MARUDHAR KESARI JAIN COLLEGE FOR WOMEN (AUTONOMOUS)
VANIYAMBADI
Department of Statistics

1st BSC STATISTICS – Semester - I

E-Notes (Study Material)

Core Course -1: Descriptive Statistics	**Code: 24USTC11**
Unit: IV – Measures of Dispersion, Measures of Skewness and Kurtosis. Introduction-Definition- Types- Range- Quartile deviation-Mean deviation- Statndard deviation- Co-efficient of variation- Karl pearson's- Bowley's- Kelly's method- Their merits and demerits, Kurtosis, Moments : Raw Moments, Central moments simple problems.	
Learning Objectives: To Understand the concept of measures of Dispersion and their importance is statistical analysis.	
Course Outcome: To understand the spread of variability of a dataset.	

Overview:

Measures of dispersion describe the spread or variability of data points in a dataset. They indicate how much the values differ from the central tendency.

Types of dispersion are (i) Absolute measure (ii) Relative measure.

Range: The difference between the maximum Values..

Inter quartile Range: The difference between the third quartile and the first quartile representing the spread of the middle 50% of the data.

Variance: The average of the squared differences from the mean.

Standard Deviation: The square root of variance, representing dispersion in the original units of the data.

Co-efficient of Variation: Expresses the standard deviation as a percentage of the mean.

Quartile co-efficient if Dispersion: Measures dispersion relative to quartiles.

UNIT-IV

19/9/24

MEASURES OF DISPERSION

Definition :

Dispersion is the measure of the variation of the items

Methods of Measuring Tendency Dispersion :-

1) Range
2) Inter quartile Range
3) Mean Deviation.
4) Standard Deviation
5) Lorenz Curve

1) Range :-

The range is the simplest measure of dispersion. It's measure depends upon the extreme items and not on all the items.

Range = Largest Value - Smallest Value

$= L - S$.

Coefficient of range $= \frac{L-S}{L+S}$.

Problem

Find the range of weights of seven students from the following:

27 30 35 36 38 40 43

Sol:

$$\text{Range} = L-S$$
$$= 43-27$$
$$= 16$$

$$\text{Coefficient of range} = \frac{L-S}{L+S}$$
$$= \frac{43-27}{43+27}$$
$$= \frac{16}{70}$$
$$= 0.228.$$

Merits of Range:

1) It is simple to compute and understand.
2) It gives a rough but quick answer.

Demerits of Range:

1) It is not reliable, because it is affected by the extreme items.
2) It cannot be applied to open ~~any~~ end cases.
3) It is not suitable for mathematical treatments.

2) Inter Quartile Range & Quartile Deviation

The distance between the first and the third quartiles called Inter-quartile range.

Inter Quartile range $= Q_3 - Q_1$

Quartile Deviation.

Semi-Inter Quartile range or quartile deviation is defined as half the distance between the third & first quartiles.

Semi Inter Quartile range (or) Quartile Deviation $= \dfrac{Q_3 - Q_1}{2}$

Coefficient of Quartile deviation $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$

$\boxed{8}$

<u>Problem</u>

Calculate the semi Inter Quartile range and Quartile coefficient from the following data.

Age	20	30	40	50	60	70	80
No. of members	3	61	132	153	140	51	3

Soln:

X	f	C.F
20	3	3
30	61	64
40	132	196
50	153	349
60	140	489
70	51	540
80	3	543
	$\Sigma f = 543$	

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

$=$ size of $\left(\frac{543+1}{4}\right)^{th}$ item

$=$ size of $\left(\frac{544}{4}\right)^{th}$ item

$=$ size of $(136)^{th}$ item

$\boxed{Q_1 = 40}$

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

$=$ size of $3\left(\frac{543+1}{4}\right)^{th}$ item

$=$ size of $3\left(\frac{544}{4}\right)^{th}$ item

$=$ size of $3\,(136)^{th}$ item

$=$ size of $(408)^{th}$ item

$\boxed{Q_3 = 60}$

Quartile Deviation $= \frac{Q_3 - Q_1}{2}$

$= \frac{60-40}{2}$

$= \frac{\cancel{20}^{10}}{\cancel{2}}$

$= 10.$

Coefficient of Quartile Deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \frac{60-40}{60+40}$

$= \frac{20}{100}$

$= 0.2.$

Problem

Calculate the Range and Semi Inter-Quartile range of wages for the below data

Wages	30-32	32-34	34-36	36-38	38-40	40-42	42-44
No. of labours	12	18	16	14	12	8	6

Also Calculate the Quartile Coefficient of Dispersion.

Soln: Range $= L-S$

$= 44-30$

$= 14$

Coefficient of range $= \frac{L-S}{L+S}$

$= \frac{44-30}{44+30}$

$= \frac{14}{74}$

$= 0.1892.$

Soly:

C.I	f	CF
30-32	12	12
32-34	18	30
34-36	16	46
36-38	14	60
38-40	12	72
40-42	8	80
42-44	6	86
	$\Sigma f = 86$	

Q_1 = size of $\left[\frac{N\cancel{+1}}{4}\right]^{th}$ item

= size of $\left[\frac{86}{4}\right]^{th}$ item

= size of $[21.5]^{th}$ item.

$Q_1 = 32-34$

$= L + \left[\frac{\frac{N}{4} - C.F}{F}\right] \times i$

$L = 32$, $N/4 = 21.5$ $C.F = 12$ $F = 18$. $i = 2$.

$= 32 + \left[\frac{21.5 - 12}{18}\right] \times 2$

$= 32 + \left[\frac{9.5}{18}\right] \times 2$

$= 32 + 0.5278 \times 2$

$= 32 + 1.0556$

$= 33.05$

$\boxed{Q_1 = 33.05}$

Q_3 = Size of $3\left(\frac{N}{4}\right)^{th}$ item

= Size of $3\left(\frac{86}{4}\right)^{th}$ item

= Size of $3(21.5)^{th}$ item

= Size of $(64.5)^{th}$ item

$Q_3 = 38-40$

$= L + \frac{\left[\frac{N}{4} - C.F\right]}{F} \times i$

$L = 38 \quad \frac{N}{4} = 64.5 \quad C.F = 60 \quad i = 2$

$= 38 + \left[\frac{64.5 - 60}{12}\right] \times 2.$

$= 38 + \left[\frac{4.5}{12}\right] \times 2$

$= 38 + 0.3750 \times 2$

$= 38 + 0.7500$

$\boxed{Q_3 = 38.75}$

Quartile Deviation $= \frac{Q_3 - Q_1}{2}$

$= \frac{38.75 - 33.05}{2}$

$= \frac{5.7}{2}$

$= 2.85$

Coefficient of Quartile Deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \frac{38.75 - 33.05}{38.75 + 33.05}$

$= \frac{5.7}{71.8} = 0.0794 = 0.07$

5/9/24

05/9

Merits of Quartile Deviation

. It is simple to understand and easy to calculate.

. It is not influenced by the extreme Values.

. It can be found out with open end distribution.

. It is not affected by presence of extreme values.

Demerits of Quartile Deviation

. It ignores the first 25% of the items and the last 25% of the items.

. It is a positional Average, hence not aminable to further mathematical treatments.

. It's value is affected by sampling fluctuations.

. It gives only a rough measure.

3. Mean Deviation

Mean Deviation is the arithmetic mean's of the deviation of a series completed from any measure of central tendency.

(i.e) Mean, Median & Mode.

All the deviations are taken has positive.

(i.e) Plus and minus signs are ignored.

$$\text{Mean Deviation} = \frac{\Sigma|D|}{N}.$$

$$\text{Coefficient of Mean deviation} = \frac{MD}{\bar{X} \text{ (or) } M \text{ (or) } Z}$$

Where.

MD → Mean deviation

N → Number of Observation (or) items.

$\Sigma|D|$ → Sum of the deviation

Problem

Calculate mean deviation from mean & median for the following.

100 150 200 250 360 490 500 600 671.

Also calculate coefficient of mean deviation.

Sol:

MD by mean

X	$\|D\| = \|X - \bar{X}\|$ ($\bar{X}$ = 369)
100	269
150	219
200	169
250	119
360	9
490	121
500	131
600	231
671	302
ΣX = 3321	$\Sigma\|D\|$ = 1570.

$\bar{X} = \frac{\Sigma X}{N} = \frac{3321}{9}$

$= 369.$

$MD = \frac{\Sigma|D|}{N} = \frac{1570}{9}$

$\boxed{MD = 174.4}$

Coefficient of MD $= \frac{MD}{\bar{X}}$

$= \frac{174.4}{369}$

$= 0.472.$

<u>MD by Median</u>.

X	$\lvert D\rvert = \lvert X - M\rvert$ 360
100	260
150	210
200	160
250	110
360	0
490	130
500	140
600	240
671	311
	$\Sigma\lvert D\rvert = 1561$.

M = size of $\left(\frac{N+1}{2}\right)^{th}$ item

= size of $\left(\frac{9+1}{2}\right)^{th}$ item

= size of $\left(\frac{10}{2}\right)^{th}$ item

= size of $(5)^{th}$ item

$\boxed{M = 360}$

Mean Deviation $= \frac{\Sigma|D|}{N}$

$= \frac{1561}{9}$

$MD = 173.4$

Coefficient of Mean Deviation $= \frac{MD}{M}$

$= \frac{173.4}{360}$

$= 0.481.$

Calculate mean deviation from the below data.

X	2	4	6	8	10
F	1	4	6	4	1

stoly

X	F	Fx	$\lvert D\rvert = \lvert X-\bar{X}\rvert$ 6	$f\lvert D\rvert$.
2	1	2	4	4
4	4	16	2	8
6	6	36	0	0
8	4	32	2	8
10	1	10	4	4
	$\Sigma f =$ N= 16	$\Sigma fx =$ 96		$\Sigma f\lvert D\rvert =$ 24

$\bar{X} = \frac{\Sigma fx}{N}$.

$= \frac{96}{16}$

$\boxed{\bar{X} = 6}$

Mean Deviation $= \frac{\Sigma f\lvert D\rvert}{N} = \frac{24}{16} = 1.5$.

Coefficient of MD $= \frac{MD}{\bar{X}} = \frac{1.5}{6} = 0.25.$

Calculate the mean deviation from the given data.

Class Int	2-4	4-6	6-8	8-10
Fre	3	4	2	1.

C.I	f	m	fm	$\lvert D\rvert = \lvert x-\bar{x}\rvert$ $\lvert m-\bar{x}\rvert$ 5.2	$f\lvert D\rvert$
2-4	3	3	9	2.2	6.6
4-6	4	5	20	0.2	0.8
6-8	2	7	14	1.8	3.6
8-10	1	9	9.	3.8	3.8
	$\Sigma f = N = 10$		$\Sigma fm = 52.$		$\Sigma f\lvert D\rvert = 14.8$

$$\bar{X} = \frac{\Sigma fm}{\Sigma f(N)}$$

$$= \frac{52}{10}$$

$$\boxed{\bar{X} = 5.2}$$

$$\text{Mean deviation} = \frac{\Sigma f\lvert D\rvert}{N}$$

$$= \frac{14.8}{10}$$

$$= 1.48.$$

5/9/24

$$\text{Coeff of M.D} = \frac{MD}{\bar{X}} = \frac{1.48}{5.2} = 0.28$$

Merits of Mean Deviation

- It is simple to understand and easy to compute.
- M.D is a calculated value.
- It is not much affected by the fluctuations of sampling.
- It is less affected by the extreme items
- It is rigidly defined.
- It is based on all the items of the series
- It is flexible.
- It is a better measure for comparison.

Demerits of Mean Deviation.

- It is non-algebraic treatment
- It is not suitable for further mathematical calculation.
- It is rarely used.

06/09/24

4. Standard Deviation.

It is defined as positive square of the Arithmetic mean of the square of the deviation of the given observation from there Arithmetic mean.

Individual series $\sigma = \sqrt{\frac{\Sigma(X-\bar{X})^2}{N}}$ (or) $\sigma = \sqrt{\frac{\Sigma d^2}{N}}$

Discrete series $\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$

Continuous series $\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i$

Problem.

1. Calculate standard deviation for the following.

14, 22, 9, 15, 20, 17, 12, 11.

Soln:

X	$d = X - \bar{X}$	d^2
14	-1	1
22	-7	49
9	-6	36
15	0	0
20	5	25
17	2	4
12	-3	9
11	-4	16
$\Sigma X = 120$		$\Sigma d^2 = 140$

$$\bar{X} = \frac{\Sigma X}{N}.$$

$$= \frac{120}{8}$$

$$\boxed{\bar{X} = 15}$$

$$\sigma = \sqrt{\frac{\Sigma d^2}{N}}$$

$$= \sqrt{\frac{140}{8}}$$

$$= \sqrt{17.5}$$

$$\boxed{\sigma = 4.18}$$

2. Problem.

Compute standard deviation for the following distribution table.

Marks	10	20	30	40	50	60
No. of. student	8	12	20	10	7	3

Soln:-

X	f	d = X-A	fd	d^2	fd^2
10	8	-20	-160	400	3200
20	12	-10	-120	100	1200
30	20	0	0	0	0
40	10	10	100	100	1000
50	7	20	140	400	2800
60	3	30	90	900	2700
	$\Sigma f = 60$		$\Sigma fd =$ 50		$\Sigma fd^2 =$ 10900

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

$$= \sqrt{\frac{10900}{60} - \left(\frac{50}{60}\right)^2}$$

$$= \sqrt{1.8166 - (0.833)^2}$$

$$= \sqrt{181.66 - 0.6944}$$

$$= \sqrt{180.96}$$

$$\boxed{\sigma = 13.45}$$

3. Determine standard deviation for the below information.

Class Interval	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45
Frequency	6	5	15	10	5	4	3	2

Soln:-

C.I	f	m	$d' = \frac{m-A^{22.5}}{i}$	fd'	fd'^2	$f(d')^2$
5-10	6	7.5	-3	-18	9	54
10-15	5	12.5	-2	-10	4	20
15-20	15	17.5	-1	-15	1	15
20-25	10	22.5	0	0	0	10
25-30	5	27.5	1	5	1	5
30-35	4	32.5	2	8	4	16
35-40	3	37.5	3	9	9	27
40-45	2	42.5	4	8	16	32
	$\Sigma f = 50$			$\Sigma fd' = 13$		$\Sigma f(d')^2 = 179.$

$$\sigma = \sqrt{\frac{\Sigma f d'^2}{N} - \left(\frac{\Sigma f d'}{N}\right)^2} \times i$$

$$= \sqrt{\frac{179}{50} - \left(\frac{-13}{50}\right)^2} \times 5$$

$$= \sqrt{3.58 - (0.26)^2} \times 5$$

$$= \sqrt{3.58 - 0.0676} \times 5$$

$$= \sqrt{3.51} \times 5$$

$$= 1.87 \times 5$$

$$\boxed{\sigma = 9.35}$$

Merits of standard deviation.

1.) It is based on arithmetic mean, it has all the merits of arithmetic mean.

2.) It is the most important and widely used measures of dispersion.

3) It is possible for further algebraic treatment.

4) It can be used to calculate the combined standard deviation of 2 (or) more groups.

Demerits.

1.) It is not easy to understand, & it is difficult to calculate.

2.) It gives more weight to extreme values because the values are squared up.

3.) It is affected by the Value of every item in the series.

4) It has not found favour with the economics & businessmen.

07/9/24 Comparison between Mean Deviation & Standard Deviation

	Mean Deviation	Standard Deviation
i)	Deviation are calculated from mean, median & mode.	Deviation are calculated only from mean.
ii)	Algebraic sign are ignored while calculating mean Deviation.	Algebraic sign are taken into Account.
iii)	It lacks mathematical properties because Algebraic sign are ignored.	It is mathematically sound because Algebraic signs are taken into Account.

Coefficient of Variation.

Variance

Square of Standard Deviation is called Variance.

$$\text{Variance} = \sigma^2$$

$$\sigma^2 = \text{Variance}$$

$$\sigma = \sqrt{\text{Variance}}.$$

Coefficient of standard deviation = $\frac{\sigma}{\bar{X}}$

The coefficient of standard deviation is multiplied by 100, it gives the Coefficient of Variation.

i.e.

Coefficient of Variation (CV) = $\frac{\sigma}{\bar{X}} \times 100$.

Coefficient of Variation - Greater.

The series ~~are~~ or groups of data for which the coefficient of variation is greater, indicates that the group is more variable, less stable, less uniform, less consistent or less homogenous.

Coefficient of Variation - Lesser.

If the coefficient of Variation is less, it indicates that the group is less variable, more stable, more uniform, more consistent or more homogenous.

Problem :- 1

The index numbers of prices of cotton and coal shares in 1992 where as under.

Month	Cotton shares	Coal shares
January	188	131
February	178	130
March	173	129
April	164	129
May	172	120
June	183	127
July	184	127
August	185	130
September	211	137
October	217	140
November	232	130
December	240	142

Soln Which of the two shares do you consider more variable in price?

Month	Cotton shares (X)			Coal shares (Y)		
	X	d_x	d_x^2	Y	d_y	d_y^2
Jan	188	6	36	131	0	0
Feb	178	16	256	130	-1	1
Mar	173	21	441	129	-2	4
April	164	30	900	129	-2	4
May	172	22	484	120	-11	121
June	183	11	121	127	-4	16
July	184	10	100	127	-4	16
Aug	185	9	81	130	-1	1
Sep	211	-17	289	137	-6	36
Oct	217	-23	529	140	9	81
Nov	232	-38	1444	130	-1	1
Dec	240	-46	2116	142	+11	121
	$\Sigma X = 2327$	$\Sigma dx = 1$	$\Sigma dx^2 = 6,797$	$\Sigma Y = 1572$	$\Sigma dy = -12$	$\Sigma dy^2 = 402$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{2327}{12}$$

$$= 193.9$$

$$\simeq 194$$

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

$$= \sqrt{\frac{6797}{12} - \left(\frac{1}{12}\right)^2}$$

$$= \sqrt{566.4 - 0.0069}$$

$$= \sqrt{566.39}$$

$$= 23.7$$

$$CV = \frac{\sigma}{\bar{X}} \times 100 = \frac{23.7}{194} \times 100$$

$$= 12.21$$

$$\bar{y} = \frac{\Sigma Y}{N}$$

$$= \frac{1572}{12}$$

$$\bar{y} = 131$$

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

$$= \sqrt{\frac{402}{12} - \left(\frac{-12}{12}\right)^2}$$

$$= \sqrt{33.5 - (-1)^2}$$

$$= \sqrt{33.5 - 1}$$

$$= \sqrt{32.5}$$

$$\sigma = 5.70$$

$$CV = \frac{\sigma}{\bar{y}} \times 100$$

$$= \frac{5.70}{131} \times 100$$

$$CV = 4.35$$

Hence, cotton shares are more variable in price than the coal shares.

Problem :- 2

Coefficient of Variation of two series are 58% and 69%. Their standard deviations are 21·2 & 15·6. What are their Arithmetic Means

Sol:- $CV(1) = 58\%$ $\quad CV(2) = 69\%$

$SD\ \sigma = 21.2$ $\quad S.D\ \sigma = 15.6$

$\bar{X} = ?$ $\quad \bar{Y} = ?$

X series

$$CV = \frac{\sigma}{\bar{X}} \times 100$$

$$58 = \frac{21.2}{\bar{X}} \times 100$$

$$\bar{X} = \frac{21.2}{58} \times 100$$

$$\bar{X} = 36.5$$

Y series

$$69 = \frac{\sigma}{\bar{Y}} \times 100$$

$$\bar{Y} = \frac{15.6}{69} \times 100$$

$$= 22.6$$

Problem :- 3

Prices of a particular commodity in 5 years in two cities are given below.

Prices in city A	Prices in city B.
	10
20	20
22	18
19	12
23	15
16	

From the above data find the city which had more stable prices.

	City A		City B.		
X	$dx=X-\bar{X}$	dx^2	Y	$dy=y-\bar{y}$	dy^2
				-5	25
		0	10		
20	0		20	5	25
22	2	4		3	9
19	1	1	18		
				-3	9
23	3	9	12		
				0	0
16	-4	16	15		
$\Sigma X=100$		$\Sigma dx^2=30$	$\Sigma y=75$		$\Sigma dy^2=68$.

$$\bar{X} = \frac{\Sigma X}{N}$$

$$= \frac{100}{5}$$

$$= 20$$

$$\sigma_x = \sqrt{\frac{\Sigma d^2}{N}}$$

$$= \sqrt{\frac{30}{5}}$$

$$= \sqrt{6}$$

$$= 2.45$$

$$CV = \frac{\sigma x}{\bar{x}} \times 100$$

$$= \frac{2.45}{20} \times 100$$

$$= 12.25$$

$$\bar{y} = \frac{\Sigma y}{N}$$

$$= \frac{75}{5}$$

$$= 15$$

$$\sigma y = \sqrt{\frac{\Sigma dy^2}{N}}$$

$$= \sqrt{\frac{68}{5}}$$

$$= \sqrt{13.6}$$

$$= 3.69$$

$$CV = \frac{\sigma y}{\bar{y}} \times 100$$

$$= \frac{3.69}{15} \times 100$$

$$= 24.6$$

City A had more stable prices than in City B, Because the Coefficient of Variations is lower in city A.

Scanned with OKEN Scanner

12/9/2024

Problem : 4

The mean of 50 items is 25 and units SD is 2. Find the sum and sum of the squares of all the items

Soln:

$$\bar{X} = \frac{\Sigma X}{N}$$

$$\Sigma X = N\bar{X}$$

$$= 50(25)$$

$$= 1250.$$

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

$$2 = \sqrt{\frac{\Sigma X^2}{50} - (25)^2}$$

Squaring both sides, we get.

$$4 = \frac{\Sigma X^2}{50} - (25)^2$$

$$4 \times 50 = \Sigma X^2 - (25)^2 \times 50$$

$$200 = \Sigma X^2 - 625 \times 50$$

$$= \Sigma X^2 - 31250.$$

$$\boxed{\Sigma X^2 = 31450}$$

Lorenz Curve.

Lorenz curve is a device used to show the measurement of economic inequalities as in the distribution of income & wealth.

Lorenz Curve can also be used for the study of distribution of profits, wages etc... Lorenz Curve also known as graphic method of dispersion.

Constructing of Lorenz Curve.

* The size of the item and their frequencies are to be cumulated.

* Percentage must be calculated for each cumulation, value of the size and frequency of items.

* Plot the percentage of the cumulated values of the Variable against the percentage of the corresponding cumulated frequencies. Join these points with a smooth free hand curve, this curve is known as Lorenz Curve.

* The Zero percentage on the X-axis must be joined with 100 percent on Y-axis. This line is called line of equal distribution.

18/09/24

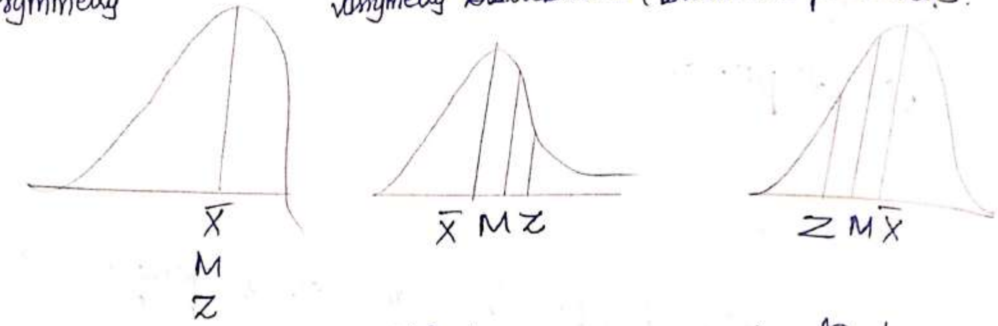
Measure of skewness

Skewness

* Skewness is a measure of the symmetry of a distribution. A distribution is assymmetrical when it's left and right side are not mirror images.

* When a series is not symmetrical it is said to be asymmetrical or skewed.

Symmetry Asymmetry Distribution (skewness present).



No skewness Positively skewed $\bar{X} > M > Z$ Negatively skewed $\bar{x} < M < Z$.

* A distribution which is not symmetrical is called skewness (i.e) Mean, Median, Mode will not ~~go~~ ~~[illegible]~~ co-inside.

Measures of skewness

Absolute skewness = Mean - Mode.

$$= \begin{bmatrix} + \text{Positive skewness} \\ - \text{Negative skewness} \end{bmatrix}$$

If the Value of the mean is greater than the mode, the skewness is positive.

If the value of the mean is less than the mode, the skewness is negative.

Relative measure of skewness.

(P)

1) Karl Pearson's Coefficient of skewness.

2) Bowley's Coefficient of skewness.

3) Kelly's coefficient of skewness.

1) Karl Pearson's Coefficient of skewness.

$$\text{Coefficient of skewness } (Sk_P) = \frac{\text{Mean} - \text{Mode}}{\text{Standard Deviation}}$$

$$= \frac{\bar{X} - Z}{\sigma}.$$

In case, the mode is illdefined. the coefficient can be determined by,

$$(Sk_P) = \frac{3(\text{Mean} - \text{Median})}{SD}$$

$$= \frac{3(\bar{X} - M)}{\sigma}.$$

Individual.

Calculate Karl Pearson's Coeff of skewness for the following data.

25 15 23 40 27 25 23 25 20

X	$d = X - A$ (27)	d^2
25	-2	4
15	-12	144
23	-4	16
40	13	169
27	0	0
25	-2	4
23	-4	16
25	-2	4
20	-7	49
	$\Sigma d = -20$	$\Sigma d^2 = 406$

-2.22

$$\bar{X} = A \pm \frac{\Sigma d}{N}.$$

$$= 27 - \left(\frac{-20}{9}\right)$$

$$= 27 - (-2.22)$$

$$= 27 + 2.22$$

$$= 29.22$$

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

$$= \sqrt{\frac{406}{9} - \left(\frac{-20}{9}\right)^2}$$

$$= \sqrt{45.1 - 4.93}$$

$$= \sqrt{40.17}$$

$$\sigma = 6.33.$$

Mode $= 25$

25 is repeated 3 times. so this is most common item.

$$\text{Coefficient of skewness } (Sk_p) = \frac{\text{Mean} - \text{Mode}}{S.D}$$

$$= \frac{\bar{X} - Z}{\sigma}$$

$$= \frac{29.22 - 25}{6.33}$$

$$= \frac{4.220}{6.33}$$

$$\boxed{Sk_p = 0.667}$$

∴ The distribution is positively skewed.

Discrete Series

2. Find the Coefficient of Skewness from the data given below.

Size	: 3	4	5	6	7	8	9	10
Frequency	: 7	10	14	35	102	136	43	8.

X	f	d = X−A	fd	d^2	fd^2.
3	7	−3	−21	9	63
4	10	−2	−20	4	40
5	14	−1	−14	1	14
6	35	0	0	0	0
7	102	1	102	1	102
8	136	2	272	4	544
9	143	3	129	9	387
10	8	4	32.	16	128
	N = 355		Σfd = 480		Σfd^2 = 1278

$$\text{Mean } \bar{X} = A \pm \frac{\Sigma fd}{N}$$

$$= 6 + \frac{480}{355}$$

$$= 6 + 1.35$$

$$= 7.35$$

Standard Deviation $\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$

$$= \sqrt{\frac{1278}{355} - \left(\frac{480}{355}\right)^2}$$

$$= \sqrt{3.6 - (1.35)^2}$$

$$= \sqrt{3.6 - 1.82}$$

$$= \sqrt{1.78}$$

$$= 1.33$$

Mode = 8

Highest frequency is 136. So mode occur on 8.

Coeff of skewness $S_{kp} = \frac{\bar{x} - z}{\sigma}$

$$= \frac{7.35 - 8}{1.33}$$

$$= \frac{-0.65}{1.33}$$

$$\boxed{S_{kp} = -0.49}$$

∴ So the distribution is negatively skewed.

Scanned with OKEN Scanner

Continuous Series

3. Find the standard deviation and co-efficient of skewness for the given distribution.

Variable Income	400-500	500-600	600-700	700-800	800-900
No. of employee	8	16	20	17	3

C.I	f	m	$d' = \frac{M-A}{i}$	d'^2	fd'	fd'^2
400-500	8	450	-2	4	-16	32
500-600	16	550	-1	1	-16	16
600-700	20	650	0	0	0	0
700-800	17	750	1	1	17	17
800-900	3	850	2	4	6	12
	$\Sigma f = 64$				$\Sigma fd' = -9$	$\Sigma fd'^2 = 77$

Mean.

$$\bar{X} = A \pm \frac{\Sigma fd'}{N} \times i$$

$$= 650 - \left[\frac{9}{64}\right] \times 100$$

$$= 650 - 14.06$$

$$= 635.94$$

Standard Deviation.

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i$$

$$= \sqrt{\frac{77}{64} - \left(\frac{-9}{64}\right)^2} \times 100$$

$$= \sqrt{1.2031 - 0.0198} \times 100$$

$$= \sqrt{1.1833} \times 100$$

$$= \sqrt{1.087}$$

$$= 108.7$$

Mode

By inspection Method.

Highest frequency 20.

$\therefore$ Modal class is 600-700.

$$Z = L_1 + \left[\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right] \times i$$

$$= 600 + \left[\frac{20 - 16}{2(20) - 16 - 17}\right] \times 100$$

$$= 600 + \left[\frac{4}{40 - 33}\right] \times 100$$

$$= 600 + 57.14$$

$$Z = 657.14$$

Coeff of skewness $S_{KP} = \dfrac{\bar{x} - Z}{\sigma}$

$$= \frac{635.94 - 657.14}{108.7}$$

$$= \frac{-21.20}{108.7}$$

$$S_{KP} = -0.195$$

$\therefore$ The distribution is negatively skewed.

4. From a moderately skewed distribution of retail prices for men's shoes, it is found that the mean price is Rs. 20 and the median price is Rs. 17. If the Coefficient of Variation is 20%. Find the Pearsonian Coeff of skewness of the distribution

Given

Mean $\bar{X} = 20$

Median $M = 17$

$CV = 20$

$$CV = \frac{\sigma}{\bar{X}} \times 100$$

$$20 = \frac{\sigma}{20} \times 100$$

$$400 = \sigma \times 100$$

$$\sigma = \frac{400}{100}$$

$$\boxed{\sigma = 4}$$

Coefficient of skewness $S_{kp} = \frac{3[\bar{X} - M]}{\sigma}$

$$= \frac{3(20-17)}{4}$$

$$= \frac{3(3)}{4}$$

$$= \frac{9}{4}$$

$\therefore$ Coeff of $S_{kp} = 2.25$

$\therefore$ The distribution is positively skewed.

19/09

2) Bowley's Coefficient of skewness.

Professor Bowley has suggested a formula based on relative position of quartiles.

In a symmetrical distribution, the quartile are equidistant from the value of the mean.

(i.e) Median $- Q_1 = Q_3 -$ Median.

Which means the value of the median is the mean of Q_1 and Q_2, but in a skewed distribution the quartiles will not be equidistant from the median.

$$\text{Bowley's coefficient of skewness } SK_B = \frac{Q_3 + Q_1 - 2\,\text{Median}}{Q_3 - Q_1}$$

Problem

From the following table compute the quartile deviation & bowley's coefficient of skewness.

Size	4-8	8-12	12-16	16-20 36-40.	20-24	24-28	28-32	32-36
Fre	6	10	18	30 2	15	12	10	6

C.I	f	cf
4-8	6	6
8-12	10	16
12-16	18	34
16-20	30	64
20-24	15	79
24-28	12	91
28-32	10	101
32-36	6	107
36-40	2	109
	$\Sigma f = 109$	

Q_1

Median

Q_3

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

= size of $\left(\frac{109}{4}\right)^{th}$ item

= size of $(27.25)^{th}$ item

Q_1 = 12 − 16. [∵ Q_1 lies between 12 to 16]

$$= L + \left[\frac{N/4 - C.F}{F}\right] \times i$$

$L = 12$, $N/4 = 27.25$, $C.F = 16$, $F = 18$, $i = 4$.

$$= 12 + \left[\frac{27.25 - 16}{18}\right] \times 4$$

$$= 12 + \left[\frac{11.25}{18}\right] \times 4$$

$$= 12 + 0.6250 \times 4$$

$$= 12 + 2.5$$

$$\boxed{Q_1 = 14.5}$$

Q_3 = size of $3\left(\frac{N}{4}\right)^{th}$ item.

= size of $3(27.25)^{th}$ item

= size of $(81.75)^{th}$ item

Q_3 lies between 24 to 28.

$= L_1 + \left[\frac{3(N/4) - C.F}{F}\right] \times i$

$L_1 = 24$, $3(N/4) = 81.75$, $C.F = 79$, $F = 12$, $i = 4$.

$= 24 + \left[\frac{81.75 - 79}{12}\right] \times 4$

$= 24 + \left[\frac{2.75}{12}\right] \times 4$

$= 24 + 0.2292 \times 4$

$= 24 + 0.9167$

$\boxed{Q_3 = 24.9167}$

Bowley's Co. Median = size of $\left(\frac{N}{2}\right)^{th}$ item

= size of $\left(\frac{109}{2}\right)^{th}$ item

= size of $(54.5)^{th}$ item.

Median lies between 16 to 20.

$= L + \left[\frac{N/2 - C.F}{F}\right] \times i$

$L_1 = 16$, $N/2 = 54.5$, $C.F = 34$, $F = 30$, $i = 4$

$= 16 + \left[\frac{54.5 - 34}{30}\right] \times 4$

$= 16 + \left[\frac{20.5}{30}\right] \times 4$

$= 16 + 0.6833 \times 4$ $\boxed{M = 18.7333}$.

$= 16 + 2.7333$

Bowley's Coefficient of skewness $Sk_B = \frac{Q_3 + Q_1 - 2\,Median}{Q_3 - Q_1}$

$$= \frac{24.9167 + 14.5 - 2(18.7333)}{24.9167 - 14.5}$$

$$= \frac{24.9167 + 14.5 - 37.4666}{10.4167}$$

$$= \frac{1.9501}{10.4167}$$

$$\boxed{Sk_B = 0.1872}$$

19/9/24

$\therefore$ So, the distribution is positively skewed.

Find the coefficient of skewness, if difference between two quartiles is 8. sum of two quartiles is 22 and Median is 10.5.

Sol: Given

$Q_3 - Q_1 = 8$

$Q_3 + Q_1 = 22$

$M = 10.5$

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

$$= \frac{22 - 2(10.5)}{8}$$

$$= \frac{22 - 21}{8}$$

$$= \frac{1}{8}$$

$\boxed{Sk_B = 0.125}$ $\therefore$ The given distribution is positively skewed.

Hw

1.) In a frequency distribution, the coeff of skewness based on quartile is 0.6. If the sum of the upper & lower quartiles is 100, And the median is 38, find the Value of the upper quartile.

2.) Find the coeff of skewness for the following information.

$Q_1 = 18$

$Q_3 = 25$

Mode $= 21$

Mean $= 18$.

1.) $Sk_b = 0.6$, $Q_3 + Q_1 = 100$

Median $= Q_2 = 38$

$$Sk_b = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

$$0.6 = \frac{100 - 2 \times 38}{Q_3 - Q_1}$$

$$Q_3 - Q_1 = \frac{100 - 76}{0.6}$$

$$Q_3 - Q_1 = \frac{24}{0.6}$$

$$Q_3 - Q_1 = \frac{24}{0.6} \times \frac{10}{10}$$

$$= \frac{240}{6}$$

$$= 40$$

$Q_3 - Q_1 = 40 \rightarrow (1)$

$Q_3 + Q_1 = 100 \rightarrow (2)$

ff of skewness
of the upper
an is 38. Find

llowing

Add ① & ②

$$Q_3 - \cancel{Q_1} = 40$$
$$Q_3 \cancel{+} Q_1 = 100$$
$$2Q_3 = 140$$
$$Q_3 = \frac{\cancel{140}\ 70}{\cancel{2}}$$
$$\boxed{Q_3 = 70}$$

Substitute the Value of Q_3 in ②

$$70 + Q_1 = 100.$$
$$Q_1 = 100 - 70$$
$$\boxed{Q_1 = 30}$$

20/9/24

$\therefore$ upper quartile = 70 & lower quartile = 30

2) Mean − Mode = 3(Mean − Median)

$$18 - 21 = 3(18 - \text{Median})$$
$$-3 = 3(18 - \text{Median})$$
$$\frac{-\cancel{3}}{\cancel{3}}\ 1 = 18 - \text{Median}$$
$$\text{Median} = 18 + 1 = 19.$$
$$Sk_b = \frac{Q_3 + Q_1 - 2(\text{Median})}{Q_3 - Q_1}$$
$$= \frac{25 + 18 - 2(19)}{25 - 18}$$
$$= \frac{43 - 38}{7}$$
$$= \frac{5}{7}$$
$$\therefore \boxed{Sk_b = 0.714}$$

20/09/24

3) Kelly's Coefficient of Skewness.

2M

To measure the skewness, it would be better to consider the entire data or more extreme items. Kelly's measures of skewness is based on two deciles. (i.e) 10^{th} & 90^{th} percentiles ($1st$ & 9^{th} deciles).

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

Also

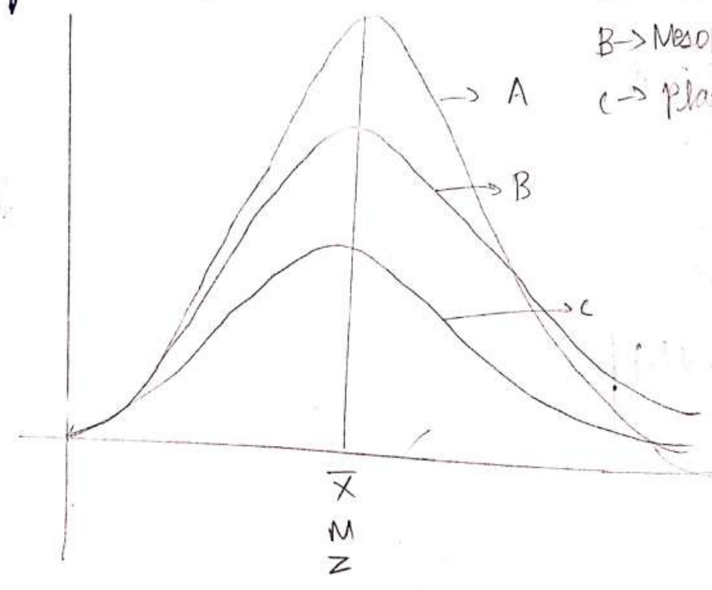
$$Sk_K = \frac{D_9 + D_1 - 2\,Median}{D_9 - D_1}$$

Kurtosis

2M

* The degree of kurtosis of a distribution is measured relative to the peakedness or a normal.
* A measure of kurtosis is used to describe the peakedness of a curve.

A -> Leptokurtic (or) Peaked
B -> Mesokurtic (or) normal
C -> Platykurtic (or) flat topped



* A normal curve which is symmetrical and bell shaped, is designed as <u>mesokurtic</u> because it is curve kurtic in the centre.

* If a curve is relatively more narrow and peaked at the top, it is defined as <u>leptokurtic</u>.

* If the frequency curve is more flat than normal curve, it is designed as <u>Platykurtic</u>

MEASURES OF KURTOSIS

The measures of Kurtosis of a frequency distribution are based upon the fourth moment about the mean of the distribution.

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

Where,

$\mu_4 \rightarrow$ Fourth moments

$\mu_2 \rightarrow$ Second moments

Lepto kurtic   Mesokurtic   Platy kurtic

* If $\beta_2 > 3$, the distribution is said to be more peaked and the curve is leptokurtic or peakedness.

* If $\beta_2 = 3$, the distribution is said to be normal and the curve is a mesokurtic or normal.

* If $\beta_2 < 3$, the distribution is said to be flat topped and the curve is platykurtic or flat topped.

CALCULATION

MOMENTS.

Moments can be defined as the arithmetic mean of various powers of deviation taken from the mean of a distribution.

RAW
CENTRAL MOMENTS. ABOUT MEAN (or) CENTRAL MOMENTS

	Individual Series I·S	D·S
First moment about the mean μ_1'	$\frac{\Sigma d}{N} = 0$	$\frac{\Sigma fd}{N}$
Second moment about the mean μ_2'	$\frac{\Sigma d^2}{N}$	$\frac{\Sigma fd^2}{N}$
Third moment about the mean μ_3'	$\frac{\Sigma d^3}{N} = 0$	$\frac{\Sigma fd^3}{N}$
Fourth moment about the mean μ_4'	$\frac{\Sigma d^4}{N} = 0$	$\frac{\Sigma fd^4}{N}$

25/10/2024

CENTRAL MOMENTS

$$\mu_1 = \mu_1' - \mu_1' = 0$$

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

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

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

Problem.

The calculation of moments about the mean is shown below.

2 4 6 8 10

X	d = X - A	d^2	d^3	d^4
2	-4	16	-64	256
4	-2	4	-8	16
6	0	0	0	0
8	2	4	8	16
10	4	16	64	256
	$\Sigma d = 0$	$\Sigma d^2 = 40$	$\Sigma d^3 = 0$	$\Sigma d^4 = 544$

$$\mu_1 = \frac{\Sigma d}{N} \text{ (or) } \frac{\Sigma |X - A|}{N}$$

$$= \frac{0}{N} = 0$$

$$\mu_2 = \frac{\Sigma d^2}{N} = \frac{40}{5} = 8$$

$$\mu_3 = \frac{\Sigma d^3}{N} = \frac{0}{N} = 0$$

$$\mu_4 = \frac{\Sigma d^4}{N} = \frac{544}{5} = 108.8$$

Kurtosis

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

$$= \frac{108.8}{(8)^2}$$

$$= \frac{108.8}{64}$$

$$\boxed{\beta_2 = 1.7}$$

$\therefore \beta_2 < 3$

So, the distribution is platykurtic

2) The first four central moments of a distribution are 0, 2.5, 0.7, 18.75. Test the skewness and kurtosis of the distribution.

Soln: Given

Central moments

$\mu_1 = 0$

$\mu_2 = 2.5$

$\mu_3 = 0.7$

$\mu_4 = 18.75$

Skewness $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$

$$= \frac{(0.7)^2}{(2.5)^3}$$

$$= \frac{0.49}{15.625}$$

$$\boxed{\beta_1 = 0.031}$$

$\therefore$ The distribution is not perfectly symmetrical

Kurtosis.

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

$$= \frac{18.75}{(2.5)^2}$$

$$= \frac{18.75}{6.25}$$

$$\boxed{\beta_2 = 3}$$

$\therefore \beta_2 = 3.$

So, the distribution is Mesokurtic (or) Normal curve.

3.) In a certain distribution, the first four moments about the ~~its~~ me are 2, 20, 40, 50. Calculate β_1 & β_2 and state whether the distribution is leptokurtic or platykurtic.

4: The first four raw moments

$\mu_1' = 2$

$\mu_2' = 20$

$\mu_3' = 40$

$\mu_4' = 50.$

Central moments.

$\mu_1 = \mu_1' - \mu_1' = 2 - 2 = 0.$

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

$= (20) - (2)^2$

$= 20 - 4$

$\mu_2 = 16.$

$$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2(\mu_1')^3$$
$$= 40 - 3(2)(20) + 2(2)^3$$
$$= 40 - 120 + 16$$
$$= 56 - 120$$
$$= -64.$$

$$\mu_4 = \mu_4' - 4\mu_1'\mu_3' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4.$$
$$= 50 - 4(2)(40) + 6(20)(2)^2 - 3(2)^4$$
$$= 50 - 320 + 120(4) - 3(16)$$
$$= 50 - 320 + 480 - 48$$
$$= 162.$$

Kurtosis $\Rightarrow \beta_2 = \dfrac{\mu_4}{(\mu_2)^2}$

$$= \frac{162}{(16)^2}$$
$$= \frac{162}{256}$$
$$\beta_2 = 0.6328$$
$$\boxed{\beta_2 = 0.6}$$

$\therefore \beta_2 < 3$

so, the given distribution is platykurtic.

Skewness

$$\beta_1 = \frac{(\mu_3)^2}{(\mu_2)^3}$$
$$= \frac{(-64)^2}{(16)^3}$$
$$= \frac{4096}{4096}$$
$$\beta_1 = 1$$

$\therefore \beta < 3$

so, the given distribution is platykurtic

5. Calculate first term moments from the following data and find out β_1 & β_3.

X	0	1	2	3	4	5	6	7	8
f	5	10	15	20	25	20	15	10	5.

Soln:

X	f	fx	d = X - A	fd	fd²	fd³	fd⁴
0	5	0	-4	-20	80	-320	1280
1	10	10	-3	-30	90	-270	810
2	15	30		-30	60	-120	240
3	20	60	-2	-20	20	-20	20
4	25	100	-1	0	0	0	0
5	20	100	0	20		20	20
6	15	90	1	30		120	240
7	10	70	2	30	90	270	810
8	5	40	3 4	20	80	320	1280
	Σf = 125	Σfx = 500		Σfd = 0	Σfd² = 500	Σfd³ = 0	Σfd⁴ = 4700.

$$\mu_1 = \frac{\Sigma fd}{N} = \frac{0}{125} = 0$$

$$\mu_2 = \frac{\Sigma fd^2}{N} = \frac{500}{125} = 4$$

$$\mu_3 = \frac{\Sigma fd^3}{N} = \frac{0}{125} = 0$$

$$\mu_4 = \frac{\Sigma fd^4}{N} = \frac{4700}{125} = 37.6$$

Skewness $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{0}{64}$

$$\boxed{\beta_1 = 0}.$$

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

$$= \frac{37.6}{4^2}$$

$$= \frac{37.6}{16}$$

$$\boxed{\beta_2 = 2.35}$$

$\beta_2 < 3$,
Hence the curve is platykurtic.

Unit-IV Verified
[illegible]

Practice Questions

Section-A

1. Define Range?

2.What is Inter-quartile range?

3.Write the formula for Range?

4.What is Standard deviation?

5.What do you mean by Mean deviation

Section-B

1.Explain about the Standard deviation with example?

2.Discuss about the merits and demerits of mean deviation?

3..Describe the co-efficient of variation with example?

Section-C

1.Explain the Measures of dispersion with example?

Reference Books:

Goom A.M.Gupta A.K and Das Gupta (1987) Fundamental of Statistics. Vol 2 World press Pvt., Kalkata.