Engineering Mathematics-4
(B.Tech,B.Sc,M.Sc)
(Part-3)

Module III : Statistical Techniques-I

Skewness: Skewness means lack of symmetry i.e., opposite symmetrical.

Skew symmetrical distribution: In case of skew symmetrical distribution the left tail and the right tail are not equal.

(a)Negative skew distribution: In case of negative skew

distribution then we consider the left tail is longer than the right tail.

(b)Positive skew distribution: In case of positive skew distribution then we consider the right tail is longer than the left tail.

Test of skewness: We consider following cases

(i)There is no skewness in the distribution if

A.M = Mode = Median

(ii)There is no skewness if

the sum of the frequencies which are less than mode = the sum of the friquencies which are greater than mode (iii)The distribution is negative skew if A.M less than mode

(iv)The curve is not symmetrical about the median if

A.M ≠ Median ≠ Mode

Measure of skewness: We consider

(i)Absolute measure i.e., mean = mode

(ii)Relative measure i.e.,

(a)Karl Pearson's coefficient of skewness

(b)Bowley's coefficient of skewness

(c)Kelly's coefficient of skewness

(d)Measure of skewness based on moments

(Mode = 3 median – 2 mean)

(a)Karl Pearson Coefficient of skewness: We consider

1. $\frac{Mean-mode}{S.D}$

2. $\frac{Mean \quad (3\,median-2\,mean)}{S.D}$ or $\frac{3(mean-median)}{S.D}$

Types of skewness in terms of mean and mode: We consider following postulates

1. There is no skewness in the distribution if $S_k = 0$

   i.e., $\frac{Mean - Mode}{S.D} = 0$ then Mean = Mode

2. The distribution is positive skewed if $S_K > 0$

i.e., $\frac{Mean - Mode}{S.D} > 0$ then Mean $> Mode$

3. The distribution is negative skewed if $S_k < 0$

i.e., $\frac{Mean - Mode}{S.D} < 0$ then Mean $< Mode$

QNO1. Compute the coefficient of S_k from the following data:

x	6	7	8	9	10	11	12
f	3	6	9	13	8	5	4

Solution

Let a = 9

Now we consider following table for the given data

x	f	d = x - a	f.d	f. d^2	c.f
6	3	- 3	- 9	27	3

7	6	- 2	-12	24	9
8	9	- 1	- 9	9	18
9	13	0	0	0	31
10	8	1	8	8	39
11	5	2	10	20	44
12	4	3	12	36	48

$\sum f$= 48 $\sum fd$= 0 $\sum fd^2$=124

Mean = a + $\frac{\sum fd}{\sum f}$ = 9 + $\frac{0}{48}$ = 9

Mode = Item of maximum frequency (13) = 9

S.D = $\sqrt{\frac{\sum fd^2}{\sum f} - (\frac{\sum fd}{\sum f})^2}$ = $\sqrt{\frac{124}{48}}$ = 1.61

Therefore Karl Pearson Coefficient of skewness

= $\frac{Mean - Mod}{S.D}$ = $\frac{9-9}{1.61}$ = 0 i.e., there is no skewness in the given data.

QNO2. Calculate the Karl Pearson coefficient of skewness from the following data

x	14-5	15.5	16.5	17.5	18.5	19.5	20.5	21.5
f	35	40	48	100	125	87	43	22

Solution

Let a = 17.5

Now we consider following table

x	f	d = x - a	f.d	$f.d^2$	c.f
14.5	35	- 3	-105	315	35
15.5	40	-2	-80	160	75
16.5	48	-1	-48	48	123
17.5	100	0	0	0	223
18.5	125	1	125	125	348
19.5	87	2	174	348	435
20.5	43	3	129	387	478
21.5	22	4	88	352	500
	500		283	1735	

Mean = $a + \frac{\sum fd}{\sum f}$ = 18.066

Mode = Item of maximum frequency = 18.5

S.D = $\sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2}$ = 1.77

Hence Karl Pearson's coefficient of Skewness

$= \frac{Mean - Mode}{S.D}$ = - 0.245

QNO3. Calculate the coefficient of skewness from the following data

x	55-58	58-61	61-64	64-67	67-70
f	12	17	23	18	11

Solution

Let a = 62.5

Now we consider following table

X	f	Mid x	d = x-a	f.d	$f.d^2$	c.f
55-58	12	56.5	-6	-72	432	12
58-61	17	59.5	-3	-51	153	**29**
61-64	**23**	62.5	0	0	0	52
64-67	18	65.5	3	54	162	70
67-70	11	68.5	6	66	396	81
	81			-3		

Mean = $a + \frac{\sum fd}{\sum f}$ = 62.46

Median = $l + \frac{\frac{N}{2}-(c.f)}{f}$ I = $61 + \frac{\frac{81}{2}-29}{23}$ 3 = 62.5

S.D = $\sqrt{\frac{\sum fd^2}{\sum f} - \{\frac{\sum fd}{\sum f}\}^2}$ = 3.76

Hence the Karl Pearson coefficient of skewness

$= \frac{3(Mean - Median)}{S.D}$ = -0.032

QNO4. Calculate the coefficient of skewness from the following data

x	0-10	10-20	20-30	30-40	40-50	50-60	60-70
f	10	12	18	25	16	14	8

Ans 0.0021

(b) Bowley's coefficient of skewness : A distribution is symmetrical if the distance between the first quartile and median is equal to the distance between the median and third quartile i.e.,

$$\text{median} - Q_1 = Q_3 - \text{median}$$

Measure of Bowley's Coefficient of S_k :

1. Bowley's absolute measure of skewness

   $= Q_3 + Q_1 - 2$ median

2. Bowley's relative measure of skewness

$$= \frac{Q_3 + Q_1 - 2\ median}{Q_3 - Q_1}$$

QNO1. Find the Bowley's coefficient of skewness.

Given that difference of quartile = 80

Sum of quartile = 120

Mode = 60 and Mean = 45

Solution Given that

$Q_3 + Q_1 = 120$

$Q_3 – Q_1 = 80$

Mode = 60 and Mean = 45

We know that Mode = 3 Median – 2 Mean

60 = 3 Median - 2 (45)

Then Median = 50

Hence, Bowley's Coefficient of $S_k = \frac{Q_3+Q_1-2M}{Q_3-Q_1}$

= 0.25

Kelly's Coefficient of Skewness : The modification of Bowley's is known as Kelly's Coefficient of Skewness.

Kelly's Coefficient of $S_k = \frac{P_{10}+P_{90}-2P_{50}}{P_{90}-P_{10}}$

Kelly's Coefficient of $S_k = \frac{D_9+D_1-2D_5}{D_9-D_5}$

Where, P denotes Percentile and D denote deciles

And Median = $P_{50} = D_5$

QNO1. Calculate Percentile Coefficient of skewness from the following: P_{90}= 110, P_{10} = 30, P_5 = 80

Solution The Kelly's Coefficient of Skewness

$$= \frac{P_{90}+P_{10}-2P_{50}}{P_{90}-P_{10}}$$

$$= \frac{110+3 \qquad (80)}{110-3} = -0.25$$

(d)Measure of skewness based on moment:

We consider

1. $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$
2. $\beta_2 = \frac{\mu_4}{\mu_2^2}$
3. $\gamma_1 = \pm\sqrt{\beta_1}$
4. $\gamma_2 = \beta_2 - 3$

Kurtosis

In case of kurtosis then we observe degree of peackedness of the given distribution.

$\mu_2 = \frac{\sum(x-\bar{x})^2}{N}$, $\mu_4 = \frac{\sum(x-\bar{x})^4}{N}$

And we consider following cases

Case 1st: If β_2= 3 then the curve is normal i.e., Mesokurtosis.

Case 2nd: If $\beta_2 > 3$ then the curve represent peacked

i.e., Leptokurtosis

Case 3rd: If $\beta_2 < 3$ then the curve represent topped i.e., Platykurtic

The general formula is $\boldsymbol{\gamma_2 = \beta_2 - 3}$

Type 1:

QNO1. Find the first four moments for the following data

X	1	3	9	12	20

Solution

We consider following table for the given distribution

S.No	x	d=x - $\bar{x}$	d^2	d^3	d^4
1	1	-8	64	-512	4096
2	3	-6	36	-216	1296
3	9	0	0	0	0
4	12	3	9	27	81
5	20	11	121	1331	14641
n = 5	=45	=0	=230	=630	=20114

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

$\mu_1 = \frac{\sum d}{n} = 0$

$\mu_2 = \frac{\sum d^2}{n} = 46$

$\mu_3 = \frac{\sum d^3}{n}$ =126

$\mu_4 = \frac{\sum d^4}{n}$ = 4022.8

Type2.

QNO1: Find the Kurtosis from the following data

X	0 -10	10 - 20	20 - 30	30 - 40
F	1	3	4	2

Solution

Let a = 25

Now we consider following table

Class	f	Mid(x)	d	f.d	f.d^2	f.d^3	f.d^4
0 -10	1	5	-20	-20	400	-8000	160000
10-20	3	15	-10	-30	300	-3000	30000
20-30	4	25	0	0	0	0	0
30-40	2	35	10	20	200	2000	20000
	=10			=-30	=900	-9000	210000

Then we consider

$\mu_1' = \frac{\sum fd}{\sum f} = \frac{-30}{10}$ = -3

μ_2' C= $\frac{\sum f d^2}{\sum f}$ = $\frac{900}{10}$ = 90

μ_3' = -900

μ_4' = 21000

Therefore, $\mu_2 = \mu_2' - \mu_1'^2$ = 90 – 9 = 81

And $\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$

= 14817

$\beta_2 = \frac{\mu_4}{\mu_2^2}$ = 2.258

Hence, $\gamma_2 = \beta_2$ - 3 = - 0.742

Type3:

QNO1. Calculate $\mu_1, \mu_2, \mu_3, \mu_4$ for the following data

X	0-10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60
F	1	6	10	15	11	7

Solution

X	f	mid	f.x	d	f.d	f.d^2	f.d^3	f.d^4
0-10	1	5	5	-30	-30	900	-27000	810000
10-20	6	15	90	-20	-120	2400	-48000	960000
20-30	10	25	250	-10	-100	1000	-10000	100000

30-40	15	35	525	0	0	0	0	0
40-50	11	45	495	10	110	1100	11000	110000
50-60	7	55	385	20	140	2800	56000	1120000
	50		1750		0	8200	-18000	3100000

Mean

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{1750}{50} = 35$$

$$\mu_1 = \frac{\sum fd}{\sum f} = \frac{0}{50} = 0$$

$$\mu_2 = \frac{\sum fd^2}{\sum f} = \frac{8200}{50} = 164$$

$$\mu_3 = -360$$

$$\mu_4 = 62000$$

Moments: The r^{th} central moment of a variable x about the mean $\bar{x}$ is denoted by μ_r and represented by

$$\mu_r = \frac{1}{N}\sum f(x-\bar{x})^r$$

And the r^{th} moment of a variable x about any point a is denoted by

$$\mu_r' = \frac{1}{N}\sum f(x-a)^r$$

Relation between μ_r and μ'_r

We consider $\mu_1 = 0$

$$\mu_2 = \mu'_2 - \mu'^2_1$$

$$\mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2\mu'^3_1$$

$$\mu_4 = \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2\mu'^2_1 - 3\mu'^4_1$$

Moment Generating function The moment generating function of a variable x about x = a is defined as the expected value of $e^{t(x-a)}$ and is denoted by $M_a(t)$ then we write

$$M_a(t) = \sum pe^{t(x-a)}$$

$$= \sum p[1 + t(x-a) + \frac{t^2}{2}(x-a)^2 + \cdots + \frac{t^r}{r!}(x-a)^r + .]$$

$$= \sum p + t\sum p(x-a) + \frac{t^2}{2}\sum p(x-a)^2 + ... + \frac{t^r}{r!}\sum p(x-a)^r + ...$$

$$= 1 + t\sum f(x-a) + \frac{t^2}{2}\sum f(x-a)^2 + ... + \frac{t^r}{r!}\sum f(x-a)^r + ..$$

Where, $\sum p = 1 = \mu_0$ & p = f

$$= \mu_0 + t\,\mu'_1 + \frac{t^2}{2!}\mu'_2 + ... + \frac{t^r}{r!}\mu'_r + ...$$

Where, μ_r' is the moment of order r about a

Hence, $\boldsymbol{\mu_r'}$ **= coefficient of** $\frac{\boldsymbol{t^r}}{\boldsymbol{r!}}$ or $\mu_r'= [\frac{d^r}{dt^r} M_a(t)\ ]_{t=0}$

Again, $M_a(t) = \sum pe^{t(x-a)}$

$= e^{-at} \sum pe^{tx}$

$= e^{-at}\ M_0(t)$

Thus the Moment Generating Function about any point =

e^{-a} Moment Generating Function about the origin.

QNO1. Find the moment generating function of the exponential distribution

$f(x) = \frac{1}{c}\, e^{-x/c}$, $0 \leq x \leq \infty$ and $c > 0$

Hence, find its mean and standard deviation.

Solution The moment generating function about the origin is $M_0(t) = \int_0^\infty e^{tx} f(x)\, dx$

$= \int_0^\infty e^{tx} . \frac{1}{c} e^{-x/c}\, dx$

$= \frac{1}{c} \int_0^\infty e^{(t-c^{-1})x}\, dx$

$= \frac{1}{c} \frac{1}{t-c^{-1}} [e^{(t-c^{-1})x} \ ]_0^\infty$

$= \frac{1}{c} \frac{1}{t-c^{-1}} [0 - 1]$

$= (1 - ct)^{-1}$

Now, Moment about origin = $[\frac{d}{dt} M_0(t) \ ]_{t=0}$

$= \frac{d}{dt} [1 + ct + c^2t^2 + \ldots\ldots\ldots\ldots\ldots]_{t=0}$

$= [c + 2c^2t + 3c^3t^2 + \ldots\ldots\ldots\ldots\ldots]_{t=0}$

$= c$

Then $\mu_1' = \bar{x} = c$

$\mu_2' = [\frac{d^2}{dt^2} M_0(t)]$

$= \frac{d^2}{dt^2} [1 + ct + c^2 t^2 + \ldots\ldots\ldots\ldots]_{t=0}$

$= \frac{d}{dt} [c + 2c^2t + 3c^3t^2 + \ldots\ldots\ldots\ldots]_{t=0}$

$= [2c^2 + 6c^3t + \ldots\ldots\ldots\ldots\ldots]_{t=0}$

$\mu_2' = 2c^2$

Thus $\mu_2 = \mu_2' - \mu_1'^2$

$= 2c^2 - c^2 = c^2$

Hence, standard deviation = $\sqrt{\mu_2}$

$= \sqrt{c^2} = c$

QNO2. Obtain the moment generating function of the random variable x having probability distribution

$$f(x) = \begin{matrix} x & for & 0 < x < 1 \\ 2 - x & for & 1 \leq x < 2 \\ 0 & for & anywhere \end{matrix}$$

Also determine μ_1' , μ_2' and μ_2

Solution We know that

$M_x(t) = \int e^{tx} f(x) dx$

$= \int_0^1 x.e^{tx} dx + \int_1^2 (2 - x) e^{tx} dx + \int_2^\infty 0.e^{tx} dx$

$= \{\frac{xe^{tx}}{t} - \frac{e^{tx}}{t^2}\}_0^1 + \{\frac{2e^{tx}}{t} - \frac{xe^{tx}}{t} + \frac{e^{tx}}{t^2}\}_1^2$

$= \frac{e^{2t} - 2e^t + 1}{t^2}$

$= \frac{1}{t^2} [\{e^{2t}\} - 2\{e^t\} + 1]$ Expand by exponential series

$= \frac{1}{t^2} + \frac{2}{t} + 2 + \frac{4t}{3} + \frac{2t^2}{3} + \ldots\ldots\ldots\ldots, - \frac{2}{t^2} - \frac{2}{t} - 1 - \frac{t}{3} - \frac{t^2}{12} + \ldots\ldots\ldots\ldots, + 1$

Now, μ_1' = coefficient of $\frac{t}{1!} = \frac{4}{3} - \frac{1}{3} = 1$

And μ_2' = coefficient of $\frac{t^2}{2!} = \frac{4}{3} - \frac{1}{6} = \frac{7}{6}$

Then $\mu_2 = \mu_2' - (\mu_1')^2 = \frac{1}{6}$

Hence, $\mu_1' = 1$, $\mu_2' = \frac{7}{6}$, $\mu_2 = \frac{1}{6}$

Curve Fitting (Method of least squares) We consider

1. St.line
2. Parabola
3. Exponential form

1.St.line

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

Be the st. line to be fitted the given data points (x_1, y_1) (x_2, y_2) (x_3, y_3) ....(x_n, y_n).

Now consider PN = PM – NM

$e_1 = y_1 - y'_1$

$e_1 = y_1 - (a + bx_1)$

$e_1^2 = (y_1 - a - bx_1\ )^2$

Similarly, $e_2^2 = (y_2 - a - bx_2\ )^2$

$e_n^2 = (y_n - a - bx_n\ )^2$

$S = e_1^2 + e_2^2 + e_3^2 + \ldots\ldots\ldots\ldots\ldots\ldots + e_n^2$

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

For S to be minimum

$\frac{\partial s}{\partial a} = \sum_{i=1}^{n} 2\ (\ y_i - a - b\ x_i\ )\ (-1) = 0$

Or $\sum(y - a - bx) = 0$

$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$

And $\frac{\partial s}{\partial b} = \sum_{i=1}^{n} 2(y_i - a - bx_i)(-x_i) = 0$

Or $\sum(xy - ax - bx^2) = 0$

$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$

From (2) & (3) we have

$\sum y = n\,a + b\sum x$

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

i.e., Normal equation

Then solving normal equations we get the values of a and b .Therefore putting these values in equation one then we have equation of st.line with respect to method of least square

Type 1:

QNO 1 Fit a st. line to the following data

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

Solution

Let $y = a + bx$ ……………………..(1)

Now we consider following table

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

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

Then

We consider normal equation i.e.,

$$\sum y = n\,a + b \sum x \text{ ....(2)}$$

$$\sum xy = a \sum x + b \sum x^2 \text{ ....(3)}$$

Or 16.9 = 5 a + 10 b

47.1 = 10 a + 30 b

Solving we get a = 0.72 and b = 1.33

Putting the values of a & b in equation one then

y = 0.72 + 1.33 x

QNO 2 By the method of least square find the st.line that best fit the following data

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

Solution

Let y = a + b x ....(1)

Now we consider following table

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

Then we using Normal equation i.e.,

$\sum y$ = n a + b $\sum x$ ....(2)

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

Or 204 = 5a + 15b

748 = 15a + 55b Solving we get

a = 0 & b = 13.6

Putting the values of a and b in equation one we get

y = 13.6 x

i.e., equation of st.line with respect to method of least square

Type 2

QNO 1:

Find the least squares fit the form $y = a + b\,x^2$ to the following data

x	-1	0	1	2
y	2	5	3	0

Solution Given that

$$y = a + b\,x^2$$

put, $x^2 = X$ then

$$y = a + b\,X \text{ ....}(1)$$

Now we consider following table

x	Y	X	Xy	X^2
-1	2	1	2	1
0	5	0	0	0
1	3	1	3	1
2	0	4	0	16
	$\sum y$= 10	$\sum X$= 6	$\sum Xy$=5	$\sum X^2$=18

Normal equation is

$$\sum y = n\,a + b \sum X$$

And $$\sum Xy = a \sum X + b \sum X^2$$

Or $10 = 4a + 6b$

$5 = 6a + 18b$

On solving $a = 25/6$ & $b = -10/9$

$18y = 75 - 20x^2$

Type 3

QNO 1: Fit a curve of the type $xy = ax + b$ to the following data

x	1	3	5	7	9	10
y	36	29	28	26	24	15

Solution

Given that $xy = ax + b$

Or $y = a + \frac{b}{x}$ ....(1)

Now we consider following table

X	Y	1/x	y/x	$1/x^2$
1	36	1	36	1
3	29	0.333	9.66	0.11
5	28	0.200	5.6	0.04
7	26	0.14	3.71	0.02

9	24	0.11	2.6	0.012
10	15	0.1	1.5	0.01
	$\sum y$=158	$\sum 1/x$=1.88	= 59.07	=1.192

Now we consider Normal equation i.e.,

$$\sum y = n\,a + b \sum 1/x$$

And $\sum y/x = a \sum 1/x + b \sum 1/x^2$

a = 21.381 & b = 15.744

xy = 21.381 x + 15.744

Type 4

QNO1 Using the method of least squares fit the curve

$y = \frac{c_0}{x} + c_1 \sqrt{x}$ for the following data

x	0.1	0.2	0.4	0.5	1	2
y	21	11	7	6	5	6

Solution

Given that $y = \frac{c_0}{x} + c_1 \sqrt{x}$ ………………… (1)

Now we consider following table

X	Y	y/x	$y\sqrt{x}$	$1/\sqrt{x}$	$1/x^2$
0.1	21	210	6.640	3.162	100
0.2	11	55	4.919	2.236	25
0.4	7	17.5	4.427	1.581	6.25
0.5	6	12	4.242	1.414	4
1	5	5	5	1	1
2	6	3	8.485	0.707	0.25
= 4.2		=302.5	=33.715	=10.100	=136.5

Then we consider Normal equation i.e.,

$$\sum y/x = c_0 \sum \frac{1}{x^2} + c_1 \sum \frac{1}{\sqrt{x}} \quad \ldots\ldots\ldots\ldots\ldots\ldots(2)$$

$$\sum y\sqrt{x} = c_0 \sum \frac{1}{\sqrt{x}} + c_1 \sum x \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

Or $302.5 = 136.5\, c_0 + 10.100\, c_1$

$33.715 = 10.100\, c_0 + 4.2\, c_1$

Solving we get

$c_0 = 1.973$ & $c_1 = 3.281$

Putting in equation one then

$$y = \frac{1.973}{x} + 3.281\sqrt{x}$$

i.e., equation of st.line with respect to method of least square

2. Fit the Parabola: Let $y = a + b x + c x^2$ be the equation of a parabola.

And normal equation $\sum y = n a + b \sum x + c \sum x^2$

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

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

Solving we get a, b & c then putting in equation one.

We get equation of parabola with respect to method of least square.

QNO1. Fit a second degree parabola for the following data

x	1	2	3	4	5
y	1090	1220	1390	1625	1915

Solution

Let equation of the parabola

$$y = a + b x + c x^2 \text{ ....(1)}$$

Now, we consider following table

x	y	xy	x^2	x^2y	x^3	x^4
1	1090	1090	1	1090	1	1
2	1220	2440	4	4880	8	16
3	1390	4170	9	12510	27	81
4	1625	6500	16	26000	64	256
5	1915	9575	25	47875	125	625
= 15	=7240	=2377	= 55	=92355	=225	=979

Then using normal equation i.e.,

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

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

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

Putting all values then we have

$$7240 = 5\,a + 15\,b + 55\,c$$

$$23775 = 15\,a + 55\,b + 225\,c$$

$$92355 = 55\,a + 225\,b + 979\,c$$

Solving we get

$a = 1024$ $\quad b = 81/2$ $\quad c = 55/2$

Hence $\quad 2y = 2048 + 81\,x + 55\,x^2$

i.e., Second degree parabola equation with respect to method of least square

QNO2. Fit a second degree parabola to the following data

x	10	15	20	25	30	35	40
y	11	13	16	20	27	34	41

Solution

Let equation of the parabola

$$y = a + b\,x + c\,x^2 \;.............................\; (1)$$

Now we consider following table

x	Y	xy	x^2y	x^2	x^3	x^4
10	11	110	1100	100	1000	10000
15	13	195	2925	225	3375	50625
20	16	320	6400	400	8000	160000
25	20	500	12500	625	15625	390625
30	27	810	24300	900	27000	810000
35	34	1190	41650	1225	42875	1500625
40	41	1640	65600	1600	64000	2560000
=175	=162	=4765	=154475	=5075	=161875	=5481875

Then using normal equation i.e.,

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

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

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

Putting all values then solving

a = 145/14 b = - 135/700 c = 17/700

Hence $700y = 7250 - 135x + 17x^2$

i.e., Second degree parabola equation with respect to method of least square

3. Fit the exponential form: In case of exponential form then taking log on both side then solving we get st.line.

QNO1. Fit a curve of the form $y = a b^x$ to the following data

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

Solution

Given that $y = a b^x$ .... (1)

Taking log on both side then we write

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

Y = A + B x ………………………… (2) i.e., st.line

Where Y = $\log_e$ y, A = $\log_e$ a, B = $\log_e$ b

Now we consider following table

x	y	Y = $\log_e$ y	xY	x^2
2	144	2.158	4.316	4
3	172.3	2.236	6.708	9
4	207.7	2.316	9.267	16
5	248.8	2.395	11.979	25
6	298.5	2.474	14.849	36
$\sum x = 20$		= 11.582	= 47.121	= 90

Using normal equation

$$\sum Y = n A + B \sum x$$

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

Or 11.582 = 5 A + 20 B

47.121 = 20 A + 90 B

Solving we get A = 1.999 & B = 0.0792

Now, A = $\log_e$ a

$1.999 = \log_e a \Rightarrow a = \text{Antilog of } (1.999)$

$= \text{Antilog of } 2$

$= 100$

And $B = \log_e b \Rightarrow 0.0792 = \log_e b$

$\Rightarrow b = \text{Antilog of } (0.0792)$

$= 1.2$

Hence $y = 100\,(1.2)^x$

QNO2. 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

Solution

Given that $y = a\,e^{bx}$ .... (1)

$\log_e y = \log_e a + b\,x\,\log_{10} e$

$Y = A + B\,x$ ....(2) i.e., st.line

Where $Y = \log_e y$, $A = \log_e a$, $B = b\,\log_{10} e$

Now we consider following table

x	y	Y= $\log_e$ 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
$\sum x$ = 21		= 8.1753	=36.693	=91

Using normal equation

$$\sum Y = n\,A + B \sum x$$

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

Or $$8.1753 = 6\,A + 21\,B$$

$$36.693 = 21\,A + 91\,B$$

Solving we get A = - 0.2534 & B = 0.4617

Now, $A = \log_{10} a \Rightarrow -0.2534 = \log_{10} a$

$$a = \text{Antilog of } (-0.2534)$$

$$= 0.5580$$

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

Hence y = 0.5580 $e^{1.0631\,x}$

Correlation : If two variables x and y represent (i) x and y both variables increases (ii) x and y both variables decreases(iii)variables x increases and y decreases(iv)variables x decreases and y increases.

It is called correlation

Types of correlation We consider (i)Positive correlation

(ii)Negative correlation (iii) Linear correlation (iv)Perfect correlation (v)Positive perfect correlation (vi)Negative perfect correlation

Karl Pearson's Coefficient of correlation: Let r be the coefficient of correlation then

$$r = \frac{\sum XY}{\sqrt{\sum x^2 \sum y^2}}$$ Where, $X = x - \bar{x}$ & $Y = y - \bar{y}$

QNO1 . Calculate the correlation coefficient between the following data

x	5	9	13	17	21
y	12	20	25	33	35

Solution

$\bar{x} = \frac{5+9+13+17+21}{5} = 13$

$\bar{y} = \frac{12+20+25+33+35}{5} = 25$

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

Now we consider following table

x	X= x-$\bar{x}$	X^2	y	Y= y-$\bar{y}$	Y^2	XY
5	- 8	64	12	-13	169	104
9	-4	16	20	-5	25	20
13	0	0	25	0	0	0
17	4	16	33	8	64	32
21	8	64	35	10	100	80
	$\sum x$= 0	=160		=0	=358	=236

Now

$$r = \frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}} = \frac{236}{\sqrt{160X3}} = 0.986$$

QNO2. Calculate the correlation coefficient between x and y for the following data

x	21	23	30	54	57	58	72	78	87
y	60	71	72	83	110	84	100	92	113

Spearman rank correlation coefficient.

Let r be the correlation coefficient then

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

QNO1. Calculate spearman's rank correlation coefficient for the following data

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

Solution

In case of spearman's rank correlation then we consider following table

X	y	d = x - y	d^2
9	1	8	64
10	2	8	64
6	3	3	9
5	4	1	1
7	5	2	4
2	6	-4	16
4	7	-3	9
8	8	0	0
1	9	-8	64

3	10	-7	49
			$\sum d^2$= 280

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

= 1 - $\frac{6X280}{10(100-1)}$ = - 0.697

QNO2. Establish the formula

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

Where, r is the correlation coefficient between x and y

Regression

If the scatter diagram represents some relationship between x & y then we gets a curve, this curve is known as curve of regression. And equation of the regression curve is said to be regression equation.

The regression line of y on x is

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

And the regression line of x on y is

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

Where, $b_{yx} = r\frac{\sigma_y}{\sigma_x}$ and $b_{xy} = r\frac{\sigma_x}{\sigma_y}$ are known as regression coefficients.

QNO1. Two lines of regression are given by 5y – 8x +17 = 0 and 2y – 5x +14 = 0.If σ_y^2 = 16 then find (i) Mean values of x and y (ii)σ_x^2 (iii) The coefficient of correlation between x and y

Solution Given that 5y – 8x + 17 = 0

And 2y – 5x +14 = 0

Let p ($\bar{x}$ $\bar{y}$) be the common point then lines of regression

We write $5\bar{y} - 8\bar{x} + 17 = 0$ & $2\bar{y} - 5\bar{x} + 14 = 0$

Solving we get $\frac{\bar{x}}{36} = \frac{\bar{y}}{27} = \frac{1}{9} \Rightarrow \bar{x} = 4$ and $\bar{y} = 3$

i.e., Mean values of x and y

Now the regression equation we write

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

Then $m_1 = b_{yx} = r\frac{\sigma_y}{\sigma_x} = \frac{8}{5}$ and $m_2 = b_{xy} = r\frac{\sigma_x}{\sigma_y} = \frac{2}{5}$

$r = \pm\frac{4}{5}$ i.e., coefficients of correlation between x and y.

And putting the value of r & σ_y (given) in m_1 then we have $\sigma_x = 2$ then $\sigma_x^2 = 4$

Hence, the mean values of x and y is 4 & 3

And the coefficients of correlation between x and y is $\pm\frac{4}{5}$

And $\sigma_x^2 = 4$

QNO2 Two lines of regression are given by 8 x – 10 y + 66 = 0 & 40 x – 18 y = 214. If $\sigma_x^2 = 9$ then find (i) Mean values of x & y (ii) Standard deviation of y (iii) Coefficients of correlation between x & y

Ans. (i) 13 and 17 (ii) 4 (iii) 0.6

QNO3 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	9

Solution

$$\bar{x} = \frac{\sum x}{n} = \frac{5+7+8+10+11+13+16}{7} = 10$$

$$\bar{y} = \frac{\sum y}{n} = \frac{33+30+28+20+18+16+9}{7} = 22$$

Now we consider following table

x	y	X	Y	XY	X^2	Y^2
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
=70	=154			=-191	= 84	=446

Coefficient of correlation $r = \frac{\sum XY}{\sqrt{\sum X^2 . \sum Y^2}} = \frac{-191}{\sqrt{84X446}} = -0.986$

Now, $m_1 = b_{yx} = r\frac{\sigma_y}{\sigma_x} = r\sqrt{\frac{\sum Y^2}{\sum X^2}} = -0.986\sqrt{\frac{446}{84}} = -2.273$

And $m_2 = b_{xy} = r\frac{\sigma_x}{\sigma_y} = r\sqrt{\frac{\sum X^2}{\sum Y^2}} = -0.986\sqrt{\frac{84}{446}} = -0.428$

Hence, regression line of y on x is

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

y – 22 = - 2.273 (x – 10)

y = - 2.273 x + 44.738

Similarly regression line of x on y is

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

x – 10 = - 0.428 (y – 22)

x = - 0.428 y + 19.422

QNO4. Calculate the coefficient of correlation and 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

Ans r = 0.95

y = .95x + 7.25 & x = .95y – 6.4

QNO5. Calculate the coefficient of correlation from the following n = 10, $\sum X$ = 100, $\sum Y$ = 150, $\sum(x-10)^2$ =180, $\sum(y-15)^2$ = 215, $\sum(x-10)(y-15)$ = 60

Solution Mean $\bar{x} = \frac{\sum X}{n} = \frac{100}{10} = 10$

$$\bar{y} = \frac{\sum Y}{n} = \frac{150}{10} = 15$$

Now, $r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2.\sum(y-\bar{y})^2}} = \frac{\sum(x-1\ \ )(y-15)}{\sqrt{\sum(x-10)^2.\sum(y-15)^2}}$

$$= \frac{60}{\sqrt{180X215}} = 0.305$$

QNO6. If θ be the acute angle between the two regression lines in the case of two variables x and y, then show that $\tan\theta = \frac{1-r^2}{r} . \frac{\sigma_x \sigma_y}{\sigma_x^2+\sigma_y^2}$, where r, σ_x, σ_y represent their usual meanings. Also explain the significance when r = 0 and r = $\pm$ 1.

Solution Lines of regression are

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

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

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

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

(a) If r = 0 then $\tan\theta = \infty$, $\theta = \frac{\pi}{2}$ so the lines of regression are perpendicular.

(b) If r = 1 or -1 then $\tan\theta = 0$ or $\theta = 0$ i.e., lines of regression coincide.

Multiple Correlation: Let three variables x_1 , x_2 and x_3 then

$R_{1.23}$ = Multiple correlation coefficient with x_1 as dependent variable and x_2 , x_3 as independent variables.

$R_{2.13}$ = Multiple correlation coefficient with x_2 as dependent variable and x_1 and x_3 as independent variables.

$R_{3.12}$ = Multiple correlation coefficient with x_3 as

dependent variable and x_1 and x_2 as independent variables.

Now, $R_{1.23} = \sqrt{\frac{r_{12}^2+r_{13}^2-2r_{12}r_{13}r_{23}}{1-r_{23}^2}}$

$$R_{2.13} = \sqrt{\frac{r_{21}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{13}^2}}$$

$$R_{3.12} = \sqrt{\frac{r_{31}^2 + r_{32}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{12}^2}}$$

Properties: We consider

1. Its value lies between 0 to 1

2. If $R_{1.23} = 0$ then $r_{12} = 0 = r_{13}$

3. $R_{1.23} \geq r_{12}$ and $R_{1.23} \geq r_{13}$

4. $R_{1.23} = R_{1.32}$

QNO1. If $r_{12} = 0.6$, $r_{23} = 0.35$ and $r_{31} = 0.4$ then find $R_{3.12}$

Solution We know that

$$R_{3.12} = \sqrt{\frac{r_{31}^2 + r_{32}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{12}^2}}$$

$$= \sqrt{\frac{(0.4)^2 + (0.35)^2 - 2(0.6)(0.4)(0.35)}{1 - (0.6)^2}} = 0.423$$

QNO2. If $r_{12} = 0.25$, $r_{13} = 0.35$ and $r_{23} = 0.45$ then find $R_{2.13}$ **Ans 0.461**

Regression equation of three variables:

$x_1 = b_{12.3}\, x_2 + b_{13.2}\, x_3$ where, $b_{12.3} = \frac{\sigma_1}{\sigma_2}\left[\frac{r_{12}-r_{13}r_{23}}{1-r_{23}^2}\right]$

$b_{13.2} = \frac{\sigma_1}{\sigma_3}\left[\frac{r_{13}-r_{12}r_{23}}{1-r_{23}^2}\right]$

$x_2 = b_{21.3}\, x_1 + b_{23.1}\, x_3$ $b_{21.3} = \frac{\sigma_2}{\sigma_1}\left[\frac{r_{12}-r_{23}r_{13}}{1-r_{31}^2}\right]$

$b_{23.1} = \frac{\sigma_2}{\sigma_3}\left[\frac{r_{23}-r_{12}r_{13}}{1-r_{13}^2}\right]$

$x_3 = b_{31.2}\, x_1 + b_{32.1}\, x_2$ $b_{31.2} = \frac{\sigma_3}{\sigma_1}\left[\frac{r_{13}-r_{23}r_{12}}{1-r_{12}^2}\right]$

$b_{32.1} = \frac{\sigma_3}{\sigma_2}\left[\frac{r_{23}-r_{12}r_{13}}{1-r_{12}^2}\right]$

QNO1. If $r_{12} = 0.6$, $r_{13} = 0.8$, $r_{23} = 0.3$, $\sigma_1 = 8$, $\sigma_2 = 9$, $\sigma_3 = 5$ then find the regression equation of x_1 on x_2 and x_3

Solution We know that

$x_1 = b_{12.3}\, x_2 + b_{13.2}\, x_3$ where, $b_{12.3} = \frac{\sigma_1}{\sigma_2}\left[\frac{r_{12}-r_{13}r_{23}}{1-r_{23}^2}\right]$

$= 0.35$

$b_{13.2} = \frac{\sigma_1}{\sigma_3}\left[\frac{r_{13}-r_{12}r_{23}}{1-r_{23}^2}\right]$

$= 1.09$

$x_1 = 0.35\, x_2 + 1.09\, x_3$

QNO2. If $r_{12} = 0.75$, $r_{13} = 0.65$, $r_{23} = 0.55$, $\sigma_1 = 9$, $\sigma_2 = 7$,

$\sigma_3 = 4$ then find regression equation of x_2 on x_1 & x_3

Ans. $x_2 = 0.528\, x_1 + 0.189\, x_3$