UNIT-IV
MEASURES OF DISPERSION & MEASURES OF SKEWNESS

Measures of Dispersion

It helps us to study the extent of the variation of the individual observations from the central value and among themselves

For example, consider the three series

	Series A	Series B	Series C
	201	200	75
	198	200	405
	199	200	150
	200	200	270
	202	200	1**00**
total	1000	1000	1000
A.M	200	200	200

Two Broad category of Measures of Dispersion:

- **Absolute measures**: Absolute measures are given in the same units as the individual data
- **Relative measures**: Relative measures are free from units of measurements and are called as co-efficients and are used to compare two or more series for their variability.

Absolute measures & their Relative measures (coefficients)

[1] Range and co-efficient of range
[2] Quartile deviation (Q.D) and co-efficient of Q.D.
[3] Mean deviation (M.D) and co-efficient of M.D.
[4] Standard deviation (S.D) and co-efficient of variation(C.V).

[1] RANGE & CO-EFFICIENT OF RANGE

Definition:

Range is defined as the difference between the highest and lowest value of x.

Range R = largest value – smallest value = L – S

Co-efficient of Range = (L - S) /(L+S)

Example:

Find the range and its co-efficient: 234, 22, 425, 325, 78, 236, 120, 422

Solution: Largest value L= 425; Smallest value S = 22;
Range = 425 – 22 = 403
Co-efficient of Range = (L - S) /(L+S)
= (425-22)/(425 +22)
= 403/ 447 = 0.90

Example: Find range and its co-efficient

X	f
10 -20	4
20-30	12
30-40	25
40-50	16
50-60	6

Solution:

Largest value of X , L = 60
Smallest value of X , S = 10
Range = L – S = 60 -10 =50
Co-efficient of range = (L - S) /(L+S) = (60 -10) /(60+10) = 50/70 =0.71

[2] QUARTILE DEVIATION (Q.D) & CO-EFFICIENT OF Q.D.

Definition:
Q.D. is the difference between upper and lower quartile divided by 2.
i.e., Q.D. = $(Q_3 - Q_1) /2$
Co-efficient of Q.D. = $(Q_3 - Q_1) /(Q_3 + Q_1)$

Example: Find the Q.D. and its co-efficient
23, 32, 46, 67, 12, 89, 32, 45, 90,15

Solution:
Arrange the data in ascending order:12,15,23,32,32,45,46,67,89,90
Q_3 = value of {3(n+1)/4}th observation
= 8.25 th value = 8^{th} value + 0.25 (9^{th} -8^{th} value)
=67 + 0.25 (89 - 67) = 67 + 0.25 x22 = 67 + 5.5 = 72.5
Q_1 = value of (n+1)/4 th observation
= 2.75 th value = 2^{nd} value + 0.75 (3^{rd} - 2^{nd} value)
=15 + 0.75 (23 - 15) = 15 + 0.75 x 8 = 15 +6 = 21

Q.D. = **(Q_3 - Q_1) /2 = (72.5 – 21)/2 = 51.5/2 =25.75**

Co-efficient of Q.D. = (Q_3 -Q_1) /(Q_3 + Q_1)
= (72.5 -21) /(72.5 + 21)
= 51.5 / 93.5 = 0.55

Example: Find Quartile deviation and its co-efficient

x	f
20	5
22	8
24	12
26	6
28	3
30	2

Solution:
Q_3 = value of x corresponding to the c.f. just greater than or = 3N/4
Q_1 = value of x corresponding to the c.f. just greater than or = N/4

x	f	c.f.
20	5	5
22	8	13
24	12	25
26	6	31
28	3	34
30	2	36
	N=36	

N/4 = 36/4 = 9
3N/4 = 3 x 9 =27

Q_1 = value of x corresponding to the c.f. just ≥ N/4 i.e. 13 = 22
Q_3 = value of x corresponding to the c.f. just ≥ 3N/4 i.e. 31 = 26

Q.D. = (Q_3 - Q_1) /2 = (26 -22)/2 = 4/2 =2
Co-efficient of Q.D. = (Q_3 -Q_1) /(Q_3 + Q_1)
= (26-22) /(26 + 22)
= 4/48 = 1/12

Example: Find the Q.D. and Its co-efficient

C.I	f
20-25	3
25-30	8
30-35	16
35-40	22
40-45	15
25-50	5

Solution:

C.I	f	c.f.
20-25	3	3
25-30	8	11
30-35	16	27
35-40	23	50
40-45	15	65
25-50	5	70
	N=70	

$N/4 = 70/4 = 17.5$; $3N/4 = 3 \times 17.5 = 52.5$

Q_1 class is 30-35, Q_3 class = 40- 45

$Q_1 = L_1 + \{(N/4 – c.f_1) \times c_1/f_1\}$

Q_1 classis 30-35, $L_1 = 30$, $c_1 = 35-30 = 5$, $f_1 = 16$, $c.f_1 = 11$

$Q_1 = 30 + \{17.5 – 11) \times 5 /16\}$

$= 30 + \{6.5 \times 5 /16\} = 30 + 2.03 = 32.03$

$Q_3 = L_3 + \{(3N/4 – c.f_3) \times c_3/f_3\}$

Q_3 class is 40-45

$L_3 = 40$, $c_3 = 45 – 40 = 5$, $f_3 = 15$, $c.f_3 = 50$

$Q_3 = L_3 + \{(3N/4 – c.f_3) \times c_3/f_3\}$

$= 40 + \{(52.5 – 50) \times 5/ 15\}$

$= 40 + \{2.5 \times 5/15\} = 40 + 0.83 = 40.83.$

Q.D. $= (Q_3 - Q_1) /2 = (40.83 – 32.03)/2 = 8.8/2 = 4.4$

Co-efficient of Q.D. $= (Q_3 - Q_1) / (Q_3 + Q_1)$

$= (40.83 – 32.03) / (40.83 + 32.03)$

$= 8.8 / 72.86 = 0.12$

[3] MEAN DEVIATION M.D. or AVERAGE DEVIATION & its Coefficient

Mean deviation is defined as the arithmetic mean of the absolute deviations of the observations from mean or median or mode. By absolute deviation we mean the positive deviation of the observations from average.

Calculation of M.D. for Raw data

M.D. about A = $\frac{\Sigma\, |x - A\,|}{n}$

Co-efficient of mean deviation = $\frac{M.D.\ about\ A}{A}$

where A is mean or median or mode

Example:

Calculate mean deviation about mean for the data: 10,20,30,40,50

Solution

A.M. = (10 + 20 + 30 + 40 + 50) /5 = 150/5 = 30

M.D. about A = $\frac{\Sigma\, |x - A\,|}{n} = \frac{\Sigma\, |x - 30\,|}{5}$ = 60 /5 =12

Co-efficient of M.D. = $\frac{M.D.\ about\ mean}{mean}$ = 12/30 = 0.4

x	$|x - A\,| =$ $|x - 30\,|$
10	20
20	10
30	0
40	10
50	20
	60

Example: Daily earnings in Rs. (x) coolies are given. Calculate all three mean deviations and its coefficient for x: 32,51,23,46,20,78,57,56,57,30.

Solution:

x	$|x - 45\,|$	$|x - 48.5\,|$	$|x - 57\,|$
32	13	16.5	25
51	6	2.5	6
23	22	25.5	34
46	1	2.5	11
20	25	28.5	37
78	33	29.5	21
57	12	8.5	0
56	11	7.5	1
57	12	8.5	0
30	15	18.5	27
	150	148	162

A.M. = $\frac{\Sigma x}{n} = \frac{450}{10}$ = 45

Median: Arranging in ascending order: 20,23,30,32,46,51,56,57,57,78

Median = Md = (n+1)/2 th item

= [(10+1)/2 =5.5th item] = (46 + 51) /2 = 48.5

Mode, Z = value which repeats more often = 57

M.D. about mean= $\frac{\Sigma |x-A.MI}{n} = \frac{\Sigma |x-45|}{10}$ = 150/10 =15

Co-efficient of M.D. = $\frac{M.D.\ about\ mean}{mean}$ = 15/45 = 0.33

M.D. about median= $\frac{\Sigma |x-md|}{n} = \frac{\Sigma |x-48.5|}{10}$ =148/10 =14.8

Co-efficient of M.D. = $\frac{M.D.\ about\ md}{md}$ = 14.8/48.5 = 0.31

M.D. about mode= $\frac{\Sigma |x-Z|}{n} = \frac{\Sigma |x-57|}{10}$ =162/10 =16.2

Co-efficient of M.D. = $\frac{M.D.\ about\ mode}{mode}$ = 16.2 /57 = 0.28

Discrete Data

M.D. about A = $\frac{\Sigma[f\ |x-A|]}{\Sigma f}$

Co-efficient of mean deviation = $\frac{M.D.\ about\ A}{A}$

where A is mean or median or mode

Example: Calculate the mean deviation

X	2	4	6	8	10
f	1	4	6	4	1

x	f	xf	|x – a.m. |= |x - 6 |	f | x - 6 |
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
	16	96		24

Mean = $\frac{\Sigma fx}{\Sigma f} = \frac{96}{16}$ = 6 ; M.D. about A = $\frac{\Sigma[f\ |x-A|]}{\Sigma f}$ = 24/16 = 1.5

Co-efficient of mean deviation = $\frac{M.D.\ about\ A}{A}$ = 1.5 /6 = 0.25

Continuous Data

M.D. about A = $\frac{\Sigma[f\ |m-A|]}{\Sigma f}$

Co-efficient of mean deviation = $\frac{M.D.\ about\ A}{A}$

where A is mean or median or mode

Example: Calculate mean deviation about mean

C.I.	0-10	10-20	20-30	30-40	40-50	50-60	60 -70
f	6	5	8	15	7	6	3

Mean = $\frac{\Sigma(fm)}{\Sigma f}$ = 1670/50 =33.4

x	f	m	mf	$\lvert m - A \rvert = \lvert m - 33.4 \rvert$	$f\lvert m - A \rvert$
0-10	6	5	30	28.4	170.4
10-20	5	15	75	18.4	92.0
20-30	8	25	200	8.4	67.2
30-40	15	35	525	1.6	24.0
40-50	7	45	315	11.6	81.2
50-60	6	55	330	21.6	129.6
60-70	3	65	195	31.6	94.8
	50		1670		659.2

M.D. about mean = $\frac{\Sigma[f \mid m-A \mid]}{\Sigma f}$ = 659.2 /50 =13.18

Co-efficient of mean deviation = $\frac{M.D.\ about\ mean}{mean}$ = 13.18/ 33.4 = 0.39

[4] STANDARD DEVIATION (SD)

It is defined as the square root of the arithmetic mean of the squares of deviations of the observations from arithmetic mean and is denoted by **σ.**

Raw data

Example: Calculate SD for the data: x: 32, 51, 23, 46, 20, 78, 57, 56, 57, 30

Solution: A.M. = $\frac{\Sigma x}{n}$ = $\frac{450}{10}$ = 45;

SD = σ = $\sqrt{\frac{\Sigma(x-a.m.)2}{n}}$ =√ (3038/10) = √303.8 = 17.43.

x	$(x-45)$	(x - A.M)= $(x-45)^2$
32	-13	169
51	6	36
23	-22	484
46	1	1
20	-25	625
78	33	1089
57	12	144
56	11	121
57	12	144
30	-15	225
450		3038

Discrete Data

$\sigma = \sqrt{\frac{\Sigma[f(x-A)2]}{\Sigma f}}$

x	2	4	6	8	10
f	1	4	6	4	1

A.M. = $\frac{\Sigma fx}{\Sigma f} = \frac{96}{16} = 6$

x	f	xf	(x – a.m.)=(x – 6)	(x-6)^2	f(x-6)^2
2	1	2	-4	16	16
4	4	16	-2	4	16
6	6	36	0	0	0
8	4	32	2	4	16
10	1	10	4	16	16
	16	96			64

SD = $\sigma = \sqrt{\frac{\Sigma[f(x-A)2]}{\Sigma f}} = \sqrt{[64/16]} = \sqrt{4} = 2$

Continuous data

Example: Calculate standard deviation

C.I.	0-10	10-20	20-30	30-40	40-50	50-60	60 -70
f	6	5	8	15	7	6	3

Solution:

Mean = $\frac{\Sigma(fm)}{\Sigma f}$ = 1670/50 =33.4

x	f	m	mf	(m – A) = (m -33.4)	(m -33.4)^2	f(m -33.4)^2
0-10	6	5	30	-28.4	806.56	4839.36
10-20	5	15	75	-18.4	338.56	1692.8
20-30	8	25	200	-8.4	70.56	564.48
30-40	15	35	525	1.6	2.56	38.4
40-50	7	45	315	11.6	134.56	941.92
50-60	6	55	330	21.6	466.56	2799.36
60-70	3	65	195	31.6	998.56	2995.68
	50		1670			13872.00

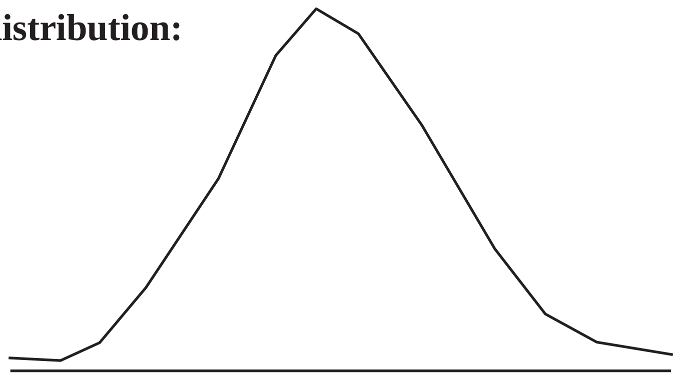
SD: $\sigma = \sqrt{\frac{\Sigma[f(m-A)2]}{\Sigma f}} = \sqrt{[13872.00/50]} = \sqrt{277.44} = 16.66.$

MEASURES OF SKEWNESS

Meaning of Skewness:

- Skewness means **lack of symmetry** .
- We study skewness to have an idea about the shape of the curve which we can draw with the help of the given data.
- If, in a distribution**,** **Mean = Median = Mode**, then that distribution is known as **Symmetrical Distribution**.
- If, in a distribution, **Mean ≠ Median ≠ Mode**, then it is not a symmetrical distribution and it is called a **Skewed Distribution** and such a distribution could either be **positively skewed or negatively skewed.**

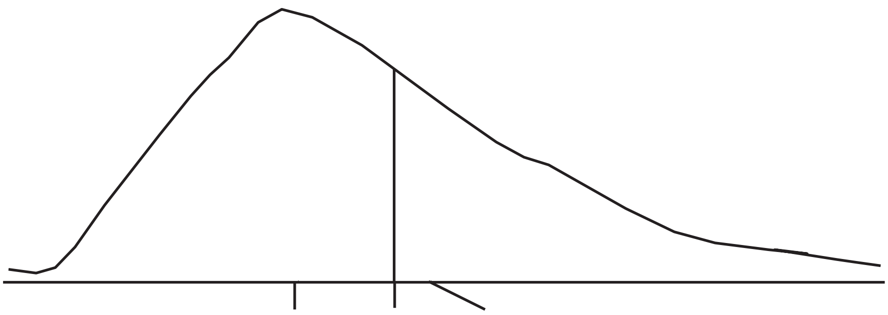
a) Symmetrical distribution:



Mean = Median = Mode

It is clear from the above diagram that in a symmetrical distribution the values of mean,median and mode coincide. The spread of the frequencies is the same on both sides of the centerpoint of the curve.

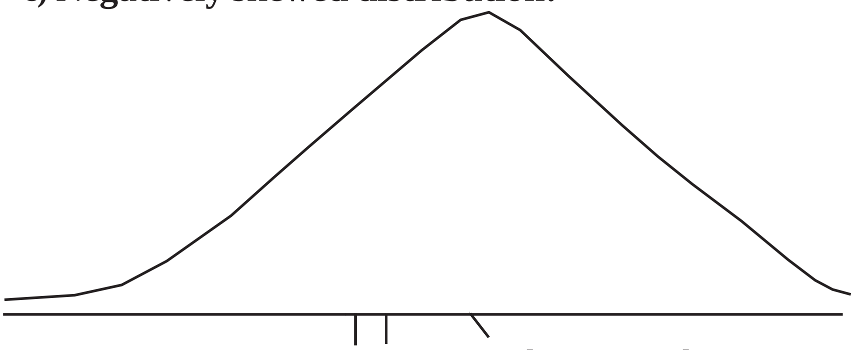
b) Positively skewed distribution:



Mode < Median < Mean

It is clear from the above diagram, in a positively skewed distribution, the value of the mean is maximum and that of the mode is least, the median lies in between the two. In the positively skewed distribution, the frequencies are spread out over a greater range of values on the right-hand side than they are on the left hand side.

c) Negatively skewed distribution:



Mean < Median < Mode

It is clear from the above diagram, in a negatively skewed distribution, the value of the mode is maximum and that of the mean is least. The median lies in between the two. In the negatively skewed distribution the frequencies are spread out over a greater range of values on the left hand side than they are on the right hand side.

Measures of skewness:

The important measures of skewness are

[1] Karl – Pearson's coefficient of skewness
[2] Bowley's coefficient of skewness
[3] Measure of skewness based on moments.

We are interested in studying only the **first two** methods only.

[1] Karl-Pearson's Coefficient of skewness:

According to Karl – Pearson, the absolute measure of skewness = mean – mode. This measure is not suitable for making valid comparison of the skewness in two or more distributions because the unit of measurement may be different in different series. To avoid this difficulty, we use relative measure of skewness, called Karl-Pearson's coefficient of skewness given by:

$$\text{Karl-Pearson's Coefficient of Skewness} = \frac{\text{Mean} - \text{Mode}}{\text{S.D.}}$$

In case of **mode is ill- defined**, the coefficient can be determined by

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

Example:

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

25, 15, 23, 40, 27, 25, 23, 25, 20

Solution:

Size	Deviation from A=25 D	d^2
25	0	0
15	-10	100
23	-2	4
40	15	225
27	2	4
25	0	0
23	-2	4
25	0	0
20	-5	25
N = 9	$\sum d = -2$	$\sum d^2 = 362$

$$\text{Mean} = A + \frac{\sum d}{n}$$

$$= 25 + \frac{-2}{9}$$

$$= 25 - 0.22 = 24.78$$

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

$$= \sqrt{\frac{362}{9} - \left(\frac{-2}{9}\right)^2}$$

$$= \sqrt{40.22 - 0.05}$$

$$= \sqrt{40.17} = 6.3$$

Mode = 25, as this size of item repeats 3 times

Karl-Pearson' s coefficient of skewness is given by

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

$$= \frac{24.78 - 25}{6.3}$$

$$= \frac{-0.22}{6.3}$$

$$= -0.03$$

Example

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

Solution :

Size	Frequency (f)	Deviation From A = 6 (d)	d^2	fd	fd^2
3	7	-3	9	-21	63
4	10	-2	4	-20	40
5	14	-1	1	-14	14
6	35	0	0	0	0
7	102	1	1	102	102
8	136	2	4	272	544
9	43	3	9	129	387
10	8	4	16	32	128
	N = 355			$\Sigma fd = 480$	$\Sigma fd^2 = 1278$

$$\text{Mean} = A + \frac{\Sigma fd}{N}$$

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

$$= 6 + 1.35$$

$$= 7.35$$

Mode = 8

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

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

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

$$= \sqrt{1.78} = 1.33$$

Example:

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

Solution :

Size	Frequency (f)	Deviation from A = 6 (d)	d^2	fd	fd^2
3	7	-3	9	-21	63
4	10	-2	4	-20	40
5	14	-1	1	-14	14
6	35	0	0	0	0
7	102	1	1	102	102
8	136	2	4	272	544
9	43	3	9	129	387
10	8	4	16	32	128
	N = 355			$\sum fd = 480$	$\sum fd^2 = 1278$

$$\text{Mean} = A + \frac{\sum fd}{N} = 6 + \frac{480}{355} = 6 + 1.35 = 7.35$$

$$\text{Mode} = 8$$

$$\sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} = \sqrt{\frac{1278}{355} - \left(\frac{480}{355}\right)^2} = \sqrt{3.6 - 1.82} = \sqrt{1.78} = 1.33$$

$$\text{Coefficient of skewness} = \frac{\text{Mean} - \text{Mode}}{\text{S.D.}} = (7.35 - 8)/1.33 = -0.65/1.33 = -0.5.$$

Example:

Find Karl – Pearson' s coefficient of skewness for the given distribution:

X :	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
F :	2	5	7	13	21	16	8	3

Solution :

Mode lies in 20-25 group which contains the maximum frequency

$$\text{Mode} = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times C$$

$$l = 20,\ f_1 = 21,\ f_0 = 13,\ f_2 = 16,\ C = 5$$

$$\text{Mode} = 20 + \frac{21 - 13}{2 \times 21 - 13 - 16} \times 5$$

$$= 20 + \frac{8 \times 5}{42 - 29}$$

$$= 20 + \frac{40}{13} = 20 + 3.08 = 23.08$$

Computation of Mean and Standard deviation

X	Mid-point M	Frequency f	Deviations $d' = \frac{m - 22.5}{5}$	fd'	d'^2	fd'^2
0-5	2.5	2	-4	-8	16	32
5-10	7.5	5	-3	-15	9	45
10-15	12.5	7	-2	-14	4	28
15-20	17.5	13	-1	-13	1	13
20-25	22.5	21	0	0	0	0
25-30	27.5	16	1	16	1	16
30-35	32.5	8	2	16	4	32
35-40	37.5	3	3	9	9	27
		N = 75		$\Sigma fd' = -9$		$\Sigma fd'^2 = 193$

$$\text{Mean} = A + \frac{\Sigma fd}{N} \times c$$

$$= 22.5 + \frac{-9}{75} \times 5$$

$$= 22.5 - \frac{45}{75}$$

$$= 22.5 - 0.6 = 21.9$$

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

$$= \sqrt{\frac{193}{75} - \left(\frac{-9}{75}\right)^2} \times 5$$

$$= \sqrt{2.57 - 0.0144} \times 5$$

$$= \sqrt{2.5556} \times 5$$

$$= 1.5986 \times 5 = 7.99$$

Karl – Pearson' s coefficient of skewness

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

$$= \frac{21.9 - 23.08}{7.99}$$

$$= \frac{-1.18}{7.99} = -0.1477$$

[2] Bowley' s Coefficient of Skewness:

In Karl- Pearson's method of measuring skewness the whole of the series is needed. Prof. Bowley has suggested a formula based on relative position of quartiles. In a symmetrical distribution, the quartiles are equidistant from the value of the median; ie., Median – Q_1 = Q_3 – Median. But in a skewed distribution, the quartiles will not be equidistant from the median. Hence Bowley has suggested the following formula:

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

Example:

Find the Bowley' s coefficient of skewness for the following series.

2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22

Solution:

The given data in order

2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22

$$Q_1 = \text{size of } \left(\frac{n+1}{4}\right)^{th} \text{ item}$$

$$= \text{size of } \left(\frac{11+1}{4}\right)^{th} \text{ item}$$

$$= \text{size of } 3^{rd} \text{ item} = 6$$

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

$$= \text{size of } 3\left(\frac{11+1}{4}\right)^{th} \text{ item}$$

$$= \text{size of } 9^{th} \text{ item}$$

$$= 18$$

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

$$= \text{size of } \left(\frac{11+1}{2}\right)^{th} \text{ item}$$

$$= \text{size of } 6^{th} \text{ item}$$

$$= 12$$

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

$$= \frac{18 + 6 - 2 \times 12}{18 - 6} = 0$$

Since sk = 0, the given series is a symmetrical data.

Example:

Find Bowley's coefficient of skewness of the following series.

Size :	4	4.5	5	5.5	6	6.5	7	7.5	8
F :	10	18	22	25	40	15	10	8	7

Solution :

Size	f	c.f
4	10	10
4.5	18	28
5	22	50
5.5	25	75
6	40	115
6.5	15	130
7	10	140
7.5	8	148
8	7	155

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

$$= \text{Size of } \left(\frac{155+1}{4}\right)^{th} \text{ item}$$

$$= \text{Size of } 39^{th} \text{ item}$$

$$= 5$$

$$Q_2 = \text{Median} = \text{Size of } \left(\frac{N+1}{2}\right)^{th} \text{ item}$$

$$= \text{Size of } \left(\frac{155+1}{2}\right)^{th} \text{ item}$$

$$= \text{Size of } 78^{th} \text{ item}$$

$$= 6$$

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

$$= \text{Size of } 3\left(\frac{155+1}{4}\right)^{th} \text{ item}$$

$$= \text{Size of } 117^{th} \text{ item}$$

$$= 6.5$$

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

$$= \frac{6.5 + 5 - 2\times 6}{6.5 - 5}$$

$$= \frac{11.5 - 12}{1.5} = \frac{0.5}{1.5}$$

$$= -0.33$$

@@@ ***End of UNIT-IV*** @@@