UNIT – II

Skewness

Def: Literally Skewness means 'lack of symmetry'. We study the skewness to have an idea about the shape of the curve which we can draw with the help of the given data. A distribution is said to be skewed if

i) Mean, median and mode fall at different points.

ii) Quartiles are not equidistant from median;

iii) The curve drawn with the help of the given data is not symmetrical but stretched more to one side than to the other.

Measures of Skewness.

i) $Sk = M - Md$

ii) $Sk = M - Mo$

iii) $Sk = (Q_3 - Md) - (Md - Q_1)$

where M, Mo, Md are mean, mode and median respectively.

Coefficient of Skewness

i) Karl Pearson coefficient of Skewness

$$Sk = \frac{\text{Mean} - \text{Mode}}{\sigma}$$

$$= \frac{3(M - Md)}{\sigma}$$

ii) Bowley's coefficient of Skewness

$$Sk = \frac{Q_3 + Q_1 - 2Md}{Q_3 - Q_1}$$

Examples.

1. Compute Pearson's coefficient of Skewness for the following data:

25 15 23 40 27 25 23 25 20

Sol:

										Total
X	25	15	23	40	27	25	23	25	20	223
X^2	625	225	529	1600	729	625	529	625	400	5887

$$\bar{x} = \frac{\Sigma x}{n} = \frac{223}{9} = 24.78$$

$$S.D = \sigma = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2}$$

$$= \sqrt{\frac{5887}{9} - (24.78)^2}$$

$$= 6.33$$

Mode = 25

$\therefore$ Pearson's coeff. of Skewness $= \dfrac{\text{Mean} - \text{Mode}}{\sigma}$

$$= \frac{24.78 - 25}{6.33} = -0.0348$$

2. Find Karl Pearson coefficient of Skewness.

x :	12	15	20	25	30	40	50
f :	10	25	40	70	32	13	10

Sol: Find Mean $= \dfrac{\Sigma xf}{N} = 25.13$

$$S.D = \sigma = \sqrt{\frac{\Sigma fx^2}{N} - \bar{x}^2} = 8.71$$

Mode = 25

$$\therefore Sk = \frac{\text{Mean} - \text{Mode}}{\sigma} = \frac{25.13 - 25}{8.71}$$

$$= 0.0149.$$

3. For the following frequency distribution, find karl Pearson's coefficient of skewness:

C.I :	10-20	20-30	30-40	40-50	50-60
f :	18	20	30	22	10

Sol: Find $\bar{x} = \dfrac{\Sigma xf}{N} = 33.6$

$$S.D = \sigma = \sqrt{\frac{\Sigma fx^2}{N} - \bar{x}^2} = 12.33$$

$$\text{Mode} = l + \frac{h(f_1 - f_0)}{2f_1 - f_0 - f_2} = 35.56$$

$$\therefore Sk = \frac{\text{Mean} - \text{Mode}}{\sigma} = \frac{33.6 - 35.56}{12.33}$$

$$= -0.1590$$

C-I	f	cf	
18-24	18	18	
24-30	22	40	
30-36	40	80	$\rightarrow Q_1$
36-42	50	130	$\rightarrow Md$
42-48	38	168	$\rightarrow Q_3$
48-54	12	180	
54-60	4	184	

Q_1: $\frac{N}{4} = \frac{184}{4}$

$$Q_1 = l + \frac{h}{f}(N/4 - c)$$

$$= 30 + \frac{6}{40}\left(\frac{184}{4} - 40\right)$$

$$= 30.9$$

Md: $\frac{N}{2} = \frac{184}{2}$

$$Md = l + \frac{h}{f}\left(\frac{N}{2} - c\right)$$

$$= 36 + \frac{6}{50}\left(\frac{184}{2} - 80\right)$$

$$= 37.44$$

Q_3: $\frac{3N}{4} = \frac{3 \times 184}{4}$

$$Q_3 = l + \frac{h}{f}\left(\frac{3N}{4} - c\right)$$

$$= 42 + \frac{6}{38}\left(3 \times \frac{184}{4} - 130\right)$$

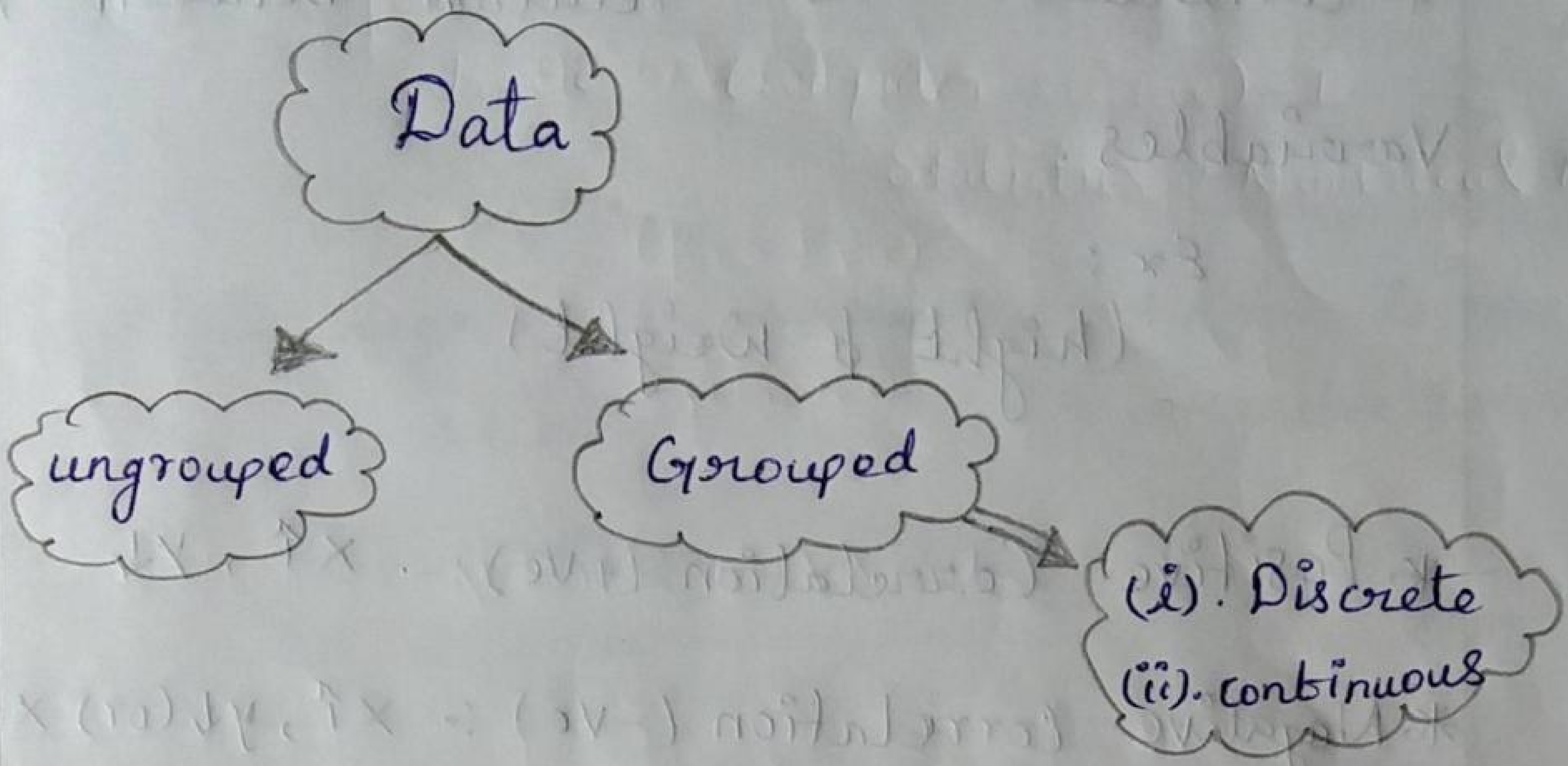
$$= 42.95$$

$$Sk = \frac{Q_3 + Q_1 - 2Md}{Q_3 - Q_1} = \frac{42.95 + 30.9 - 2(37.44)}{42.95 - 30.9}$$

$$= -0.0085$$

---o---

Statistics & Numerical Methods.



Ungrouped :

Data is without any alternation.

Ex :
Marks obtained in a particular subject of I - Bsc. Computer Science.

50, 60, 40, 15, 20, 25 . . .

Grouped :

(i). Discrete:

Values (x)	10	20	30
frequency (F)	1	5	6

(ii). continuous

class	10-20	20-30	30-40
Frequency	3	12	7

Correlation

* Correlation is relation between two Varriables.

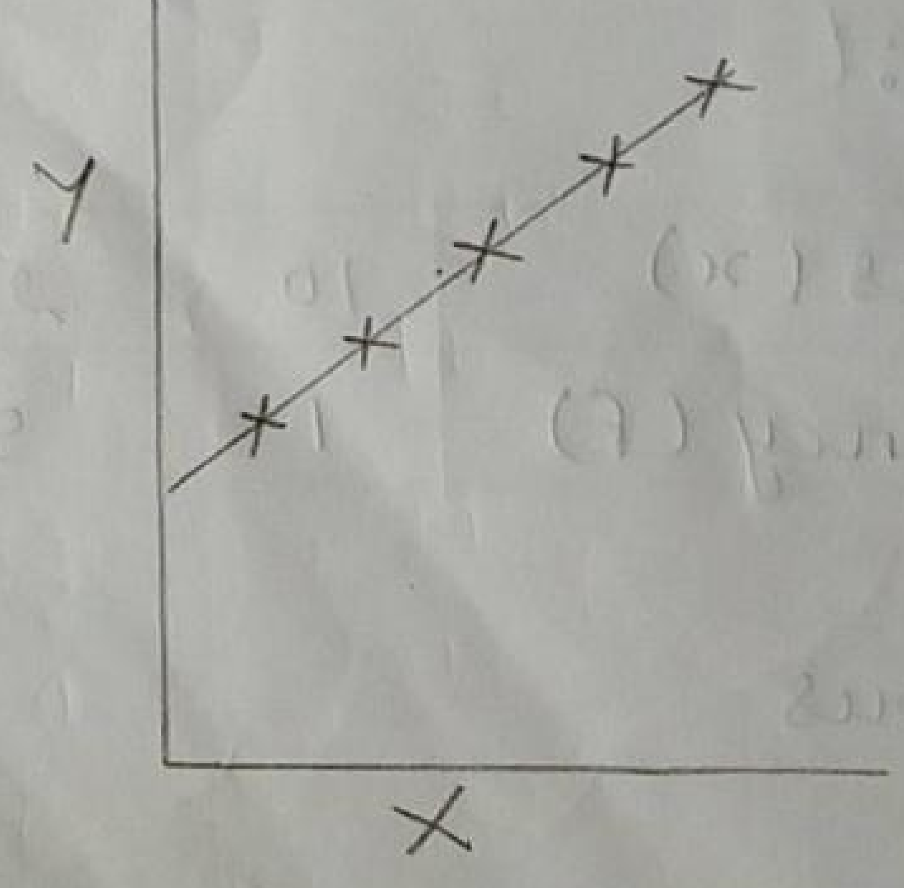
  Ex:
  (hight & Weight)

* Positive Correlation (+Ve) :- X↑, Y↓
* Negative correlation (-ve) :- X↑, Y↓ (or) X↓, Y↑

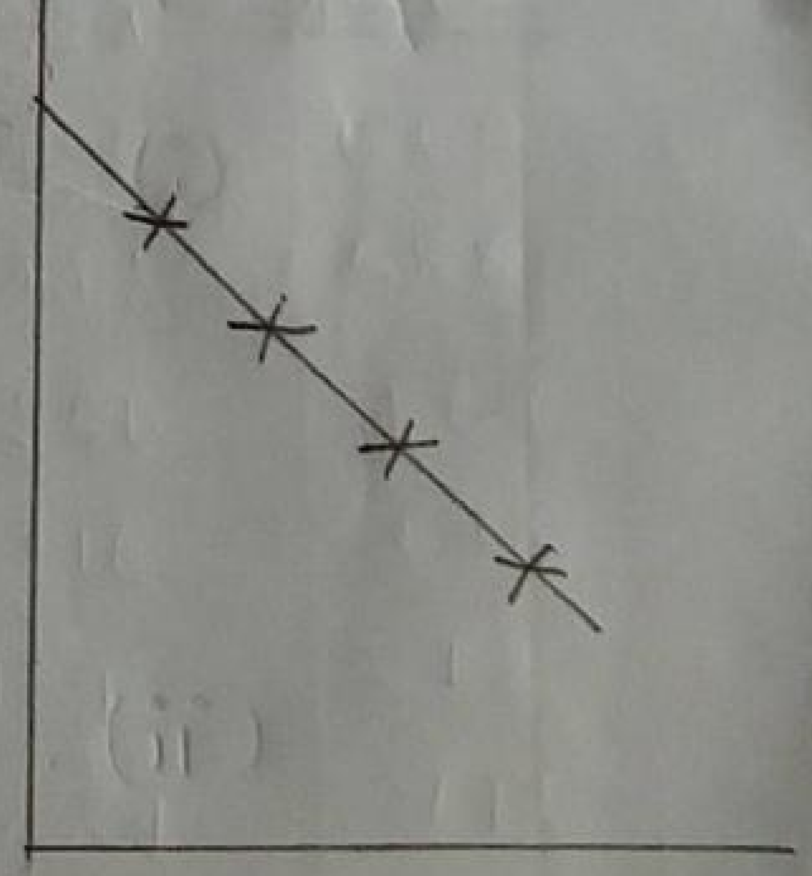
  Ex:
  Price & demand

* Correlation can be Studied using Various Methods they are :-
  1). Scatter diagram
  2). Karl pearson's Coeffcient of correlation
  3). Spearman's rank Correlation Coffcient.

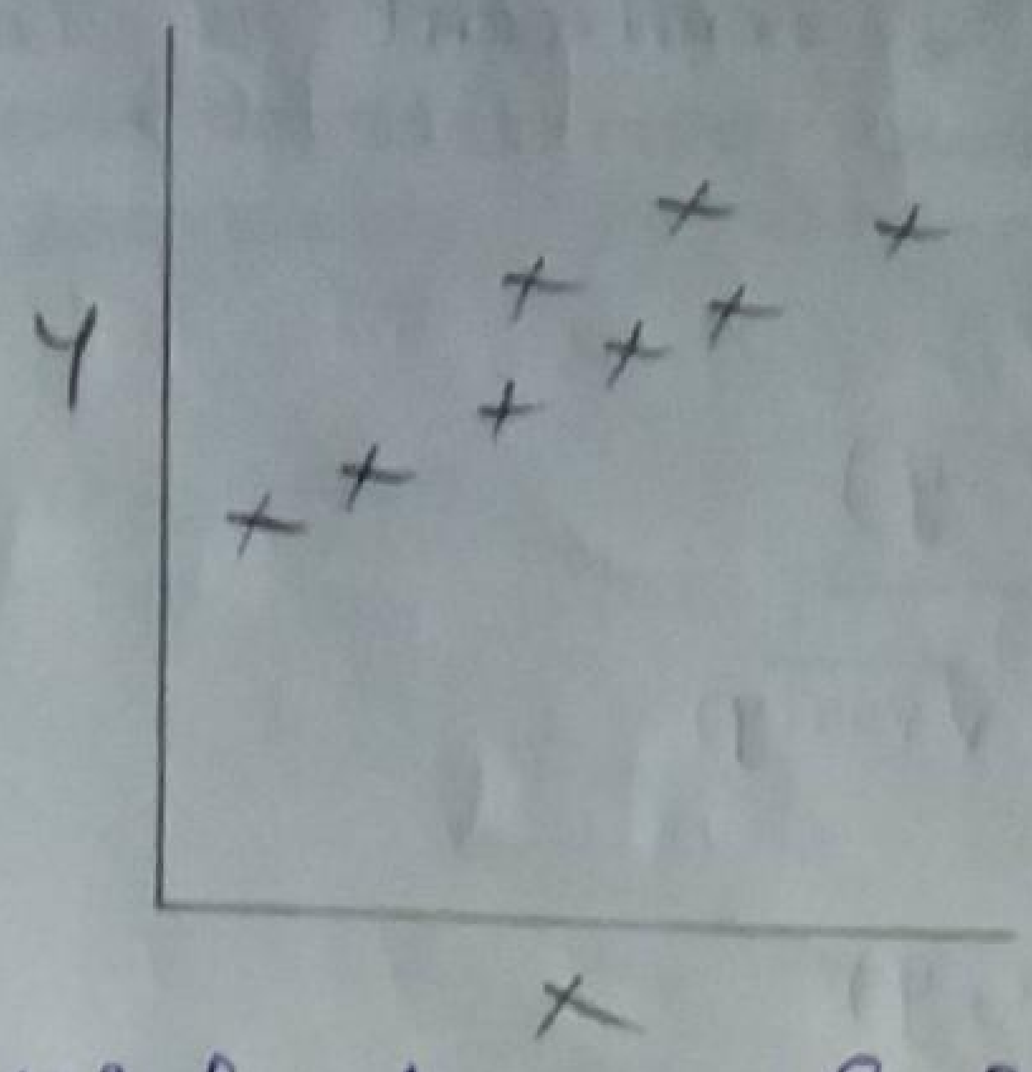
Scatter Diagram :



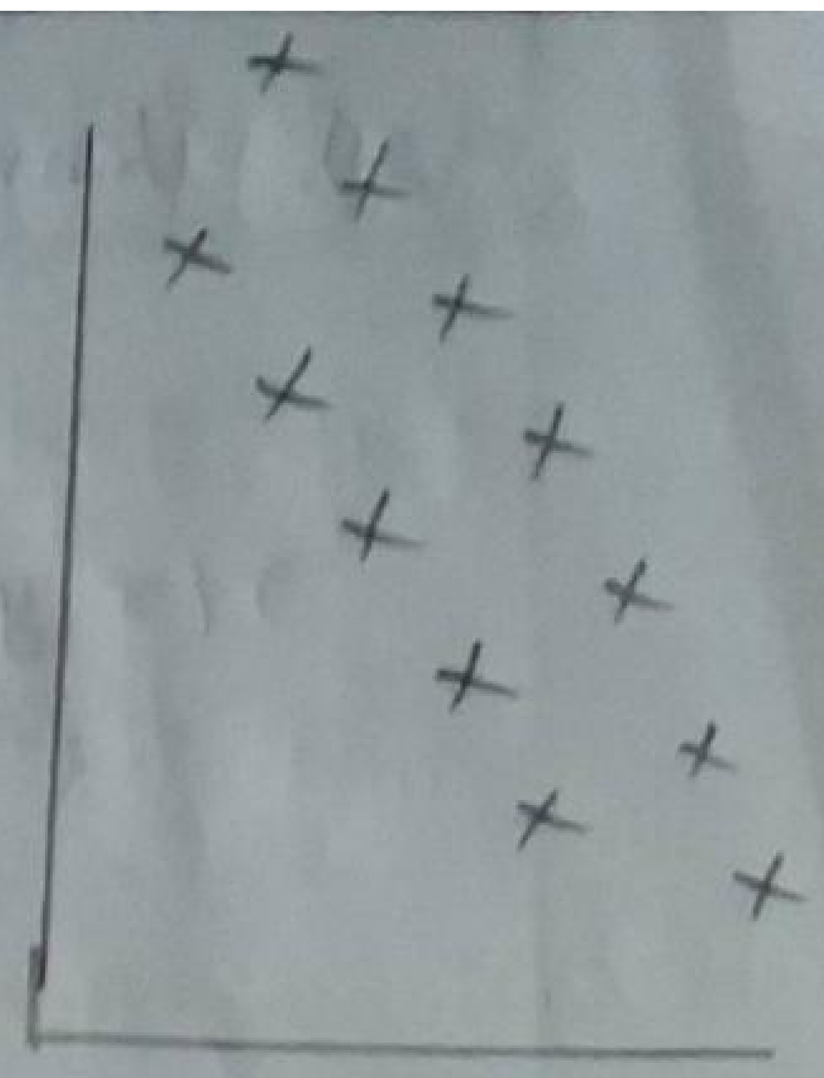
Perfect Positive Correlation



Perfect Negative Correlation.



High degree of Positive Correlation



high degree of Negative Correlation

No correlation ($r=0$)

<u>Properties of Correlation Coefficent :</u>

The coefficient of correlation lies between -1 and $+1$ or $|r| \leq 1$.

* If $r=1$, then there is a Perfect (+ve) Correlation.
* If $r=-1$, then there is a Perfect Negative correlation.
* If $r=0$, then the Variable are uncorrelated.

2). Karl Pearson's Coefficent of Correlation

$$r(x \cdot y) = \frac{Cov(x \cdot y)}{\sqrt{var(x)}\sqrt{var(y)}}$$

$$= \frac{Cov(x, y)}{\sigma x \cdot \sigma y}$$

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

Solved Problem 5.5

Calculate Coefficent of Correlation between x and y Karl Pearson's method :

x : 25 30 28 29 32 24 36 28 27 21

y : 18 20 21 16 14 13 22 15 19 12

Solution:

x	y	x^2	y^2	xy
25	18	625	324	450
30	20	900	400	600
28	21	784	441	588
29	16	841	256	464
32	14	1024	196	448
24	13	576	169	312
36	22	1296	484	792
28	15	784	225	420
27	19	729	361	513
21	12	441	144	252
$\Sigma x = 280$	$\Sigma y = 140$	$\Sigma x^2 = 8000$	$\Sigma y^2 = 3000$	$\Sigma xy = 4839$

Scanned by TapScanner

$n = 10$

According to karl Pearson's Co-efficent of Correlation :

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

$\Rightarrow$ By Substituting the know Values.

$$r = \frac{10(4839) - ((280)(170))}{\sqrt{10(8000 - (280)^2}\sqrt{10(3000) - (170)^2}}$$

$$r = \frac{48390 - 47600}{\sqrt{10(8000) - 78400}\sqrt{10(3000) - 28900}}$$

$$r = \frac{790}{\sqrt{1600}\sqrt{1100}}$$

$$r = \frac{790}{40 \times 33.16}$$

$$r = \frac{790}{1326.4}$$

$$\boxed{r = 0.5955.}$$

Scanned by TapScanner

Calculate the Correlation co-efficent for the following heights (in inches) of fathers (x) and their sons (y).

X :	65	66	67	67	68	69	70	72
Y :	67	68	65	68	72	72	69	71

Solution :

x	y	x^2	y^2	xy
65	67	4225	4489	4355
66	68	4356	4624	4488
67	65	4489	4225	4355
67	68	4489	4624	4556
68	72	4624	5184	4896
69	72	4761	5184	4968
70	69	4900	4761	4830
72	71	5184	5041	5112
$\Sigma x = 544$	$\Sigma y = 552$	$\Sigma x^2 = 37028$	$\Sigma y^2 = 38132$	$\Sigma xy = 37560$

$n = 8$

Applying karl pearson's co-efficent of correlation

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

By substuting the know values we applied.

$$r = \frac{8(37560) - (544 \times 552)}{\sqrt{8(37028) - (544)^2}\ \sqrt{8(38132) - (552)^2}}$$

$$r = \frac{300480 - 300288}{\sqrt{296224 - 295936}\ \sqrt{305056 - 304704}}$$

$$r = \frac{196}{\sqrt{288}\ \sqrt{352}}$$

$$r = \frac{196}{16.97 \times 18.76}$$

$$r = \frac{196}{318.3} \Rightarrow r = 0.6030.$$

Rank Correlation :

The correlation co-efficent between the ranks is called the rank correlation between the two characteristics A and B for that group of individuals.

The Spearman's co-efficient of rank correlation is given by :-

$$\boxed{\rho_s = 1 - \frac{6 \Sigma d_i^2}{n(n^2-1)}} \quad (d_i = x_i - y_i)$$

* The rank are the same $\rho = 1$
* lies between -1 and 1

i.e :- $-1 \leq \rho \leq 1$

Repeated Ranks :

If there is more than one item .

Then :- $\frac{m(m^2-1)}{12}$

$$\therefore \rho_s = 1 - \frac{6(\Sigma d_i^2 + \text{Correlation factor}))}{n(n^2-1)}.$$

$$\rho_s = 1 - \frac{6\left(\Sigma d_i^2 + \frac{m(m^2-1)}{12}\right)}{n(n^2-1)}$$

Problem :- 4 : 34

The following are the ranks obtained by 10 students in Statistics and Mathematics to what extent is knowledge of students in statistics related to know in Mathematics

Statistics :	1	2	3	4	5	6	7	8	9	10
Mathematics :	2	4	1	5	3	9	7	10	6	8

Solution :

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

$\Sigma d^2 = 40$

By Appleying Spearman's coefficient of rank

we get,

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

$$= 1 - \frac{6 \times 40}{10(100-1)}$$

$$= 1 - \frac{240}{990}$$

$$\rho_s = 0.76$$

There is high correlation between knowledge in the two subjects.

Problem 4.35

Ten competitors in a beauty contest are ranked by 3 judges in the following order.

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

Find which pair of judges have the nearest approach to common taste of beauty.

Solution:

A	B	C	$d_1 = A-B$	$d_2 = B-C$	$d_3 = A-C$	d_1^2	d_2^2	d_2^3
1	3	6	-2	-3	-5	4	9	2
6	5	4	1	1	2	1	1	4
5	8	9	-3	-1	-4	9	1	1
10	4	8	6	-4	2	36	16	4
3	7	1	-4	6	2	16	36	4
2	10	2	-8	8	0	64	64	0
4	2	3	2	-1	1	4	1	1
9	1	10	8	-9	-1	64	81	1
7	6	5	+1	1	2	1	1	4
8	9	7	-1	2	-1	1	4	1
						200	214	60

$n = 10$ $\quad \rho_s = 1 - \frac{6 \Sigma d^2}{n(n^2-1)}$

$$\rho_{AB} = 1 - \frac{6 \times 200}{10(100-1)} = -0.212$$

$$\rho_{BC} = 1 - 6 \times 214 / 10(100-1) \Rightarrow -0.297$$

$$\rho_{AC} = 1 - 6 \times 60 / 10(100-1) \Rightarrow \boxed{0.636}$$

Hence judges A and C have the nearest approach to common tastes of beauty.

Solved Problem 4.37.

Calculate the coefficients of rank correlation from the following data.

X : 48 34 40 12 16 16 66 25 16 57

Y : 15 15 24 8 13 6 20 9 9 15

Solution

A - B

X	Y	rank X	rank Y	d	d^2
48	15	3	4	-1	1
34	15	5	4	1	1
40	24	4	1	3	9
12	8	10	9	1	1
16	13	8	6	2	4
16	6	8	10	-2	4
66	20	1	2	-1	1
25	9	6	7.5	-1.5	2.25
16	9	8	7.5	0.5	0.25
57	15	2	4	-2	4

$\Sigma d^2 = 27.50$

In X-series, the values 16 is repeated three times $m_1 = 3$

In Y-series, the value 15 is repeated 3 times $m_2 = 3$ and the value 9 is reapted 2 times.

$$\therefore \rho_s = 1 - \frac{6\left[\Sigma d^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \frac{1}{2}(m_3^3 - m_3)\right]}{n(n^2-1)}$$

$$= 1 - \frac{6\left[27.5 + \frac{1}{12}(3^3 - 3) + \frac{1}{12}(3^3 - 3) + \frac{1}{12}(2^3 - 2)\right]}{10(100-1)}$$

$$= 1 - \frac{6 \times 32}{990} \Rightarrow \rho_s = 0.806$$

There is high Positive correlation.

Scanned by TapScanner

V. Concurrent Deviation Method

* This method of studying correlation is the simplest of all the methods.

* The only thing that is required under this method is to find out the direction of change of X variable and Y variable.

* The formula applicable is:

$$r_c = \pm\sqrt{\pm\left(\frac{2C-n}{n}\right)}$$

* Where r_c stands for coefficient of correlation by the concurrent method

* C stands for the number of concurrent deviations or the number of positive signs obtained after multiplying D_x withy D_y

n* = Number of pairs of observations compared.

Steps(i):

Find out the direction of change of X variable as compared withe the 1st value, whether the 2nd value is incressing or decreasing or is constant.

* If it is increasing put a + sign

* If it is decreasing put a − sign

* If it is constant put a zero (0).

* Similarly as compared to 2^{nd} value find out whether the 3^{rd} value is increasing, decreasing or constant. Repeat the same process for other values.

* Denote this column by D_x.

Step (ii) :

* In the same manner as discussed above.

* Find out the direction of change of Y variable and denote this column by D_y.

Step (iii) :

* Multiply D_x with D_y and determine the value of C; the number of positive signs.

Step (iv):

* Apply the above formula,

$$r_c = \pm\sqrt{\pm\left(\frac{2C - n}{n}\right)}$$

Note:

The significance of ± signs, both inside the under-root and outside the under-root is that we cannot take the under-root of minus sign.

Therefore, if $\frac{2c-n}{n}$ is negative, this negative value multiplied with the minus sign inside would make it positive and we can take the under-root. But the ultimate result would be negative.

If $\frac{2c-n}{n}$ is positive then, of course we get a positive value of the coefficient of correlation.

Illustration 21.

Calculate Coefficient of Concurrent deviation from the following data:

Price	Imports	Price	Imports
368	22	384	26
384	21	395	24
385	24	403	29
361	20	400	28
347	22	385	27

Scanned by TapScanner

Solution:

Calculate of Coefficient of Concurrent deviation

Price X	Direction of change of varriable X D_x	Imports Y	Direction of change of variable Y D_y	$D_x D_y$
368		22		
384	+	21	−	−
385	+	24	+	+
361	−	20	−	+
347	−	22	+	−
384	+	26	+	+
395	+	24	−	−
403	+	29	+	+
400	−	28	−	+
385	−	27.	−	+

C = 6

$$\boxed{r_c = \pm\sqrt{\pm\left(\frac{2c-n}{n}\right)}} \quad (1) \qquad c = 6,\ n = 9$$

$$r_c = \pm\sqrt{\pm\left(\frac{2\times 6 - 9}{9}\right)}$$

$$= \pm\sqrt{\pm\left(\frac{12-9}{9}\right)}$$

$$= \pm\sqrt{\pm\left(\frac{+3}{9}\right)}$$

$$= \pm\sqrt{\pm 0.333}$$

$$r_c = +0.577 //$$

Illustration 20

Calculate the Coefficient of Concurrent deviation from the following:

Year	1979	1980	1981	1982	1983
Supply	140	154	160	140	140
Price (Rs)	180	160	190	200	210

Solution:

Calculation of Coefficient of Concurrent deviation

Year	Supply	Direction of Change D_x	Price Y	Direction of Change D_y	$D_x D_y$
1979	140		180		
1980	154	+	160	−	−
1981	160	+	190	+	+
1982	140	−	200	+	−
1983	140	+	210	+	+

$c = 2$

$$r_c = \pm\sqrt{\pm\left(\frac{2c-n}{n}\right)} \qquad c = 2,\ n = 4$$

$$= \pm\sqrt{\pm\left(\frac{2\times 2-4}{4}\right)}$$

$$= \pm\sqrt{\pm\left(\frac{4-4}{4}\right)} \Rightarrow \pm\sqrt{\pm\left(\frac{0}{4}\right)}$$

$$r_c = 0$$