Unit 3. Descriptive Statistics:

Measure of Dispersion

Introduction A scatteredness of observation from a central value is called ~~disperssi~~ dispersion

for example:

following table contains distribution of marks of the three players.

Players	Runs	A.M.
x	94 98 100 102 106 100	100
y	60 75 100 125 140 106	106
z	80 90 100 110 170 100	100

If we observe the above score all the players are differ in variation. We can say that player x is more consistent than y. y is more consistent. The is rans are scattered from control value 100.

Characterictics for an ideal measure of dispersion

1) It should be based on all observations on dataset
2) It should be easy to understand.
3) It should be capable of further mathemetical treatements.
4) It should be affected by fuctuations of sampling.
5) It should be affected by extreom observations.

Measure of dispersion:

Measure of dispersion are enlisted as below.

1) Range.
2) Quartile deviation.
3) Mean deviation.
4) Standard deviation.

These measures have the same units as that of the observations eg. cm. hours etc. and measures are called as absolute measures of dispersion.

Absolute measure of dispersion:

Absolute measure of dispersion possesses units and hence create diffuculty in comparison of dispersion for two are more frequency distribution.

As they uses original units ~~and~~ of data and most useful for understanding the dispersion within the context of experiment.

Relative measure of dispersion:

Relative measure of dispersion are calculated as ratios or percentages.

ex. ratio of standard deviation to the mean.

This dispersion are always dimensionless and useful for making comparisons between

seperate datasets or experiments.
They are called as coefficient of dispersion.

Range:

Range is the size of the data.
Difference between the highest and the lowest value is called as range.

Range = L - S
L = largest value, S = Smallest value

Before calculating the range, arrange the data in ascending or descending order.

ex. The range of the data

2, 8, 10, 11, 12, 18, 21, 26, 27, 28, 33, 35, 38

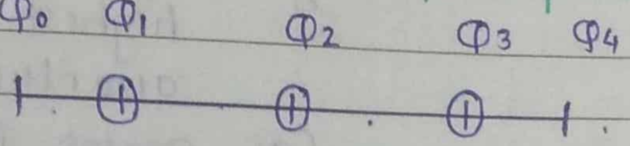
∴ Range = L - S
= 38 - 2
= 36

Limitations:

1. The range doesn't tell us the number of data points.
2. Range can't be used to find mean, median or mode.
3. The range is affected by extream values.
4. The range can't be used for open ended distribution.

$$\boxed{\text{coefficient of range} = \frac{L-S}{L+S}}$$

Quartile Deviation: (Q_1, Q_2, Q_3 are three quartiles divide our series in four equal parts.)



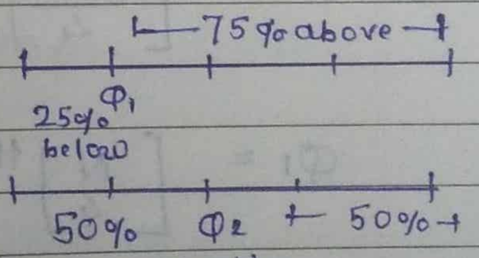
Division of dataset in four groups.

The quartile deviation are convenient to assess the spread of a distribution. This distribution measures the spread from its central tendency.

Interquartile deviation can be defined as the difference between the first quartile and the third quartile in the frequency distribution table.

When this difference is divided by two then this is called as quartile deviation or semi-interquartile range.

$$\boxed{\text{Quartile deviation OR Semi-interquartile Range} = \frac{Q_3 - Q_1}{2}}$$



Coefficient of quartile deviation = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

Quartile deviation - difference betn the quartile Q_1 & Q_3

Interquartile Range = $Q_3 - Q_1$
highest quartile − lowest quartile
(as series is in ascending order)

For individual observations or discrete obs.

I. When n = number of observations are o

Q_1 = Size of $\left(\frac{n+1}{4}\right)^{th}$ observation

Q_3 = The value of $\left(\frac{3(n+1)}{4}\right)^{th}$ observation

Continuous series

$Q_3 = \left[\frac{3N}{4}\right]^{th}$ item

$Q_1 = \left[\frac{N}{4}\right]^{th}$ item

Quartile deviation:

Quartile deviation is half of the difference of upper quartile (Q_3) and lower quartile (Q_1).

$$\text{Quartile deviation} = \frac{Q_3 - Q_1}{2} = \frac{\text{Interquartile range}}{2}$$

OR

↓

Semiquartile Range.

* Quartile deviation for individual series.

Q: Find the interquartile range, quartile deviation and coefficient of quartile deviation from the given data below

200, 210, 208, 160, 220, 250, 300

Ans ⇒ Ascending order

$$160, \underset{Q_1}{\boxed{200}}, 208, 210, 220, \underset{6^{th}}{\overset{Q_3}{\boxed{250}}}, 300$$

$Q_3 = \left[\frac{3(n+1)}{4}\right]^{th}$ item — gives position.

$Q_1 = \left[\frac{n+1}{4}\right]^{th}$ item.

here $n = 7$.

$Q_1 = \left[\frac{7+1}{4}\right]^{th}$ item $= 2^{nd}$ item

$Q_3 = 3 \times 2 = 6^{th}$ item

Interquartile Range $= Q_3 - Q_1$
$= 250 - 200$
$= 50.$

Quartile deviation $= \frac{Q_3 - Q_1}{2} = \frac{50}{2} = 25$

Coefficient of quartile deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \frac{250 - 200}{250 + 200}$

$= \frac{50}{450} = 0.1.$

② Calculate the lower and upper quartile when the quartile deviation = 10; and coefficient of quartile deviation = 0.5.

Soln →

$Q_1 = ?$; $Q_3 = ?$

Quartile deviation $= \frac{Q_3 - Q_1}{2} = 10.$

$Q_3 - Q_1 = 20 \cdots ①$

Coefficient of Q.d $= \frac{Q_3 - Q_1}{Q_3 + Q_1} = 0.5.$

$= \frac{20}{Q_3 + Q_1} = 0.5.$

So;

$Q_3 + Q_1 = 40$

$\Rightarrow Q_3 = 40 - Q_3$

Put in ①

$40 - Q_1 - Q_1 = 20$

$-2Q_1 = -20$

Page No.
Date

$Q_1 = 10$.
from ① $Q_3 = 30$.

Lower quartile = 10.
Upper quartile = 30.

Quartile deviation for discreate and continuous Series:

Q. From the following table giving height of students find interquartile range, quartile deviation and coefficient of quartile deviation

Height in cm	153	155	157	159	161	163	165	167	169
No. of students	25	21	28	20	18	24	22	18	23

⇒

	Height	No. of students	C.F
	153	25	25
	155	21	46
Q_1 -	$\boxed{157}$	28	74 ✓
	159	20	94
	161	18	112
	163	24	136
Q_3 -	$\boxed{165}$	22	158
	167	18	176
	169	23	$\boxed{199}$ → इथे N हा असतो
		$\Sigma f = 199$	

Page No.
Date

$Q_1 = \left(\frac{N+1}{4}\right)^{th}$ item $= \left(\frac{199+1}{4}\right)^{th} = 50^{th}$ term

Just greater than or equal to c.f is 74 50 corresp x is Q_1.

$Q_3 = \left[\frac{3(N+1)}{4}\right]^{th}$ item

$= 3\left(\frac{199+1}{4}\right)^{th}$ item

$= 150^{th}$ item $\longrightarrow$ c.f 158 is just greater so, corresponding x is $Q_3 = 165$.

Interquartile Range $= Q_3 - Q_1$
$= 165 - 157$
$= 8$.

Quartile deviation $= \frac{Q_3 - Q_1}{2} = \frac{8}{2} = 4$

Coeff. of quartile deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$
$= \frac{165-157}{165+157}$
$= \frac{8}{322} = 0.02$

Q:2) Calculate the range of marks obtained by middle 50% students. Also quartile deviatio

Marks	2	4	6	8	10	12
No. of Students	3	5	10	12	6	4
	3	8	18	30	36	40

$\frac{7}{4} = 1.75$ 2

$\frac{21}{4} = 5.25$

Q2

25%	25%	25%	25%

Q_1 Q_2 Q_3

middle 50%

Ans ⇒ $Q_1 = 6$; $Q_3 = 10$.

Range $= 4$

deviation $= 2$

Ans: Range of marks obtained by middle 50% of the students = 4 Marks.

Quartile deviation = 2 Marks

Quartile deviation for continuous series.

In continuous series.

Q_1 = size of $\left[\frac{N}{4}\right]^{th}$ item

Exact value $= l + \left(\frac{N}{4} - c\right) \times \frac{h}{f}$

Q_3 = size of $\left[\frac{3N}{4}\right]^{th}$ item

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

Where;

l = lower limit of quartile class

c = cumultive frequency of previous class

h = Internal width (Upper limit − lower limit)

f = simple frequency of quartile class.

Q: Find the quartile range; quartile deviation and coefficient of quartile deviation for the following figures.

Size	0-5	5-10	10-15	15-20	20-25	25-30
Frequency	3	9	15	23	30	20

Soln ⇒

	C.I	f	C.f
	0-5	3	3
quartile	5-10	9	$\boxed{12}$ c
class Q_1	$\boxed{10-15}$	$\boxed{15}$ f	27
	15-20	23	50
class Q_3	$\boxed{20-25}$	30	80
	25-30	20	100

$Q_1 = \left(\frac{N}{4}\right)^{th}$ item

$= \left(\frac{100}{4}\right)^{th}$ item

$= 25^{th}$ item.

Value just greater than or equal to 25 is 27 so corresponding x is C.I = 10-15.

Q_1 = lies in the group 10-1

$N = \Sigma f = 100$

$Q_1 = l + \frac{h}{f}\left[\frac{N}{4} - c\right]$

$= 10 + \frac{5}{15}[25 - 12]$

$= 10 + \frac{1}{3} \times 13$

$= 14.33$

Q_3 = Size of $\left[\frac{3 \times 100}{4}\right]^{th}$ item = size of 75^{th} i

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

$$= 20 + \frac{5}{30}\left[\frac{3 \times 100}{4} - 50\right]$$

$= 24.17.$

Range $= Q_3 - Q_1 = 9.84$, deviation $= \frac{9.84}{2} = 4.92$

coefficient $= \frac{9.84}{38.50} = 0.25$

Q: 2) Calculate the value of interquartile range, quartile deviation or coefficient of quartile deviation.

marks.	11 - 15	16 - 20	21 - 25	26 - 30	31 - 35	36 - 40
No. of Students	10	17	22	31	42	32

41 - 45	46 - 50	51 - 55
26	19	14

(inclusive series to exclusive series)

Ans ⇒ Q_1 = class 25.5 - 30.5

$Q_1 = 26.18$

Q_3 class 40.5 - 45.5

$Q_3 = 41.6.$

Range = 15.42 = marks

Deviation = 7.71 marks

coefficient = 0.22 marks

Q: Calculate the appropriate measure of dispersion for the following data.

Marks	below 20	20 - 30	30 - 40	Above 40
No. of Students	7	10	14	9

Open class intervals → calculate directly as we exclude 25% above & below.

Q: Find the range & coeff. of range
ii) Quartile deviation & coeff. of quartile deviation for the following data.

22, 17, 20, 07, 12, 19, 23, 21, 27, 30.

Soln ⇒

Smallest Observation (S) = 7
Largest Observation (L) = 50

Range = L - S = 30 - 7 = 23

Coeff. of Range = $\frac{L-S}{L+S} = \frac{23}{37} = 0.6216.$

To find quartile deviation, we arrange the observation in ascending order.

Q1 = The size of $\left(\frac{n+1}{4}\right)^{th}$ observation

$= \left(\frac{10+1}{4}\right)^{th}$ item

= 2.75^{th} observation.

2.75th observation lies in between 2nd & 3rd observation.

Q_1 = 2nd observation + 0.75 (3rd observation − 2nd observation)

$= 12 + 0.75(17-12)$

$= 15.75$

Q_3 = The size of $\frac{3(n+1)}{4}^{th}$ observation

= size of $\left(\frac{33}{4}\right)^{th}$ observation

= 8.25th observation.

Q_3 = 8th observation 0.25 (9th − 8th)

$= 23 + 0.25\ (27-23)$

$= 24$

Quartile deviation $= \frac{Q_3 - Q_1}{2} = \frac{24 - 15.75}{2} = 4.125$

Coeff. of Q.D $= \frac{Q_3 - Q_1}{Q_3 + Q_1} = 0.2075$

Mean deviation:

The arithmetic mean of absolute deviations from any average is called mean deviation about the resp. average.

Page No.
Date

① Mean deviation about mean ($\bar{x}$)

$$M.D = \sum_{i=1}^{n} \frac{(x_i - \bar{x})}{n}$$

Q: Calculate mean deviation about mean of th following data
5, 8, 11, 12, 14.

⇒

x_i	$\lvert x_i - \bar{x} \rvert$
5	5
8	2
11	1
12	2
14	4

$\Sigma \lvert x_i - \bar{x} \rvert = 14$

$$\bar{x} = \frac{5+8+11+12+14}{5}$$

$\bar{x} = 10$

$$M.D = \sum_{i=1}^{n} \frac{\lvert x_i - \bar{x} \rvert}{n}$$

$= 14/5$

$= 2.8$

* For median formula is

$$M.D = \sum_{i=1}^{n} \frac{\lvert x_i - \text{median} \rvert}{n}$$

median = 11

for median = 2.6.

Q: Calculate the mean deviation about the mean and about median for the following freq. dis.

class int.	1-3	3-5	5-7	7-9
freq.	3	4	2	1

⇒

class int	freq. (f_i)	midvalue (x_i)	$f_i x_i$	c.f
1-3	3	2	6	3
3-5	4	4	16	7
5-7	2	6	12	9
7-9	1	8	8	10
	$\Sigma f_i = 10$		$\Sigma f_i x_i = 42$	

$$\text{Mean} = \bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{42}{10} = 4.2$$

$$\text{Median} = l + \frac{h}{f}\left[\frac{N}{2} - c\right]$$

$N = 10$ Just greater than / equal to c.f. i.e 7.

$\frac{N}{2} = 5$; $l = 3$; $h = 2$; $f = 4$; $\frac{N}{2} = 5$

$c = 3$

$$\text{Median} = 3 + \frac{2}{4}[5 - 3]$$

Median = 4.

x_i	f_i	$\lvert x_i-\bar{X}\rvert$	$f_i\lvert x_i-\bar{X}\rvert$	$\lvert x_i-M\rvert$	$f_i\lvert x_i-M\rvert$
2	3	2.2	6.6	2	6
4	4	0.2	0.8	0	0
6	2	1.8	3.6	2	4
8	1	3.8	3.8	4	4
			$\Sigma f_i\lvert x_i-\bar{X}\rvert = 14.8$		$\Sigma f_i\lvert x_i-M\rvert = 14$

M.D about mean $= \dfrac{\Sigma f_i\lvert x_i-\bar{X}\rvert}{N}$

$= \dfrac{14.8}{10}$

M.D about mean $= 1.48$

Coefficient of M.D about mean

$= \dfrac{\text{M.D about mean}}{\text{mean}}$

$= \dfrac{1.48}{4.2}$

$= 0.35238$

~~M.D~~ Mean deviation about median $= \dfrac{\Sigma f_i\lvert x_i-M\rvert}{N}$

$= \dfrac{14}{10}$

$= 1.4$

Coefficient of mean deviation about median $= \dfrac{\text{M.D about median}}{\text{median}}$

$= \dfrac{1.4}{4}$

$= 0.35.$

Page No.
Date

Mean deviation about mean for cont. freq. dis

Mean deviation about mean is

$$= \frac{\Sigma f_i |x_i - \bar{x}|}{N}$$

$$\text{Coeff. of M.D. about mean} = \frac{\text{M.D. abot mean}}{\text{mean}}$$

Q: Cal. mean deviation about mean & coefficient for the following data.

C.I.	2-4	4-6	6-8	8-10
f.	3	4	2	1

Soln ⇒ $\bar{x} = \frac{\Sigma f_i x_i}{N} = \frac{52}{10} = 5.2$

C-I	fi	$\lvert x_i - \bar{x}\rvert$	x_i midpoint	$f_i \lvert x_i - \bar{x}\rvert$	$f_i x_i$
2-4	3	$\lvert 3-5.2\rvert = 2.2$	3	66	9
4-6	~~5~~ 4	$\lvert 5-5.2\rvert = 0.2$	5	~~0.8~~ 1.0	25
6-8	2	$\lvert 7-5.2\rvert = 2.2$	7	4.4	14
8-10	1	$\lvert 9-5.2\rvert = 4.2$	9	4.2	9
	$\Sigma f_i = 10$			$\Sigma f_i \lvert x_i - \bar{x}\rvert = 16.2$	$\Sigma f_i x_i = 52.$

∴ Meandeviation about mean

$$= \frac{\Sigma f_i |x_i - \bar{x}|}{N}$$

$$= \frac{16.2}{10}$$

$$= 14.8 / 10 = 1.48$$

coeff. of mean deviation = 0.2846.

Q! Calculate mean deviation about mean.

CI	20-40	40-80	80-100	100-120	120-140
fi	3	6	20	12	9

Ans ⇒

C.I	fi	xi (midpoint)	fixi	$\lvert xi-\bar{x}\rvert$	$fi\lvert xi-\bar{x}\rvert$
20-40	3	30	90	64.8	194.4
40-80	6	60	360	34.8	208.8
80-100	20	90	1800	4.8	96
100-120	12	110	1320	15.2	182.4
120-140	9	130	1170	35.2	316.8.
	$\Sigma fi = 50$		$\Sigma fixi = 4740$		$\Sigma fi\lvert xi-\bar{x}\rvert = 998.4.$

$$\bar{x} = \frac{\Sigma fixi}{\Sigma fi} = \frac{4740}{50} = 94.8$$

$$\frac{\Sigma fi\lvert xi-\bar{x}\rvert}{N}$$

$$= \frac{998.4}{50}$$

$$= 19.968$$

$$= \frac{19.968}{94.8}$$ C.O. m.d = 0.21.

Page No.
Date

Q. Calculate M.D. about median & its coefficient.

C.I	0-20	20-40	40-60	60-80	80-100
f.	10	16	30	32	12

Ans ⇒

C.I.	f.	c.f	midpoint (x_i)	$\lvert x_i - M \rvert$	$f_i \lvert x_i - M \rvert$
0-20	10	10	10	46	460
20-40	16	26	30	26	416
40-60	30	56	50	6	180
60-80	32	88	70	14	448
80-100	12	100	90	34	408
	$\Sigma f_i = 100$				$\Sigma f_i \lvert x_i - M \rvert = 1912$

$$\text{Median} = l + \frac{h}{f}\left[\frac{N}{2} - c\right]$$

$$\text{median class} = \left(\frac{N}{2}\right)^{th} \text{term}$$

= 50^{th} item just greater.

$$\text{median} = 40 + \frac{20}{30}[50 - 26]$$

$$= 56.$$

Ans = 19.

C.O.m.d = 0.34

Variance: The arithmetic mean of squares of deviation taken from its arithmetic mean is called variance.

$$\Rightarrow \quad Var(x) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n} \quad \text{(for individual series)}$$

$$\Rightarrow \quad Var(x) = \frac{\sum f_i (x_i - \bar{x})^2}{N} \quad \text{(for discrete \& continuous freq. distribution)}$$

Standard deviation

Standard deviation is the positive squareroot of variance. In other words, standard deviation is the squareroot of the arithmetic mean of squares of deviations taken from its mean.

Q: Calculate std. deviation of 5, 8, 7, 11, 14.

$$S.D = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}} \quad \text{(for individual series)}$$

$$S.D = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{N}} \quad \text{(for discrete \& cont. freq. dist)}$$

by actual mean method.

Q. Cal. S.D. of 5, 8, 7, 11, 14

$= \frac{\Sigma x_i}{n}$

$\boxed{\overline{X} = 9}$

Soln ⇒

x	$(x - \overline{x})$	$(x - \overline{x})^2$
5	-4	16
8	-1	1
7	-2	4
11	2	4
14	5	25
		$\Sigma(x - \overline{x})^2 = 50$

$$\therefore S.D = \sqrt{\frac{\Sigma(x - \overline{x})^2}{n}}$$

$$= \sqrt{\frac{50}{5}}$$

$$= \sqrt{10}$$

$$= 3.1622.$$

Note:
Standard deviation can be calculated using the following methods.

1. Actual mean method.
2. Direct method
3. Assumed mean method
4. Step deviation.

Q: Calculate the S.D of 5, 10, 15, 20, 25.

$\overline{X}$

X	$(X-\overline{X})$	$(X-\overline{X})^2$
5.	-10	100
10	-5	25
15	0	0
20	5	25
25	10	100
		$\Sigma(X-\overline{X})^2 = 250$

$$S.D = \sqrt{\frac{\Sigma(X-\overline{X})^2}{n}}$$

$$= \sqrt{\frac{250}{5}}$$

$$= \sqrt{50}$$

$$S.D. = 7.07.$$

Direct method:

$$\sigma = \sqrt{\frac{\Sigma x_i^2}{n} - (\overline{X})^2}$$

where σ = standard deviation

Σx_i^2 = Sum of total squares of observ

n = no. of observations

$\overline{X}$ = mean.

Page No.
Date

Q: Calculate S.D by direct method 5, 8, 7, 11, 14.

x_i =	5	8	7	11	14	
x_i^2 =	25	64	49	121	196	$\Sigma x_i^2 = 455$

$$\therefore \sigma = \sqrt{\frac{455}{5} - (9)^2}$$

$$= 3.16.$$

Q: Calculate S.D by direct method 26, 5, 15, 20, 24

x_i =	26.	5	15	20	24	
x_i^2 =	676	25	225	400	576.	$\Sigma x_i^2 = 1902$

$$\therefore \sigma = \sqrt{\frac{1902}{5} - (18)^2}$$

$$= 7.50.$$

Assumed mean method:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} u_i^2}{n} - \left(\frac{\Sigma u_i}{n}\right)^2}$$

Where σ = Std. deviation
u_i = x_i − A (Assumed mean)
n = no. of observations.

Q: Calculate S.D. with assumed mean method
5, 8, 7, 11, 14.

$\underline{A = 7}$

x_i	$U_i = x_i - A$	U_i^2
5	-2	4
8	1	1
7	0	0
11	4	16
14	7	49
	$\Sigma U_i = 10$	$\Sigma U_i^2 = 70$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} U_i^2}{n} - \left(\frac{\Sigma U_i}{n}\right)^2}$$

$$\sigma = \sqrt{\frac{70}{5} - \left(\frac{10}{5}\right)^2}$$

$$= 3.16.$$

Q: Calculate S.D by direct method

Marks	0-10	10-20	20-30	30-40	40-50
No. of Students	4	3	6	5	2

Soln ⇒

$$\sigma = \sqrt{\frac{\Sigma f_i x_i^2}{N} - (\bar{x})^2}$$

x_i = midpoint of C-I

$\bar{X}$ = mean

$N = \Sigma f_i$

Marks	f_i	x_i (midpoint	$f_i x_i$	x_i^2	$f_i x_i^2$
0-10	4	5	20	25	100
10-20	3	15	45	225	675
20-30	6	25	150	625	3750
30-40	5	35	175	1225	6125
40-50	2	45	90	2025	4050
	$\Sigma f_i = 20$		$\Sigma f_i x_i = 480$		$\Sigma f_i x_i^2 = 14700$

$$\bar{X} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{48\cancel{0}}{2\cancel{0}} = 24.$$

$$\therefore \sigma = \sqrt{\frac{14700}{20} - (24)^2}$$

$$\sigma = 12.60.$$

Assumed mean method for continuous frequency distribution:

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

Where $u_i = x_i - A$ (A - Assumed mean)

$N = \Sigma f$

Q: Calculate S.D by Assumed mean method.

Marks	0-10	10-20	20-30	30-40	40-50
No. of stu	4	3	6	5	2

Soln ⇒

$A = 25$

C.I.	f	x_i (midpoints)	$u_i = x_i - A$	u_i^2	fu_i	fu_i^2
0-10	4	5	-20	400	-80	-~~16~~1600
10-20	3	15	-10	100	-30	-300
20-30	6	25	0	0	0	0
30-40	5	35	10	100	50	500
40-50	2	45	20	200	40	800
	$\Sigma f_i = 20$				$\Sigma f_i u_i = -20$	$\Sigma f_i u_i^2 = -60$

Page No.
Date

$$\sigma = \sqrt{\frac{-60\cancel{0}}{2\cancel{0}} - \left(\frac{-20}{20}\right)^2}$$

$$= \sqrt{-30 - 1}$$

Please check calculation Ans is correct.

$= 12.60$.

Q: Calculate S.D by Assumed mean method,

x	0-10	10-20	20-30	30-40
f	2	5	4	1

Soln ⇒ Assumed mean = 15

C.I	f	Xi (midpoints)	Ui = xi - A	Ui²	fui	fui²
0-10	2	5	10	100	200	200
10-20	3	15	6	0	0	0
20-30	4	25	10	100	400	400
30-40	1	35	20	400	400	400
	Σfi = 10				Σfui = 1000	Σfui² = 1000

$$\sigma = \sqrt{\frac{\Sigma f(u_i)^2}{N} - \left(\frac{\Sigma f_i u_i}{N}\right)^2}$$

<u>Ans</u> $= 9.16.$

Step deviation method:

$$\sigma = \sqrt{\frac{\Sigma f_i(u')^2}{N} - \left(\frac{\Sigma f_i u'}{N}\right)^2} \times h$$

$U_i = x_i - A$

$U' = \frac{x_i - A}{h}$

$U' = \frac{U_i}{h}.$

Q: Calculate S.D by step deviation method,

marks	0-10	10-20	20-30	30-40	40-50
No. of Students	4	3	6	5	2

Soln ⇒

$h = 10$

C.I	f	x_i (midpoint)	$U_i = X_i - A$, $A = 25$	$U' = \frac{X_i - A}{h}$	$(U')^2$	f_iU'	$f_i(U')^2$
0-10	4	5	-20	-2	4	-8	16
10-20	3	15	-10	-1	1	3	3
20-30	6	25	0	0	0	0	0
30-40	5	35	10	1	1	5	5
40-50	2	45	20	2	2	4	4
	$\Sigma f = 20$					$\Sigma f_iU' = 4$	$\Sigma f_i(U')^2 = 28$

$$\sigma = \sqrt{\frac{\Sigma f_i(U')^2}{N} - \left(\frac{\Sigma f_iU'}{N}\right)^2} \times h$$

$$= \sqrt{\frac{28}{20} - \left(\frac{4}{20}\right)^2} \times 10$$

$$= 9.16.$$

Q: Calculate step deviation method to calculate S.D.

marks	0 - 10	10 - 20	20 - 30	30 - 40
f	2	3	4	1

$h = 10$

Soln ⇒

Marks	f	x_i	$U_i = x_i - A$, $A = 15$	$U' = \frac{x_i - A}{h}$	$(U')^2$	f_iU'	$f_i(U')^2$
0-10	2	5	-10	-1	1	-2	2
10-20	3	15	0	0	0	0	0
20-30	4	25	10	1	1	4	4
30-40	1	35	20	2	4	2	4
	$\Sigma f = 10$					$\Sigma f_iU' = 4$	$\Sigma f_i(U')^2 = 10$

$$\sigma = \sqrt{\frac{\Sigma f_i(U')^2}{N} - \left(\frac{\Sigma f_iU'}{N}\right)^2} \times h$$

$$= \sqrt{\frac{10}{10} - \left(\frac{4}{10}\right)^2} \times 10$$

$$= \sqrt{1 - \frac{16}{100}} \times 10$$

$$= \sqrt{\frac{100 - 16}{100}} \times 10$$

$$= \sqrt{\frac{84}{100}} \times 10$$

$$= 0.916 \times 10$$

$$= 9.16.$$

(IMP)

Moments :

1] Mean
2] Variance
3] Skewness (deviation)
4] Kurtosis (nature)

Moment about mean: r^{th} moment of observation x_i about the mean $\bar{X}$ is given by

$$\mu_r = \frac{1}{n} \Sigma (x_i - \bar{X})^r \quad \text{(for individual series)}$$

$$\mu_r = \frac{1}{N} \Sigma f_i(x_i - \bar{X})^r \quad \text{(for discrete/cont. series)} \quad -① $$

It is also known as central moments.

put $r = 0, 1, 2, 3, 4$ in ①.

$\mu_0 = 1$ $(r = 0)$

$\mu_1 = 0.$ $(r = 1)$

$\Rightarrow$ $\mu_2 =$ Variance $\quad \mu_2 = 1/N \; \Sigma f_i (x_i - \bar{X})^2$

$S.D = \sqrt{\cancel{\mu_2}}.\ \sqrt{\mu_2}$

$$\mu_3 = \frac{1}{N} \Sigma f_i (x_i - \overline{x})^3$$

$$\mu_4 = \frac{1}{N} \Sigma f_i (x_i - \overline{x})^4.$$

Moment about any point A:

rth moment of observation X_i about the any point A is given by.

$$\mu_r' = \frac{1}{n} \Sigma (x_i - A)^r \quad \text{(for individual series)}$$

$$\mu_r' = \frac{1}{N} \Sigma f_i (x_i - A)^r \quad \text{(for discrete series)}. \quad -①.$$

put $r = 0, 1, 2, 3, 4.$

$\mu_0' = 0$

$\mu_r' = 1$

$$\mu_1' = \frac{1}{N} \Sigma f_i (x_i - A)$$

$$= \frac{\Sigma f_i x_i}{N} - \frac{\Sigma f_i A}{\cancel{N}}$$

$$\boxed{\mu_1 = \overline{x} - A}$$

Imp $\boxed{\therefore \overline{x} = \mu_1 + A}$

$\mu_2' = \frac{1}{N} \sum f_i (x_i - A)^2$

$\mu_3' = \frac{1}{N} \sum f_i (x_i - A)^3$

$\mu_4' = 1/N \sum f_i (x_i - A)^4$.

Moment about origin (Raw moment):

rth moment of observation x_i about the point A is given by

$\mu_r' = \frac{1}{n} \sum x_i^r$ (for individual series)

$\mu_r' = \frac{1}{N} \sum f_i (x_i)^r$ (for discrete/cont. series)

Relation between moment about mean (μ_r) and moment about any point (μ_r')

$\mu_1 = 0$

$\mu_2 = \mu_2' - (\mu_1')^2$

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

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

V Imp

Q: The first four moments of a distribution about the value 5 are 2, 20, 40, 50. From the given information obtain first 4 central moments, mean and Std. deviation.

Soln ⇒

$A = 5$, $\mu_1' = 2$, $\mu_2' = 20$, $\mu_3' = 40$, $\mu_4' = 50$

$\mu_1, \mu_2, \mu_3, \mu_4 = ?$

$\boxed{\mu_1 = 0}$

$\mu_2 = \mu_2' - (\mu_1')^2$

$= 20 - (2)^2$

$= 20 - 4$

$\boxed{\mu_2 = 16}$

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

$= 40 - 3(20)(2) + 2(2)^3$

$= 40 - 120 + 16$

$= 40 - 136$

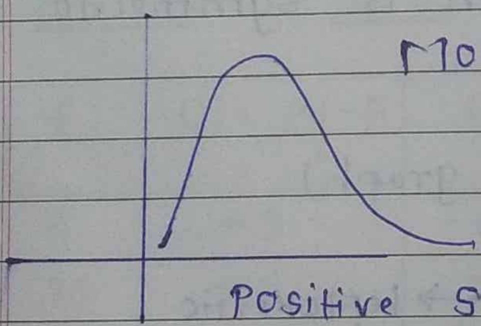
$\boxed{\mu_3 = -64.}$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4$$
$$= 50 - 4(40)(2) + 6(20)(2)^2 - 3(2)^4$$
$$= \cancel{50 - 8(20) + 24 \times 20 - 48}$$
$$= 162.$$

$$\overline{X} = \mu_1' + A$$
$$= 2 + 5$$
$$\boxed{\overline{X} = 7}$$

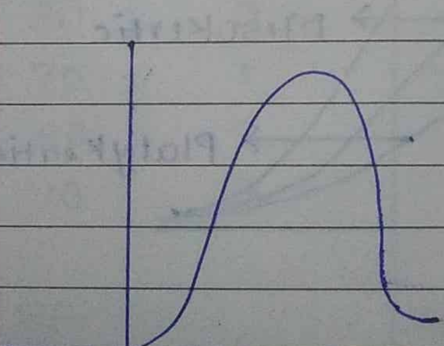
$$S.D = \sqrt{\mu_2}$$
$$= \sqrt{16}$$
$$\boxed{S.D = 4}$$

Skewness: (lack of symmetry)



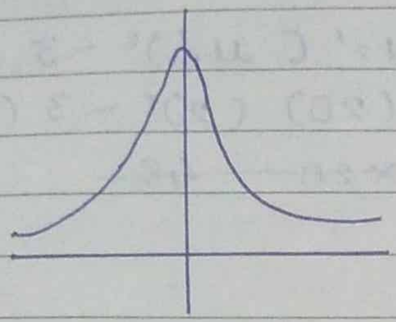
Mode < Median < Mean

Positive skewness



Mode > Median > Mean

negative skewness



Mean = Median = Mode

Normal distribution

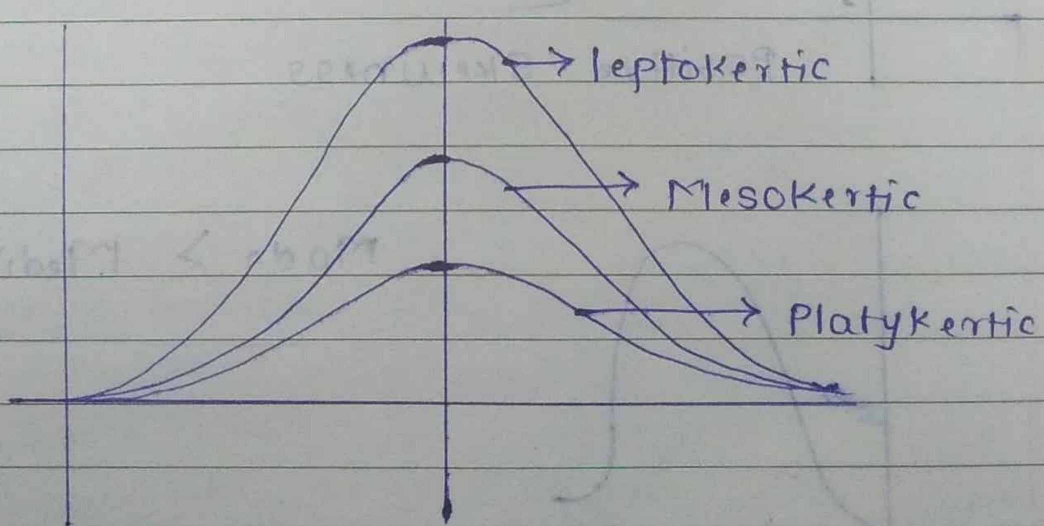
Measure of skewness:

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

$$\text{Coefficient of skewness} = \frac{\mu_3^2}{\mu_2^2}$$
$$= \beta_1$$

if $\beta_1 = 0$ then graph is symmetric.

Kurtosis: (Nature of graph)



1) Kertosis is the property of distribution which expresesses peakness or flatness of the curve

2) Kertosis is measured by β_2 which is given by

$$\beta_2 = \frac{\mu_4}{(\mu_2)^2}$$

3) When $\beta_2 = 3$; the distribution is mesokertic
When $\beta_2 < 3$; then it is platykertic.
When $\beta_2 > 3$; then it is leptokertic.

Q: Calculate the first 4 moments about mean of the given distribution. Also find β_1 & β_2.

x	2.0	2.5	3.0	3.5	4.0	4.5	5.0
f	4	36	60	90	70	40	10

Soln ⇒ $A = 3.5$ $h = 0.5$

x	f	$U_i = \frac{x_i - A}{h}$	f_iu_i	$f_iu_i^2$	$f_iu_i^3$	$f_iu_i^4$
2.0	4	−3	−12	36	−108	324
2.5	36	−2	−72	144	−288	576
3.0	60	−1	−60	60	−60	60
3.5	90	0	0	0	0	0
4.0	70	1	70	70	70	70
4.5	40	2	80	160	320	640
5.0	10	3	30	90	270	810
$\Sigma f =$ 310		$\Sigma u_i = 0$	Σf_iu_i = 36	$\Sigma f_iu_i^2$ = 560	$\Sigma f_iu_i^3$ = 204	$\Sigma f_iu_i^4$ = 2480

$\cancel{\mu_1' = \frac{h}{n} \Sigma u_i}$

$\boxed{\mu_1' = 0}$

$\mu_1' = \frac{h}{\Sigma f_i} \Sigma f_i u_i$

$= \frac{0.5}{310} \times 36$

$\boxed{\mu_1' = 0.05}$

$\mu_2' = \frac{h^2}{\Sigma f_i} \Sigma f_i u_i^2$

$\boxed{\mu_2' = 0.451}$

$\mu_3' = \frac{h^3}{\Sigma f_i} \Sigma f_i u_i^3$

$\boxed{\mu_3' = 0.0822}$

$\boxed{\mu_4' = 0.5}$

$\boxed{\mu_1 = 0}$

$\mu_2 = \mu_2' - (\mu_1')^2$

$= 0.451 - (0.05)^2$

$\boxed{\mu_2 = 0.4485}$

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

$= 0.0822 - 3(0.451) \times (0.05) + 2(0.05)^3$

$\boxed{\mu_3 = \cancel{0.6148}}$

0.00398

Page No.
Date

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4$$
$$= 0.5 - 4\times 0.0822\times 0.05 + 6\times 0.451\times(0.05)^2 - 3(0.05)^4$$

$$\boxed{\mu_4 = 0.490}$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^2} = 7.88537 \times 10^{-5}.$$

$$\beta_2 = \frac{\mu_4}{(\mu_2)^2} = 2.4391, \text{ (Platikertic).}$$

Q2) Calculate the first four central moments for the following frequencies.

No of Jobs	0 – 10	10 – 20	20 – 30	30 - 40	40 - 50
No. of workers	6	26	47	15	6

($h = 10$; $A = 25$)

x_i	f_i	$u_i = x_i - A / h$	f_iu_i	$f_iu_i^2$	$f_iu_i^3$	$f_iu_i^4$
5	6	-2	-12	24	-48	96
15	26	-1	-26	26	-26	26
25	47	0	0	0	0	0
35	15	1	15	15	15	15
45	6	2	12	24	48	96
	$\Sigma f_i =$	$\Sigma u_i = 0.$	$\Sigma f_iu_i =$	$\Sigma f_iu_i^2 =$	$\Sigma f_iu_i^3 =$	$\Sigma f_iu_i^4 =$

Correlation:

In many situations we come across variables are interrelated. If two variables x & y changes in such a way that the change in one variable affects the change in another variable then variables are correlated.

Linear retn betn correlated variables is called as correlation.

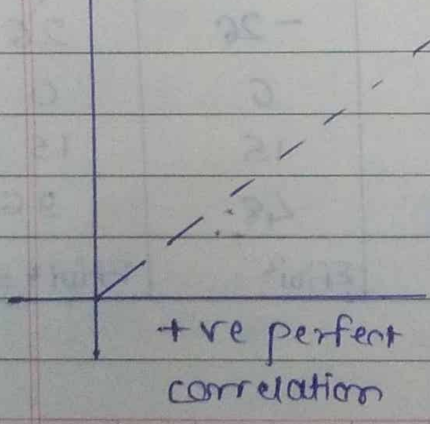
Types of Correlation:

① **Positive correlation**- If the variable x increases, y also increases. OR vice versa. ex: Work & production

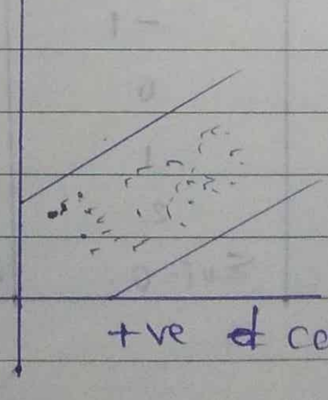
② **Negative correlation**: If the increase in one variable (or dicrease) corresponds to decrease in other variable (or increase) then it is -ve Correction

③ **Perfect Correlation**: If two variables vary in such a way that their ratio is always constant. then it is perfect correlation.

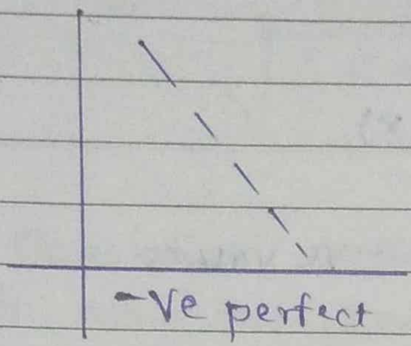
Scattered diagram:



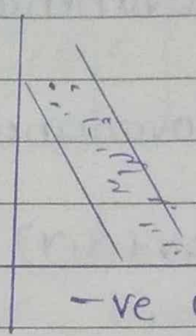
+ve perfect correlation



+ve ~~of~~ correlation



-ve perfect correlation



-ve correlation.

Covarience: $(x, y) = (x_i, y_i)$

$(x_i - \bar{x} ; y_i - \bar{y})$

$$\text{cov}(x, y) = \frac{1}{n} \Sigma (x_i - \bar{x})(y_i - \bar{y})$$

$$= \frac{1}{n} \Sigma (x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x}\bar{y})$$

$$= \frac{1}{n} \Sigma x_i y_i - \bar{y}\left(\frac{\Sigma x_i}{n}\right) - \bar{x} \frac{\Sigma y_i}{n} + \bar{x}\bar{y} \frac{\Sigma 1}{n}$$

$$= \frac{1}{n} \Sigma x_i y_i - \bar{y}\bar{x} - \bar{x}\bar{y} + \bar{x}\bar{y}$$

$$\boxed{\text{cov}(x, y) = \left(\frac{1}{n} \Sigma x_i y_i\right) - \bar{x}\bar{y}}$$

Properties of Covarience:

1] $\text{cov}(x, x) = \text{Var}(x)$

2] $\text{cov}(x-a ; y-b) = \text{cov}(x, y)$

3] $\text{cov}\left(\frac{x-a}{h}, \frac{y-b}{k}\right) = \frac{1}{hk} \text{cov}(x, y)$

If x and y are independent variables then covarience of $(x, y) = 0$

5] Covariance $(x, y) = \text{cov}(y, x)$

6] $\text{cov}(x, y)$ may have the -ve values.

Imp. **Karl - person's coefficient of correlation:**

$$\boxed{r(x, y) = \frac{\text{cov}(x, y)}{\sigma x \cdot \sigma y}}$$

Correlation coeff is calculated to study the degree of relationship betn x and y. it is denoted by $r(x, y)$ OR r_{xy} OR $\text{corr}(x, y)$

$$Var = \sigma x^2 = \frac{1}{n} \Sigma (x_i - \bar{x})^2$$

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

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

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

$$\boxed{\sigma x^2 = \frac{1}{n} \Sigma x_i^2 - (\bar{x})^2}$$

$\Rightarrow$ $\sigma x = \sqrt{1/n \, \Sigma x_i^2 - (\bar{x})^2}$

$\Rightarrow$ $\sigma y = \sqrt{1/n \, \Sigma y_i^2 - (\bar{y})^2}$

$$\boxed{r(x,y) = \frac{1/n \sum x_i y_i - \bar{x}\bar{y}}{\sqrt{\frac{1}{n}\sum x_i^2 - (\bar{x})^2}\sqrt{\frac{1}{n}\sum y_i^2 - (\bar{y})^2}}}$$

Note: Correlation coefficient $r(x,y)$ doesnot change in magnitude under the change of origin and scale.

Coefficient of correlation lies betn -1 to 1

$$-1 \leq r(x,y) \leq 1$$

Method of step deviation

$$U_i = \frac{x_i - A}{h} \; ; \quad V_i = \frac{y_i - B}{k}$$ (for equal length ~~deviat~~ interval)

or

$U_i = x_i - A; \; V_i = y_i - B.$

$$\text{Cov}(x,y) = \text{cov}(U,V)$$

$$\text{cor}(x,y) = \frac{\frac{1}{n}\sum U_i V_i - \bar{U}\bar{V}}{\sqrt{\frac{1}{n}\sum U_i^2 - (\bar{U})^2}\sqrt{\frac{1}{n}\sum V_i^2 - (\bar{V})^2}}$$

$$\bar{U} = \frac{\sum U_i}{n} \qquad \bar{V} = \frac{\sum V_i}{n}$$

For continious F.D.

$$\text{cor}(x,y) = \text{cor}(u,v) = \frac{1/N \sum f_i u_i v_i - \bar{U}\bar{V}}{\sqrt{1/N \sum f_i u_i^2 - (\bar{U})^2}\sqrt{1/N \sum f_i v_i^2 - (\bar{V})^2}}$$

Page No.
Date

Remark: 1) If $r(x,y) > 0$; corriliation is +ve
2) If $r(x,y) < 0$; " -ve.
3) If $r(x,y) = 0$; no " corrilation.

~~r(x,y)~~

Q: Find the coefficient of corrilation for the data $n = 25$, ; $\Sigma x_i = 100$; $\Sigma y_i = 125$; $\Sigma x_i^2 = 250$; $\Sigma y_i^2 = 500$; $\Sigma x_i y_i = 522$.

Soln ⇒

$$r(x,y) = cor(x,y) = \frac{1/n \, \Sigma x_i y_i - \bar{x}\bar{y}}{\sqrt{1/n \, \Sigma x_i^2 - (\bar{x})^2}\sqrt{1/n \, \Sigma y_i^2 - (\bar{y})^2}}$$

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{100}{25} = 4$$

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

$$\therefore cor(x,y) = \frac{1/25 \times 522 - 4\times 5}{\sqrt{\frac{1}{25}\times 100 - 16}\sqrt{\frac{1}{25}\times 500 - 25}}$$

$$= \cancel{\frac{22}{\sqrt{1850}}} = \frac{22}{\sqrt{-6}\times\sqrt{-5}}$$

$$= \frac{22/25}{\sqrt{30}}$$

$$= \cancel{0.16.}$$

$$= 0.160665.$$

Q) Compute the coefficient of correlation for the following data.

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

Soln ⇒ $n = 6$ $\quad \bar{x} = 20. \quad \bar{y} = 21$

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
10	18	180	100	324
14	12	168	196	144
18	24	432	324	576
22	6	132	484	36
26	30	780	676	900
30	36	1080	900	1296
		$\Sigma x_i y_i = 2772$	$\Sigma x_i^2 = 2680$	$\Sigma y_i^2 = 3276$

$$\cancel{\mathrm{cov}(x,y)} = \left(\frac{1}{n} \Sigma x_i y_i\right) - \bar{x}\,\bar{y}$$

$r(x,y)$ Calculate

$$= \frac{1}{6} \times 2772 - 21 \times 20$$

Correct Answer $= \underline{\underline{0.6}}$

$U_i = x_i - 22$

$V_i = y_i - 24$

Same Sum with Stepdeviation	x_i	y_i	u_i	v_i	v_i^2	u_i^2	$u_i v_i$
	10	18	~~144~~ -12	-6	~~144~~ 36	~~36~~ 144	72
	14	12	-8	-12	144	64	96
	18	24	-4	0	6	16	0
	22	6	0	-18	324	6	D
	26	30	4	6	36	16	24
	30	36	8	12	144	64	96
			$\Sigma u_i = -12$	$\Sigma v_i = -18$	$\Sigma v_i^2 = 304$	Σu_i^2 684	$\Sigma u_i v_i = 288$

Page No.
Date

$$r(x,y) = r(u,v) = \frac{\frac{1}{n}\sum u_i v_i - \bar{U}\bar{V}}{\sqrt{\frac{1}{n}\sum u_i^2 - (\bar{U})^2}\sqrt{\frac{1}{n}\sum v_i^2 - (\bar{V})^2}}$$

$$\bar{U} = \frac{\sum u_i}{n} = -2 \qquad \frac{\sum f_i u_i}{\sum f_i}$$

$$\bar{V} = \frac{\sum v_i}{n} = -3. \qquad \frac{\sum f_i v_i}{\sum v_i}$$

$$\boxed{r(u,v) = 0.638}$$

Q: Calculate the coefficient of correlation for the following distribution.

x_i	5	9	15	19	24	28	32
y_i	7	9	14	21	23	29	30
f_i	6	9	13	20	16	11	7

Soln ⇒

x_i	y_i	f_i	u_i	v_i	$u_i v_i$	$f_i u_i v_i$	u_i^2	$f_i u_i^2$	v_i^2	$f_i v_i^2$

$$r(x,y) = \cancel{1.048}$$
$$= 1.048$$
$$\approx 1$$

Regression: A process of predicting a unknown value of one variable using known value of other variables.

Note: To find regression, two variables need to be correlated.

Regression lines:

y on x | x on y

x is independent variable & y is dependent variable so we can find the values of y variable from known value x.

Eqn of Regression line for y on x = $(y-\bar{y}) = b_{yx}(x-\bar{x})$

Where b_{yx} ⇒ is coeff. of regression.

$(x-\bar{x}) = b_{xy}(y-\bar{y})$

Where b_{xy} ⇒ is coeff. of regression

Where $b_{yx} = r(x,y)\dfrac{\sigma y}{\sigma x}$ → S.D. on y, → S.D. on x

coeff. of ~~correlation~~ regression

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

$$\boxed{b_{yx} = \frac{cov(x,y)}{\sigma x^2}}$$

$$\boxed{b_{xy} = \frac{cov(x,y)}{\sigma y^2}}$$

Remarks:

Imp 1) $r = \sqrt{b_{xy}\, b_{yx}}$ coefficient of correlation is the geometric mean of regression coefficie

$$b_{xy} \cdot b_{yx} = \frac{cov(x,y)}{\sigma x^2} \cdot \frac{cov(x,y)}{\sigma y^2} = \left[\frac{cov(x,y)}{\sigma x \sigma y}\right]^2$$

$$= [r(x,y)]^2$$

$$b_{xy} = \frac{cov(x,y)}{\sigma x^2}$$

Imp $$\boxed{b_{xy} = \frac{1/n \sum x_i y_i - \bar{X}\bar{Y}}{1/n \sum x_i^2 - (\bar{X})^2}}$$

Imp $$\boxed{b_{yx} = \frac{1/n \sum x_i y_i - \bar{X}\bar{Y}}{1/n \sum y_i^2 - (\bar{Y})^2}}$$

PAGE :
DATE : / /

Q Determine the eqns of regression lines for the following data. Find the value of

① y for $x = 4.5$ ② x for $y = 1.3$.

x	2	3	5	7	9	10	12	15
y	2	5	8	10	12	14	15	16

Soln ⇒

x_i	y_i	$x_i y_i$	y_i^2	x_i^2
2	2	4	4	4
3	5	15	25	9
5	8	40	64	25
7	10	70	100	49
9	12	108	144	81
10	14	140	196	100
12	15	180	225	144
15	16	240	256	225
$\Sigma x_i = 63$	$\Sigma y_i =$ 82	$\Sigma x_i y_i =$ 797	$\Sigma y_i^2 =$ 1014	$\Sigma x_i^2 =$ 637

∴ $\bar{x} = \frac{\Sigma x_i}{8} = \frac{63}{8}$

$= 7.875$

$\bar{y} = \frac{\Sigma y_i}{8} = \frac{82}{8}$

$= 10.25$

$$b_{yx} = \frac{1/n \, \Sigma x_i y_i - \bar{x}\bar{y}}{1/n \, \Sigma y_i^2 - (\bar{y})^2}$$

$$= \frac{1/8 \times 797 - 7.875 \times 10.25}{1/8 \times (\cancel{82}^{1014})^2 - (10.25)^2}$$

$$b_{yx} = \frac{18.90625}{\cancel{735.4375}} = 1.0736.$$

$$b_{xy} = \frac{1/n \, \Sigma x_i y_i - \bar{x}\bar{y}}{1/n \, \Sigma x_i^2 - (\bar{x})^2}$$

$$= \frac{1/8 \times 797 - 7.875 \times 10.25}{1/8 \times 637 - (7.875)} = 0.8718$$

PAGE :
DATE : / /

$(y-\bar{y}) = b_{yx}(x-\bar{x})$

$(x-\bar{x}) = b_{yx}(y-\bar{y})$

Q: In a partially distroyed labouratory record of an analysis of a correlation data following results are eligible.

Varience of $x = 9$

Regression eqns $= 8x - 10y + 66 = 0$

$40x - 18y = 244$

Find ① the mean values of x & y

② S.D. of y.

③ The correlation coefficient.

$\bar{x} = ?$ $\bar{y} = ?$ $\sigma y = ?$ $r = ?$

Ans ⇒

Given data : $\sigma x^2 = 9$.

Regression eqn

Replace x by $\bar{x}$ & y by $\bar{y}$

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

$40\bar{x} - 18\bar{y} = +244$

∴ Solve simultaneous eqn for $\bar{x}$ & $\bar{y}$

∴ $\bar{x} = 14.1718$ $\bar{y} = 17.967$

Regression lines

$$8x - 10y + 66 = 0 \quad \cdots (1)$$
$$40x - 18y = 244 \quad \cdots (2)$$

(1) eqn regression line y on x.

$$10y = 8x + 66$$
$$y = \underline{0.8}x + 6.6.$$

$$\boxed{b_{yx} = 0.8}$$

considering eqn (2) in regression line x on y

$$x = \frac{18}{40}y + \frac{244}{40} = 0.45y + 6.1$$

$$\boxed{b_{xy} = 0.45}$$

$$r = \sqrt{0.45 \times 0.8}$$
$$r = 0.6$$

Now $\sigma_x^2 = 9$

$\sigma_x = 3$; $\sigma_y = ?$; $r = 0.6$.

$$b_{xy} = \frac{r\sigma_x}{\sigma_y}$$

$$0.45 \times \sigma_y = 0.6 \times 3$$
$$\sigma_y = \frac{0.6 \times 3}{0.45}$$

$$\boxed{\sigma_y = 4}$$

PAGE:
DATE: / /

Remark: If we assume eq^n (1) as regression line x on y & eq^n (2) as y on x

$$x = \frac{10}{8}y - \frac{66}{8} = 1.25y - 8.25$$

$$\boxed{b_{xy} = 1.25}$$

$$y = \frac{40}{18}x + \frac{244}{18} = 2.22x + 13.5$$

$$\boxed{b_{yx} = 2.22}$$

Here, both the values of b_{xy} & b_{yx} can't be greater than 1 simultaneously. Hence the asumption is wrong.

Q: If two lines of regression are $9x + y - \lambda = 0$, $4x + y = u$ the means of x & y are 2 and -3 resp. Find the values of λ, u coefficient of correlation, coeff. of regression.

Soln ⇒

Mean satisfies x & y. replace it by $\bar{x}$ & $\bar{y}$

$$9\bar{x} + \bar{y} - \lambda = 0$$
$$4\bar{x} + \bar{y} = u$$
$$9 \times 2 + (-3) = \lambda$$
$$\boxed{\lambda = 15}$$

$$4 \times (2) + (-3) = u$$
$$\boxed{u = 5}$$

PAGE :
DATE : / /

$\therefore \quad 9x + y - 15 = 0$

$4x + y = 5$

$y = -9x + 15$ $\qquad\qquad$ $4x = -y + 5$

$\cancel{y = -4x + 5}$ $\qquad\qquad$ $x = \frac{-1}{4}y + \frac{5}{4}$

$\therefore \boxed{b_{yx} = -9}$ $\qquad$ $x = -0.25y + 1.25.$