Answer.

Ans. $\pi/2$

Ans. $\pi(b-a)$

Ans. $\frac{-1}{2}$

Ans. $\frac{\pi}{\sqrt{3}}$

Ans. $\pi/2$.

UNIT - II

STATISTICAL TECHNIQUE-I

Frequency Distribution

Frequency Distribution is the arranged data which is summarised by distributing it into classes or categories with their frequencies. The following is a frequency distribution marks obtained by 75 students

Marks obtained	0-10	10-20	20-30	30-40	40-50
No. of students	7	18	30	15	5

The above frequency distribution has exclusive class-intervals. But following frequency distribution has inclusive class-intervals.

Income Rs.	80-89	90-99	100-109	110-119	120-129	130-139
No. of workers	6	11	14	20	7	5

Mean of Frequency Distribution

$$\text{Mean } \bar{x} = \frac{\Sigma fx}{\Sigma f}$$

or

$$\bar{x} = a + \frac{\Sigma fd}{\Sigma f}$$

Median of a grouped data

$$\text{Median} = \ell + \frac{\left(\frac{N}{2} - C\right)i}{f}$$

where ℓ is the lower limit of median class, f is the frequency of median class, $N = \Sigma f$ and i is the length of interval and C is the cumulative frequency preceeding the median class.

Percentiles

$$P_x = \ell + \left(\frac{N \cdot x}{100} - C\right)\frac{i}{f},\ x = 1, 2, 3, 4, \ldots$$

$$\text{for } x = 1,\ P_1 = \ell + \left(\frac{N}{100} - C\right)\frac{i}{f}$$

$$\text{for } x = 10,\ P_{10} = \ell + \left(\frac{N}{10} - C\right)\frac{i}{f}.$$

$$P_{50} = \text{Median} = \ell + \left(\frac{N}{2} - C\right)\frac{i}{f}$$

Quartiles :

Lower quartile, $Q_1 = P_{25} = \ell + \left(\frac{N}{4} - C\right)\frac{i}{f}$

Upper quartile, $Q_3 = P_{75} = \ell + \frac{i}{f}\left(\frac{3N}{4} - C\right)$

Deciles : First Decile, $D_1 = P_{10} = \ell + \left(\frac{N}{10} - C\right)\frac{i}{f}$ and Second decile, $D_2 = P_{20} = \ell + \left(\frac{N}{5} - C\right)\frac{i}{f}$ and so on.

Mode

$$\text{Mode} = \ell + \frac{f_m - f_1}{2f_m - f_1 - f_2} \cdot i$$

where ℓ is the lower limit of modal class, i, the length of class-interval, f_m is the modal frequency, f_1 & f_2 are frequencies preceeding and succeeding the modal frequency respectively.

Standard Deviation

$$\text{S.D.} = \sigma = \sqrt{\frac{\Sigma f(x - \bar{x})^2}{\Sigma f}}$$

or $$\sigma = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd}{\Sigma f}\right)^2}$$

Symmetric Frequency Distribution

A frequency distribution is said to be symmetric when the mode, mean, median are identical (equal). A symmetrical distribution when plotted on a graph will give a perfectly bell shaped curve called normal curve.

Skewness : The lack of symmetry of a frequency distribution is called skewness. A frequency distribution having skewness is called a skew-symmetrical distribution which has either its tail on the left or tail on the right. If the distribution has tail on the right, it has +ve skewness and if it has tail on the left then it is negatively skewed.

Types of Frequency Distribution

(1) Perfectly Symmetrical Distribution

(2) Positively Skewed Distribution

(3) Negatively Skewed Distribution

Moments : The rth moment of a frequency Distribution about mean $\bar{x}$,

$$\mu_r = \frac{\Sigma f_i(x_i - \bar{x})^r}{\Sigma f_i}, \; r = 0, 1, 2, 3, \ldots$$

$$\therefore \quad \mu_0 = \frac{\Sigma f_i(x_i - \bar{x})^0}{\Sigma f_i} = 1$$

Scanned by CamScanner

$$\mu_1 = \frac{\Sigma fi(x_i - \bar{x})}{\Sigma fi} = \frac{\Sigma fixi}{\Sigma fi} - \bar{x} \cdot \frac{\Sigma fi}{\Sigma fi} = \bar{x} - \bar{x} = 0$$

$$\mu_2 = \frac{\Sigma fi(x_i - \bar{x})^2}{\Sigma fi}$$

The rth moment of a distribution about any non-zero arbitrary number A is defined by

$$\mu'_r = \frac{\Sigma fi(x_i - A)^r}{\Sigma fi}, \ r = 0, 1, 2, 3, \ldots$$

$$\therefore \ \mu'_0 = \frac{\Sigma fi\,(x_i - A)^0}{\Sigma fi} = 1$$

$$\mu'_1 = \frac{\Sigma fi(x_i - A)}{\Sigma fi} = \frac{\Sigma fixi}{\Sigma fi} - A \cdot \frac{\Sigma fi}{\Sigma fi} = \bar{x} - A$$

$$\mu'_2 = \frac{\Sigma fi(x_i - A)^2}{\Sigma fi} \text{ and so on.}$$

Relation between μ_r and μ'_r

We have, $\mu_r = \dfrac{\Sigma fi(x_i - \bar{x})^r}{\Sigma fi} = \dfrac{\Sigma fi[x_i - A - (\bar{x} - A)]^r}{\Sigma fi}$

$$= \frac{\Sigma fi[(x_i - A) - \mu'_1]^r}{\Sigma fi}$$

$$= \frac{\Sigma fi}{N}\left[(x_i - A)^r - {}^rc_1 (x_i - A)^{r-1}\mu'_1 + {}^rc_2 (x_i - A)^{r-2}(\mu'_1)^2 + \ldots + (-1)^r \cdot (\mu'_1)^r\right]$$

$$= \frac{\Sigma fi(x_i - A)^r}{N} - {}^rc_1 \frac{\Sigma fi(x_i - A)^{r-1}}{N} \mu'_1 \qquad [\because N = \Sigma fi \ \ \bar{x} - A = \mu'_1]$$

$$+ \ldots + (-1)^r \cdot (\mu'_1)^r - \frac{\Sigma fi}{N}.$$

$$\Rightarrow \mu_r = \mu'_r - {}^rc_1 \ \mu'_{r-1} \cdot \mu'_1 + {}^rc_2 \mu'_{r-2}\mu'^2_1 + \ldots + (-1)^r \cdot (\mu'_1)^r$$

which is required relation.

for, $r = 2$, above relation given

$$\mu_2 = \mu'_2 - {}^2C_1\, \mu'_1 \cdot \mu'_1 + {}^2C_2 \cdot \mu'_0 \cdot \mu'^2_1$$

$$\mu_2 = \mu'_2 - 2\mu'^2_1 + \mu'^2_1 \Rightarrow \mu_2 = \mu'_2 - \mu'^2_1.$$

Karl Pearson's β and γ Coefficients

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}, \qquad \gamma_1 = \pm\sqrt{\beta_1},$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2}, \qquad \gamma_2 = \beta_2 - 3$$

Methods of Measurements of Skewness

(1) Karl Pearson's Method
(2) Bowley's Method
(3) Kelly's Method
(4) Method of Moment

(1) Karl Pearson's Method

Karl Pearson's coefficient of skewness,

$$Sk_p = \frac{\text{Mean} - \text{Mode}}{\sigma} = \frac{3\,(\text{Mean} - \text{Median})}{\sigma}$$

$$\frac{3(\text{Mean}-\text{Median})}{\cancel{\sigma}} = \text{Mean}$$

$2\,\text{Mean} = 3\,\text{Median} -$

σ is S.D.

Generally, $-1 < Sk_p < 1$.

For example, if skp = + 0·67 it means the skewness = 0·67 × 100% = 67%

Therefore distribution is 67% positively skewed.

(2) Bowley's Method

Bowley's coefficient of skewness,

$$Sk_B = \frac{Q_1 + Q_3 - 2\text{Median}}{Q_3 - Q_1}$$

Here $-1 \le Sk_B \le +1$.

This method is applicable for continuous distribution with exclusive classes.

Kelly's Method

Kelly's coefficient of skewness,

$$Sk_K = \frac{P_{10} + P_{90} - 2\text{Median}}{P_{90} - P_{10}}$$

This method is only thoretical. Generally Kerl Pearson's method is used.

Example 1. Calculate Karl Pearson Coefficient of skewness from table given below.

Wages of day	55-58	58-61	61-64	64-67	67-70
No. of workers	12	17	23	18	11

Solution. Let A = 62·5

Wages of day	Midvalue x	No. of workers	d = x − A	fd	fd^2	C
55-58	56·5	12	−6	−72	432	12
58-61	59·5	17	−3	−51	153	$\boxed{29}$
$\boxed{61\text{-}64}$	62·5	$\boxed{23}$	0	0	0	53
64-67	65·5	18	3	54	162	70
67-71	68·5	11	6	66	396	81
		Σf = 81		Σfd = − 3	1143	

-Mode / 6

Mean $\bar{x} = a + \frac{\Sigma fd}{\Sigma f} = 62{\cdot}5 + \frac{-3}{81} = 62{\cdot}46$

Mode

Median $= l + \left(\frac{N}{2} - C\right)\frac{i}{f} = 61 + \frac{\frac{81}{2} - 29}{23}(3) = 62{\cdot}5$

S.D., $\sigma = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd}{\Sigma f}\right)^2} = \sqrt{\frac{1143}{81} - \left(\frac{-3}{81}\right)^2} = 3{\cdot}75$ $\quad \sigma = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}}$

∴ Karl Pearson coefficient of skewness,

$$Sk_p = \frac{3\,(\text{Mean} - \text{Median})}{\sigma}$$

$$= \frac{3\,(62{\cdot}46 - 62{\cdot}5)}{3{\cdot}75} = -0{\cdot}032$$

Answer.

Example 2. Calculate Bowley's coefficient of skewness for following table.

C.I.	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of students	2	5	10	15	10	4	2

Solution.

C.I.	f	C
1-10	2	2
10-20	5	7
20-30	10	$\boxed{17}$
30-40	15	32✓
40-50	10	42
50-60	4	46
60-70	2	48
	Σf = 48	

$N = 48,\ \frac{3N}{4} = \frac{3}{4} \times 48 = 36,\ \frac{N}{2} = 24$

∴ Median class is 30-40

$$\text{Median} = l + \left(\frac{N}{2} - C\right)\frac{i}{f}$$

$$= 30 + (24 - 17) \times \frac{10}{15} = \frac{104}{3}$$

$$Q_1 = l + \left(\frac{N}{4} - C\right)\frac{i}{f} = 20 + \left(\frac{12-7}{10}\right)10 = 25$$

$$Q_3 = l + \left(\frac{3N}{4} - C\right)\frac{i}{f} = 40 + \frac{36-32}{10} \times 10 = 44$$

$$\therefore S_{k_B} = \frac{Q_1 + Q_3 - 2\text{Median}}{Q_3 - Q_1}$$

$$= \frac{25 + 44 - 2 \times 104/3}{44 - 25}$$

$$= -0\cdot0175$$

Answer.

Example 3. Calculate Kelly's coefficient of skewness from data given below

Marks obtd.	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
No. of students	3	8	9	14	16	18	8	4

Solution.

Marks	No. of students (f)	Cumulative Frequency (C)
20-30	3	3
30-40	8	11 8 P_{10}
40-50	9	20
50-60	14	34
60-70	16	50 40 P_{50}
70-80	18	68
80-90	8	76 72 P_{90}
90-100	4	80

$N = 80$, $\frac{N}{10} = 8$, C.I. is 30-40 for P_{10}

$\frac{90N}{100} = \frac{80 \times 90}{100} = 72$, C.I. is 80-90 for P_{90}

$\frac{50N}{100} = \frac{N}{2} = \frac{80}{2} = 40$, C.I. is 60-70 for P_{50}

$$\therefore P_{10} = l + \left(\frac{N}{10} - C\right)\frac{i}{f} = 30 + (8-3) \times \frac{10}{8} = 36\cdot25$$

value of l in kelly's method depends upon (N/2) value consists in ~~cumulative frequency~~ (C) class interval

$$P_{90} = l + \left(\frac{9N}{10} - C\right)\frac{i}{f} = 80 + \frac{(72-68)\times 10}{8} = 85$$

$$\text{median } P_{50} = l + \left(\frac{N}{2} - C\right)\frac{i}{f} = 60 + \frac{(40-34)\cdot 10}{16} = 63\cdot 75$$

$$\therefore \text{ Coefficient of skewness, } Sk_k = \frac{P_{10} + P_{90} - 2P_{50}}{P_{90} - P_{10}} = \frac{36\cdot 25 + 85 - 2\times 63\cdot 75}{85 - 36\cdot 25}$$

$$\Rightarrow Sk_k = -0\cdot 128$$

Answer.

EXERCISE

1. Calculate Karl Pearson's coefficient of skewness from the following data

Marks obtd.	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of students	10	40	20	0	10	40	16	14

Ans. 0·75

2. Calculate Karl Pearson's coefficient of skewness from data given below

Scores	0	10	20	30	40	50	60	70	80
No. of Players	150	140	100	80	80	70	30	14	0

Ans. – 0·462

3. Calculate Bowley's coefficient of skewness from following table

Marks obtd.	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of studetns	10	25	20	15	10	35	25	10

Ans. – 0·25

4. Calculate Bowley's coefficient of skewness from following table

Age (in years)	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of participants	358	2417	976	129	62	17	10

Ans. 0·131

5. Calculate Kelly's coefficient of skewness from table below

Age in yrs.	10-12	12-14	14-16	16-18	18-20
No. of studetns	100	16	15	18	8

Ans. 0·373

Kurtosis

Kurtosis : The relative flatness of the top of the frequency distribution curve is called Kurtosis. Actually; Kurtosis measures the degree of peakedness of a frequency distribution.

The coefficient $\gamma_2 = \beta_2 - 3$ and $\beta_2 = \frac{\mu_4}{\mu_2^2}$.

$$\mu_4 = \frac{\Sigma fi(x_i - \bar{x})^4}{\Sigma fi}, \gamma_2 = \frac{\mu_4}{\mu_2^2} - 3, \mu_2 = \frac{\Sigma fi(x_i - \bar{x})^2}{\Sigma fi}$$

$p v^r = k$

$\log p + r \log v = \log k$

$v =$

Answer.

ata given below

0	80-90	90-100
	8	4

e Frequency (C)

3

11 (8) P_{10}

20

34

50 (40) P_{50}

68

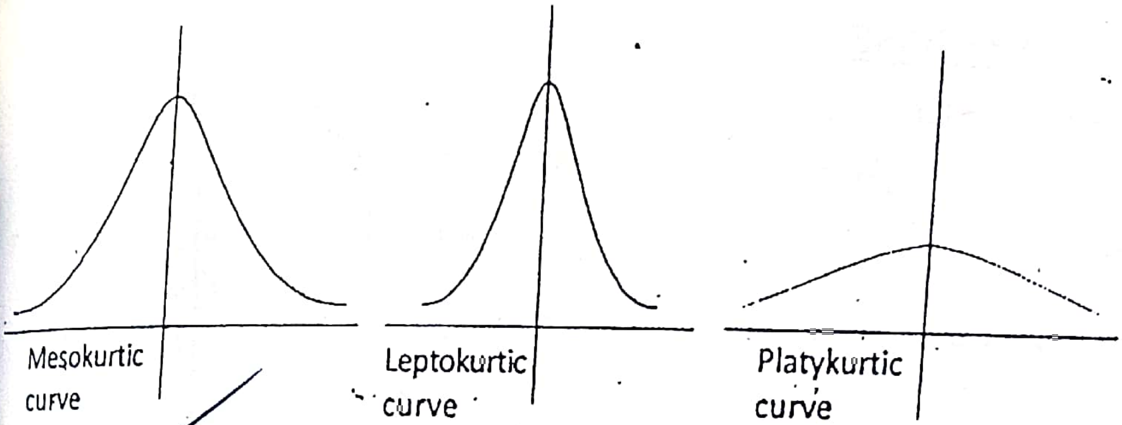
76 (72) P_{90}

80

Case I. If $\gamma_2 = 0$ i.e., if $\beta_2 = 3$ i.e. if $\dfrac{\mu_4}{\mu_2^2} = 3$ then the frequency distribution curve is mesokurtic or normal.

Case II. If $\gamma_2 > 0$ i.e. if $\beta_2 > 3$ i.e. if $\dfrac{\mu_4}{\mu_2^2} > 3$ then the curve is sharply peaked or leptokurtic.

Case III. If $\gamma_2 < 0$ i.e. if $\beta_2 < 3$ i.e. if $\dfrac{\mu_4}{\mu_2^2} < 3$ the curve is flat topped or platykurtic



Example 1. Calculate the coefficient γ_2 and find the nature of following frequency distribution.

Class Interval	0-10	10-20	20-30	30-40
Frequency	1	3	4	2

Solution. Let a = 25

C.I.	f	Mid value (x)	x − a	f (x − a)	f (x − a)^2	f (x − a)^3	f (x − a)^4
0-10	1	5	− 20	− 20	400	− 8000	16000
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
	Σf = 10			− 30	900	− 9000	21000

$$\mu'_1 = \frac{\Sigma f(x-25)}{\Sigma f} = \frac{-30}{10} = -3$$

$\bar{x} = \dfrac{\Sigma x}{\Sigma f_i} = 70$

$$\mu'_2 = \frac{\Sigma f(x-25)^2}{\Sigma f} = \frac{900}{10} = 90$$

$$\mu'_3 = \frac{\Sigma f(x-25)^3}{\Sigma f} = \frac{-9000}{10} = -900$$

$$\mu'_4 = \frac{\Sigma f(x-25)^4}{\Sigma f} = \frac{210000}{10} = 21000$$

Now, $\mu_2 = \mu'_2 - (\mu'_1)^2 = 90 - (-3)^2 = 81$

Moment
Th
about x =

$$\mu_4 = \mu'_4 - 4\,\mu'_3\,\mu'_1 + 6\mu'_2\,\mu_1'^2 - 3\,(\mu'_1)^4 = 21000 - 4\,(-900)\,(-3)$$
$$+ 6\,(90)\,(-3)^2 - 3\,(-3)^4 = 14817$$

$$\therefore\ \gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{14817}{81 \times 81} - 3 = 2{\cdot}258 - 3$$

$$= -0{\cdot}742$$

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

$\therefore\ \gamma_2 = -0{\cdot}742 < 0$ i.e. γ_2 is $-$ve

Therefore, distribution is platykurtic.

Moment Generating Function

The moment generating function, $M_a(t)$ of x about a is defined as

$$M_a(t) = \sum_{i=0}^{\infty} p_i e^{(x_i - a)t}$$

$$= \sum_{i=0}^{\infty} p_i \left[1 + (x_i - a)t + \frac{1}{2!}(x_i - a)^2 t^2 + \ldots\right]$$

$$\Rightarrow M_a(t) = \Sigma p_i + \Sigma p_i\,(x_i - a)\,t + \frac{1}{2}\Sigma p_i\,(x_i - a)^2\,t^2 + \ldots + \frac{1}{r!}\Sigma p_i\,(x_i - a)^r\,t^r + \ldots$$

$$= \frac{\Sigma fi}{\Sigma fi} + \frac{\Sigma fi(x_i - a)t}{\Sigma fi} + \frac{1}{2!}\frac{\Sigma fi(x_i - a)^2}{\Sigma fi}t^2 + \ldots + \frac{1}{r!}\cdot\Sigma fi(x_i - a)^r t^r + \ldots$$

$$= 1 + \mu'_1 \cdot t + \mu'_2\,\frac{t^2}{2!} + \ldots + \mu'_r \cdot \frac{t^r}{r!} + \ldots$$

$$M_a(t) = \mu'_0 + \mu'_1 t + \mu'_2 \cdot \frac{t^2}{2!} + \ldots + \mu'_r \cdot \frac{t^r}{r!} + \ldots$$

15

Now, μ'_r = coefficient of $\frac{t^r}{r!}$

$$\Rightarrow \mu'_r = \left[\frac{d^r}{dt^r} M_a(t)\right]_{t=0}$$

$$\text{Also } M_a(t) = \frac{\Sigma fi \cdot e^{(x_i - a)t}}{\Sigma fi}$$

$$= e^{-at}\,\frac{\Sigma fi e^{x_i t}}{\Sigma fi}$$

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

Moment generating function of a continuous variable function

The moment generating function of the continuous probability distribution about x = a is defined as

$$M_a(t) = \int_{-\infty}^{\infty} e^{t(x-a)} f(x)\, dx$$

Example 1. Find the moment generating function of the function $f(x) = 2e^{-2x}$, $0 \le x \le \infty$.

Solution. The moment generating function about origin is

$$M_0(t) = \int_0^{\infty} e^{tx} \cdot f(x)\, dx$$

$$= \int_0^{\infty} e^{tx} \cdot 2e^{-2x}\, dx$$

$$= 2\int_0^{\infty} e^{(t-2)x}\, dx$$

$$= 2 \cdot \frac{1}{t-2}\left[e^{(t-2)x}\right]_0^{\infty} = \frac{2}{t-2}\{0-1\} = \frac{-2}{-2\left[1-\frac{t}{2}\right]}$$

$$= \left(1-\frac{t}{2}\right)^{-1} = 1 + \frac{1}{2}t + \frac{1}{4}t^2 + \frac{1}{8}t^3 + \dots$$

$$\Rightarrow m_0(t) = 1 + \frac{1}{2}t + \frac{1}{4}t^2 + \frac{1}{8}t^3 + \dots$$

It is required generating function.

Example 2. Find moment generating function of probability distribution

$$f(x) = \begin{cases} k, & 0 < x < k \\ k - x, & k \le x < 2k \\ 0, & \text{otherwise} \end{cases}$$

Solution. $M_x(t) = \int_0^{\infty} e^{xt} \cdot f(x)\, dx$

$$= \int_0^{k} ke^{xt}\, dx + \int_k^{2k} (k-x)e^{xt}\, dx + \int_{2k}^{\infty} 0 \cdot e^{xt}\, dx$$

$$= \left[k\frac{e^{xt}}{t}\right]_0^k + \left[\frac{ke^{xt}}{t} - \frac{xe^{xt}}{t} + \frac{e^{xt}}{t^2}\right]_k^{2k}$$

$$\Rightarrow M_x(t) = \frac{k}{t}\left[e^{kt} - 1\right] + \left[\frac{ke^{2kt}}{t} - \frac{2ke^{2kt}}{t} + \frac{e^{2kt}}{t^2}\right] - \left[\frac{ke^{kt}}{t} - \frac{ke^{kt}}{t} + \frac{e^{kt}}{t^2}\right]$$

$$\Rightarrow M_x(t) = \frac{k}{t}\left[e^{kt}-1\right]+\frac{e^{2kt}}{t^2}-\frac{ke^{2kt}}{t}-\frac{e^{kt}}{t^2}$$

$$= \frac{k}{t}(e^{kt}-1)-\frac{e^{kt}}{t^2}(1-e^{kt})-\frac{k}{t}e^{2kt}$$

$$= (e^{kt}-1)\left(\frac{k}{t}+\frac{e^{kt}}{t^2}\right)-\frac{k}{t}e^{2kt}$$ *Answer.*

EXERCISE

1. Find moment generating function for

$f(x) = xe^{-x}, x > 0$

$= 0,$ otherwise *Ans.* $\frac{1}{1-t}$

2. Find moment generating function for

$f(x) = \frac{1}{k}, k < x < 2k, k > 0$

$= 0,$ otherwise *Ans.* $\frac{1}{kt}\cdot e^{kt}(e^{kt}-1)$.

3. Find moment generating function for density function

$f(x) = 2e^{-2x}, \; x \geq 0$

$= 0, \quad x < 0$ *Ans.* $\frac{2}{2-t}$, for $2 < t$

4. Find moment generating function for

$f(x) = \frac{1}{2x}, \quad$ for $x = 1, 2, 3, \ldots$

$= 0, \quad$ otherwise *Ans.* $\frac{e^t}{2-e^t}$.

Method of Least Squares

Principle of Least Squares

Suppose $y = f(x)$ be the equation of curve to be fitted to the given points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. Let at $x = x_1$, y_1 be observed value of y and $f(x_1)$ is actual or theoretical value of y then difference e_1 is called error given by

$$e_1 = y_1 - f(x_1)$$

Similarly, $e_2 = y_2 - f(x_2)$

$$e_n = y_n - f(x_n)$$

Let $E = e_1^2 + e_2^2 + \ldots + e_n^2$.

The curve of the best fit is the curve for which the sum of the squares for errors, E is minimum. This is called the principle of least squares.

To fit a straight line to the given data using method of least squares :

Let $y = a + bx$ ...(1) be equation of straight line to be fitted to the given points $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$. Let y_r, be the actual or theoretical value of y and y_1 be observed value of y. Then error at $x = x_1$,

$$e_1 = y_1 - y_{r1} = y_1 - (a + bx_1)$$
$$e_2 = y_2 - (a + bx_2)$$
$$\ldots\ldots\ldots\ldots$$
$$e_n = y_n - (a + bx_n)$$

Let $E = e_1^2 + e_2^2 + \ldots + e_n^2 = \sum_{i=1}^{n} e_i^2$

$$\Rightarrow E = \sum_{i=1}^{n}[y_i - (a + bx_i)]^2 \quad \text{or} \quad E = \Sigma\,[y - (a + bx)]^2$$

For E to be minimum,

$$\frac{\partial E}{\partial a} = 2\Sigma\,[y - (a + bx)]\,(-1) = 0$$

and $$\frac{\partial E}{\partial b} = 2\Sigma\,[y - (a + bx)]\,(-x) = 0$$

$$\Rightarrow \quad \Sigma\,[y - (a + bx)] = 0$$

and $$\Sigma\,[xy - (a + bx)\,x] = 0$$

$$\Rightarrow \quad \Sigma y = an + b\Sigma x \quad \ldots(2)$$

and $$\Sigma xy = a\Sigma x + b\Sigma x^2 \quad \ldots(3)$$

Equations (2) and (3) are called normal equtions of the straight line (1).

On solving equation (2) and (3), we get the values of a and b. After putting the values of a and b in eqn. (1), we get the equation of straight line of best fit.

Example. Find the equation of straight line of best fit to the data given in the table :

x	0	1	2	3	4	5
y	2·1	4	6·2	9	12·3	16

Solution. Let $y = a + bx$ be straight line ...(1)

x	y	xy	x^2
0	2·1	0	0
1	4	4	1
2	6·2	12·4	4
3	9	27	9
4	12·3	49·2	16
5	16	80	25
$\Sigma x = 15$	$\Sigma y = 34·6$	$\Sigma xy = 172·6$	$\Sigma x^2 = 55$

$\therefore$ Normal equations are

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

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

which reduce to

$$34{\cdot}6 = 6a + 15b \quad ...(2)$$

$$172{\cdot}6 = 15a + 55b \quad ...(3)$$

solving (2) and (3) for a and b we get

$a = -6{\cdot}53,$ $b = 4{\cdot}92$

Therefore, required straight line of best fit is

$y = -6{\cdot}53 + 4{\cdot}92x$ *Answer.*

EXERCISE

1. Find the equation of straight line of best fit to the following data :

x	– 1	0	1	2	3	4	5	6
y	10	9	7	5	4	3	0	–1

Ans. $y = 8{\cdot}643 - 1{\cdot}607x$

2. Fit a straight line to the following data :

x	1	2	3	4	5	6
y	1200	900	600	200	109	49

Ans. $y = 1362 - 243{\cdot}33\ x$

3. Find a line of best fit to the data below :

x	– 2	– 1	0	1
y	6	3	2	2

Ans. $y = 2{\cdot}61 - 1{\cdot}3x$

4. Fit a line to the data given in table below :

x	50	70	100	120
y	12	15	21	25

Ans. $y = 2{\cdot}28 - 0{\cdot}188x$

5. The pressure and volume of a gas are related by $PV^r = k$, r and k are constants fit this equation to the table below :

P (kg/cm^2)	0·5	1·0	1·5	2·0	2·5	3·0
V (liters)	1·62	1·00	0·75	0·62	0·52	0·45

Ans. $PV^{1{\cdot}275} = 1{\cdot}04$

6. Fit a curve $y = a{\cdot}b^x$ to the data below :

x	2	3	4	5	6
y	8·3	15·4	33	65·2	127·4

Ans. $y = (2{\cdot}04){\cdot}(1{\cdot}995)^x$

7. Fit the eqn. $y = ae^{bx}$ to the data below :

x	1	1·2	1·4	1·6
y	40	73·2	133·4	243

Ans. $\frac{y}{2} = e^{3x}$.

To fit a parabola of second degree tothe given data points using least squares method:

Let $y = a + bx + cx^2$ be the equation of parabola

$\therefore$ Error e_i in the observed and actual value of y is given by

$e_i = y_i - (a + bx_i + cx_i^2)$

$$\Rightarrow e_i^2 = \left[y_i - (a + bx_i + cx_i^2)\right]^2$$

$$\Rightarrow E = \Sigma e_i^2 = \Sigma\, [y - (a + bx + cx^2)]^2$$

For E to be minimum,

$$\frac{\partial E}{\partial a} = 2\Sigma\, [y - (a + bx + cx^2)]\,(-1) = 0$$

$$\frac{\partial E}{\partial b} = 2\Sigma\, [y - (a + bx + cx^2)]\,[-x] = 0$$

$$\frac{\partial E}{\partial c} = 2\Sigma\, [y - (a + bx + cx^2)]\,(-x^2) = 0$$

$$\Rightarrow \Sigma y = an + b\Sigma x + c\Sigma x^2$$

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

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

These last 3 equations are called the normal equations which on solving give the values of a, b and c. Putting the values of a, b and c, $y = a + bx + cx^2$, we get the required equation of parabola.

Example 1. Find the equation of parabola of best fit to the following data :

x	– 2	–1	0	1	2
y	10	2	1	4	20

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

be eqn. of second degree parabola.

x	y	xy	x^2	x^2y	x^3	x^4
– 2	10	– 20	4	40	– 8	16
– 1	2	– 2	1	2	– 1	1
0	1	0	0	0	0	0
1	4	4	1	4	1	1
2	20	40	4	80	8	16
$\Sigma x = 0$	$\Sigma y = 37$	22	10	126	0	34

The normal equations are :

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

$\Rightarrow$ $37 = 5a + 0 + 10c$ ...(2)

$22 = 10b$ ...(3)

$126 = 10a + 0 + 34c$ ...(4)

(3) $\Rightarrow$ $b = 2{\cdot}2$

(2) & (4) $\Rightarrow$ $a = \dfrac{-8}{35}$, $c = \dfrac{26}{7}$

$\therefore$ Required eqn. is $y = \dfrac{-8}{35} + 2{\cdot}x + \dfrac{26}{7}x^2$ *Ans.*

EXERCISE

1. Fit a parabola to the following data :

x	0	1	2	3	4
y	−4	−1	4	11	19

Ans. $y = x^2 + 2x - 4$.

2. Find eqn. of parabola of second degree to the data given below :

x	1	2	3	4
y	6	11	18	27

Ans. $y = x^2 + 2x + 3$

3. Fit a second degree parabola to the data given below using least squares method.

x	0	1	2	3	4
y	1	1·8	1·4	2·5	6·4

Ans. $y = 1{\cdot}42 - 1{\cdot}69x + {\cdot}55x^2$

4. Fit a second degree parabola to the data given below :

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

Ans. $y = -1{\cdot}1 + 3{\cdot}56x - 0{\cdot}267x^2$.

Correlation and Regression

Covariance

Let $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ be the ordered pairs corresponding to variables x and y then covariance of x, y is defined by

$$\text{Cov}(x, y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n}$$

$a = \sqrt{\frac{\Sigma f d^2}{N} - \left(\frac{\Sigma f d}{N}\right)^2}$

Correlation : There is a relation between variables x and y such that an increase in x correspond to increase or decrease in y then x and y are said to be correlated.

Types of Correlation

(1) Positive Correlation : If an increase in x results in a corresponding increase in y then it is called positive correlation. Also, if decrease in x corresponds to decrease in y then it is positive correlation.

(2) Negative Correlation : If a decrease in x corresponds to increase in y and vice-versa then it is negative correlation.

Karl Pearson Coefficient of Correlation

Karl Pearson coefficient of correlation r between x and y in defined as

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

Also, $(1) \Rightarrow r = \dfrac{\frac{1}{n}\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\frac{1}{n}\cdot\Sigma(x-\bar{x})^2}\sqrt{\frac{1}{n}\cdot\Sigma(y-\bar{y})^2}}$

$$\Rightarrow r = \frac{\text{cov}(x,y)}{\sigma x \cdot \sigma y} \quad ...(2)$$

Again, $\Sigma(x-\bar{x})(y-\bar{y}) = \Sigma(xy + \bar{x}\bar{y} - x\bar{y} - y\bar{x})$

$= \Sigma xy + \bar{x}\bar{y}\Sigma - \bar{y}\Sigma x - \bar{x}\Sigma y$

$= \Sigma xy + n\bar{x}\bar{y} - n\bar{x}\bar{y} - n\bar{x}\bar{y}$

$= \Sigma xy - n\bar{x}\bar{y}$

$= \Sigma xy - n\cdot\dfrac{\Sigma x}{n}\cdot\dfrac{\Sigma y}{n}$

$$\Rightarrow \Sigma(x-\bar{x})(y-\bar{y}) = \frac{n\Sigma xy - \Sigma x\Sigma y}{n} \quad ...(3)$$

Also, $\Sigma(x-\bar{x})^2 = \Sigma(x^2 + \bar{x}^2 - 2x\bar{x})$

$= \Sigma x^2 + \bar{x}^2\cdot n - 2\bar{x}\cdot n\bar{x}$

$= \Sigma x^2 - n\bar{x}^2$

$= \Sigma x^2 - n\cdot\left(\dfrac{\Sigma x}{n}\right)^2$

$$\Rightarrow \quad \Sigma(x-\bar{x})^2 = \frac{n\Sigma x^2 - (\Sigma x)^2}{n} \quad ...(4)$$

Similarly, $\Sigma(y-\bar{y})^2 = \dfrac{n\Sigma y^2 - (\Sigma y)^2}{n}$ ...(5)

$$\therefore\ r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}}$$

$$\Rightarrow r = \frac{n\Sigma xy - \Sigma x\Sigma y}{\sqrt{n\Sigma x^2-(\Sigma x)^2}\sqrt{n\Sigma y^2-(\Sigma y)^2}}, \text{ using (3), (4), (5)} \quad \ldots(6)$$

Example 1. Calculate correlation coefficient between x and y using following data :

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

Solution. Here $\bar{x} = \frac{\Sigma x}{n} = \frac{65}{5} = 13$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{125}{5} = 25$$

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
5	12	−8	−13	64	169	104
9	20	−4	−5	16	25	20
13	25	0	0	0	0	0
17	33	4	8	16	64	32
21	35	8	10	64	100	80
65	125	0	0	160	358	236

$$\therefore\ r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}} = \frac{236}{\sqrt{160}\sqrt{358}}$$

$r = 0{\cdot}9861$ ***Ans.***

Spearman's Rank Correlation Coefficient

Spearmen's rank correlation coefficient,

$r = 1 - \frac{6\Sigma d^2}{n(n^2-1)}$, where n is the number of individuals and d is the difference in ranks in two characteristics.

Proof. Let (x_1, y_1) (x_2, y_2), ..., (x_n, y_n) be the ranks of n individuals for two characteristics. Obviously in each characteristic every individuals gains the values 1, 2, 3, ..., n as their ranks and therefore their arithmetic mean $\bar{x}$ & $\bar{y}$ are equal.

$$\therefore\ \bar{x} = \frac{\Sigma x}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2}, \bar{y} = \bar{x} = \frac{n+1}{2}.$$

Also, $\Sigma x^2 = 1^2 + 2^2 + \ldots, n^2 = \frac{n(n+1)(2n+1)}{6} = \Sigma y^2$

$$\Sigma (x-\bar{x})^2 = \Sigma [x^2 + (\bar{x})^2 - 2x\bar{x}]$$

$$= \Sigma x^2 + n(\bar{x})^2 - 2\bar{x}\,\Sigma x$$

$$= \Sigma x^2 + n(\bar{x})^2 - 2\bar{x}\cdot n\bar{x}$$

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

$$= \frac{n(n+1)(2n+1)}{6} - n\cdot\left(\frac{n+1}{2}\right)^2$$

$$= \frac{(n+1)n}{2}\left[\frac{2n+1}{3} - \frac{n+1}{2}\right]$$

$$= \frac{n(n+1)}{2}\left[\frac{n-1}{12}\right] = \frac{1}{12}n(n^2-1) \quad \ldots(1)$$

Similarly, $\Sigma (y-\bar{y})^2 = \frac{1}{12}n(n^2-1)$ ...(2)

Also, $\frac{1}{2}\Sigma d^2 = \frac{1}{2}\Sigma (x-y)^2 = \frac{1}{2}\Sigma [(x-\bar{x})-(y-\bar{y})]^2$

$$= \frac{1}{2}\Sigma \left[(x-\bar{x})^2 + (y-\bar{y})^2 - 2(x-\bar{x})(y-\bar{y})\right]$$

$$= \frac{1}{2}\Sigma(x-\bar{x})^2 + \frac{1}{2}\Sigma(y-\bar{y})^2 - \Sigma(x-\bar{x})(y-\bar{y})$$

$$= \frac{1}{2}\cdot\frac{n(n^2-1)}{12} + \frac{1}{2}\cdot\frac{n(n^2-1)}{12} - \Sigma(x-\bar{x})(y-\bar{y})$$

$$\therefore \Sigma (x-\bar{x})(y-\bar{y}) = \frac{n(n^2-1)}{12} - \frac{1}{2}\Sigma d^2 \quad \ldots(3)$$

using (1), (2) and (3) in $r = \dfrac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}}$

We get, $$r = \frac{\dfrac{n(n^2-1)}{12} - \dfrac{1}{2}\Sigma d^2}{\sqrt{\dfrac{n(n^2-1)}{12}}\sqrt{\dfrac{n(n^2-1)}{12}}}$$

$$\Rightarrow r = \frac{\dfrac{n(n^2-1)}{2} - \dfrac{1}{2}\Sigma d^2}{\dfrac{n(n^2-1)}{12}}$$

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

Example 1. Calculate Spearman's rank correlation coefficient for the table below.

Individual	A	B	C	D	E	F
Rank in Beauty	1	2	3	4	5	6
Rank in Honesty	3	1	2	6	4	5

Solution. Here n = 6.

Individual	Rank in Beauty (R_1)	Rank in Honesty (R_2)	$d = R_1 - R_2$	d^2
A	1	3	−2	4
B	2	1	1	1
C	3	2	1	1
D	4	6	2	4
E	5	4	1	1
F	6	5	1	1
	21	21		$\Sigma d^2 = 12$

$$\therefore \quad r = 1 - \frac{6\Sigma d^2}{n(n^2-1)} = 1 - \frac{6 \times 12}{6(36-1)}$$

$$= 1 - \frac{12}{35}$$

$$= \frac{23}{35} = 0.657$$ ***Ans.***

Result 1. Show that the correlation coefficient

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

Proof. We have $\sigma_x = \sqrt{\dfrac{\Sigma(x-\bar{x})^2}{n}}$

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

Similarly, $$\sigma_y^2 = \frac{\Sigma(y-\bar{y})^2}{n}$$

Now, $$\sigma_{x-y} = \sqrt{\frac{\Sigma[(x-y)-(\overline{x-y})]^2}{n}}$$

$$\Rightarrow \sigma_{x-y}^2 = \frac{\Sigma[(x-y)-(\overline{x-y})]^2}{n}$$

$$= \frac{\Sigma[(x-y)-(\bar{x}-\bar{y})]^2}{n}$$

$$= \frac{\Sigma[(x-\bar{x})-(y-\bar{y})]^2}{n}$$

$$\Rightarrow \sigma_{x-y}^2 = \frac{\Sigma(x-\bar{x})^2}{n} + \frac{\Sigma(y-\bar{y})^2}{n} - \frac{2\Sigma(x-\bar{x})(y-\bar{y})}{n}$$

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

$$\left[\because r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n\sigma_x\sigma_y}\right]$$

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

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

Proved.

EXERCISE

1. Calculate coefficient of correlation between x and y from table given below :

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

Ans. 0·986

2. Compute the correlation coefficient from the data given below :

x	1	2	3	4	5	6	7	8
y	8	7	10	12	11	13	14	17

Ans. 0·947

3. Calculate coefficient of correlation from following data :

x	3	4	6	8	10	12
y	1	2	3	4	6	10

Ans. 0·948

4. Find the rank coefficient between beauty and honesty of 10 ladies to the following data :

Ladies	A	B	C	D	E	F	G	H	I	J
Rank in Beauty	6	4	3	1	2	7	9	8	10	5
Rank in Honesty	4	1	6	7	5	8	10	9	3	2

Ans. 0·224

Regression

Regression Curve : The dots of the scatter diagram are concentrated round a curve which is called regression curve

Line of Regression : A line of regression in the straight line which gives the best fit in the least square sence of the given data points.

To find equation of line of regression:

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

be eqn. of line regression of y on x.

Let E be the sum of distances between points of scatter diagram and points on line.

$$E = \Sigma (y - a - bx)^2$$

By principle of least squares

$$\frac{\partial E}{\partial a} = 2\Sigma (y - a - bx) = 0$$

and $$\frac{\partial E}{\partial b} = -2\Sigma [y - (ax + bx^2)] = 0$$

$$\Rightarrow \quad \Sigma y = an + b\Sigma x \qquad ...(2)$$

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

By (2), $$\frac{\Sigma y}{n} = a + b\frac{\Sigma x}{n}$$

$$\Rightarrow \quad \bar{y} = a + b\bar{x}$$

$\Rightarrow$ $(\bar{x}, \bar{y})$ lies on regression line $y = a + bx$.

Now shifting the origin to $(\bar{x}, \bar{y})$, eqn. (3) becomes,

$$\Sigma (x - \bar{x})(y - \bar{y}) = a\Sigma (x - \bar{x}) + b\Sigma (x - \bar{x})^2$$

$$\Rightarrow \quad \Sigma (x - \bar{x})(y - \bar{y}) = a [\Sigma x - \Sigma \bar{x}] + b\Sigma (x - \bar{x})^2$$

$$\Rightarrow \quad \Sigma (x - \bar{x})(y - \bar{y}) = a [n\bar{x} - n\bar{x}] + b\Sigma (x - \bar{x})^2$$

$$\Rightarrow \quad \Sigma (x - \bar{x})(y - \bar{y}) = b\,\Sigma (x - \bar{x})^2$$

$$\Rightarrow \quad r \cdot n\, \sigma_x \cdot \sigma_y = b \cdot n \cdot \sigma_x^2$$

$$\Rightarrow \quad r \cdot \sigma_y = b \cdot \sigma_x \left[\because r = \frac{\Sigma (x - \bar{x})(y - \bar{y})}{n\sigma_x \sigma_y} \text{ and } \sigma_x^2 = \frac{\Sigma (x - \bar{x})^2}{n}\right]$$

$\Rightarrow$ $b = r \cdot \dfrac{\sigma_y}{\sigma_x}$, b is slope of regression line.

Now, equation of regression line through $(\bar{x}, \bar{y})$ is given by

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

$\Rightarrow$ $y - \bar{y} = r \cdot \dfrac{\sigma_y}{\sigma_x}(x - \bar{x})$ which is required line of regression.

where $b_{yx} = r \dfrac{\sigma_y}{\sigma_x}$ is called the coefficient of regression of y on x.

Similarly, $b_{xy} = r \cdot \dfrac{\sigma_x}{\sigma_y}$.

Now $b_{yx} \cdot b_{xy} = r \cdot \dfrac{\sigma_y}{\sigma_x} \cdot r \cdot \dfrac{\sigma_x}{\sigma_y}$

$\Rightarrow \quad b_{yx} \cdot b_{xy} = r^2$

$\Rightarrow \quad r = \sqrt{b_{xy} \cdot b_{yx}}$

Result : If θ is angle between regression line of y on x and regression line of x on y then $\tan\theta = \dfrac{1-r^2}{r} \cdot \dfrac{\sigma_x \cdot \sigma_y}{\sigma_x^2 + \sigma_y^2}$.

Proof. The equation of line of regression of y on x is

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

slope of line (1) is $m_1 = r \cdot \dfrac{\sigma_y}{\sigma_x}$.

The equation of line of regression of x on y is

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

$$\Rightarrow \quad y - \bar{y} = \frac{1}{r} \cdot \frac{\sigma_y}{\sigma_x}(x - \bar{x}) \qquad ...(2)$$

slope of (2) is $m_2 = \dfrac{1}{r} \cdot \dfrac{\sigma_y}{\sigma_x}$

$$\therefore \quad \tan\theta = \frac{\frac{1}{r} \cdot \frac{\sigma_y}{\sigma_x} - r\frac{\sigma_y}{\sigma_x}}{1 + \frac{1}{r} \cdot \frac{\sigma_y}{\sigma_x} \cdot r \cdot \frac{\sigma_y}{\sigma_x}}$$

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

$$\Rightarrow \tan\theta = \left(\frac{1-r^2}{r}\right)\cdot\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2} \qquad ...(3)$$

Case I. If r = 0 then (3) $\Rightarrow \tan\theta = \infty$

$\Rightarrow \theta = \pi/2$

$\Rightarrow$ Two lines are perpendicular and there is no relation between x and y.

Case II. If $r = \pm 1$, eqn. (3) gives

$$\tan\theta = 0$$

$$\Rightarrow \quad \theta = 0$$

$\Rightarrow$ regression lines are coincident or parallel and the correlation between x and y is perfect.

Example 1. The regression line of y on x is $2x - 9y = -6$ and line of x on y is $x - 2y = -1$. Calculate correlation coefficient r between x and y.

Solution. $2x - 9y = -6$

$$\Rightarrow y = \frac{2x+6}{9} \quad \Rightarrow y = \frac{2}{9}x + \frac{2}{3} \quad \Rightarrow b_{yx} = \frac{2}{9}$$

The second line $x - 2y = -1$

$\Rightarrow x = 2y - 1$

$\Rightarrow b_{xy} = 2$

$$\therefore \; r = \sqrt{b_{yx}\cdot b_{xy}} = \sqrt{\frac{2}{9}\cdot 2} = \frac{2}{3} \qquad \textbf{Ans. } r = \frac{2}{3}$$

Example 2. Two lines of regression are

$4y - 7x + 18 = 0$ and $y - 3x + 11 = 0$

Find $\bar{x}$ and $\bar{y}$ and r.

Solution. Since $(\bar{x}, \bar{y})$ lie on both lines of regression therefore

$$4\bar{y} - 7\bar{x} = -18$$

and
$$\bar{y} - 3\bar{x} = -11$$

$$\Rightarrow \quad \bar{x} = 5\cdot 2, \bar{y} = 4\cdot 6$$

$$4y = 7x - 18 \qquad \Rightarrow y = \frac{7}{4}x - 4\cdot 5 \qquad ...(1)$$

$$y - 3x + 11 = 0 \qquad \Rightarrow x = \frac{y}{3} + 1\frac{11}{3} \qquad ...(2)$$

$$\therefore \quad b_{yx} = \frac{7}{4}, b_{xy} = \frac{1}{3}$$

$$\therefore \quad r = \sqrt{b_{yx}\cdot b_{xy}} = \sqrt{\frac{7}{4}\times\frac{1}{3}} = \sqrt{\frac{17}{12}} = \sqrt{\cdot 5833} = 0\cdot 76$$

Example 3. The two regression lines are given as

$$8x - 10y + 66 = 0 \text{ and } 40x - 18y = 214$$

If $\sigma_x = 3$ then find $\bar{x}, \bar{y}$ and r.

Solution. Since $(\bar{x}, \bar{y})$ is common point of regression lines,

$$\therefore \quad 8\bar{x} - 10\bar{y} = -66$$

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

which on solving yield $\bar{x} = 13$, $\bar{y} = 17$.

Now, $8x - 10y = -66 \quad \Rightarrow 10y = 8x + 66$

$$\Rightarrow b_{yx} = \frac{8}{10}$$

$$40x - 18y = 214$$

$$\Rightarrow x = \frac{18}{40} y + \frac{214}{40}$$

$$\Rightarrow b_{xy} = \frac{9}{20}$$

$$\therefore \quad r = \sqrt{b_{yx} \cdot b_{xy}} = \sqrt{\frac{8}{10} \times \frac{9}{20}} = \sqrt{\frac{36}{100}} = 0{\cdot}6$$

Also, $b_{yx} = 0{\cdot}8$

$$\Rightarrow r \cdot \frac{\sigma_y}{\sigma_x} = 0{\cdot}8 \qquad \left[\because b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x}\right]$$

$$\Rightarrow 0{\cdot}6 \times \frac{\sigma_y}{3} = 0{\cdot}8$$

$$\Rightarrow \sigma_y = \frac{{\cdot}8}{{\cdot}6} \times 3 = 4$$

Ans. $\bar{x} = 13$, $\bar{y} = 17$, $\sigma_y = 4$, $r = 0{\cdot}6$

EXERCISE

Find the coefficient of correlation and regression lines from following data :

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

Ans. $r = -0{\cdot}919$, $y = -0{\cdot}65x + 11{\cdot}9$, $x = -1{\cdot}3y + 16{\cdot}4$

Find the eqn. of regression line of x on y using data given below :

x	2	4	6	8	10
y	12	10	8	6	4

Ans. $y = \frac{1}{3}x - \frac{1}{3}$

3. Find regression line of y on x :

x	1	2	3	4	5
y	3	2	5	1	4

Ans. $x - 10y + 270 = 0$

4. Find regression line of y on x :

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

Ans. $6y = 3{\cdot}4 + 7x$

5. Obtain the line of regression :

x	10	12	13	16	17	20	25
y	10	22	24	27	29	33	38

Ans. $y = {\cdot}83 + 1{\cdot}55x$

6. If regression coefficients are 0·9 and 0·4. Find correlation coefficient r.

Ans. 0·6

7. If regression lines are $7x - 16y = -9$ and $5y - 4x = 3$. Find r, $\bar{x}, \bar{y}$.

Ans. $r = 0{\cdot}75$, $\bar{x} = 0{\cdot}1$, $\bar{y} = 0{\cdot}5$

8. Two regression lines are $20x - 9y = 107$ and $4x - 5y = -33$. Variance of $x = 9$. Calculate $\bar{x}, \bar{y}$ and σ_y.

Ans. $\bar{x} = 13$, $\bar{y} = 17$, $\sigma_y = 4$.

9. The two regression lines are $3x + 3y = 26$ and $6x + y = 31$. Find $\bar{x}, \bar{y}$ and r.

Ans. $\bar{x} = 4$, $\bar{y} = 7$, $r = \frac{1}{2}$.

10. The two regression lines are $x = 19{\cdot}13 - {\cdot}87y$ and $y = 11{\cdot}64 - {\cdot}50x$. Calculate $\bar{x}, \bar{y}$ and correlation coefficient

Ans. $\bar{x} = 15{\cdot}94$, $\bar{y} = 3{\cdot}67$, $r = \pm 0{\cdot}66$.

11. Prove that airthmetic mean byx and bxy in greater then r.

12. Find the regression line of y on x :

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

Ans. $11y = 7x + 6$.