Chapter - IV
MEASURES OF DISPERSION

Meaning and Definition

In a hypothetical one day series between India and another country, the runs scored by 2 batsmen of India, V.S and R.D, are stated below:

V.S	5	146	12	48	29	48	48
R.D	48	43	48	51	50	60	36

A keen student of statistics calculates the measures of Central tendency for the above scores and is amused to note that the Arithmetic Mean, Median and Mode scores for both the batsmen works out to 48. Does that mean that there is nothing to choose from between the two batsman and they are more or less the same?

Any effort to analyze the scores would reveal that the batsman V.S. has scored as low as 5 but has also scored a high score of 146. Batsman R.D has always scored in excess of 36 but has never been able to cross 60. Hence, there appears to be a lot of variation in the scores of V.S whereas scores of R.D appear to be consistent. If the captain needs a player who is consistent, his choice would be R.D whereas if he is willing to take the risk but is aiming for a high scoring batsman, his choice would be V.S. Obviously, this choice is not made by looking at the averages of the two batsmen.

Simpson and Kafka have highlighted this issue with respect to data stating that an average does not reveal the full story. An average may not provide us insight into the characteristics of any data unless we understand the manner in which the individual data items scatter around the average. This phenomenon is called dispersion. It measures the extent to which the items vary from some central value.

According to Bowley, "Dispersion is the measure of the variation of items".

According to Spiegel, "The degree to which numerical data tend to spread about an average value is called the variation or dispersion of the data".

Thus, a measure of dispersion describes the extent by which observations vary or scatter from a central value and is measured as an average of the deviations of the observations from the central value. Measures of dispersion show the average of the differences of various items from an average. For this reason they are also called averages of the second order.

Objectives of Dispersion: The main objectives of dispersion are

1. To gain a better understanding of a given data series by understanding the average variation and the dispersion of values on either side of the measure of central tendency
2. To establish whether a measure of central tendency of a data series is truly representative of the series. Small measures of dispersion imply true representation while high measures of dispersion indicate that the central tendency measure may not be representative of the data series.
3. To understand the range of values that the given data series can take.
4. A practical application of the measures of dispersion is to measure and control undesirable variability. This is particularly true with respect to quality control wherein variation in excess of desired levels lead to rejection of the product or unit.
5. To enable comparison of different data series, compare the disparity between them and estimate the degree of variation.

Properties of a good measure of dispersion: A good measure of dispersion must have the following characteristics or properties:

1. It must be easy to calculate.
2. It should be simple enough to understand the message the measure is trying to convey. Even a layman should be able to understand what it is trying to convey.
3. It must be rigidly defined. While every value in the series might have an impact on the measure, it must not be capable of completely changing the measure such that the conclusions are totally different. In other words, it should be based on all values in a given series but it must not be unduly affected by extreme values.
4. It should be amenable to further algebraic or statistical treatment.
5. It should not be significantly impacted by size or composition of the sample.

Measures of dispersion: There are different measures of dispersion. These measures can be classified into Absolute measures and relative measures. An 'Absolute' measure is one that is expressed in terms of the same unit in which the variable (or given data) is measured. A 'Relative" measure of dispersion is expressed as a pure number (without any units) which enables comparison of the levels of dispersion from a central tendency across different series (stated in different units). These measures are also called as "Co-efficient(s) of Dispersion". The important methods of studying variation are listed as under:

Absolute measures

(1) The Range
(2) Inter Quartile Range and Quartile Deviation.
(3) The Mean Deviation or Average Deviation.
(4) The Standard Deviation and Variance
(5) The Lorenz Curve

Relative measures:

(1) Coefficient of Range
(2) Coefficient of Quartile Deviation.
(3) Coefficient of Mean Deviation.
(4) Coefficient of Variation

Let us now study each of the measures.

Range

Range is the difference between the values of the extreme items of a series. It is the difference between the values of the largest item and the value of the smallest item in the distribution.

Range = L–S
L = Largest item
S= Smallest item

$$\text{Co-efficient of Range} = \frac{L-S}{L+S}$$

Students may note that n calculating the Range of a frequency distribution, frequencies are completely ignored. In calculating the range of a continuous series, the lower limit of the lower class is subtracted from the upper limit of the highest class. Range cannot be calculated for open ended distributions.

Merits of Range

1. It is the simplest to calculate of all measures.
2. It is very easy to understand.
3. It gives a high level or broad picture of the data at one glance.

4. Range is used for a variety of applications. Range is used to study variations in prices of stocks, commodities, Gold, etc. Meteorological department uses this measure to forecast the temperature and other weather details. It is also extensively used for Quality control purposes.

Demerits of Range

1. It is based only on the smallest and largest values of a series and hence is unduly impacted by extreme values. In other words, it is not based on all the terms and hence, not rigidly defined.
2. Range is impacted by sample size and composition of the sample.
3. Range cannot be calculated for open ended distributions.
4. It is not amenable to algebraic or statistical treatment.

Illustration 1: (Individual Observations). The following are the wages of 10 workers of a factory. Find the range of variation and also compute the co-efficient of range.

120, 170, 240, 100, 105, 205, 300, 160, 150, 180

Solution: Range = L–S

L = Largest Value = 300

S = Smallest Value = 100

Range = 300–100 = 200

$$\text{Co-efficient or Range} = \frac{L-S}{L+S} = \frac{300-100}{300+100}$$

$$= \frac{200}{400} = 0.5$$

Illustration 2: (Continuous series) The following are the marks obtained by students of a class. Find the range of variation of marks and also compute co-efficient of range.

Marks	0–10	10–20	20–30	30–40	40–50	50–60	60–70
No. of students	5	8	12	20	15	7	3

Solution:

Range = L–S = 70–0= 70

$$\text{Co-efficient of Range} = \frac{L-S}{L+S} = \frac{70-0}{70+0} = 1$$

QUARTILE DEVIATION

Quartile Deviation shows the average amount by which the two quartiles differ from Median.

Inter quartile Deviation = $Q_3 - Q_1$

$$\text{Semi-Inter quartile Range or Quartile Deviation or Q.D.} = \frac{Q3-Q1}{2}$$

$$\text{Co-efficient of Q.D.} = \frac{Q3-Q_1}{Q3+Q_1}$$

Merits of Quartile Deviation

1. It is the simple to calculate and very easy to understand.
2. It is not impacted by extreme values.
3. It can be computed for open ended distributions and for data containing unequal classes

Demerits of Quartile Deviation

1. It is impacted by sample size and composition of the sample.
2. It is not amenable to algebraic or statistical treatment.

3. It does not tell us anything about the spread of the various data items across the measure of central tendency.

Illustration 3: (Individual Observations) Find out the value of quartile deviation and its co-efficient from the following data:

Roll No.	1	2	3	4	5	6	7
Marks	25	33	45	17	35	20	55

Solution: Calculation of Quartile Deviation

Marks arranged in ascending order 17 20 25 33 35 45 55

Lower Quartile (Q_1) = Size of $\frac{N+1}{4}$ th item

N = No. of observations.

= Size of $\frac{7+1}{4}$ th item = 2^{nd} item = 20

Upper Quartile (Q3) = Size of $\frac{3(N+1)}{4}$ th item.

= Size of $\frac{3x8}{4}$ = 6^{th} item. = 45

Quartile Deviation (Q.D.) = $\frac{Q_3-Q_1}{2} = \frac{45-20}{2} = 12.5$

Co-efficient of Q.D. = $\frac{Q_3-Q_1}{Q_3+Q_1} = \frac{45-20}{45+2} = \frac{25}{65} = 0.38$

Illustration 4: (Discrete Series) Compute Quartile Deviation and its co-efficient from the following data.

Height in inches	58	59	60	61	62	63	64	65	66
No. of Students	15	20	32	35	33	22	20	10	8

(B.Com Agra)

Solution: Calculation of Quartile Deviation

Height	***f***	**C. *f***
58	15	15
59	20	35
60	32	67
61	35	102
62	33	135
63	22	157
64	20	177
65	10	187
66	8	195

Lower Quartile (Q_1) = Size of $\left(\frac{N+1}{4}\right)^{th}$ item = $\left(\frac{195+1}{4}\right)^{th}$ item= 49^{th} item.

Size of 49^{th} item = 60, Hence $Q_1 = 60$

Upper Quartile (Q_3) = Size of $\frac{3(N+1)th}{4}$ item = $\frac{3(195+1)}{4}$ = 147^{th} item.

Size of 147^{th} item = 63, Hence Q3 = 63

Quartile Deviation (Q.D.) = $\frac{Q_3-Q_1}{2} = \frac{63-60}{2} = 1.5$

Co-efficient of Q.D. = $\frac{Q3-Q1}{Q3+Q1} = \frac{63-60}{63+60} = 0.024$

Illustration 5: (Continuous Series) Calculate Quartile Deviation and its coefficient of the marks of 59 students in statistics as given below.

Marks	No. of Students	Marks	No. of Students
0–10	4	40–50	11
10–20	8	50–60	7
20–30	10	60–70	3
30–40	16		

Solution:

Calculation of Quartile Deviation

Marks	f	Cf
0–10	4	4
10–20	8	12
20–30	10	22
30–40	16	38
40–50	11	49
50–60	7	56
60–70	3	59

Lower Quartile = Size of $\frac{N}{4}$ th item $= \frac{59}{4} = 14.75^{th}$ item

Q_1 group = 20–30

$Q_1 = L_1 + \frac{\frac{N}{3} - C.f}{f} \times I = 20 + \frac{14.75-12}{10} \times 10 = 20 + \frac{2.75}{10} \times 10 = 22.75$

Upper Quartile (Q_3) = Size of $\frac{3N}{4}$ th item = Size of $\frac{3 \times 59}{4}$ = 44.25 th item.

Q_3 group = 40-50

$Q_3 = L_1 + \frac{\frac{3N}{4} - C.f}{f} \times i$

$L_1 = 40, \frac{3N}{4} = 44.25, C.f = 38, f = 11, i = 10$

$= 40 + \frac{44.25-38}{1110} \times 10 = 40+5.68 = 45.68$

Quartile Deviation = $\frac{Q3-Q1}{2} = \frac{45.68-22.75}{2} = \frac{22.93}{2} = 11.465$

Quartile Deviation = $\frac{Q3-Q1}{Q3+Q1} = \frac{22.93}{68.43} = 0.335$

MEAN DEVIATION

"Mean Deviation of a series is the arithmetic average of the deviations of various items from a measure of central tendency (either mean, median or mode)" Theoretically, deviations can be taken from any of the three averages mentioned above, but in actual practice it is calculated either from mean or from Median. While Calculating deviations algebraic signs are not taken into account.

Merits of Mean Deviation

1. It is rigidly defined

2. It is not least impacted by sampling fluctuations
3. It takes into account every single value in the series
4. Mean deviation from Median is least impacted due to extreme values
5. It is extensively used in multiple fields such as Economics, Commerce, etc as it is the best measure for comparison of two or more series.

Demerits of Mean Deviation

1. It is relatively difficult to compute
2. It is not amenable to further algebraic or statistical treatment.
3. It is difficult for a layman to understand as to why or when a particular average should be considered for calculation of Mean Deviation. The Mean Deviations obtained by taking the Mean, Median and Mode as average differ widely
4. It is not effective for open ended series, particularly when the average is Arithmetic Mean.

Individual Observations

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

Steps: (1) Compute Median or Mean of the series.

(2) Compute the deviations of items from the average, ignoring ± signs, denote them as |D|. Denote the total of these deviations as $\sum|D|$.

(3) Divide the total by no of items.

$$\text{Co-efficient of Mean Deviation} = \frac{\text{Mean Deviation}}{\text{Mean or Median or Mode}}$$

Illustration 6: Calculate the mean deviation and its coefficient from (i) arithmetic mean and (ii) median in respect of marks obtained by nine students given below.

Marks (out of 25) 7, 4, 10, 9, 15, 7, 9, 7 *(B.Com. Bombay)*

Solution: For calculating median the item have to be rearranged in the ascending order.

4, 7, 7, 7, 9, 9, 10, 12, 15

$$\text{Median} = \text{Size of } \left(\frac{N+1}{2}\right)^{th} \text{ item} = \text{Size of } \left(\frac{9+1}{2}\right)^{th} \text{ item} = 5\text{th item}$$

Size of 5th item = 9

$$\text{Mean} = \frac{4+7+7+7+9+9+10+12+15}{9} = \frac{80}{9} = 8.9$$

Marks	Deviations from Mean (8.9)	Deviations from Median (9)
4	4.9	5
7.	1.9	2
7.	1.9	2
7	1.9	2
9	0.1	0
9	0.1	0
10	1.1	1
12	3.1	3
15	6.1	6
	$\sum|D|=21.1$	$\sum|D|=21$

Mean Deviation from Mean = $\frac{\sum|D|}{N} = \frac{21.1}{9} = 2.34$

Mean Deviation from Median = $\frac{\sum|D|}{N} = \frac{21}{9} = 2.33$

Co-efficient of Mean Deviation (through Mean) = $\frac{\text{Mean Deviation}}{\text{Mean}}$

$= \frac{2.34}{8.9} = 0.263$

Co-efficient of Mean Deviation (through Median) = $\frac{\text{Mean Deviation}}{\text{Median}}$

$= \frac{2.33}{9} = 0.259$

Discrete Series: Mean Deviation = $\frac{\sum f|D|}{N}$

$\sum f|D|$ = Total of Frequency multiplied with deviations from the average
N = Total of Frequency

Steps: (1) Compute Mean or Median (as required)
(2) Calculate deviation of the size from the average.
(3) Multiply the deviations with respective frequencies and obtain the total and denote it as $\sum f|D|$. Apply the formula.

Illustration 7: Calculate (a) Mean co-efficient of dispersion and (b) Median co-efficient of dispersion from the following data.

Marks	10	15	20	30	40	50
Frequency	8	12	15	10	3	2

(B.Com. Agra)

Solution: Calculation of Mean co-efficient of dispersion

Marks	***f***	***fX***	**Deviation from Mean 21.6** $\lvert D\rvert$	$f\lvert D\rvert$
10	8	80	11.6	92.8
15	12	180	6.6	79.2
20	15	300	1.6	24.0
30	10	300	8.4	84.0
40	3	120	18.4	55.2
50	2	100	28.4	56.8
	N=50	$\sum fX = 1080$		$\sum f\lvert D\rvert = 392.0$

$\overline{X} = \frac{\sum fX}{N}$, $\sum fX = 1080$, N= 50 $= \frac{1080}{50} = 21.6$

Mean Deviation = $\frac{\sum f|D|}{N} = \frac{392}{50} = 7.84$

Mean Co-efficient of Dispersion = $\frac{\text{Mean Deviation}}{\text{Mean}} = \frac{7.84}{21.6} = 0.363$

Calculation of median coefficient of dispersion

Marks	*f*	C*f*	Deviation from Median 20 |D|	*f*|D|
10	8	8	10	80
15	12	20	5	60
20	15	35	0	0
30	10	45	10	100
40	3	48	20	60
50	2	50	30	60
	N=50			$\Sigma f\|D\|=360$

Median =Size $\left(\frac{N+1}{2}\right)^{th}$ item = Size of $\frac{51}{2}$ th = 25.5th item.

Median = 20

Mean Deviation = $\frac{\Sigma f|D|}{N} = \frac{360}{50} = 7.2$

Median Co-efficient of dispersion

$= \frac{\text{Mean Deviation}}{\text{Median}} = \frac{7.2}{20} = 0.36$

Continuous Series:

For calculating mean deviation in continuous series steps are same as in case of discrete series. The only difference is that mid-points of the various classes is obtained and deviations of these mid-points are taken from mean or median. The formula is same, i.e.

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

Illustration 8: Calculate Mean Coefficient of Dispersion.

Marks	No. of Students	Marks	No. of Students
0–10	5	40–50	28
10–20	8	50–60	20
20–30	7	60–70	10
30–40	12	70–80	10

Solution:

Arithmetic Mean ($\overline{X}$) = $A + \frac{\Sigma fd'}{N} \times C$

A= 45, $\Sigma fd'$ =0 N = 100, C=10

$= 45 + \frac{0}{100} \times 10 = 45$

Mean Deviation = $\frac{\Sigma f|D|}{N}$; $\Sigma f|D| = 1400$, N = 100

$= \frac{1400}{100} = 14$

Mean Co-efficient of dispersion = $\frac{\text{Mean Deviation}}{\text{Mean}} = \frac{14}{45}$ 0.31

Calculation of Mean Deviation

Marks	m.v.	freq.	Deviation from Mean (A=45) D	Step Deviation D/10 D'	fd'	Deviations from Mean D (±ignored)	f|D|
0-10	5	5	-40	-4	-20	40	200
10-20	15	8	-30	-3	-24	30	240
20-30	25	7	-20	-2	-14	20	140
30-40	35	12	-10	-1	-12	10	120
40-50	45	28	0	0	0	0	0
50-60	55	20	+10	+1	+20	10	200
60-70	65	10	+20	+2	+20	20	200
70-80	75	10	+30	+3	+30	30	300
			N=100		Σfd'=0		f|D|=1400

STANDARD DEVIATION

"Standard Deviation is the square root of the arithmetic average of the squares of the deviations measured from mean". Standard Deviation is denoted by the small Greek letter σ (read as sigma)

Calculation of Standard Deviation

Individual Observation

(1) By taking deviation of the items from the actual mean.

(2) By taking deviation of the items from an assumed mean.

(1) Deviation from actual mean

$$\sigma = \sqrt{\frac{\sum x^2}{N}}$$

Where $\sum x^2$ = Sum of the squares of the deviations taken from mean.

N = No. of items

Steps: (1) Calculate arithmetic mean of the series.

(2) Take the deviations of the items from the mean denote it as x.

(3) Square the deviations and obtain the total, denote is as $\sum x^2$

(4) Divide the total by total number of items and find the square root

Deviations from assumed mean:

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

Where $\sum d^2$ = Total of the squares of the deviations from assumed mean.

$\sum d$ = Total of the deviations from assumed mean.

N = Total number of items.

Steps: (1) Take the deviations from assumed mean, denote the deviation as d, obtain the total, and denote the total as $\sum d$.

(2) Square these deviations and obtain the total. Denote the total as $\sum d^2$ and apply the formula.

Note: In actual practice, the second formula is used widely, because the first formula becomes cumbersome if the actual mean is in fractions.

Merits of Standard Deviation

1. It is rigidly defined
2. It takes into account every single value in the series
3. It is amenable to further algebraic or statistical treatment.
4. It is extensively used in various other statistical calculations such as correlation, regression, sampling, etc

Demerits of Standard Deviation

1. It is relatively difficult to compute
2. It is calculated with only Arithmetic Mean as the average. Standard deviation from other averages such as Median is not an effective measure of dispersion.

Illustration 9: Compute Standard Deviation from the following data of income of 10 employees of a firm by (i) taking deviations from actual mean (ii) taking deviations from assumed mean 140.

Income (Rs.) 120, 100, 160, 100, 220, 130, 150, 170, 150, 200

Solution: (i) Computation of Standard Deviation (taking the deviations from mean).

Income X	x $(X-\overline{X})$	x^2
120	-30	900
100	-50	2500
160	+10	100
100	-50	2500
220	+70	4900
130	-20	400
150	0	0
170	+20	400
150	0	0
200	+50	2500
$\sum X = 1500$	$\sum X = 0$	$\sum X^2 = 14200$

$$\overline{X} = \frac{\sum X}{N} = \frac{1500}{10} = 150$$

$$\sigma = \sqrt{\frac{\sum x^2}{N}} \quad \text{where } \sum x^2 = 14200,\ N = 10$$

$$= \sqrt{\frac{14200}{10}} = \sqrt{1420} = 37.68$$

Solution: Computation of Standard Deviation (taking deviations from 140)

Income X	(X–140) d	d^2
120	-20	400
100	-40	1600
160	+20	400
100	-40	1600

220	+80	6400
130	-10	100
150	+10	100
170	+30	900
150	+10	100
200	+60	3600
N=10	**Σd+100**	**Σd²=15200**

$$\sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$$ where $\Sigma d^2 = 15200$, $\Sigma d = 100$, $N=10$

$$= \sqrt{\frac{15200}{10} - \left(\frac{100}{10}\right)^2} = \sqrt{1520-100} = \sqrt{1420} = 37.68$$

Discrete Series:

(1) Actual Mean Method

(2) Assumed Mean Method

(3) Step Deviation Method.

1) Actual Mean Method:

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

Steps: (1) Calculate the arithmetic mean.

(2) Find out the deviations of the various values from mean value, denote it as x.

(3) Square these deviations and multiply with respective frequencies and obtain the total, denote it as Σfx^2 and apply the formula.

2) Assumed Mean Method:

$$\sigma = \sqrt{\frac{\Sigma fd2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$$

Steps: (1) Take the deviations of the items from an assumed mean and denote these deviations as d

(2) Multiply these deviations by the respective frequencies and obtain the total, denote the total as Σfd.

(3) Square the deviations (d^2) and multiply them with respective frequencies denote the total as Σfd^2 and apply the formula.

3) Step Deviation Method:

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

$\Sigma fd'2$ = Total of frequency multiplied with square of the deviations from assumed mean.

$\Sigma fd'$ = Total of frequency multiplied with deviations from assumed mean.

N = Total of frequency

C = Common factor

Note: The first method (taking deviations from actual mean) is rarely followed because if the actual mean is in fractions it involves lot calculations. The third method (step deviation method) is usually followed when it is possible to take common factor.

Illustration 10: Find the standard deviation for the following distribution.

x	5	15	25	35	45	55	65
f	3	10	20	30	15	12	10

Calculation of Standard Deviation

x	f	d A=35 (X-A)	d/10 d'	fd'	d'^2	fd^2
5	3	-30	-3	-9	9	27
15	10	-20	-2	-20	4	40
25	20	-10	-1	-20	1	20
35	30	0	0	0	0	0
45	15	+10	+1	+15	1	15
55	12	+20	+2	+24	4	48
65	10	+30	+3	+30	9	90
	N=100			$\sum fd'=+20$		$\sum fd'^2=240$

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

$\sum fd'2 = 240, \sum fd' = 20, N = 100, C = 10$

$$\sigma = \sqrt{\frac{240}{100} - \left(\frac{20}{100}\right)^2} \times 10 = \sqrt{2.4 - 0.04} \times 10 = \sqrt{2.36} \times 10 = 15.36$$

Continuous Series: Step Deviation Method

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

$\sum fd'^2$ = Total of frequency multiplied with square of the deviations from assumed mean (after taking common factor)

$\sum fd'$ = Total of frequency multiplied with deviations of the size from assumed mean (after taking common factor)

N = Total of frequency

C = Common factor

Steps: (1) Find the mid-points of the various classes.

(2) Take deviation of these mid-points from an assumed mean and denote it as d.

(3) Take common factor of the deviations and denote it as d'.

(4) Multiply frequency with d' obtain the total and denote it as $\sum fd'$

(5) Square the deviations (d') and denote it as d'^2

(6) Multiply the frequencies with d'^2 obtain the total and denote it as $\sum fd^2$ and apply the formula.

Illustration 11: Find the arithmetic mean and standard deviation of the following data.

Age under	No. of Persons dying
10	15
20	30
30	53
40	75
50	100
60	115
70	125
80	

(B.Com. Nagarjuna)

Solution:

Calculation of Mean and Standard Deviation.

Age	f	mid points	m–A d	d/10 d'	fd'	d'2	fd'2
0-10	15	5	-30	-3	-45	9	135
10-20	15	15	-20	-2	-30	4	60
20-30	23	25	-10	-1	-23	1	23
30-40	22	35	0	0	0	0	0
40-50	25	45	+10	+1	+25	1	25
50-60	10	55	+20	+2	+20	4	40
60-70	5	65	+30	+3	+15	9	45
70-80	10	75	+40	+4	+40	16	160
	N=125				Σfd'+2		Σfd'2=488

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times C$$

$$A = 35, \Sigma fd' = +2, N = 125, C = 10 = 35 + \frac{2}{125} \times 10 = 35 + 0.16 = 35.16$$

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

$\Sigma fd'^2 = 488, \Sigma fd' = +2, N = 125, C = 10$

$$\sigma = \sqrt{\frac{488}{125} - \left(\frac{2}{125}\right)^2} \times 10 = \sqrt{3.904 - 0.0003} \times 10$$

$$= \sqrt{3.9037} \times 10 = 1.976 \times 10 = 19.76$$

Variance:

Variance is the square of Standard Deviation. Variance helps in isolating the impact of different factors. Lesser the variance, lower is the variability or dispersion of the series. Variance is calculated exactly as standard deviation, without the square root.

Thus, Variance = σ^2 ; $\sigma = \sqrt{\text{Variance}}$

Co-efficient of Variation:

The standard deviation is an absolute measure of dispersion. The relative measure is known as the co-efficient of variation. This concept was developed by Karl Pearson. It is used in such problems where variability of two or more series is to be compared. The series for which the co-efficient of variation is greater is said to be more variable or less consistent. The series for which co-efficient of variation is less, is said to be less variable or more consistent. Co-efficient of variation is denoted by C.V. and is obtained as follows.

$$C.V. = \frac{\sigma}{\overline{X}} \times 100$$

Illustration 12: Two Cricketers scored the following runs in the several innings. Find who is better run-getter and who is more consistent player?

A	42	17	83	59	72	76	64	45	40	32
B	28	70	31	0	59	108	82	14	3	95

(B.Com. Madras)

Solution: In order to find out who is better run-getter we have to compare the average runs scored and to find who is more consistent, we have to compare the co-efficient of variation.

Calculation of Mean and Co-efficient of Variation.

Cricketer A			**Cricketer B**		
X	**$X-\overline{X}$ ($\overline{X}$ =53)** x	x^2	**X**	**$X-\overline{X}$ ($\overline{X}$ =49)** x	x^2
42	-11	121	28	-21	441
17	-36	1296	78	+21	441
83	+30	900	31	-18	324
59	+6	36	0	-49	2401
72	+19	361	59	+10	100
76	+23	529	108	+59	3481
64	+11	121	82	+33	1089
45	-8	64	14	-35	1225
40	-13	169	3	-46	2116
32	-21	441	95	+46	2116
ΣX=530		**Σx^2=4038**	**ΣX=490**		**ΣX^2=13734**

Cricketer A = $\overline{X} = \frac{\Sigma X}{N}$ ΣX=530, N=10

$= \frac{530}{10} = 53$ runs

$\sigma = \sqrt{\frac{\Sigma x^2}{N}}$ $\Sigma x^2 = 4038$, N=10

$= \sqrt{\frac{4038}{10}} = \sqrt{403.8} = 20.09$

Co-efficient of variation $= \frac{\sigma}{\overline{X}} \times 100 = \frac{20.09}{53} \times 100 = 37.91$

Cricketer B $\overline{X} = \frac{\Sigma X}{N} = \Sigma x = 490$, N = 10

$= \frac{490}{10} = 49$ runs

$$\sigma = \sqrt{\frac{\Sigma fx^2}{N}}, \Sigma x^2 = 13734, N = 10$$

$$= \sqrt{\frac{13734}{10}} = \sqrt{1373.4} = 37.06$$

$$\text{Co-efficient of variation} = \frac{\sigma}{\overline{X}} \times 100, \frac{37.06}{49} \times 100 = .75.63$$

Since average is more in case of A, he is a better rein getter. The C0-efficient of variation is less in case of A, he is more consistent player.

Combined Standard Deviation:

$$\sigma 12 = \sqrt{\frac{N1\sigma_1^2 + N2\,\sigma_2^2 + N_1 d_1^2 + N_2 d_2^2}{N_1 + N_2}}$$

Where σ_{12} = Combined standard Deviation, σ_1 = Standard Deviation of first group, σ_2 = Standard deviation of second group,

$$d_1 = (\overline{X}\,1 - \overline{X}\,12)\ d2 = (\overline{X}\,2 - \overline{X}\,12)$$

Illustration 13. You are given

	A	B
Number of items	100	150
Arithmetic Mean	50	40
Standard Deviation	5	6

Find combined mean and combined Standard Deviation. **(I.C.W.A. Inter)**

Solution:

Combined Mean:

$$\overline{X}\,12 = \frac{N1\overline{X}1 + N2\,\overline{X}2}{N1 + N2}$$

$$N\,1 = 100, N2 = 150, \overline{X}\,1 = 50, \overline{X}\,2 = 40$$

$$= \frac{(100 \times 50) + (150 \times 40)}{250} = \frac{5000+6000}{250} = \frac{11000}{250} = 44$$

Combined Standard Deviation:

$$\sigma 12 \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1 d_1^2 + N_2 d_2^2}{N_1 + N_2}}$$

$$d_1 = \overline{X}_1 - \overline{X}_{12} = 50\text{-}44 = 6$$

$$d2 = \overline{X}_2 - \overline{X}_{12} = 40\text{–}44 = -4$$

$$= \sqrt{\frac{100x(5)^2 + 150 \times (6)^2 + 100 \times (6)^2 + 150x\,(-4)^2}{100+150}}$$

$$= \sqrt{\frac{(100 \times 25) + (150 \times 36) + (100 \times 36) + (150 \times 16)}{250}}$$

$$= \sqrt{\frac{2500+5400+3600+2400}{250}} = \sqrt{\frac{13900}{250}} = \sqrt{55.6} = 7.46$$

Hence combined mean = 44, combined standard Deviation = 7.46

SHORT - ANSWER QUESTIONS

Illustration 1: Calculate coefficient of Range for the following:

Marks	10	20	30	40	50	60	70	80

(B.Com. Osmania)

Solution:

Range = L–S L=80, S=0, = 80-0=80

$$\text{Co-efficient of Range} = \frac{L-S}{L+S} = \frac{80-0}{80+0}\ 1$$

Illustration 3: Calculate Inter-Quartile Deviation, Semi-inter quartile range and co-efficient of Quartile Deviation from the following:

Lower Quartile = 24, Upper Quartile = 60

Solution: Inter-Quartile Range = $Q_3 - Q_1$

$Q_3=60$, $Q_1=24$, 60-24 = 36

$$\text{Semi-inter quartile Range} = \frac{Q3-Q1}{2}$$

$$\text{or Quartile Deviation} = \frac{60-24}{2} = \frac{36}{2} = 18$$

$$\text{Coefficient of Quartile Deviation} = \frac{Q3-Q1}{Q3+Q1} = \frac{60-24}{60+24} = \frac{36}{84} = 0.429$$

Illustration 4: Given Mean deviation from Mean 2.67 from Median 2.69,

Mean 8.9, Median 9

Calculate (i) mean co-efficient of dispersion and (ii) Median co-efficient of Dispersion.

Solution: (1) Mean co-efficient of dispersion $= \frac{\text{Mean Devition}}{\text{Mean}} = \frac{-2.67}{8.9} = 0.3$

(ii) Median co-efficient of dispersion $\frac{\text{Mean Deviation}}{\text{Median}} = \frac{2.69}{9} = 0.299$

Illustration 5: Given

Standard Deviation = 148.2, Mean = 919.167

Calculate co-efficient of Variation

Solution: $C.V. = \frac{\sigma}{\overline{X}} \times 100, \frac{148.2}{919.167} \times 100 = 16.12$

Illustration 6: The arithmetic mean of runs scored by three batsmen, Vijay, Subhash and Kumar in the same series of 10 innings are 50, 48 and 12 respectively. The standard deviation of their runs is respectively, 15, 12, and 2. Who is the most consistent of the three? If one of the three is to be selected who will be selected?

(B.Com. Bombay)

Solution:

Co-efficient of variation of runs scored by Vijay.

$$= \frac{\sigma}{\overline{X}} \times 100, \text{ Where } \sigma = 15, \overline{X} = 50 = \frac{15}{50} \times 100 = 30$$

Co-efficient of variation of runs scored by Subhash.

$$= \frac{\sigma}{\overline{X}} \times 100, \text{ Where } \sigma = 12, \overline{X} = 48 = \frac{12}{48} \times 100 = 25$$

Co-efficient of variation of runs scored by Kumar

$$= \frac{\sigma}{\overline{X}} \times 100, \text{ Where } \sigma = 2, \overline{X} = 12 = \frac{2}{12} \times 100 = 16.67$$

Kumar is most consistent as the co-efficient of variation off runs scored by him is least. If we want to select a player who is expected to score highest, then Vijay should be selected. However, if most consistent batsman is to be selected than Kumar should be selected.

Illustration 7: For a distribution, the co-efficient of variation is 22.5 and the value of arithmetic average is 7.5. Find out the value of standard deviation.

(B.Com. Delhi)

Solution:

$$C.V. = \frac{\sigma}{\overline{X}} \times 100$$

$$\text{i.e. } \sigma = \frac{C.V. \times \overline{X}}{100} \quad C.V. = 22.5 \; \overline{X} = 7.5$$

$$\sigma = \frac{22.5 \times 7.5}{100} = 1.6875$$

Standard Deviation = 1. 6875

Illustration 8: Co–efficient of variation of two series is 75 and 90 and their standard deviations 15 and 18 respectively. Find their means **(B.A. Delhi)**

Solution: $C.V. = \frac{\sigma}{\overline{X}} \times 100$; Thus, $\overline{X} = \frac{\sigma \times 100}{C.V.}$

Mean of First series = $\frac{15 \times 100}{75}$ = 20; Mean of Second series= $\frac{18 \times 100}{90}$ = 20

Illustration 9: Two workers on the same job show the following results over a long period of time:

	Worker A	Worker B
Meantime of completing the job (minutes)	30	25
Standard Deviation (minutes)	6	4

(i) Which worker appears to be more consistent in the time he required to complete the job?

(ii) Which worker appears to be faster in completing the job

(B.Com. Osmania.)

Solution: For ascertaining the consistency of the time taken by the workers, we have to compare co-efficient of variation.

$$C.V. = \frac{\sigma}{\overline{X}} \times 100$$

$$\text{C.V. of Worker A: } \sigma = 6, \overline{X} = 30 = \frac{6}{30} \times 100 = 20$$

$$\text{C.V. of Worker B: } \sigma = 4, \overline{X} = 25, = \frac{4}{25} \times 100 = 16$$

Since the C.V. is less in case of B, he appears to be more consistent. Since the average time taken to complete the work is less in case of B, he appears to be faster in completing the job.

Illustration 10: From the following information calculate variance.

Standard Deviation = 9, Arithmetic mean = 25,

Solution: Variance = σ^2, $\sigma = 9$, Hence Variance = $9^2 = 81$

EXERCISES

PROBLEMS

1. Compute range and co – efficient of range :

Marks (x):	20	23	25	28	32	48	49	62

(SVU – Spet. – 2010)

2. Find out the range and co-efficient of range from the following :

X:	22	24	30	32	35	37	40	42

(Satavahana University – March – 2014)

3. Calculate the range and the coefficient of range of A's monthly earnings of a year.

Month	1	2	3	4	5	6	7	8	9	10	11	12
Monthly Income	150	165	175	190	200	210	220	240	250	260	270	300

(Ans. Range 150, Coefficient of Range = 0.333)

4. Calculate Semi-inter Quartile Range and its coefficient:

Height in inches	53	55	57	59	61	63	65	67	69
No.of students	25	20	29	20	17	25	20	20	23

(Ans. Semi-inter quartile Range= 4, Coefficient= 0.066)

5. The following are the marks of 100 students of a class. Find the range of variation of marks and also compute coefficient of range.

Marks	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of students	8	16	26	20	12	10	8

(Ans. Range 70, Co-efficient of Range = 0.78)

6. Compute Quartile deviation for the data given below of 7 persons, whose yearly earnings are:

480 650 370 600 300 240 1200

(SVU – March – 2010)

7. Calculate the Quartile-Deviation and it's co-efficient from the following data:

Month	1	2	3	4	5	6	7	8	9	10	11	12
Monthly earnings	39	40	40	41	41	42	42	43	43	44	44	45

(B.Com.Meerut) (Ans. Q.D. = 1.75; Coefficient of Q.D. = 0.042)

8. Compute Quartile Deviation and its Co-efficient from the following data.

Marks:	10	20	30	40	50	60
No. of Students:	5	8	16	9	7	2

9. Compute Quartile Deviation and its coefficient from the following data:

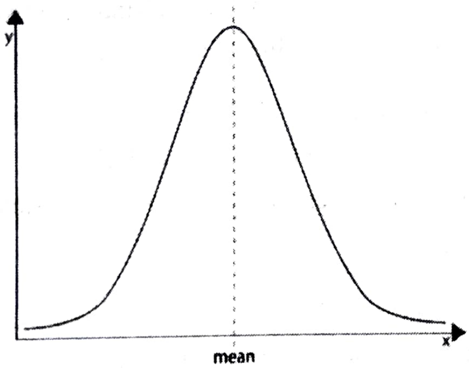
Chapter - V

SKEWNESS AND KURTOSIS

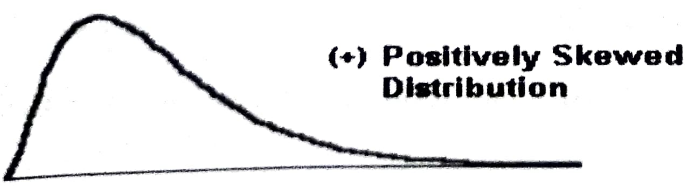
The measures of central tendency and dispersion do not adequately describe distribution in the sense that there could be two distributions with the same Mean and Standard Deviation but still different from each other in respect of shape or pattern of distribution.

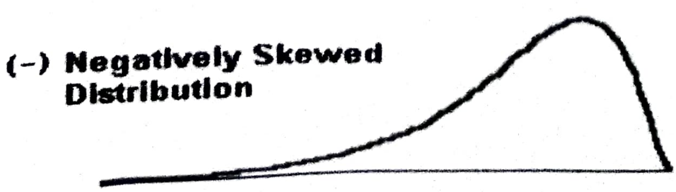
Skewness means lack of symmetry in a frequency distribution. Symmetry is implied when data values are distributed in the same way above and below a middle value. According to Croxton and Cowden "When a series is not symmetrical it is said to be asymmetrical or skewed". Skewness is defined by Spiegel as "the degree of asymmetry or departure from symmetry of a distribution". Values on one side of the distribution tend to be further from the 'middle' than values on the other side. Skewness tells us about the asymmetry of the frequency distribution.

A frequency distribution could be a symmetrical distribution or a skewed distribution. In a symmetrical distribution, the values of Mean, Median and Mode coincide. $\overline{x} = M = Z$. The spread of the frequencies is the same on both sides of the central point of the curve. A symmetrical or Normal" distribution would look as under:



A skewed distribution could be positively skewed or negatively skewed. If the longer tail of the frequency curve of the distribution lies to the right of the central point, it is said to be positively skewed. In such a case, $\overline{x} > M > Z$. If the longer tail of the frequency curve of the distribution lies to the left of the central point, it is said to be negatively skewed. In such a case, $\overline{x} < M < Z$.





Tests of Skewness:

In a distribution, Skewness is present if

(1) The values of mean, median and mode do not coincide.
(2) The sum of the positive deviations from the median is not equal to the sum of the negative deviations.
(3) Quartiles are not equidistant from the median.
(4) Data when plotted on a graph paper will not give normal bell-shaped curve.

Difference between Dispersion and Skewness

1. Dispersion is concerned with measuring the amount of variation in a series. Skewness is concerned with direction of variation or the departure from symmetry.
2. Dispersion gives us the extent to which the values of a given series are scattered. Skewness explains the extent and direction in which the distribution differs from symmetry.
3. Dispersion tells us about the composition of the series. Skewness tells us about the shape of the series.
4. Dispersion helps us to ascertain the extent to which a central value is representative of the series. Skewness deals with the nature of variations on either side of the central value.
5. Measures of Dispersion are based on averages of the first order (i.e. measures of central tendency). They are averages of the second order. Measures of Skewness are based on averages of first and second order (i.e. measures of central tendency and dispersion). Measures of Skewness are not averages at all.

Measures of Skewness

Absolute measure of Skewness = Mean-Mode

Relation measures of Skewness:

(1) **Karl Pearson's Co-efficient of Skewness.**

$$= \frac{\text{Mean - Mode}}{\text{Standard Deviation}}$$

If mode is ill-defined then = $\frac{3(\text{Mean-Median})}{\text{Standard Deviation}}$

Theoretically the value of this co-efficient varies between ±3, however, in practice it lies between ±1.

(2) Bowley's Co-efficient of Skewness:

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

This measure is also called Quartile measure of Skewness. The value obtained by this formula varies between ±1.

Note: The results given by the two formulae given above need not be same.

Illustration 1: Calculate Karl Pearson's measure of Skewness on the basis of Mean, Mode and Standard Deviation.

Size	1.45	15.5	16.5	17.5	18.5	19.5	20.5	21.5
f	35	40	48	100	125	87	43	22

(C.A. Inter)

Solution: Calculation of co-efficient of Skewness.

X	f	d (X–A)	fd	d2	fd2
14.5	35	-3	-105	9	315
15.5	40	-2	-80	4	160
16.5	48	-1	-48	1	48
17.5	100	0	0	0	0
18.5	125	+1	+125	1	125
19.5	87	+2	+174	4	125
20.5	43	+3	+129	9	387
21.5	22	+4	+88	16	352
	N= 500		**∑fd=+283**		**∑fd2=1735**

$\overline{X} = A + \frac{\sum fd}{N}$, A = 17.5, ∑fd = 283, N= 500

$17.5 + \frac{283}{500} = 17.5+0.566 = 18.066$

$s = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} = \sqrt{\frac{1735}{500} - \left(\frac{283}{500}\right)^2} = \sqrt{3.47 - 0.32}$

$= \sqrt{3.15} = 1.775$

By inspection it is clear that mode is 18.5

Karl Pearson's

$\text{Co-efficient of Skewness} = \frac{\text{Mean-Mode}}{\text{Standard Deviation}}$

$= \frac{18.006-18.5}{1.755} = -0.245$

Illustration 2: Calculate Pearson's Co-efficient of Skewness from the table given below:

Life time (hrs.)	No. of tubes	Life time (hrs.)	No. of tubes
300-400	14	800-900	62
400-500	46	900-1000	48
500-600	58	1000-1100	22
600-700	76	1100-1200	5
700-800	68		

(B.Com. Bombay)

Solution:

Calculation of Pearson's Co-efficient of Skewness

X	f	m.p.	d	d'	fd'	d'2	fd'2
300-400	14	350	-400	-4	-56	16	224
400-500	46	450	-300	-3	138	9	224
500-600	58	550	-200	-2	-116	4	232
600-700	76	650	-100	-1	-76	1	76
700-800	68	750	0 0	0	0	0	
800-900	62	850	+100	+1	62	1	62
900-1000	48	950	+200	+2	+96	4	192
1000-1100	22	1050	+300	+3	+66	9	198
110-1200	5	1150	+400	+4	+20	16	80
	N=399				**$\sum fd^1$=-142**		**$\sum fd^2$=1478**

$\overline{X} = A + \frac{\sum fd}{N} \times C$

A= 750, $\sum fd^1$ = -142, N=399, C=100

$= 750 + \frac{-142}{399} \times 1000 = 750 - 35.59 = 714.41$

$s = \sqrt{\frac{\sum fd^2}{n} - \left(\frac{\sum fd}{N}\right)^2} \times C$

$\sum fd^2$= 1478, $\sum fd'$ = –142, N = 399, C= 100

$= \sqrt{\frac{1478}{399} - \left(\frac{-142}{399}\right)^2} \times 100 = \sqrt{3.704} - 0.127 = \sqrt{3.577} \times 100$

= 1.891 x 100 = 189.1

By inspection, it is clear that mode lies in 600-700 class

$\text{Mode} = L_1 + \frac{f_1 - f_0}{2 f_1 - f_0 - f_2} \times i$

$L_1 = 600, f_1 = 76, f_0 = 58, f_2 = 68, i = 100$

$= 600 + \frac{76-58}{22 \times 76-58-68} \times 100$

$= 600 + \frac{18}{26} \times 100$

= 600 + 69.23 = 669.23

$\text{Co-efficient of Skewness} = \frac{\text{Mean--Mode}}{\text{Standard Deviation}}$

$= \frac{714.41 - 669.23}{189.1} = \frac{45.18}{189.1} = +0.239$

Illustration 3: Calculate the co-efficient of Skewness based on mean, median and standard deviation from the following data:

Variable	Frequency	Variable	Frequency
100-110	4	140-	64
110-	16	150-	40
120-	36	160-	32
130-	52	170-180	(C.A. Inter)

Solution:

Variable	f	m.p.	m.p. -A A -135	d/10 d	fd'	d'2	fd'2	C.f
100-110	4	105	-30	-3	-12	9	36	4
110-120	16	115	-20	-2	-32	4	64	20
120-130	36	125	-10	-1	-36	1	36	56
130-140	32	135	0	0	0	0	0	108
140-150	64	145	+10	+1	+64	1	64	172
150-160	40	155	+20	+2	+80	4	160	212
160-170	32	165	+30	+3	+96	9	288	244
170-180	11	175	+40	+4	+44	16	176	255
	N=255				**Σfd' = +204**		**Σfd'2=824**	

$\overline{X} = A + \frac{\sum fd'}{N} \times C$

A = 135, $\sum fd' = 204$, N=255, C=10

$$=135 + \frac{204}{255} \times 100 = 135+8 = 143$$

$$s = \sqrt{\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N}\right)^2} \times C = \sqrt{\frac{824}{255} - \left(\frac{204}{255}\right)^2} \times 10$$

$$= \sqrt{3.231-0.64} \times 10 = \sqrt{2.591} \times 10 = 1.61 \times 10 = 16.1$$

Median = Size of $\frac{N}{2}$ th item = Size of $\frac{255}{2}$ th item

= Size of 127.5th item

Median lies in 140-150 class

$$\text{Median} = L1 + \frac{\frac{N}{2} - C.f}{f}$$

L1 = 140, $\frac{N}{2}$ = 127.5, f=64, C.f=108, i=10

$$= 140 + \frac{127.5-108}{64} \times 10 = 140 + \frac{195}{64} = 140+3.05 = 143.05$$

$$\text{Karl Pearson's Co-efficient of Skewness} = \frac{3(\text{Mean-Median})}{\text{Standard Deviation}}$$

$\overline{X}$ = 143, Median = 143.05, s = 16.1

$$= \frac{3(143-143.05)}{16.1} = \frac{-0.15}{16.1} = -0.0093$$

Illustration 4: Find Bowley's Coefficient of Skewness for the following data:

Solution: First the data has to be rearranged in ascending order.

S.No.	Marks (in ascending order)
1	11
2	12
3	14
4	18
5	22
6	26
7	30
8	32
9	35
10	41

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

$$= \text{Size of } \frac{10+1}{2} = 5.5^{th} \text{ item.} = \frac{\text{Size of 5th item+Size of 6th item}}{2}$$

$$= \frac{22+26}{2} = \frac{48}{2} = 24$$

Median = 24

Lower Quartile (Q1) = Size of $\left(\frac{N+1}{4}\right)$ the item.

$$= \text{Size of } \left(\frac{10+1}{4}\right) = 2.75^{th} \text{ item.}$$

+7.5 (Size of 3rd item-size of second item)
= 12+.75(14-12) = 12+1.5 = 13.5
Q1 = 13.5

Upper Quartile (Q3) = Size of $\frac{3(N+1)}{4}$ th item.

=Size of $\frac{3(10+1)}{4}$ th item

= Size of 8.25 the item.
= Size of 8th item + 0.25 (9th item=8th item)
= 32+.25 (35-32) = 32.75
Bowley's Coefficient of Skewness

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

Where Q_3 = 32.75, Q_1 = 13.5, Median = 24

$$= \frac{32.75+13.5-2(24)}{32.75-13.5} = \frac{-175}{19.25} = -0.09$$

Illustration 5: Find Bowley's coefficient of Skewness for the following data.

Wages	100	110	120	130	140	150	160	170	180
No. of Workers	10	18	22	25	40	15	10	8	7

Solution: Calculation of Bowley's coefficient of Skewness.

Wages	**No. of Workers**	**C.f.**
100	10	10
110	18	28
120	22	50
130	25	75
140	40	115
150	15	130
160	10	140
170	8	148
180	7	155

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

= Size of $\frac{156}{2}$ = 78th item

Median = 140

Lower Quartile (Q_1) = Size of $\left(\frac{N+1}{4}\right)$ th item = Size of $\left(\frac{155+}{4}\right)$ th

= Size of 39th item
Size of 39th item = 120; Hence Q_1 = 120

Upper Quartile (Q_3) = Size of $\frac{3(N+1)}{4}$ th item = Size of $\frac{3(155+1)}{4}$ th

= Size of 117th item.
Size of 117th item = 150; Hence Q_3 =150

Bowley's Co-efficient of Skewness = $\frac{Q_3+Q_1-2Med}{Q_3-Q_1}$

Q_3 = 150, Q_1=120, Median = 140

$$= \frac{150+120-2(140)}{150-120} = \frac{270-280}{30} = \frac{-10}{30} = -0.33$$

Illustration 6: Calculate co-efficient of Quartile Deviation and Bowley's coefficient Skewness from the following data.

Profits (Rs. Lakhs)	No. of Companies
Below 10	5
10-20	12
20-30	20
30-40	16
40-50	2

(M.Com. H.P.U.)

Solution: Calculation of Co-efficient of Quartile Deviation and Co-efficient of Skewness.

Profits (Rs. Lakhs)	No. of Companies	C.f.
Below 10	5	5
10-20	12	17
20-30	20	37
30-40	16	53
40-50	5	58
Above 50	2	60
	N=60	

Median $= \text{Size of } \frac{N}{2} \text{ th item} = \text{Size of } \frac{60}{2} = 30\text{th item}$

Median lies in the class 20–30

$$\text{Median} = L_1 + \frac{\frac{N}{2} - C.f.}{f} \times i$$

$L_1 = 20 \; \frac{N}{2} = 30, C.f=17, f=20, i=10$

$$= 20 + \frac{30-17}{20} \times 10 = 20 + \frac{13}{20} \times 10 = 20+6.5 = 26.5$$

$Q_1 = \text{Size } \frac{N}{4} \text{ th item} = \text{Size of } \frac{60}{4} \text{ th} = 15\text{th item}$

Q_1 lies in the class 10–20

$$Q_1 = L_1 + \frac{\frac{N}{4} - C.f}{f} \times i$$

$L_1 = 10, \frac{N}{4} = 15, C.f=5, f=12, i=10$

$$Q_1 = 10 + \frac{15-5}{12} \times 10 = 10 + \frac{10}{12} \times 10 = 10+8.33$$

$Q_1 = 18.33$

$Q_3 = \text{Size of } \frac{3N}{4}^{th} \text{ item, Size of } \frac{3 \times 60}{4} = 45^{th} \text{ item}$

Q_3 lies in the class 30–40

$$Q_3 = L_1 + \frac{\frac{3N}{4} - C.f}{f} \times i$$

$L_1=30, \frac{3N}{4}=45, C.f=37, f=16, i=10$

$$= 30 + \frac{45-37}{16} \times 10 = 30 + \frac{8}{16} \times 10 = 30+5=35$$

$Q_3 = 35$

$$\text{Co-efficient of Q.D.} = \frac{Q_3-Q_1}{Q_3+Q_1} = \frac{35-18.33}{35+18.33} = 0.313$$

$$\text{Bowley's co-efficient of Skewness.} = \frac{Q_3+Q_1-2\text{ Med}}{Q_3-Q_1}$$

$$\frac{35+18.33-2\,(26.5)}{35-18.33} = \frac{0.33}{16.67} = 0.0198 = 0.02$$

Illustration 7: For the frequency distribution given below, calculate coefficient of skewness based on quartiles.

Class limits	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89
Frequency	5	9	14	20	25	15	8	4

(I.C.W.A. Inter)

Solution:

Calculation of Bowley's Co-efficient of Skewness

Size	Frequency	C.f.
10-19	5	5
20-29	9	14
30-39	14	28
40-49	20	48
50-59	25	73
60-69	15	88
70-79	8	96
80-89	4	100

$Q_1 = \text{Size of } \frac{N}{4} \text{th item} = \text{Size of } \frac{100}{4} = 25\text{th item.}$

Q_1 class is 30-39 but the real class limits are 29.5-39.5

$$Q_1 = L_1 + \frac{\frac{N}{4} - C.f}{f} \times i$$

$L_1=29.5, \frac{N}{4}=25, C.f=14, f=14, i=10$

$$Q_1 = 29.5 + \frac{25-14}{14} \times 10 = 29.5 + \frac{11}{14} \times 10 = 29.5+7.86=37.36$$

$\text{Median} = \text{Size of } \frac{N}{2} \text{th item} = \text{Size of } \frac{100}{2} = 50\text{th item}$

Median class = 50-59 i.e. 49.5-59.5

$$\text{Median} = L_1 + \frac{\frac{N}{2} - C.f}{f} \times i$$

$$49.5 + \frac{50-48}{25} \times 10 = 49.5 + 0.8 = 50.3$$

$$Q_3 = \text{Size of } \frac{3N}{4}^{th} \text{ item} = \frac{3 \times 100}{4} = 75^{th} \text{ item.}$$

Q_3 Class =60-69 i.e. 59.5-69.5

$$Q_3 = L_1 + \frac{\frac{3N}{4} - C.f}{f} \times i$$

$L_1 = 59.5, \frac{3N}{4} = 75, C.f = 73, f = 15, i = 10$

$$= 59.5 + \frac{75-73}{15} \times 10 = 59.5 + \frac{2}{15} \times 10 = 59.5 + 1.33 = 60.83$$

$$\text{Coefficient of Skewness} = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1} = \frac{60.83 + 37.36 - 2(50.3)}{60.83 - 37.36}$$

$$= \frac{-2.41}{23.47} = -0.103$$

Illustration 8: From the information give below, Calculate Karl Pearson's co-efficient of skewness and also quartile coefficient of skewness.

Measure	Place A	Place B
Mean	150	140
Median	142	155
Standard Deviation	30	55
Third Quartile	195	260
First Quartile	62	80

(B.Com. Osmania.)

Place A

Karl Pearson's co-efficient of Skewness

$$= \frac{3(\text{Mean} - \text{Median})}{\text{Standard Deviation}} = \frac{3(150-142)}{30} = \frac{3 \times 8}{30} = 0.8$$

Bowley's Co-efficient of Skewness

$$= \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1} = \frac{195 + 62 - 2(142)}{195 - 62} = \frac{-27}{133} = -0.203$$

Place B

Karl Pearson's Coefficient of Skewness

$$\frac{3(\text{Mean} - \text{Median})}{\text{Standard Deviation}} = \frac{3(140-155)}{55} = \frac{3x-15}{55} = 0.818$$

Bowley's Co-efficient of Skewness.

$$= \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1 1} = \frac{260 + 80 - 2(155)}{260 - 80} = 0.167$$

SHORT-ANSWER QUESTIONS

Illustration 1: Calculate Karl Pearson's Coefficient of Skewness and coefficient of variation from the following data.

Mode=33.5, Mean=30.08, Standard Deviation = 13.405

Solution: Co-efficient of Skewness

$$= \frac{\text{Mean-Mode}}{\text{Standard Deviation}} = \frac{30.08-33.5}{13.405} = -0.255$$

Coefficient of Variation

$$= \frac{\sigma}{\overline{X}} \times 100 = \frac{13.405}{30.08} \times 100 = 44.56$$

Illustration 2: Calculate Karl Pearson's Coefficient of Skewness when $\overline{X}$ =20, Mode=20, s=13.62

Solution: Co-efficient of Skewness

$$= \frac{\text{Mean-Mode}}{\text{Standard Deviation}} = \frac{20020}{13.62} = 0$$

Illustration3: Calculate Co-efficient of Skewness when defined values of Mean, Median and Standard Deviation 140, 146 and 16.4 respectively

Karl Pearson's Coefficient of Skewness (when mode is ill-defined)=

$$\frac{3(\text{Mean-Median})}{\text{Standard Deviation}} = \frac{3(140-146)}{16.4} = \frac{-18}{16.4} = -1.098$$

Illustration 4: Calculate Coefficient of Skewness from the following information

First Quartile (Q_1) = 14 cm

Third Quartile (Q_3) = 25 cm

Median = 18cm *(B.Com. Kakatiya)*

Solution:

Bowley's Co-efficient of Skewness

$$= \frac{Q_3+Q_1-2M}{Q_3-Q_1} = \frac{25+14-2(18)}{25-14} = \frac{3}{11} = 0.273$$

Illustration5: From the data given below, calculate Karl Pearson's and Bowley's Coefficients of Skewness.

Mean= 160, Median 152, Q1=72, Q3=205

and Standard Deviation =40

Solution:

Karl Pearson's Coefficient of Skewness

$$= \frac{3(\text{Mean-Median})}{\text{Standard Deviation}} = \frac{3(160-152)}{40} = \frac{3 \times 8}{40} = 0.6$$

Bowley's Co-efficient of Skewness

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

$$= \frac{205+62-2(152)}{205-72} = \frac{-27}{133} = -203$$

Illustration 6: You are given the following information SK = 0.8, Arithmetic mean = 30, Mode=24 Find the value of Standard Deviation. ***(B.Com. Osmania)***

Solution:

$$\frac{\overline{X} - \text{Mode}}{\sigma} = SK$$

$$= \frac{30\text{-}24}{s} = 0.8 \Rightarrow \frac{6}{\sigma} = 0.8; \Rightarrow 0.8\ \sigma = 6$$

$$\sigma = \frac{6}{0.8} = 7.5$$

Illustration 7: From the data given below calculate Coefficient of Variation .

Pearson's measures of Skewness = 0.42

Arithmetic mean = 86

Median = 80 ***(C.A. Inter)***

Solution:

Karl Pearson's Coefficient of Skewness

$$\frac{3(\overline{X} - \text{Median})}{\sigma} = SK$$

$$= \frac{3(86\text{-}80)}{\sigma} = 0.42 \Rightarrow \frac{18}{\sigma} = 0.42, \Rightarrow 0.42,\ \sigma = 18$$

$$\sigma = \frac{18}{0.42} = 42.86$$

$$\text{Co-efficient of Variation} = \frac{\sigma}{\overline{X}} \times 100 \qquad = \frac{42.86}{86} \times 100 = 49.84$$

Illustration 8: For a distribution, Karl Pearson's Coefficient of Skewness is 0.40. Its standard deviation is 8 and mean is 30. Find the mode of the distribution.

(B.Com. Kakatiya)

Solution:

Karl Pearson's Coefficient of Skewness

$$= \frac{\overline{X} - \text{Mode}}{\sigma} = SK = \frac{30\text{-Mode}}{8} = 0.40 = 30\text{–Mode} = 3.2$$

Mode = 26.8

Illustration 9: For a distribution, the arithmetic mean is 100, standard deviation is 35 and the Karl Pearson's Coefficient of Skewness is 0.2 Find the median.

(B.Com. Kakatiya)

Solution:

$$= \frac{3\ (\overline{X} - \text{Median})}{s} = SK = \frac{3\ (100\text{-Median})}{35} = 0.2$$

= 300-3 Median = 7, = 3 Median = 293, Median = 97.67

Illustration 10: Find the value of standard deviation of Mean = 50, Mode=25, and Coefficient of Skewness = 0.4 ***(B.Com. Kakatiya)***

Solution:

$$= \frac{\overline{X} - \text{Mode}}{\sigma} = \text{Coefficient of Skewness.} = \frac{50\text{-}25}{\sigma} = 0.4 = \frac{25}{\sigma} = 0.4$$

$= 0.4,\ \sigma = 25,\ \sigma = 62.5$

Illustration 11: In a certain distribution the following results were obtained:

$\overline{X} = 45$, Median = 48, Coefficient of SK =–0.4

The person who gave you the data failed to give the value of the standard deviation and you are required to estimate it with the help of the available information.

(C.A. Inter)

Solution:

$$= \frac{3(\overline{X} - \text{Median})}{\sigma} = \text{Co-efficient of Skewness.}$$

$$= \frac{3(45-48)}{\sigma} = -0.4 = -0.4, \sigma = -9, \sigma = 22.5$$

Illustration 12: For a distribution, the coefficient of Quartile Deviation is 0.33, Q_1 is 16 and Median is 19. Find out Q_3 and Bowley's Coefficient of Skewness.

(B.Com. Osmania)

Solution:

$$\frac{Q_3-Q_1}{Q_3+Q_1} = \text{Co-efficient of Q.D.}$$

$$\frac{Q_3-16}{Q_3+16}\ 0.33$$

$$0.33\ Q_3+5.28 = Q_3-16$$

$$0.33\ Q_3-Q_3 = -16-5.28$$

$$0.67\ Q_3 = 21.28, Q_3 = 31.76$$

Bowley's Coefficient of Skewness

$$\frac{Q_3+Q_1-2\ \text{Median}}{Q_3-Q_1} = \frac{72.67-2(36.18)}{2.05} = \frac{0.31}{2.05} = 0.151$$

Illustration 14: In a frequency distribution, the coefficient of skewness based on quartiles in 0.6. If the sum of the Upper and Lower Quartiles is 100 and Median is 38, find the value of Upper Quartile.

(B.Com. Delhi)

Solution:

Given S.K. = 0.6, $Q_3+Q_1 = 100$, Median = 38

Coefficient of Skewness

$$= \frac{Q_3+Q_1-2\ \text{Median}}{Q_3-Q_1} = \frac{100-76}{Q_3-Q_1} = 0.6 = \frac{24}{Q_3-Q_1} = 0.6$$

$$= \frac{24}{Q_3-Q_1} = 0.6 = 0.6\,(Q_3-Q_1) = 24 = Q_3-Q_1 = \frac{24}{0.6} = 40$$

$= Q_3+Q_1 = 100$, $Q_3-Q_1=40$, by adding the two we get

$2\,Q_3 = 140$, $Q_3 = 70$

Illustration 15: The measure of skewness for a certain distribution is -0.8 If the lower and upper quartiles are 44.1 and 56.6 respectively. find the median.

(I.C.W.A.)

Solution:

Given S.K. = –0.8, $Q_1 = 44.1$ $Q_3 = 56.6$

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