2.4. Geometric Mean

The geometric mean of a set of n observations is the nth root of their product. If there are n observations, x1, x2, .., xn such that xi > 0 for each i, their geometric mean GM is given by

$$\text{GM} = (x_1 \cdot x_2 \cdot \cdots \cdot x_n)^{\frac{1}{n}}$$

The nth root is calculated with the help of logarithms. Taking logarithms of both the sides,

$$\begin{aligned}
\log \text{GM} &= \log(x_1 \cdot x_2 \cdots x_n)^{\frac{1}{n}} \\
&= \frac{1}{n}\log(x_1 \cdot x_2 \cdots x_n) \\
&= \frac{1}{n}(\log x_1 + \log x_2 + \cdots + \log x_n) \\
&= \frac{\sum \log x}{n}
\end{aligned}$$

$$\text{GM} = \text{antilog}\left(\frac{\sum \log x}{n}\right)$$

In case of a frequency distribution consisting of n observations x1, x2, .., xn with respective frequencies f1, f2, .., fn, the geometric mean is given by

$$\text{GM} = (x_1^{f_1} \cdot x_2^{f_2} \cdots x_n^{f_n})^{\frac{1}{N}}, \text{ where } N = \sum f$$

Taking logarithms of both the sides,

$$\log \text{GM} = \frac{1}{N}(f_1 \log x_1 + f_2 \log x_2 + \cdots + f_n \log x_n)$$

$$= \frac{\sum f \log x}{N}$$

$$\text{GM} = \text{antilog}\left(\frac{\sum f \log x}{N}\right)$$

Thus, the geometric mean is the antilog of the weighted mean of the different values of log xi whose weights are their frequencies fi.

In case of a continuous or grouped frequency distribution, x is taken to be the value corresponding to the midpoints of the class intervals.

Example 1: Calculate the geometric mean of the following data:

10, 110, 120, 50, 52, 80

Solution

$$\text{GM} = \text{antilog}\left(\frac{\sum \log x}{n}\right)$$

$$= \text{antilog}\left(\frac{\log 10 + \log 110 + \log 120 + \log 50 + \log 52 + \log 80}{6}\right)$$

$$= \text{antilog}\left(\frac{2.3026 + 4.7005 + 4.7875 + 3.9120 + 3.9512 + 4.3820}{6}\right)$$

$$= \text{antilog}\,(4.006)$$

$$= 54.9267$$

Example 2: Find the geometric mean of the following data:

x	5	10	15	20	25	30
f	13	18	50	40	10	6

Solution

x	f	$\log x$	$f \log x$
5	13	1.6094	20.9227
10	18	2.3026	41.4465
15	50	2.7081	135.4025
20	40	2.9957	119.8293
25	10	3.2189	32.1888
30	6	3.4012	20.4072
	$\sum f = 137$		$\sum f \log x = 370.197$

$$N = \sum f = 137$$

$$\begin{aligned} \text{GM} &= \text{antilog}\left(\frac{\sum f \log x}{N}\right) \\ &= \text{antilog}\left(\frac{370.197}{137}\right) \\ &= 14.912 \end{aligned}$$

Example 3 : Find the geometric mean of the following data:

Marks	0–10	10–20	20–30	30–40
No. of Students	5	8	3	4

Solution

Marks	No. of Students f	Midvalue x	$\log x$	$f \log x$
0–10	5	5	1.6094	8.047
10–20	8	15	2.7081	21.6648
20–30	3	25	3.2189	9.6567
30–40	4	35	3.5553	14.2212
	$\sum f = 20$			$\sum f \log x = 53.5897$

$$N = \sum f = 20$$

$$\text{GM} = \text{antilog}\left(\frac{\sum f \log x}{N}\right)$$

$$= \text{antilog}\left(\frac{53.5897}{20}\right)$$

$$= 14.5776$$

2.5. Harmonic Mean:

The harmonic mean of a number of observations, none of which is zero, is the reciprocal of the arithmetic mean of the reciprocals of the given values.

The harmonic mean of n observations x1, x2, ..., xn is given by

$$\text{HM} = \frac{1}{\frac{1}{n}\sum\left(\frac{1}{x}\right)}$$

$$= \frac{1}{\frac{1}{n}\left(\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}\right)}$$

$$= \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$$

For example, the harmonic mean of 2, 4 and 5 is

$$\text{HM} = \frac{3}{\frac{1}{2} + \frac{1}{4} + \frac{1}{5}} = 3.16$$

In case of a frequency distribution consisting of n observations x1, x2, ..., xn with respective frequencies f1, f2, ..., fn, the harmonic mean is given by

$$\text{HM} = \frac{f_1 + f_2 + \cdots + f_n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \cdots + \frac{f_n}{x_n}}$$

$$= \frac{\sum f}{\sum\left(\frac{f}{x}\right)}$$

If x1, x2, ..., xn are n observations with weights w1, w2, ..., wn respectively, their weighted harmonic mean is given by

$$\text{HM} = \frac{\sum w}{\sum\left(\frac{w}{x}\right)}$$

Example 4 : Calculate the harmonic mean of the following data:

x	20	21	22	23	24	25
f	4	2	7	1	3	1

Solution

x	f	$\frac{f}{x}$
20	4	0.2
21	2	0.095
22	7	0.318
23	1	0.043
24	3	0.125
25	1	0.04
	$\sum f = 18$	$\sum\left(\frac{f}{x}\right) = 0.821$

$$\text{HM} = \frac{\sum f}{\sum\left(\frac{f}{x}\right)} = \frac{18}{0.821} = 21.924$$

Example 5

Find the harmonic mean of the following distribution:

Class interval	0–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80
Frequency	5	8	11	21	35	30	22	18

Solution

Class Interval	Frequency f	Midvalue x	$\frac{f}{x}$
0–10	5	5	1
10–20	8	15	0.533
20–30	11	25	0.44
30–40	21	35	0.6
40–50	35	45	0.778
50–60	30	55	0.545
60–70	22	65	0.338
70–80	18	75	0.24
	$\sum f = 150$		$\sum\left(\frac{f}{x}\right) = 4.474$

$$\text{HM} = \frac{\sum f}{\sum\left(\frac{f}{x}\right)} = \frac{150}{4.474} = 33.527$$

- **Relation between arithmetic Mean, Geometric Mean, and harmonic Mean**

The arithmetic mean (AM), geometric mean (GM), and harmonic mean (HM) for a given set of observations of a series are related as

$$\text{AM} \geq \text{GM} \geq \text{HM}$$

For two observations x1 and x2 of a series

$$\text{AM} = \frac{x_1 + x_2}{2}$$

$$\text{GM} = \sqrt{x_1\ x_2}$$

$$\text{HM} = \frac{2}{\frac{1}{x_1} + \frac{1}{x_2}} = \frac{2x_1\ x_2}{x_1 + x_2}$$

$$\text{AM} \cdot \text{HM} = \left(\frac{x_1 + x_2}{2}\right)\left(\frac{2x_1\ x_2}{x_1 + x_2}\right) = x_1\ x_2 = (\text{GM})^2$$

$$\therefore \qquad \text{GM} = \sqrt{\text{AM} \cdot \text{HM}}$$

Example 6

If the AM of two observations is 15 and their GM is 9, find their HM and the two observations.

Solution

$$\text{GM} = \sqrt{\text{AM} \cdot \text{HM}}$$

$$9 = \sqrt{15 \times \text{HM}}$$

$$\therefore \quad \text{HM} = 5.4$$

Let the two observations be x_1 and x_2.

$$\text{AM} = \frac{x_1 + x_2}{2} = 15$$

$$x_1 + x_2 = 30 \qquad \text{...(1)}$$

$$\text{GM} = \sqrt{x_1\ x_2} = 9$$

$$x_1\ x_2 = 81 \qquad \text{...(2)}$$

Solving Eqs (1) and (2),

$$x_1 = 27,\ x_2 = 3$$

3. Dispersion Measurements:

For comparing two sets of data, Central Tendency Measurements can be used . But the use of these methods alone is not sufficient for comparing, the measure of the central tendency of the two groups may be equal , , but there may be a significant difference between the two groups the convergence and divergence of the data from each other **. for example, the following two set are**

Set one 63 70 78 81 85 67 88

Set two 73 78 77 78 75 74 77

The arithmetic mean of each group is equal to 76 degrees, however, the scores of the second group are more homogeneous than the scores of the first group. Therefore, the statistical science takes another indicator to find more the pater understanding for the data from this indicator dispersion measurements which include the following: -

3.1. Standard Deviation:

Standard deviation is the positive square root of the arithmetic mean of the squares of the deviations of the given values from their arithmetic mean. It is denoted by the Greek letter s. Let X be a random variable which takes on values, viz., x1, x2, ..., xn. The standard deviation of these n observations is given by

$$\sigma = \sqrt{\frac{\sum (x-\overline{x})^2}{n}}$$

where $\overline{x} = \frac{\sum x}{n}$ is the arithmetic mean of these observations.

Or

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

In case of a frequency distribution consisting of n observations x1, x2, ..., xn with respective frequencies f1, f2, ..., fn, the standard deviation is given by

$$\sigma = \sqrt{\frac{\sum f(x-\bar{x})^2}{N}}$$

Or

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

3.2. Variance :-

The variance is the square of the standard deviation and is denoted by σ^2.

The method for calculating variance is same as that given for the standard deviation.

Example 7

Calculate the standard deviation of the weights of ten persons.

Weight (in kg)	45	49	55	50	41	44	60	58	53	55

Solution:

$$\bar{x} = \frac{\sum x}{n} = \frac{510}{10} = 51$$

x	$x-\overline{x}$	$(x-\overline{x})^2$
45	–6	36
49	–2	4
55	4	16
50	–1	1
41	–10	100
44	–7	49
60	9	81
58	7	49
53	2	4
55	4	16
		$\sum(x-\overline{x})^2 = 356$

$$\sigma = \sqrt{\frac{\sum(x-\overline{x})^2}{n}}$$
$$= \sqrt{\frac{356}{10}}$$
$$= 5.967$$

Example 8

Calculate the standard deviation of the following data:

x	10	11	12	13	14	15	16	17	18
f	2	7	10	12	15	11	10	6	3

Solution

$$N = \sum f = 76$$

$$\bar{x} = \frac{\sum fx}{N} = \frac{1064}{76} = 14$$

x	f	$x-\bar{x}$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
10	2	–4	16	32
11	7	–3	9	63
12	10	–2	4	40
13	12	–1	1	12
14	15	0	0	0
15	11	1	1	11
16	10	2	4	40
17	6	3	9	54
18	3	4	16	48
				$\sum f(x-\bar{x})^2 = 300$

$$\sigma = \sqrt{\frac{\sum f(x-\bar{x})^2}{N}}$$

$$= \sqrt{\frac{300}{76}}$$

$$= 1.987$$

3.3 Coefficient of Variation

The standard deviation is an absolute measure of dispersion. The coefficient of variation is a relative measure of dispersion and is denoted by CV.

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

Example 9

The arithmetic mean of the runs scored by three batsmen Ahmed, Suha, and Ali in the series are 50, 48, and 12 respectively. The standard deviations of their runs are 15, 12, and 2 respectively. Who is the more consistent of the three?

Solution

Let $\bar{X}_1$, $\bar{X}_2$, $\bar{X}_2$ be the arithmetic means and **and σ_1 , σ_2, σ_3** be the standard deviations of the runs scored by Ahmed, Suha, and Ali .

$$\bar{x}_1 = 50, \bar{x}_2 = 48, \bar{x}_3 = 12, \sigma_1 = 15, \sigma_2 = 12, \sigma_3 = 2$$

$$\begin{aligned} CV_1 &= \frac{\sigma_1}{\bar{x}_1} \times 100 \\ &= \frac{15}{50} \times 100 \\ &= 30\% \end{aligned}$$

$$\begin{aligned} CV_2 &= \frac{\sigma_2}{\bar{x}_2} \times 100 \\ &= \frac{12}{48} \times 100 \\ &= 25\% \end{aligned}$$

$$\begin{aligned} CV_3 &= \frac{\sigma_3}{\bar{x}_3} \times 100 \\ &= \frac{2}{12} \times 100 \\ &= 16.67\% \end{aligned}$$

Since the coefficient of variation of Ali is least, he is the most consistent.

Example 10

The number of matches played, and goals scored by two teams A and B in World Cup Football 2002 were as follows. Find which team may be considered more consistent.

Matches played by Team *A*	27	9	8	5	4
Matches played by Team *B*	17	9	6	5	3
No. of goals scored in a match	0	1	2	3	4

Solution

For Team *A*,

$$N_A = 27 + 9 + 8 + 5 + 4 = 53$$

$$\sum fx_A = (27 \times 0) + (9 \times 1) + (8 \times 2) + (5 \times 3) + (4 \times 4) = 56$$

$$\sum fx_A^2 = (27 \times 0^2) + (9 \times 1^2) + (8 \times 2^2) + (5 \times 3^2) + (4 \times 4^2) = 150$$

$$\sigma_A = \sqrt{\frac{\sum fx_A^2}{N_A} - \left(\frac{\sum fx_A}{N_A}\right)^2}$$

$$= \sqrt{\frac{150}{53} - \left(\frac{56}{53}\right)^2}$$

$$= 1.31$$

$$\bar{x}_A = \frac{\sum fx_A}{N_A} = \frac{56}{53} = 1.06$$

$$\text{CV}_A = \frac{\sigma_A}{\bar{x}_A} \times 100$$

$$= \frac{1.31}{1.06} \times 100$$

$$= 123.58\%$$

For Team *B*,

$$N_B = 17 + 9 + 6 + 5 + 3 = 40$$

$$\sum fx_B = (17 \times 0) + (9 \times 1) + (6 \times 2) + (5 \times 3) + (3 \times 4) = 48$$

$$\sum fx_B^2 = (17 \times 0^2) + (9 \times 1^2) + (6 \times 2^2) + (5 \times 3^2) + (3 \times 4^2) = 126$$

$$\sigma_B = \sqrt{\frac{\sum fx_B^2}{N_B} - \left(\frac{\sum fx_B}{N_B}\right)^2}$$

$$= \sqrt{\frac{126}{40} - \left(\frac{48}{40}\right)^2}$$

$$= 1.31$$

$$\bar{x}_B = \frac{\sum fx_B}{N_B} = \frac{48}{40} = 1.2$$

$$\text{CV}_B = \frac{\sigma_B}{\bar{x}_B} \times 100$$

$$= \frac{1.31}{1.2} \times 100$$

$$= 109.17\%$$

Since CVB < CVA, Team B is more consistent in performance.

3.4. skewness:

Skewness is a measure that refers to the extent of symmetry or asymmetry in a distribution. A distribution is said to be symmetrical when its mean, median, and mode are equal, and the frequencies are symmetrically distributed about the mean. A symmetrical distribution when plotted on a graph will give a perfectly bell-shaped curve which is known as a normal curve (Fig. 3.1).

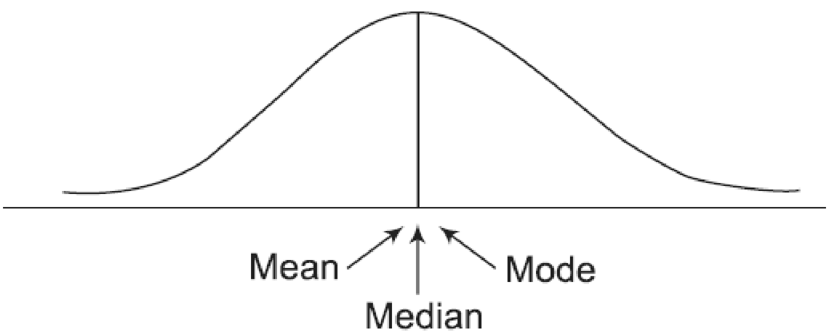

Fig. 3.1

A distribution is said to be asymmetrical or skewed when the mean, median, and mode are not equal, i.e., the mean median, and mode do not coincide. If the curve has a longer tail towards the left, it is said to be a negatively skewed distribution (Fig. 3.2a). If the curve has a longer tail towards the right, it is said to be positively skewed (Fig. 3.2b).

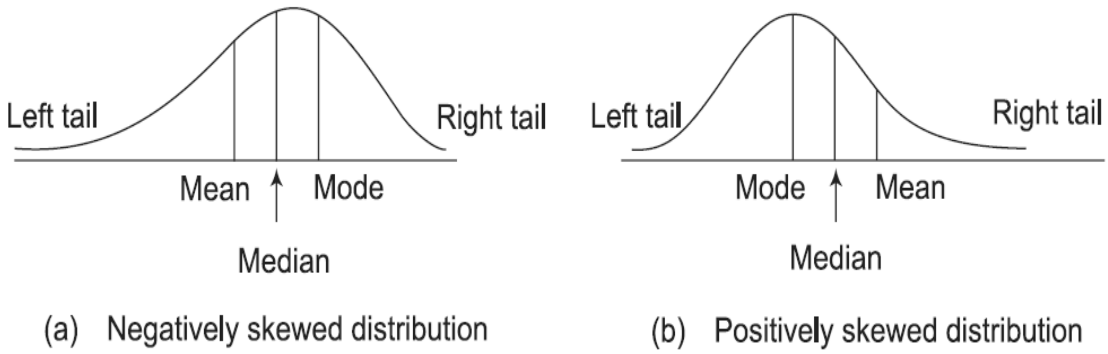

Fig. 3.2

3.4.1 Measures of skewness

A measure of skewness gives the extent and direction of skewness of a distribution. These measures can be absolute or relative. The absolute measures are also known as measures of skewness.

$$\text{Absolute skewness} = \text{Mean} - \text{Mode}$$

If the value of the mean is greater than the mode, the skewness will be positive and if the value of the mean is less than the mode, the skewness will be negative. The relative measures of skewness is called the coefficient of skewness.

3.4.2 Karl Pearson's Coefficient of Skewness

Karl Pearson's coefficient of skewness denoted by Sk, is given by

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

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

When the mode is ill-defined and the distribution is moderately skewed, the averages have the following relationship:

$$\text{Mode} = 3\ \text{Median} - 2\ \text{Mean}$$

$$S_k = \frac{\text{Mean} - (3\ \text{Median} - 2\ \text{Mean})}{\text{Standard Deviation}}$$

$$= \frac{3(\text{Mean} - \text{Median})}{\text{Standard Deviation}}$$

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

The coefficient of skewness usually lies between –1 and 1.

For a positively skewed distribution, Sk > 0.

For a negatively skewed distribution, Sk < 0.

For a symmetrical distribution, Sk = 0.

Example 11

Calculate Karl Pearson’s coefficient of skewness for the following data:

x	0	1	2	3	4	5	6	7
y	12	17	29	19	8	4	1	0

Solution

Let $a = 4$ be the assumed mean.

$d = x - a = x - 4$

x	f	d	d^2	fd	fd^2
0	12	–4	16	–48	192
1	17	–3	9	–51	153
2	29	–2	4	–58	116
3	19	–1	1	–19	19
4	8	0	0	0	0
5	4	1	1	4	4
6	1	2	4	2	4
7	0	3	9	0	0
	$\sum f = 90$			$\sum fd = -170$	$\sum fd^2 = 488$

$$N = \sum f = 90$$

$$\bar{x} = a + \frac{\sum fd}{N} = 4 + \left(\frac{-170}{90}\right) = 2.11$$

$$\sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} = \sqrt{\frac{488}{90} - \left(\frac{-170}{90}\right)^2} = 1.36$$

Since the maximum frequency is 29, the mode is 2.

$$S_k = \frac{\text{Mean} - \text{Mode}}{\sigma} = \frac{2.11 - 2}{1.36} = 0.08$$

Example 12

Calculate Karl Pearson's coefficient of skewness from the following data:

Wages (₹)	10–15	15–20	20–25	25–30	30–35	35–40	40–45	45–50
No. of Workers	8	16	30	45	62	32	15	6

Solution

Let $a = 32.5$ be the assumed mean and $h = 5$ be the width of the class interval.

$$d = \frac{x-a}{h} = \frac{x-32.5}{5}$$

Wages (₹)	No. of workers f	Midvalue x	$d = \frac{x-32.5}{5}$	d^2	fd	fd^2
10–15	8	12.5	–4	16	–32	128
15–20	16	17.5	–3	9	–48	144
20–25	30	22.5	–2	4	–60	120
25–30	45	27.5	–1	1	–45	45
30–35	62	32.5	0	0	0	0
35–40	32	37.5	1	1	32	32
40–45	15	42.5	2	4	30	60
45–50	6	47.5	3	9	18	54
	$\sum f = 214$				$\sum fd = -105$	$\sum fd^2 = 583$

$$N = \sum f = 214$$

$$\bar{x} = a + h\frac{\sum fd}{N}$$

$$= 32.5 + 5\left(\frac{-105}{214}\right)$$

$$= 30.05$$

$$\sigma = h\sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}$$
$$= 5\sqrt{\frac{583}{214} - \left(\frac{-105}{214}\right)^2}$$
$$= 7.88$$

Since the maximum frequency is 32, the mode lies in the interval 30–35.

Here, $l = 30,\quad h = 5,\quad f_m = 62,\quad f_1 = 45,\quad f_2 = 32$

$$\text{Mode} = l + h\left(\frac{f_m - f_1}{2f_m - f_1 - f_2}\right)$$
$$= 30 + 5\left[\frac{62 - 45}{2(62) - 45 - 32}\right]$$
$$= 31.81$$

$$S_k = \frac{\text{Mean} - \text{Mode}}{\sigma}$$
$$= \frac{30.05 - 31.81}{7.88}$$
$$= -0.223$$

3.5. Range:

It is the simpler Dispersion Measurements, it ca be calculated as the (Range = Maximum value – Minimum value) . but in continuous frequency distribution can be calculated as (Range = midvalue of last interval – midvalue of first interval).

Example 13: calculate the Range of the following data?

1- 4.8 6.21 5.4 5.18 5.29 5.18 5.08 4.63 5.03

Range = 6.21 – 4.63 = 1.58

2- continuous frequency distribution

Area	15-20	20-25	25-30	30-35	35-40	40-45
No.	3	9	15	18	12	3

The Range = 42.5 - 17.5 = 25

Homework two :

Two workers on the same job show the following results over a long period of time. Which one is more consistent ?

	Worker *A*	Worker *B*
Mean time (in minutes)	30	25
Standard deviation (in minutes)	6	4

Find Karl Pearson's coefficient of skewness for the following data:

Marks	0–10	10–20	20–30	30–40	40–50	50–60	60–70
No. of students	10	12	18	25	16	14	8

[Ans.: 0.013]

Calculate the geometric and harmonic means of the following series of monthly expenditure of a batch of students:

₹	125	130	75	10	45	0.5	0.4	500	1505

[**Ans.:** ₹ 22.98, ₹ 2.06]

Calculate the geometric mean of the following distribution:

Class intervals	5–15	15–25	25–35	35–45	45–55
Frequency	10	22	25	20	8

[**Ans.:** 26.65]

Calculate the standard deviation from the following data:

Heights in cm	150	155	160	165	170	175	180
No. of students	15	24	32	33	24	16	6

[**Ans.:** 8.038 cm]

Find the standard deviation of the following data:

Size of items	10	11	12	13	14	15	16
Frequency	2	7	11	15	10	4	1

[**Ans.:** 1.342]