6.8	1.6094	1.9169	3.0851	2.5902
6	7.5	1.7918	2.0149	3.6103	3.2105
		6.5792	9.9319	12.02	9.4098

The normal equations are,

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

$$\left.\begin{aligned} 6A + 6.5792B &= 9.9319 \\ 6.5792A + 9.4098B &= 12.02 \end{aligned}\right\} \Rightarrow \begin{aligned} &A = 1.0913 \Leftrightarrow a = e^A \\ &\therefore a = 2.978 \\ &B = b = 0.5144 \end{aligned}$$

$\therefore y = 2.978\, x^{0.5144}$ is the required fit.

Assignments (Practice Problems) :-

1. Fit a St. line to the following data.

year :	1961	1971	1981	1991	2001
Production in tons :	8	10	12	10	16

Also find the expected production in the year 2006.

Hint: Let $X = x - 1981 \Leftrightarrow y = a + bX$; $a = 11.2$, $b = 0.16$

at $x = 2006$, $y = -305.76 + 0.16x = 15.2$.

---

2. Fit a st. line to the following data.

year :	1911	1921	1931	1941	1951
Production in thousand tons :	8	10	12	10	6

Hint: Let $X = x - 1931$; $a = 9.2$, $b = -0.04$

$\therefore y = 86.44 - 0.04x$ is the required fit.

---

3. Fit a parabola $y = ax^2 + bx + c$ by the method of least squares.

a)

x :	2	4	6	8	10
y :	3.07	12.85	31.47	57.38	91.29

$a = 0.99196$
$b = -0.85507$
$c = 0.696$

b)

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

Ans:- $a = -0.92857$, $b = 3.52316$; $c = -0.26732$

4. An experiment on lifetime 't' of cutting tool at different cutting speeds 'V' (units) are given below to fit a relation of the form; $V = at^b$

speed (V) :	350	400	500	600
life (t) :	61	26	7	2.6

$A = 6.5539$
$a = e^A = 701.9766$
$B = b = -0.1709$

5. Fit a st. line for the following data.

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

Ans: $y = 3.23x + 11.096$

6. Fit a st. line for the following data:

x:	62	64	65	69	70	71	72
y:	65.7	66.8	67.2	69.3	69.8	70.5	70.9

Ans: $y = 0.52x + 33.46$.

7. Fit a st. line for the following data:

x:	1	2	3	4	5	6	7
y:	80	90	92	83	94	99	92

Ans: $y = 2x + 82$.

8. Fit a second degree polynomial or a parabola for the following data.

x:	0	1	2	3	4	5	6
y:	14	18	23	29	36	40	46

Ans: $y = 0.083x^2 + 4.96x + 13.46$.

9. Fit a parabola or second degree polynomial for the following data :

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

Ans : $y = 2.23x^2 + 0.18x + 1.46$

10. Fit a parabola or second degree polynomial for the following data :

x :	1	2	3	4	5
y :	25	28	33	39	46

Ans : $y = 0.64x^2 + 1.46x + 2.78$

Correlation and Regression :-

Correlation : Covariation of two independent magnitudes is known as correlation. If two variables x & y are related in such a way that increase or decrease in one of them corresponds to increase or decrease in other. Then we say that the variables are positively correlated. On the other hand, if increase or decrease in one of them corresponds to decrease or increase in the other hand. Then we say that the variables are negatively correlated.

Formulae:

The numerical measure of 2 variables x and y is known as Pearson's co-efficient or co-efficient of correlation and is denoted by 'r' and is given by:

$$r = \sum_{n=1}^{n} \frac{(x-\bar{x})(y-\bar{y})}{n\,\sigma_x\,\sigma_y}.$$

but we consider an alternative formula for the above as:

$$r = \frac{\overline{x}^2 + \overline{y}^2 - \overline{x-y}^2}{2\,\sigma_x\,\sigma_y}.$$

The range of r lies between $(-1, 1)$

Regression:-

It is an estimation of 1 independent variable in terms of other.

If x and y are correlated, the best fitting of a straight line by the method of least squares give the good relation between x and y.

The best fitting of a straight line is of the form $y = ax+b$ (x being independent variable, y is dependent) is called the regression line of y on x, which is given by

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

Let the straight line fit be $x = ay+b$ (where y is independent variable, x is dependent variable) is called regression line of x on y, which is given by:

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

There two are known as regression lines for the above forms of straight line, which can also be expressed in the form:

$$Y = \frac{\sum XY}{\sum X^2} \cdot (X) \quad \text{where} \quad X = x - \bar{x} \quad \text{and} \quad Y = y - \bar{y}$$

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

After finding regression lines the co-efficient of x and y from $y = ax + b$ and $x = ay + b$ is considered as:

$$r = \pm \sqrt{(\text{co-eff of } x)(\text{co-eff of } y)}$$

If both the co-efficients are positive then 'r' is positively correlated and if both the co-efficients are negative, 'r' is negatively correlated.

Problems:-

Q: Compute the co-efficient of correlation and the equation of lines of regression for the following data:

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

Sol:- WKT the correlation co-efficient or Carl Pearson's co-efficient is given by:

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

$$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_z^2}{2\,\sigma_x\,\sigma_y}$$

We have $\sigma_x^2 = \frac{\sum x^2}{n} - (\bar{x})^2$ where $\bar{x} = \frac{\sum x}{n}$.

$\sigma_y^2 = \frac{\sum y^2}{n} - (\bar{y})^2$ where $\bar{y} = \frac{\sum y}{n}$.

$\sigma_z^2 = \frac{\sum z^2}{n} - (\bar{z})^2$ where $\bar{z} = \frac{\sum z}{n}$.

$n = 7$.

x	y	$z = x - y$	x^2	y^2	z^2
1	9	−8	1	81	64
2	8	−6	4	64	36
3	10	−7	9	100	49
4	12	−8	16	144	64
5	11	−6	25	121	36
6	13	−7	36	169	49
7	14	−7	49	196	49
28	77	−49	140	875	347.

$\sigma_x^2 = \frac{\sum x^2}{n} - (\bar{x}) = \frac{140}{7} - \left(\frac{28}{7}\right)^2 = 20 - 16 = 4.$

$\sigma_x = \sqrt{4} = 2$

$\sigma_y^2 = \frac{\sum y^2}{n} - (\bar{y})^2 = \frac{875}{7} - \left(\frac{77}{7}\right)^2 = 125 - 121 = 4.$

$\sigma_y = \sqrt{4} = 2.$

$\sigma_z^2 = \frac{\sum z^2}{n} - (\bar{z})^2 = \frac{347}{7} - \left(\frac{-49}{7}\right)^2 = \frac{347}{7} - 49 = 0.57$

Q: Apply Lagrange's interpolation formula find $f(9)$ for.

$$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_z^2}{2\sigma_x\sigma_y}.$$

$$r = \frac{4+4-0.57}{2\times 2\times 2} = \frac{0.93}{}$$

WKT that the regression lines of y on x and x on y:-

$$(y-\bar{y}) = r\frac{\sigma_y}{\sigma_x}(x-\bar{x}) \Rightarrow y-11 = (0.93)\cdot\frac{2}{2}(x-4)$$

$$\Rightarrow y-11 = 0.93x - 3.72$$

$$\Rightarrow y = 0.93x - 7.28$$

$$(x-\bar{x}) = r\frac{\sigma_x}{\sigma_y}(y-\bar{y}) \Rightarrow x-4 = (0.93)\cdot\frac{2}{2}(y-11)$$

$$\Rightarrow x-4 = (0.93)y - 10.23$$

$$\Rightarrow x = 0.93y - 6.23.$$

These two lines are known as regression lines.

Q:- Obtain the lines of regression and hence find the correlation co-efficient for the following data

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

Sol:- WKT the regression lines of y on x and x on y are:

$$Y = \frac{\sum(XY)}{\sum X^2}\cdot X \quad ; \quad X = \frac{\sum XY}{\sum Y^2}\cdot Y$$

where ~~$Y = \frac{\sum XY}{\sum X^2}\cdot X$~~ are $X = x-\bar{x}$ and $Y = y-\bar{y}$ and $n=7$

x	y	X	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	0
5	11	1	0	0	1	1
6	13	2	2	4	4	4
7	14	3	3	9	9	9
28	77			26	28	28

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

$Y = \frac{\sum XY}{\sum X^2} \cdot X \Rightarrow y - \bar{y} = \frac{26}{28}(x - \bar{x}) \Rightarrow y - 11 = \frac{26}{28}(x-4)$

$\Rightarrow y - 11 = 0.9286x - 3.7$

$\Rightarrow y = 0.93x + 7.3$

$X = \frac{\sum XY}{\sum Y^2} \cdot Y \Rightarrow x - \bar{x} = \frac{26}{28}(y - \bar{y}) \Rightarrow x - 4 = 0.93y - 10.23$

$\Rightarrow x = 0.93y - 6.23$

∴ These two are the lines of regressions of y on x and x on y. wrt the straight line fitting $y = ax + b$ and $x = ay + b$.

From the above regression lines WKT, the co-efficient of x is 0.93 and co-efficient of y is 0.93.

$\therefore r = \pm \sqrt{(\text{co-eff of } x)(\text{coeff of } y)}$

$r = \pm \sqrt{(0.93)(0.93)}$

$r = \underline{+0.93}$. (∵ both co-efficients are positive, x & y positively correlated).

Q:- Find the correlation co-efficients for the 2 groups A and B.

A(x) :	92	89	87	86	83	77	71	63	53	50
B(y) :	86	83	91	77	68	85	52	82	37	57

Solution:- The correlation co-efficient is given by:

$$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x \sigma_y}, \quad n = 10.$$

A	B	$Z = x - y$.	x^2	y^2	z^2
92	86	6	8464	7396	36
89	83	6	7921	6889	36
87	91	-4	7569	8287	16
86	77	9	7396	5929	81
83	68	15	6889	4624	225
77	85	-8	5929	7225	64
71	52	19	5041	2704	361
63	82	-19	3969	6724	361
53	37	16	2809	1369	256
50	57	-7	2500	3249	49
751	718	33	58487	54390	1485

$\sigma_x^2 = \frac{\sum x^2}{n} - (\bar{x})^2$ $\qquad \bar{x} = \frac{\sum x}{n} = \frac{751}{10} = 75.1$

$= \frac{54487}{10} - \left(\frac{751}{10}\right)^2 = 5448.7 - 5640.01 = 208.69$

$\sigma_y^2 = \frac{\sum y^2}{n} - (\bar{y})^2$ $\qquad \bar{y} = \frac{\sum y}{n} = \frac{718}{10} = 71.8$

$= \frac{54390}{10} - (71.8)^2 = 5439 - 5155.24 = 283.76$

$\sigma_z^2 = \frac{\sum z^2}{n} - (\bar{z})^2$ $\qquad \bar{z} = \frac{\sum z}{n} = \frac{33}{10} = 3.3$

$= \frac{1485}{10} - (3.3)^2 = 148.5 - 10.89 = 137.61$

$$r = \frac{208.69 + 283.76 - 137.61}{2 \times 14.4461 \times 16.8451} = 0.729$$

∴ The correlation co-efficient is $r = 0.729$.

Q: Find the correlation coefficient and the equation of lines of regression for the following data:

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

Sol:- The correlation co-efficient is given by:

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

x	y	$z = x-y$	x^2	y^2	z^2
1	2	−1	1	4	1
2	5	−3	4	25	9
3	3	0	9	9	0
4	8	−4	16	64	16
5	7	−2	25	49	4
15	25	−10	55	151	30

$$\sigma_x^2 = \frac{\Sigma x^2}{n} - (\bar{x})^2 = \frac{55}{5} - \left(\frac{15}{5}\right)^2 = 11 - 9 = 2.$$

$$\sigma_y^2 = \frac{\Sigma y^2}{n} - (\bar{y})^2 = \frac{151}{5} - \left(\frac{25}{5}\right)^2 = 30.2 - 25 = 5.2$$

$$\sigma_z^2 = \frac{\Sigma z^2}{n} - (\bar{z})^2 = \frac{30}{5} - \left(\frac{-10}{5}\right)^2 = 6 - 4 = 2$$

$$\sigma_x = \sqrt{2} = 1.414 \quad ; \quad \sigma_y = \sqrt{5.2} = 2.280$$

$$r = \frac{2 + 5.2 - 2}{2 \times 1.414 \times 2.280} = 0.81$$

∴ The co-efficient of correlation is $r = 0.81$

The equation of regretion lines are given by:

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

$y-5 = 0.81 \times \frac{2.28}{1.414}(x-3)$; $(x-3) = 0.81 \times \frac{1.414}{2.28}(y-5)$

$\Rightarrow y-5 = 1.31x - 3.92$; $x-3 = 0.5023y - 2.512$

$\Rightarrow y = 1.31x - 1.08$; $x = 0.5023y + 0.49$

∴ These two are the lines of regression.

Q:- Obtain the regression lines and hence find the correlation co-efficient 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.

Sol:- The regression lines can be considered in the form:

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

where $X = x - \bar{x}$, $Y = y - \bar{y}$.

$\bar{x} = \frac{\sum x}{n}$

x	y	X	Y	XY	X^2	Y^2
1	8	-6	-7	42	36	63
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	35	4	9
8	16	1	1	6	1	1
9	16	2	1	1	4	1
10	10	3	-15	2	9	25
13	32	6	17	-15	36	289
15	32	8	17	102 136	64	289
70	150			360	204	818

The regression line can be considered in the form as:

$Y = \frac{\Sigma XY}{\Sigma X^2} \cdot X$; $X = \frac{\Sigma XY}{\Sigma Y^2} \cdot Y$.

where $X = x - \bar{x}$, $Y = y - \bar{y}$.

$\bar{x} = \frac{\Sigma x}{n} = \frac{70}{10} = 7$.

$\bar{y} = \frac{\Sigma y}{n} = \frac{150}{10} = 15$.

$y - 15 = \frac{360}{204}(x - 7)$; $(x - 7) = \frac{360}{818}(y - 15)$

$y - 15 = 1.765x - 12.353$; $x - 7 = 0.44y - 6.6$

$y = 1.765x + 2.68$; $x = 0.44y + 0.4$

∴ These 2 lines are the regression lines.

From the above expression WKT, the co-efficient of x is 1.76 and y is 0.44.

∴ $r = \pm \sqrt{(1.76)(0.44)}$

$r = \pm 0.88$

$r = +0.88$ (∵ both co-eff are positive, r is positively correlated).

Q:- $8x - 10y + 66 = 0$, $40x - 18y = 214$ are two regression lines. Find the mean of x and y and the correlation co-eff. Find $\bar{y}$ if $\bar{x}$ is 3.

Sol:- WKT the regression lines passes through the mean of x and y and the correlation co-efficient as $(\bar{x}, \bar{y})$

Hence the above lines can be written as:

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

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

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

$10y = 8x + 66$; $40x = 18y + 214$

$y = 0.8x + 6.6$ (y on x) ; $x = 0.45y + 5.35$ (x on y).

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

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

$$r = +0.6$$ (∵ both the co-efficients are positive, r is positively correlated).

For y on x:

$$r \frac{\sigma_y}{\sigma_x} = 0.8 \qquad (\text{at } \bar{x} = 3).$$

$$0.6 \frac{\sigma_y}{3} = 0.8$$

$$\sigma_y = \frac{0.8 \times 3}{0.6} = 4.$$

$\therefore \sigma_y = 4$

Q:- Compute $\bar{x}$, $\bar{y}$ and r from the following eqn of regression lines $2x+3y+1=0$ and $x+6y-4=0$.

Sol:- WKT the regression lines pass through the mean of x and y as $(\bar{x}, \bar{y})$.

Hence the above lines can be written as:

$$2\bar{x}+3\bar{y}+1=0$$
$$\bar{x}+6\bar{y}-4=0.$$
$$\bar{x}=-2, \quad \bar{y}=1$$

$3\bar{y} = -2\bar{x} - 1$ (y on x) ; $x = 4 - 6y$

$y = -0.667x - 0.33$

$$r = \pm \sqrt{(\text{coeff of } x)(\text{co-eff of } y)}$$
$$r = \pm \sqrt{(-0.667)(-6)}$$

$r = -2$. (∵ both the co-efficients are negative, r is negatively correlated).

$r = 2 \neq (-1, 1)$

Hence 1st eqn cannot be considered as y on x and 2nd eqn as x on y. We take the alternative ways as:-

1st eqn : x on y

2nd eqn : y on x

y on x :-

$6y = -x + 4$

$y = \left(\frac{-1}{6}\right)x + \frac{4}{6}$

x on y :-

$2x = -3y - 1$

$x = \left(\frac{-3}{2}\right)y - \frac{1}{2}$

$r = \pm \sqrt{\left(\frac{-1}{6}\right)\left(\frac{-3}{2}\right)} = -0.5$. (∵ the co-efficients are negative, r is negatively correlated)

∴ $r = -0.5$

Q:

	x series	y series
mean	18	100
SD	14	20

, $r = 0.8$

Write down the equation of lines of regression and hence find the most probable value of y at $x = 70$.

Sol:- Given, x series mean $= \bar{x} = 18$

y series mean $= \bar{y} = 100$

SD of x, $\sigma_x = 14$

SD of y, $\sigma_y = 20$

also $r = 0.8$.

WKT, the regression lines of y on x and x on y is given by:

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

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

$$y - 100 = 1.143x - 20.57 \quad ; \quad (x - 18) = 0.56y - 56$$

$$y = 1.143x + 79.43 \quad ; \quad x = 0.56y - 38$$

$$y = (1.143)(70) + 79.43 \quad \text{(at } x = 70\text{)}$$

$$y = 159.23$$

Q: If θ is the angle between the lines of regression then ST $\tan\theta = \frac{\sigma_x \cdot \sigma_y}{\sigma_x^2 + \sigma_y^2}\left[\frac{1-r^2}{r}\right]$

Sol:- We have, if θ is acute then the angle between the lines : $y = m_1x + c_1$; $y = m_2x + c_2$ is given by:

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

WKT the regression lines of y on x and x on y are:

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

$$m_1 = r\frac{\sigma_y}{\sigma_x}.$$

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

$$m_2 = \frac{\sigma_y}{r\sigma_x}.$$

substitute m_1, m_2 in $\tan\theta$, we gd,

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

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

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

$$\tan\theta = \frac{\sigma_y \cdot \sigma_x}{\sigma_x^2 + \sigma_y^2}\left[\frac{1-\mu^2}{\mu}\right]$$

$$\tan\theta = \frac{\sigma_x \cdot \sigma_y}{\sigma_x^2 + \sigma_y^2}\left[\frac{1-\mu^2}{\mu}\right]$$

Hence the proof

Assignment (Practice problems) :-

1. Find the correlation coefficient between x and y for the following data. Also obtain the regression lines.

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

Ans: $r = 0.96$

Regression lines are, $y = 4.686x + 4.927$

$x = 0.197y - 0.548$

2. The following data gives the age of husband(x) and the age of wife (y) in years. Form the two regression lines and calculate the age of husband corresponding to 16 years age of wife.

x :	36	23	27	28	28	29	30	31	33	35
y :	29	18	20	22	27	21	29	27	29	28

Ans: $r = 0.82$.

The regression lines are, $y = 0.894x - 1.82$

$x = 0.752y + 11.2$

when $y = 16$, $x = 23$.

Thus husband's age is 23 years corresponding to wife's age of 16 years.

Rank correlation coefficient :-

The coefficient of correlation in respect of the ranks of some two characteristics of an individual (or) an observation is called the rank correlation coefficient usually denoted by ρ.

$$\therefore \rho = 1 - \frac{6\Sigma d^2}{n(n^2-1)} \quad ; \quad d = R_x - R_y \text{ on } x-y \text{ if ranks given.}$$

Note :- The value of R lies b/n ± 1.

i) If $R = +1$, there is a complete agreement in the order of ranks and move in the same direction (perfect direct correlation)

ii) If $R = -1$, there is a complete agreement in the order of ranks but are in opposite directions. (perfect inverse correlation)

iii) $R = 0$, there is no association in the ranks.

Result : If ρ in case of repeated ranks :

If two or more magnitudes are repeated then there will be a 'tie' for a particular rank. In such cases we assign the average rank to all those magnitudes and use the correction factor $\frac{m(m^2-1)}{12}$ along with Σd^2 in the formula for ρ where m denotes the number of times a magnitude is repeated. The correction factor must be added every time the 'tie' occurs for a particular rank.

$$\rho = 1 - \frac{6\left[\Sigma d^2 + \frac{m(m^2-1)}{12} + \dots\right]}{n(n^2-1)}$$

Examples :-

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

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

Sol. we have $\rho = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$

For the given data, $n=10$ and we have $d = x-y$

$$\Sigma d^2 = (1-6)^2 + (6-4)^2 + (5-9)^2 + (3-8)^2 + (10-1)^2 + (2-2)^2 + (4-3)^2 + (9-10)^2 + (7-5)^2 + (8-7)^2$$

$$\Sigma d^2 = 25 + 4 + 16 + 25 + 81 + 0 + 1 + 1 + 4 + 1 = 158$$

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

$$= 1 - \frac{948}{990} = 0.042$$

$\therefore \rho = 0.042$

2. 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

Sol 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^2_{AB}	d^2_{BC}	d^2_{CA}
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
			$\Sigma d^2_{AB} = 200$	$\Sigma d^2_{BC} = 214$	$\Sigma d^2_{CA} = 60$

we have $\rho = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$ and $n=10$ for the given data

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

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

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

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) Thus we conclude that the judges C and A have the nearest approach to common taste of music.

3. 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

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

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

we have $\rho = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$; $n = 10$

$= 1 - \frac{6(30)}{10(10^2-1)} = 0.81818 \simeq 0.82$

$\therefore \rho = 0.82$

=

4. The coefficient of rank correlation obtained by ten students in statistics and Accountancy was 0.2. It was later discovered that the difference in ranks in the two subjects of one of the students was wrongly taken as 9 instead of 7. Find the correct rank correlation coefficient.

Sol) we have $\rho = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$

Here $n=10$ and $\rho=0.2$.

Hence, $0.2 = 1 - \frac{6\Sigma d^2}{10(10^2-1)}$ (or) $\frac{6\Sigma d^2}{990} = 1-0.2$

$= 0.8$

$\therefore \Sigma d^2$ (incorrect) $= \frac{990 \times 0.8}{6} = 132.$

Now, Correct $\Sigma d^2 = 132 - 9^2 + 7^2 = 100$

$\therefore$ Correct $\rho = 1 - \frac{6(100)}{10(10^2-1)} = 0.394.$

$\therefore$ Thus correct $\rho = 0.394$.

5. Compute the rank correlation coefficient 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

Sol. The data along with the assigned ranks is as follows

x	y	Rank x	Rank 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
					$\Sigma d^2 = 72$

i) Here in x the magnitude 75 is repeated twice for rank 2 and hence the average rank of 2 and 3 $= 2.5$ is assigned to both. $(m=2)$.

ii) Also in x the magnitude 64 is repeated thrice for rank 5 and hence the average rank of

$5, 6, 7 = \frac{5+6+7}{3} = 6$ is assigned to the magnitudes 64 ($m=3$).

(iii) Further in y the magnitude 68 in repeated twice for rank 3 and hence the average rank of 3 & 4 is 3.5 is assigned to both of them. ($m=2$).

We have the coefficient of rank correlation for repeated ranks as, $\rho = 1 - \frac{6\left[\Sigma d^2 + \frac{m(m^2-1)}{12} + \cdots\right]}{n(n^2-1)}$

Given $n = 10$; $m = 2, 3, 2$

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

$$= 1 - \frac{6\left[72 + 0.5 + 2 + 0.5\right]}{990} = 0.545$$

$$\therefore \rho = 0.545$$

=

*****: Module-3: completed :*****

=

Curve fitting & optimization

Curve fitting :-

This is a method of finding the specific relation connecting the dependent & independent variables for a given data as so as to satisfy the data as accurately as possible. Such a curve is called curve of best fit. (The process of determining a curve of best fit is called curve fitting).

Optimization :-

This is a method of obtaining the best results under the given circumstances.

Curve fitting by least squares method :

Curve fitting is a method of finding a suitable relation or law in the form $y = f(x)$ for a set of observed values (x_i, y_i), $i = 1, 2, \ldots$

The relation connecting x & y is known as empirical law.

The method of least squares is as follows.

(Suppose $y = f(x)$ is an approximate relation that fits into a given data comprising (x_i, y_i) $i = 1, 2, \ldots, n$ then y_i's are called observed values and $Y_i = f(x_i)$ are called expected values. Their difference $R_i = y_i - Y_i$ are called the residuals or estimate errors.

The method of least squares provides a relationship $y = f(x)$ such that the sum of the squares of the residuals is least.)

$$S = \sum_{i=1}^{n} R_i^2 = \sum_{i=1}^{n} [y_i - Y_i]^2$$

* **Fitting of a straight line $y = ax + b$**

Let $y = ax + b$ be a straight line, where a & b are parameters to be determined. Consider a set of n values (x, y) for fitting the straight line.

Let the Residual $R = y - (ax + b)$ is the difference b/w the observed and estimated values of y.

By the method of least squares we find parameters a & b such that the sums of squares of the residuals is minimum (least).

Let $S = \sum_1^n R^2$

ie $S = \sum_1^n [y - (ax+b)]^2$

Treating S as a function of two parameters a & b the necessary conditions for S to be minimum are $\frac{\partial s}{\partial a} = 0$ & $\frac{\partial s}{\partial b} = 0$

ie $\frac{\partial s}{\partial a} = 2\sum_1^n [y-(ax+b)](-x) = 0$

& $\frac{\partial s}{\partial b} = 2\sum_1^n [y-(ax+b)](-1) = 0$

Dividing both the equations by 2 we have

$$-\sum_1^n xy + \sum_1^n ax^2 + \sum_1^n bx = 0$$

$$-\sum_1^n y + \sum_1^n ax + \sum_1^n b = 0$$

But $\sum_1^n b = b+b+\cdots+n \text{ times} = nb$

$$\Rightarrow \Sigma xy = \sum_1^n ax^2 + \Sigma bx$$

$$\Sigma y = \sum_1^n ax + nb$$

These equations are called normal equations for fitting the straight line $y = ax+b$ in the least square sense. By solving these we obtain the values of a & b.

* Fitting of a second degree parabola $y = ax^2+bx+c$

Let $y = ax^2+bx+c$ be a parabola where a, b & c are parameters to be determined.

The residuals $R = y-(ax^2+bx+c)$ is the difference b/w the observed and estimated values of y.

We have to find parameters a, b, c such that the sum of the squares of the residuals is the least.

Let $S = \sum_1^n [y-(ax^2+bx+c)]^2$

Treating S as the function of three parameters a, b, c the necessary conditions for S to be minimum are $\frac{\partial s}{\partial a} = 0$, $\frac{\partial s}{\partial b} = 0$, $\frac{\partial s}{\partial c} = 0$

i.e. $\frac{\partial S}{\partial a} = 2\sum_1^n [y-(ax^2+bx+c)](-x^2) = 0$

$\frac{\partial S}{\partial b} = 2\sum_1^n [y-(ax^2+bx+c)](-x) = 0$

$\frac{\partial S}{\partial c} = 2\sum_1^n [y-(ax^2+bx+c)](-1) = 0$

dividing all these eqns by 2 we have

$$-\sum_1^n x^2y + \sum_1^n ax^4 + \sum_1^n bx^3 + \sum_1^n cx^2 = 0$$

$$-\sum_1^n xy + \sum_1^n ax^3 + \sum_1^n bx^2 + \sum_1^n cx = 0$$

$$-\sum_1^n y + \sum_1^n ax^2 + \sum_1^n bx + \sum_1^n c = 0$$

$y = a+bx$

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

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

But $\sum c = c+c+\cdots+c$ n times $= nc$ and hence we have

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

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

$$a\Sigma x^2 + b\Sigma x + nc = \Sigma y$$

These equations are called normal equations for fitting the second degree parabola $y = ax^2+bx+c$ in the least square sense.

By solving these we obtain the values of a, b & c.

* Fitting of a curve of the form $y = ae^{bx}$

Consider $y = ae^{bx}$. Taking logarithms on both sides we get

$$\log y = \log ae^{bx}$$

$$\log_e y = \log a + \log_e e^{bx}$$

$$\log_e y = \log a + bx\log_e e \qquad \log_e e = 1$$

i.e. $\log_e y = \log a + bx$

or $Y = A + BX$ ..... (1)

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

It is evident that (1) is the equation of a straight line and the associated normal eqns are as follows:

$$\Sigma Y = nA + B\Sigma X \quad \cdots (2)$$

$$\Sigma XY = A\Sigma X + B\Sigma X^2 \cdots (3)$$

Solving (2) & (3) we obtain A & B. But

$\log_e a = A \Rightarrow a = e^A$ Also $b = B$.

Substituting these value in $y = ae^{bx}$ we get the curve of best fit in the required form.

* Fitting of a curve of the form $y = ax^b$

consider, $y = ax^b$ Taking log on both sides

$$\log y = \log ax^b$$
$$= \log_e a + b \log_e x$$
$$Y = A + BX$$

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

The normal eqns associated with (1) are same as in the previous case and we can obtain a, b. Substituting a, b in $y = ax^b$ we get the curve of the best fit in the required form.

Note:

Desired eqn	Modification	Substitution	Reduced form
1. $y = ax^n + b$	-	$X = x^n$	$y = aX + b$
2. $y = a + \frac{b}{x}$	-	$X = \frac{1}{x}$	$y = a + bX$
3. $y = ax + \frac{b}{x}$	Multiply by x to obtain $xy = ax^2 + b$	$Y = xy$, $X = x^2$	$Y = aX + b$
4. $y = \frac{x}{a+bx}$	$\frac{1}{y} = \frac{a+bx}{x}$ or $\frac{1}{y} = \frac{a}{x} + b$	$Y = \frac{1}{y}$, $X = \frac{1}{x}$	$Y = aX + b$
5. $xy^n = a$	$\log_e x + n \log_e y = \log_e a$ or $\log_e y = \frac{1}{n}[\log_e a - \log_e x]$	$Y = \log_e y$, $A = -1/n$, $B = \frac{1}{n} \log_e a$	$Y = AX + B$