CHAPTER # 3

MEASURE OF CENTRAL TENDENCY

3.1 INTRODUCTION

An average is a single value, which represents the set of data as whole. Since the average tends to lie in the center of distribution they are also called measure of central tendency.

3.1.1 Qualities of Good Average

An average that possesses all or most of the following qualities is considered a good average.

1. It should be rigidly defined.
2. It should be easy to understand.
3. It should be easy to calculate.
4. It should be based on all the observations of the data.
5. It should be unaffected by extreme observations.
6. It should have sampling stability.

3.1.2 Types of Averages

The commonly used averages are
(i)   Arithmetic mean
(ii)   The Median
(iii)   The Mode
(iv)   Geometric mean
(v)   Harmonic mean

3.2 ARITHMETIC MEAN

It is defined as the sum of all the observations divided by the number of observations. It is denoted by $\bar{X}$.

Let $X_1, X_2, \ldots\ldots X_n$ be the 'n' observations then arithmetic mean is defined as

$$\bar{X} = \frac{X_1 + X_2 + \ldots\ldots X_n}{n}$$

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

EXAMPLE 3.1

The marks obtained by 5 students are given below.

70, 50, 40, 35, 55

Calculate the arithmetic mean.

SOLUTION

$$\bar{X} = \frac{\Sigma X}{n} \qquad ,$$

$$= \frac{70 + 50 + 40 + 35 + 55}{5}$$

$$= \frac{250}{5} = 50 \text{ Marks}$$

3.2.1 Mean from Group Data

When the number of observations are very large, they are grouped into a frequency distribution.

Let $X_1, X_2, \ldots\ldots X_n$ be the mid points of different class intervals and let $f_1, f_2, \ldots\ldots, f_n$ be their corresponding frequencies, then arithmetic mean is given by

$$\text{A.M.} = \frac{f_1 X_1 + f_2 X_2 + \ldots\ldots + f_n X_n}{f_1 + f_2 + \ldots\ldots + f_n}$$

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

EXAMPLE 3.2

The marks of 100 students in statistics are given below. Calculate the arithmetic mean.

Marks	30-35	35-40	40-45	45-50	50-55	55-60
No. of Stud.	14	16	18	23	18	11

SOLUTION

Marks	No. of Students (f)	Mid Points (x)	fx
30 - 35	14	32.5	455.00
35 - 40	16	37.5	600.00
40 - 45	18	42.5	765.00
45 - 50	23	47.5	1092.50
50 - 55	18	52.5	945.00
55 - 60	11	57.5	623.50
Σ	100	—	4490

$$\bar{X} = \frac{\Sigma fX}{\Sigma f}$$

$$= \frac{4490}{100} = 44.90$$

$$\bar{X} = 45 \text{ marks}$$

3.2.2 Mean from Frequency Distribution (Discrete Data)

In discrete series the values of each of the observation is multiplied by the corresponding frequency. These products are added. This sum is divided by the total frequency.

Let X_1, X_2, ... X_n be the different values and let f_1, f_2, ... f_n are corresponding frequencies, then mean is defined as

$$\bar{X} = \frac{f_1X_1 + f_2X_2 + \text{.........} + f_nX_n}{f_1 + f_2 + \text{......} + f_n}$$

$$\bar{X} = \frac{\Sigma fX}{\Sigma f}$$

EXAMPLE 3.3

Find the arithmetic mean from the following frequency distribution.

x	2	3	4	5	6
f	5	7	8	3	2

SOLUTION

x	f	fx
2	5	10
3	7	21
4	8	32
5	3	15
6	2	12
Σ	25	90

$$\bar{X} = \frac{\Sigma fX}{\Sigma f} = \frac{90}{25} = 3.6 \approx 4$$

EXAMPLE 3.4

The following table gives the marks obtained by a batch of 5 candidates in an examination in History, Statistics and Economics.

Roll No.	History	Statistics	Economics
1	41	46	50
2	35	50	52
3	38	39	41
4	34	50	46
5	30	38	39

In which subject is the level of knowledge highest?

SOLUTION

Roll No.	History	Statistics	Economics
1	41	46	50
2	35	50	52
3	38	39	41
4	34	50	46
5	30	38	39
Σ	178	223	228

$$\text{Mean marks in History} = \bar{X}_h = \frac{\Sigma X}{n}$$

$$= \frac{178}{5} = 35.6$$

Mean marks in Statistics $= \bar{X}_s = \frac{\Sigma X}{n}$

$$= \frac{223}{5} = 44.6$$

Mean marks in Economics $= \bar{X}_e = \frac{\Sigma X}{n}$

$$= \frac{228}{5} = 45.6$$

Mean marks in Economics are higher than Statistics and History. So the level of knowledge in Economics is highest.

EXAMPLE 3.5

Which class is better on the average?

Marks	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
Class A	100	125	86	45	18°	12
Class B	90	140	75	50	15	10

SOLUTION

The given data is group data first of all we calculate the average marks of class A and B separately.

Average marks of class A

Marks	Frequency (f)	X	fx
10 - 20	100	15	1500
20 - 30	125	25	3125
30 - 40	86	35	3012
40 - 50	45	45	2025
50 - 60	18	55	990
60 - 70	12	65	780
Σ	386	—	11430

$$\bar{X}_a = \frac{\Sigma fX}{\Sigma f}$$

$$= \frac{11430}{386} = 29.61$$

Average Marks of Class B

Marks	Frequency (f)	X	fx
10 - 20	90	15	1350
20 - 30	140	25	3500
30 - 40	75	35	2625
40 - 50	50	45	2250
50 - 60	15	55	825
60 - 70	10	65	650
Σ	380	—	11200

$$\bar{X}_2 = \frac{\Sigma fX}{\Sigma f}$$

$$= \frac{11200}{380} = 29.47$$

Average marks of class A is more than Class B. So class A is **better** on the average.

3.2.3 Shortcut Method for Computing Mean

$$\bar{X} = A + \frac{\Sigma D}{n} \qquad \text{(For Discrete Values)}$$

Where $D = X - A$

$\qquad A$ = Assumed mean

$$\bar{X} = A + \frac{\Sigma fD}{\Sigma f} \qquad \text{(Group Data)}$$

Where $D = X - A$

$\qquad \Sigma f$ = Total frequency

EXAMPLE 3.6

Calculate arithmetic mean by using shortcut method.

20, 25, 35, 45, 60

SOLUTION

Let $D = X - A$

$D = X - 35$

X	$D = X - A$ $(X - 35)$
20	-15
25	-10
	$\underline{-25}$
35	0
45	10
60	25
	$\underline{35}$
Σ	10

$$\bar{X} = A + \frac{\Sigma D}{n}$$

$$= 35 + \frac{10}{5}$$

$$= 35 + 2 = 37$$

EXAMPLE 3.7

The scores of students in a cricket tournament are given below. Calculate the arithmetic mean by using shortcut method.

Scores	0 - 9	10 - 19	20 - 29	30 - 39	40 - 49
No. of Students	5	15	12	10	8

SOLUTION

Scores	f	X	$D = X - A$ $(x - 24.5)$	f.D
0 - 9	5	4.5	−20	−100
10 - 19	15	14.5	−10	−150
				−250
20 - 29	12	24.5	0	0
30 - 39	10	34.5	10	100
40 - 49	8	44.5	20	160
				260
Total	50	—	0	10

$$\bar{X} = A + \frac{\Sigma fD}{\Sigma f} \qquad \text{Let A} = 24.5$$

$$= 24.5 + \frac{10}{50} = 24.5 + 0.20$$

$$= 24.70 \approx 25 \text{ scores}$$

3.2.4 Step Deviation or Coding Method for Computing Mean

It is a very short method and should always be used for group data where class interval sizes are equal. The formula for computing mean is given below.

$$\bar{X} = a + \frac{\Sigma fU}{\Sigma f} \times h$$

Where a = Assumed mean

$$U = \frac{X - a}{h}$$

h = Class interval size.

EXAMPLE 3.8

Calculate A.M. from the following distribution by coding method.

Groups	1.5-2.0	2.0-2.5	2.5-3.0	3.0-3.5	3.5-4.0	4.0-4.5
Freq.	3	7	10	15	9	6

SOLUTION

Group	f	X	$u = \dfrac{X - 2.75}{0.5}$	fu
1.5 - 2.0	3	1.75	−2	−6
2.0 - 2.5	7	2.25	−1	−7
				$\overline{-13}$
2.5 - 3.0	10	2.75	0	0
3.0 - 3.5	15	3.25	1	15
3.5 - 4.0	9	3.75	2	18
4.0 - 4.5	6	4.25	3	18
				$\overline{51}$
Total	50	—	—	38

$$\bar{X} = a + \frac{\Sigma fu}{\Sigma f} \times h \qquad \text{Let } u = \frac{X - a}{h}$$

$$= 2.75 + \frac{38}{50} \times 0.5 \qquad a = 2.75$$

$$= 2.75 + \frac{19}{50} \qquad h = 0.5$$

$$= 2.75 + 0.38 = 3.13$$

3.2.5 Weighted Arithmetic Mean

Simple arithmetic mean gives equal importance to all the observations of the date, when the observations are not of equal importance, we assign them weights according to their relative importance.

Let $x_1, x_2, \ldots, x_n$ be 'n' observations with corresponding weights $w_1, w_2, \ldots, w_n$ respectively, then weighted mean is denoted by $\bar{X}_w$ and defined as

$$\bar{X}_w = \frac{w_1 x_1 + w_2 x_2 + \ldots + w_n x_n}{w_1 + w_2 + \ldots + w_n}$$

$$= \frac{\Sigma wx}{\Sigma w}$$

EXAMPLE 3.9

A student obtained the following marks at a certain examination English = 52, Urdu = 73, Physics = 84, Chemistry = 65 and Biology = 79. Find the weighted mean if weights of 2, 1, 3, 3 and 4 respectively are allotted to the subjects.

SOLUTION

Let   X = Marks obtained

Subjects	x	w	wx
English	52	2	104
Urdu	73	1	73
Physics	84	3	252
Chemistry	65	3	195
Biology	79	4	316
Total	—	13	940

$$\bar{X}_w = \frac{\Sigma wx}{\Sigma w}$$

$$= \frac{940}{13} = 72.3 \approx 72 \text{ marks}$$

3.2.6 Properties of Arithmetic Mean

(i)   The sum of deviations of the values X_i from their mean $\bar{X}$ is zero.

$\Sigma(X_i - \bar{X}) = 0$ (Ungroup Data)

$\Sigma f(X_i - \bar{X}) = 0$ (Group Data)

(ii)   The sum of the squared deviations of values X_i from their mean is minimum.

i.e. $\Sigma(X_i - \bar{X})^2 \leq \Sigma(X_i - A)^2$ (Ungroup Data)

$\Sigma f(X_i - \bar{X})^2 \leq \Sigma f(X_i - A)^2$ (Group Data)

Where A is any value other than mean.

This is called minimal property of mean.

(iii)   If $\bar{X}_1$, $\bar{X}_2$, ..., $\bar{X}_k$ be the means of k groups with respective frequencies n_1, n_2,
         ......, n_k. Then the combined mean $\bar{X}_c$ for the whole distribution is given by

$$\bar{X}_c = \frac{n_1\bar{X}_1 + n_2\bar{X}_2 + ... + n_k\bar{X}_k}{n_1 + n_2 + ... + n_k}$$

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

(iv)   If $Y_i = ax_i + b$ (i = 1, 2, 3, ..., n)

         then $\bar{Y} = a\bar{X} + b$

         Where a and b are constants.

EXAMPLE 3.10

A distribution consists of three components with sizes 100, 150 and 200
having their means 16, 19 and 22 respectively. Find the combined mean.

SOLUTION

We have given that

$n_1 = 100$, $n_2 = 150$, $n_3 = 200$

$\bar{X}_1 = 16$, $\bar{X}_2 = 19$, $\bar{X}_3 = 22$

Combined mean is given by

$$\bar{X}_c = \frac{n_1\bar{X}_1 + n_2\bar{X}_2 + n_3\bar{X}_3}{n_1 + n_2 + n_3}$$

$$= \frac{100(16) + 150(19) + 200(22)}{100 + 150 + 200}$$

$$= \frac{8850}{450} = 19.67$$

3.2.7 When to Use Arithmetic Mean

We use arithmetic mean, when we are required to study social, economic and
commercial problems like production, price, export and import. It helps in getting
average income, average price, average production etc.

3.2.8 Advantages and Disadvantages of Arithmetic Mean

Advantages

(i)   It is easy to calculate and simple to follow.

(ii)   It is based on all the observations.

(iii) It can be determined for almost every kind of data.

(iv) It is commonly used average.

(v) It provides a good basis for comparison.

Disadvantages

(i) It is highly effected by extreme values.

(ii) It can not be accurately calculated for open end frequency distribution.

(iii) It can not be calculated accurately if any observation is missing.

3.3 MEDIAN

Median is the middle most value of a set of data when the data is arranged in order of magnitude.

If the number of observations in the array is odd, then median is the middle value and if the number of observations in the array is even, then median is the average of two middle values. It is denoted by $\tilde{X}$.

Mathematically

Median $= $ Value of $\left(\dfrac{n+1}{2}\right)$ the item.

EXAMPLE 3.11

The scores of a cricket player in 7 matches is given below. Calculate median.

15, 45, 10, 16, 40, 19, 32

SOLUTION

Arranging the score in order we have

10, 15, 16, 19, 32, 40, 45

Median $= $ Value of $\left(\dfrac{n+1}{2}\right)$ th item

Here $n = 7$, i.e. odd therefore

$= $ value of $\left(\dfrac{n+1}{2}\right)$ th item

$= $ value of 4th item

4th item corresponds to 19. Therefore Median $= 19$.

EXAMPLE 3.12

Compute the median form the following data.

12, 13, 37, 14, 57, 48, 29, 27

SOLUTION

After arranging the data in ascending order we have

12, 13, 14, 27, 29, 37, 48, 57

Here $n = 8$, i.e. even, therefore

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

$$= \text{value of} \left(\frac{8 + 1}{2}\right) \text{th item}$$

$$= \text{value of 4.5 th item}$$

$$= \frac{\text{4th item} + \text{5th item}}{2}$$

$$= \frac{27 + 29}{2} = \frac{56}{2} = 28$$

Alternative Method

After arranging the data in ascending order, we have

12, 13, 14, 27, 29, 37, 48, 57 $n = 8$

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

$$= \text{Value of} \left(\frac{8 + 1}{2}\right) \text{th item.}$$

$$= \text{value of 4.5 th item}$$

$$= \text{the value of 4th item} + 0.5 \text{ (The value of 5th item} - \text{value of 4th item)}$$

$$\text{Median} = 27 + 0.5 (29 - 27)$$

$$= 27 + 0.5(2)$$

$$= 27 + 1 = 28$$

3.3.1 Median for Discrete Frequency Distribution

Let X_1, X_2, ... X_n be the different values and let f_1, f_2, ..., f_n are corresponding frequencies. First of all we calculate the cumulative frequencies and then we see the median number $\left(\frac{n + 1}{2}\right)$ under the cumulative frequency column. the item, which corresponds to the median number is called median.

Mathematically

$$\text{Median} = \text{Value corresponding to} \left(\frac{n + 1}{2}\right) \text{th cumulative frequency.}$$

EXAMPLE 3.13

The following table shows the number of heads in an experiment of 5 coins 100 times.

No. of heads	0	1	2	3	4	5
Freq.	10	25	30	20	10	5

Calculate median.

SOLUTION

X	f	C.f
0	10	10
1	25	25 + 10 = 35
2	30	30 + 25 = 65 ⟶ MEDIAN GROUP
3	20	20 + 65 = 85
4	10	10 + 85 = 95
5	5	5 + 95 = 100

$\Sigma f = 100 = n$

$$\text{Median} = \text{Value corresponding} \left(\frac{n+1}{2}\right) \text{th C.f}$$

$$= \text{Value corresponding to} \left(\frac{100+1}{2}\right) \text{th C.f}$$

$$= \text{Value of 50.5 th item}$$

50.5 th item corresponds to 2, therefore Median = 2

3.3.2 Median In Case Of Continuous Frequency Distribution (Group Data)

In case of group data, we form the cumulative frequencies and then calculate the median number $\left(\frac{n}{2}\right)$. The group, which corresponds to median number, is called the median group. The median lies in this group. Before calculating the median we convert the class limits into class boundaries if these are not given.

Median is given by the formula

$$\text{Median} = L + \frac{h}{f}\left(\frac{n}{2} - c\right)$$

Where

Median class = value of $\left(\frac{n}{2}\right)$ th item

L = Lower class boundary of the median class

h = Size of the class interval

f = Frequency of the median class

n = Σf = Total frequency

c = Cumulative frequency preceding the median class

EXAMPLE 3.14

The heights (in inches of 40 students of II years class are given below. Calculate the median.

Height (inch)	54 - 56	56 - 58	58 - 60	60 - 62	62 - 64	64 - 66
No. of Student	5	7	10	9	6	3

SOLUTION

Height	f	C.f	
54 - 56	5	5	
56 - 58	7	12	
58 - 60	10	22	→ Median Class
60 - 62	9	31	
62 - 64	6	37	
64 - 66	3	40	

$$\text{Median} = \text{Value of}\left(\frac{n}{2}\right) \text{th item}$$

$$= \text{Value of}\left(\frac{40}{2}\right) \text{th item}$$

$$= \text{Value of 20th item}$$

The 20th item lies in the class 58 - 60. So 58 - 60 is the median class.

Therefore

$$\text{Median} = L + \frac{h}{f}\left(\frac{n}{2} - C\right)$$

Where L = 58, h = 2, f = 10, n = 40, c = 12

$$\text{Median} = 58 + \frac{2}{10}\left(\frac{40}{2} - 12\right)$$

$$= 58 + \frac{2}{10}(20 - 12)$$

$$= 58 + \frac{2}{10}(8) = 58 + \frac{16}{10}$$

$$= 58 + 1.6 = 59.6 \text{ inch}$$

EXAMPLE 3.15

Calculate median from the following frequency distribution.

Classes	10 - 19	20 - 29	30 - 39	40 - 49	50 - 59	60 - 69
f	10	15	26	24	13	• 12

SOLUTION

Classes	C.B.	f	C.f	
10 – 19	9.5 - 19.5	10	10	
20 – 29	19.5 - 29.5	15	25	
30 – 39	29.5 - 39.5	26	51	→ Median Class
40 – 49	39.5 - 49.5	24	75	
50 – 59	49.5 - 59.5	13	88	
60 – 69	59.5 - 69.5	12	100	

$$\text{Median} = \text{Value of}\left(\frac{n}{2}\right)\text{th item}$$

$$= \text{Value of}\left(\frac{100}{2}\right)\text{th item}$$

$$= \text{Value of 50 the item}$$

The 50th item lies in the group class 29.5 – 39.5. Therefore 29.5 – 39.5 is the median class.

$$\text{Median} = L + \frac{h}{f}\left(\frac{n}{2} - C\right)$$

Where L = 29.5, h = 10, f = 26, n = 100, c = 25

$$\text{Median} = 29.5 + \frac{10}{26}\left(\frac{100}{2} - 25\right)$$

$$= 29.5 + \frac{10}{26}(50 - 25)$$

$$= 29.5 + \frac{10}{26}(25)$$

$$= 29.5 + \frac{250}{26}$$

$$= 29.5 + 9.62 = 39.12$$

3.3.3 Graphic Location of Median

Median can be located by the graph of a cumulative frequency polygon (Ogive).

First of all a cumulative frequency curve is drawn for the given data. The values of the upper class boundaries are taking along x-axis and cumulative frequency along y-axis. Then the position of the median is located by the formula.

$$\text{Median} = \left(\frac{n}{2}\right) \text{th cumulative frequency}$$

A horizontal line is drawn along x-axis corresponding to the $\left(\frac{n}{2}\right)$ th position. That line intersects the ogive at certain point. Then from that point a perpendicular is drawn along x-axis which touches the x-axis on a certain point. This point on x-axis is the required value of the median.

EXAMPLE 3.16

Draw a cumulative frequency curve (Ogive) and locate median.

Marks.	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
Freq.	5	10	20	10	5

Check your result by actual calculations.

SOLUTION
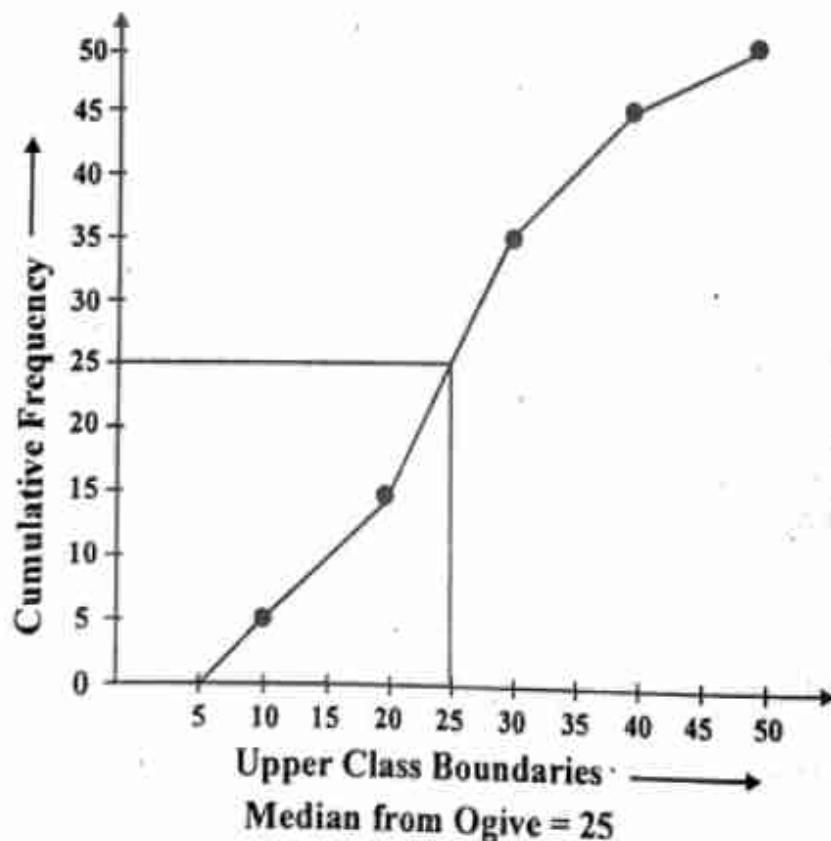
Marks	f	C.f	upper class boundaries
0 - 10	5	5	10
10 - 20	10	15	20
20 - 30	20	35	3
30 - 40	10	45	40
40 - 50	5	50	50

Median = Value corresponding to $\left(\dfrac{n}{2}\right)$ th c.f

= Value corresponding to $\left(\dfrac{50}{2}\right)$ th c.f

= Value corresponding to 25th c.f.

OGIVE

Median from Ogive = 25

Check

$$\text{Median} = \text{Value of the } \left(\frac{n}{2}\right) \text{th item}$$

$$= \frac{50}{2} \text{ th item} = \text{25th item}$$

The 25th item lies in the class 20 – 30. So 20 – 30 is the median class.

$$\text{Median} = L + \frac{h}{f}\left(\frac{n}{2} - C\right)$$

Where L = 20, h = 10, f = 20, n = 50, c = 15

$$\text{Median} = 20 + \frac{10}{20}\left(\frac{50}{2} - 15\right)$$

$$= 20 + \frac{10}{20}(25 - 15)$$

$$= 20 + \frac{10}{20}(10)$$

$$= 20 + \frac{10}{20}$$

$$= 20 + 5 = 25$$

This is as accurate as we can get with the graph.

3.3.4 Quantiles

When the number of observations is sufficiently large, the principle by which a distribution is divided into two equal parts may be extended to divide the distribution into four, ten, or hundred equal parts.

3.3.5 Quartiles

Quartiles are the values which divide the set of data into four equal parts. These are the first quartile Q_1, second quartile Q_2 (median) and third quartile Q_3.

The formula for the position of the quartiles are

$$Q_1 = \frac{n+1}{4} \text{ the value}$$

$$Q_2 = \frac{2(n+1)}{4} \text{ the value}$$

$$Q_3 = \frac{3(n+1)}{4} \text{ th value}$$

For the grouped data the quartiles are calculated as

$$Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - c\right)$$

$$Q_2 = l + \frac{h}{f}\left(\frac{2n}{4} - c\right)$$

$$Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - c\right)$$

3.3.6 Deciles

Deciles are the values which divide the set of data into 10 equal parts. These are denoted by $D_1, D_2, D_3, \ldots, D_9$. The formula for the position of deciles is

$$D_1 = \frac{n + 1}{10} \text{ th value}$$

$$D_2 = \frac{2(n + 1)}{10} \text{ the value}$$

$$D_3 = \frac{3(n + 1)}{10} \text{ the value}$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$D_9 = \frac{9(n + 1)}{10} \text{ the value}$$

In case of grouped data, the deciles are calculated as

$$D_1 = l + \frac{h}{f}\left(\frac{n}{10} - c\right)$$

$$D_2 = l + \frac{h}{f}\left(\frac{2n}{10} - c\right)$$

$$D_3 = l + \frac{h}{f}\left(\frac{3n}{10} - c\right)$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$D_9 = l + \frac{h}{f}\left(\frac{9n}{10} - c\right)$$

3.3.7 Percentiles

Percentiles are the value which divides the set of data into 100 equal parts. These are denoted by $P_1, P_2, \ldots, P_{99}$.

The formula for the position value of percentiles are

$$P_1 = \frac{n+1}{100} \text{ th value}$$

$$P_2 = \frac{2(n+1)}{100} \text{ th value}$$

$$P_3 = \frac{3(n+1)}{100} \text{ th value}$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$P_{99} = \frac{99(n+1)}{100} \text{ th value}$$

In case of grouped data, the percentiles are calculated as

$$P_1 = l + \frac{h}{f}\left(\frac{n}{100} - c\right)$$

$$P_2 = l + \frac{h}{f}\left(\frac{2n}{100} - c\right)$$

$$P_3 = l + \frac{h}{f}\left(\frac{3n}{100} - c\right)$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$P_{99} = l + \frac{h}{f}\left(\frac{99n}{100} - c\right)$$

EXAMPLE 3.17

From the data given below.

25, 40, 32, 62, 56, 38, 23, 63, 56, 42, 35, 50, 39, 47, 53

Find   (i) Q_1 (ii) Q_2 (iii) Q_3 (iv) D_3

            (v) D_8 (vi) P_{20} (vii) P_{80}

SOLUTION

Arranging the data in an order, we get

23, 25, 32, 35, 38, 39, 40, 42, 47, 50, 53, 56, 56, 62, 63

(i) $Q_1 = \dfrac{(n+1)}{4}$ the value

     $= \dfrac{(15+1)}{4}$ the value $= 4$th value

   $Q_1 = 35$

(ii) $Q_2 = \frac{2(n+1)}{4}$ the value

$$= \frac{2(15+1)}{4} \text{ the value} = \frac{32}{4} \text{ th value} = \text{8th value}$$

$$Q_2 = 42$$

(iiii) $Q_3 = \frac{3(n+1)}{4}$ the value

$$= \frac{3(15+1)}{4} \text{ the value} = \frac{48}{4} \text{ th value} = 12 \text{ th value}$$

$$Q_3 = 56$$

(iv) $D_3 = \frac{3(n+1)}{10}$ the value

$$= \frac{3(15+1)}{10} \text{ the value} = \frac{48}{10} \text{ th value}$$

$$= \text{4th value} + 0.8 (5^{th} - 4^{th} \text{ value})$$

$$= 35 + 0.8 (38 - 35)$$

$$= 35 + 0.8(3) = 37.4$$

$$D_3 = 4.8\text{th value}$$

(v) $D_8 = \frac{8(n+1)}{10}$ the value

$$= \frac{8(15+1)}{10} \text{ the value} = 12.8 \text{ th value}$$

$$D_8 = 12.8 \text{ th value}$$

$$= 12\text{th value} + 0.8 (13^{th} - 12^{th} \text{ value})$$

$$= 56 + 0.8 (56 - 56)$$

$$= 35 + 0.8(0) = 56$$

(vi) $P_{20} = \frac{20(n+1)}{100}$ the value

$$= \frac{20(15+1)}{100} \text{ the value} = \frac{320}{100} \text{ th value}$$

$$= 3.2\text{th value}$$

$$= 3\text{rd value} + 0.2 (4\text{th} - 3\text{rd value})$$

$$= 32 + 0.2(35 - 32)$$

$$= 32 + 0.2(3)$$

$$= 32 + 0.6 = 32.6$$

(vii) $P_{80} = \dfrac{80\,(n+1)}{100}$ the value

$= \dfrac{80\,(15+1)}{100}$ the value

$= 12.8$ th value

$= 12$th value $+ 0.8\,(13$th $- 12$th value)

$= 56 + 0.8(56 - 56) = 56$

EXAMPLE 3.18

Marks of students in Mathematics are given below.

Marks	0 – 9	10 – 19	20 – 29	30 – 39
Frequency	2	5	10	20
Marks	40 – 49	50 – 59	60 – 69	70 – 79
Frequency	15	12	8	3

Calculate

(i) Q_1 (ii) Q_2 (iii) Q_3

(iv) D_3 (v) D_8 (vi) P_{30}

(vii) P_{75}

SOLUTION

Marks	f	Class boundaries	c.f
0 – 9	2	0 – 9.5	2
10 – 19	5	9.5 – 19.5	7
20 – 29	10	19.5 – 29.5	17
30 – 39	20	29.5 – 39.5	37
40 – 49	15	39.5 – 49.5	52
50 – 59	12	49.5 – 59.5	64
60 – 69	8	59.5 – 69.5	72
70 – 79	3	69.5 – 79.5	75
Total	75	—	—

(i) $Q_1 = 1 + \dfrac{h}{f}\left(\dfrac{n}{4} - c\right)$ ..... (i)

$Q_1 = \dfrac{n}{4}$ th value $= \dfrac{75}{4}$ th value $= 18.75$ th value

18.75 value lies in the group 29.5 – 39.50.

Therefore, $1 = 29.5$, $h = 10$, $f = 20$, $n = 75$, $c = 17$

Putting the values in (i), we have

$Q_1 = 29.5 + \dfrac{10}{20}(18.75 - 17)$

$= 29.5 + \dfrac{1}{2}(1.75) = 30.38$

(ii) $Q_2 = 1 + \dfrac{h}{f}\left(\dfrac{2n}{4} - c\right) = $ Median ...... (ii)

$Q_2 = \dfrac{2n}{4}$ th value $= \dfrac{2(75)}{4} = 37.5$ th value

$= 37.5$th value

37.5th value lies in the group 39.5 – 49.5.

Therefore $1 = 39.5$, $h = 10$, $f = 15$, $c = 37$

Putting the value in (ii), we have

$Q_2 = 39.5 + \dfrac{10}{15}(37.5 - 37)$

$= 39.5 + \dfrac{10}{15}(0.5)$

$= 39.5 + 0.33 = 39.83$

(iii) $Q_3 = 1 + \dfrac{h}{f}\left(\dfrac{3n}{4} - c\right)$ .... (iii)

$Q_3 = \dfrac{3n}{4}$ th value

$= \dfrac{3(75)}{4}$ th value $= 56.25$ th value

56.25 th value lies in the group 49.5 – 59.5

So $1 = 49.5$, $h = 10$, $f = 12$, $c = 52$

Putting the values in (iii), we get

$Q_3 = 49.5 + \dfrac{10}{12}(56.25 - 52)$

$= 49.5 + \dfrac{10}{12}(4.25)$

$$= 49.5 + 3.54$$

$$= 53.04$$

(iv) $D_3 = 1 + \dfrac{h}{f}\left(\dfrac{3n}{10} - c\right)$ .... (iv)

$D_3 = \dfrac{3n}{10}$ th value

$\qquad = \dfrac{3 \times 75}{10}$ th value $= 22.5$ th value

22.5 th value lies in the group $29.5 - 39.5$
So $1 = 29.5,\ h = 10,\ f = 20,\ c = 17$
Putting the values in (iv), we get

$D_3 = 29.5 + \dfrac{10}{20}(22.5 - 17)$

$\qquad = 29.5 + \dfrac{1}{2}(5.5)$

$\qquad = 29.5 + 2.75 = 32.25$

(v) $D_8 = 1 + \dfrac{h}{f}\left(\dfrac{8n}{10} - c\right)$ .... (v)

$D_8 = \dfrac{8n}{10}$ th value

$\qquad = \dfrac{8 \times 75}{10}$ th value $= 60$ th value

60 th value lies in the group $49.5 - 59.5$
So $1 = 49.5,\ h = 10,\ f = 12,\ c = 52$
Putting the values in (v), we get

$D_8 = 49.5 + \dfrac{10}{12}(60 - 52)$

$\qquad = 49.5 + \dfrac{10}{12}(8)$

$\qquad = 49.5 + 6.67 = 56.17$

(vi) $P_{30} = 1 + \dfrac{h}{f}\left(\dfrac{30n}{100} - c\right)$ .... (vi)

$P_{30} = \dfrac{30n}{100}$ th value

$\qquad = \dfrac{30 \times 75}{100}$ th value

$\qquad = 22.5$ th value

22.5 th value lies in the group $29.5 - 39.5$

So $1 = 29.5$, $h = 10$, $f = 20$, $c = 17$

$P_{30} = 29.5 + \dfrac{10}{20}(22.5 - 17)$

$\quad = 29.5 + \dfrac{1}{2}(5.5)$

$\quad = 29.5 + 2.75 = 32.25$

(vii) $P_{75} = 1 + \dfrac{h}{f}\left(\dfrac{75n}{100} - c\right)$ ...... (vii)

$P_{75} = \dfrac{75n}{100}$ th value

$\quad = \dfrac{75 \times 75}{100}$ th value

$\quad = 56.25$ th value

56.25 th value lies in the group $49.5 - 59.5$

So $1 = 49.5$, $h = 10$, $f = 12$, $c = 52$

$P_{75} = 49.5 + \dfrac{10}{12}(56.25 - 52)$

$\quad = 49.5 + \dfrac{10}{12}(4.25)$

$\quad = 49.5 + 3.54 = 53.04$

3.3.8 When we Apply Median

We apply median to the situations, when the direct measurements of the variables are not possible like poverty, beauty and intelligence etc.

3.3.9 Advantages and Disadvantages of Median

Advantages

(i)   It is easy to calculate and understand.

(ii)   It is not effected by extreme values.

(iii)   It can be computed even in open end frequency distribution.

(iv)   It can be used for qualitative data.

(v)   It can be located graphically.

Disadvantages

(i)   It is not rigorously defined.

(ii)   It is not based on all the observations.

(iii)   It is not suitable for further algebraic treatment.

3.4 THE MODE

Mode is the value, which occurs <u>maximum number of times</u> in a set of data.

Or

It is the value which occurs most frequently in a set of data. A set of data may have more than one mode or no mode.

Mode = $\hat{X}$ = The value appearing maximum no. of times.

EXAMPLE 3.19

Calculate the mode for the following data.

(i)   10, 12, 6, 8, 7, 15, 10

(ii)   2, 7, 5, 2, 6, 9, 10, 6

(iii)   50, 60, 70, 60, 80, 60, 50

(iv)   20, 25, 28, 12, 18, 23

(v)   Letters of PAKISTAN.

(vi)   The days of the week for marriages in Pakistan.

(vii)   Shoe sizes sold at a store

         32, 34, 32, 35, 36, 35, 38, 40, 36, 35

SOLUTION

(i)   Mode = 10, because it is repeated 2 times.

(ii)   There are two modes

         i.e.   2 and 6, because both are repeated 2 times.

(iii)   Mode = 60, because it is repeated 3 times.

(iv)   There is no mode, because all the values occur only once.

(v)   Modal letter = A, because it is repeated 2 times.

(vi)   Modal day = Sunday, because maximum number of marriages are held on Sunday.

(vii)   Mode = 35, because it is repeated 3 times.

EXAMPLE 3.20

Calculate A.M. and Mode from the following values.

48, 34, 69, 25, 22, 34, 56, 68

SOLUTION

$$\bar{X} = \frac{\Sigma x}{n}$$

$$= \frac{48 + 34 + 69 + 25 + 22 + 34 + 56 + 68}{8}$$

$$= \frac{356}{8} = 44.50$$

Mode

Mode = 34, because it is repeated 2 times.

Mode for Discrete Frequency Distribution

Mode = The value corresponding to maximum frequency

EXAMPLE 3.21

The following are the sizes of shoes sold at a Bata shop. Find the modal size of the shoes.

Size of the shoes	4	5	6	7	8	9	10
Freq.	10	12	18	10	25	7	2

SOLUTION

Size of the Shoe	Frequency
4	10
5	12
6	18
7	10
8	25
9	7
10	2

Mode = The value corresponding to maximum frequency

Mode = 8   *i.e.* Modal size of the shoes is 8.

3.4.1 Mode in case of Group Data or Mode for Continuous Frequency Distribution

In case of group data, the mode lies in the class, which has maximum frequency. This class is called modal class. The mode is given by the formula

$$\text{Mode} = L + \frac{(f_m - f_1)}{(f_m - f_1) + (f_m - f_2)} \times h$$

Where

L = Lower class boundary of the modal class.

f_m = Frequency of the modal class.

f_1 = Frequency of the class preceding the modal class.

f_2 = Frequency of the class following the modal class

h = Size of the class interval

Modal class = A class which has maximum frequency.

EXAMPLE 3.22

Calculate the mode form the following frequency distribution.

Classes	0 - 8	8 - 16	16 - 24	24 - 32	32 - 40
Frequency	5	7	12	8	6

SOLUTION

Classes	Frequency (f)
0 - 8	5
8 - 16	7 = f_1
16 - 24	12 = f_m
24 - 32	8 = f_2
32 - 40	6

$$\text{Mode} = L + \frac{(f_m - f_1)}{(f_m - f_1) + (f_m - f_2)} \times h$$

Modal class is 16 - 24, because it has maximum frequency.

Therefore

$l = 16$, $f_m = 12$, $f_1 = 7$, $f_2 = 8$, $h = 8$

$$\text{Mode} = 16 + \frac{(12 - 7)}{(12 - 7) + (12 - 8)} \times 8$$

$$= 16 + \frac{5}{5 + 4} \times 8$$

$$= 16 + \frac{5}{9} \times 8$$

$$= 16 + \frac{40}{9}$$

$$= 16 + 4.44 = 20.44$$

3.4.2 Graphic Location of Mode

Mode can be located graphically by the following process.
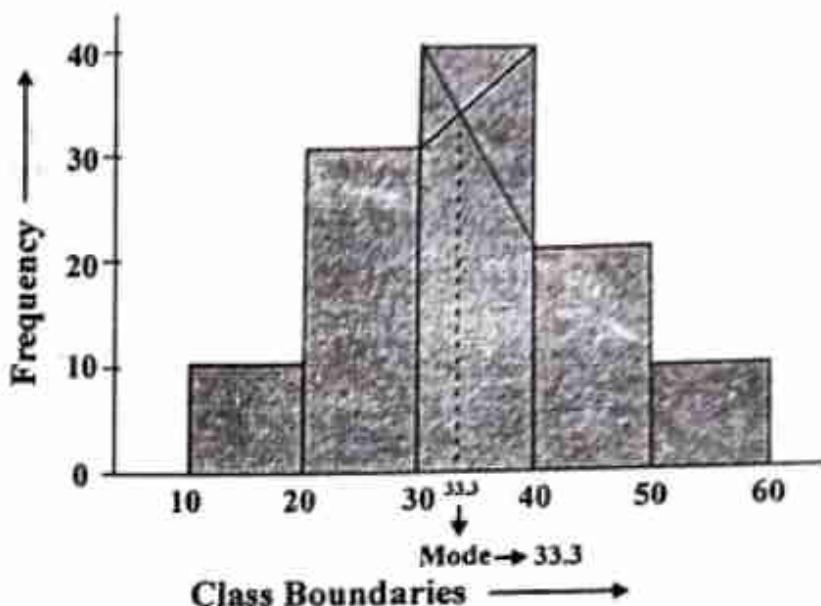
(i)   Draw histogram for the given data, taking class boundaries along x-axis and frequencies along y-axis.

(ii)   The highest rectangle represents the modal class.

(iii)   The top right and left corners of the modal class rectangle are joined with the top right and left corner of the proceeding rectangles.

(iv)   From the point of intersection of both the lines a perpendicular on the x-axis is drawn. The point where the perpendicular meets the x-axis, will be the required value of the mode.

EXAMPLE 3.23

From the following frequency distribution.

Groups	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
f	10	30	40	20	10

Find approximate value of the mode from the histogram and check your answer by calculation.

SOLUTION

HISTOGRAM

Mode → 33.3

Class Boundaries ⟶

Check by Calculations

Groups	f
10 – 20	10
20 – 30	$30 = f_1$
30 – 40	$40 = f_m$
40 – 50	$20 = f_2$
50 – 60	10

Applying the formula

$$\text{Mode} = L + \frac{(f_m - f_1)}{(f_m - f_1) + (f_m - f_2)} \times h$$

Modal class is 30 – 40 therefore

$L = 30,\ f_m = 40,\ f_1 = 30,\ f_2 = 20,\ h = 10$

$$\text{Mode} = 30 + \frac{(40 - 30)}{(40 - 30) + (40 - 20)} \times 10$$

$$= 30 + \frac{10}{10 + 20} \times 10$$

$$= 30 + \frac{100}{30}$$

$$= 30 + 3.33 = 33.33$$

The value of the mode from graphic method is also 33.3.

3.4.3 When to apply Mode

We apply mode when it is required to study the problems like average size of shoes, average size of ready made garments, average size of agriculture holding etc. This average is widely used in Biology and Meteorology.

3.4.4 Advantages and Disadvantages of Mode

Advantages

(i)   It is easy to understand.

(ii)   It is not effected by extreme values.

(iii)   It can be computed even in open-end classes.

(iv)   It can be useful in qualitative data.

Disadvantages

(i)   It is not clearly defined.

(ii)   It is not suitable for further algebraic treatment.

(iii)   It is not based on all the observations.

(iv)   It may not exist in some cases.

3.4.5 Empirical Relation Between Mean, Median and Mode

In a symmetrical distribution, mean, median and mode are coincide.

i.e.   Mean = Median = Mode

In a moderately skewed distribution

Mode = 3 Median − 2 Mean.

EXAMPLE 3.24

In a moderately skewed distribution

(i)   Mean = 15, Median = 16

Find the value of Mode.

(ii)   Median = 65 and Mode = 85

Find the value of Mean.

SOLUTION

(i) Mode is given by the relation

Mode = 3 median − 2 mean

Mean = 15

Median = 16

Mode = 3(16) − 2(15)

= 48 − 30 = 18

(ii) Median = 65, Mode = 85

Mean = ?

Mode = 3 Median − 2 Mean

85 = 3(65) − 2 Mean

85 = 195 − 2 Mean

2 Mean = 195 − 85

2 Mean = 110

$\therefore$ Mean = $\dfrac{110}{2}$ = 55

EXAMPLE 3.25

Compute mean, median and mode for the data given below.

Classes	f	Classes	f
0.7312 – 0.7313	10	0.7320 – 0.7321	30
0.7314 – 0.7315	15	0.7322 – 0.7323	8
0.7316 – 0.7317	20	0.7324 – 0.7325	2
0.7318 – 0.7319	25		

SOLUTION

(i) **Arithmetic Mean**

Classes	f	x	fx
0.7312 – 0.7313	10	0.73125	7.31250
0.7314 – 0.7315	15	0.73145	10.97175
0.7316 – 0.7317	20	0.73165	14.63300
0.7318 – 0.7319	25	0.73185	18.29615
0.7320 – 0.7321	30	0.73205	21.96150
0.7322 – 0.7323	8	0.73225	5.85800
0.7324 – 0.7325	2	0.73245	1.46490
Total	110	—	80.49790

$$\bar{X} = \frac{\Sigma fx}{\Sigma f}$$

$$= \frac{80.49790}{110} = 0.73180$$

Calculation for Median and Mode

Classes	f	c.f	C.B.
0.7312 – 0.7313	10	10	0.73115 – 0.73135
0.7314 – 0.7315	15	25	0.73135 – 0.73155
0.7316 – 0.7317	20	45	0.73155 – 0.73175
0.7318 – 0.7319	25	70	0.73175 – 0.73195
0.7320 – 0.7321	30	100	0.73195 – 0.73215
0.7322 – 0.7323	8	108	0.73215 – 0.73235
0.7324 – 0.7325	2	110	0.73235 – 0.73255

(i) **Median**

Median $= \left(\dfrac{n}{2}\right)$ th item

$= \dfrac{110}{2}$ th item $=$ 55th item

55th item lies in the class 0.713175 – 0.73195. So this is median class.

Where L = 0.73175,   h = 0.0002,

   f = 25,   n = 110, c = 45

Median $= L + \dfrac{h}{f}\left(\dfrac{n}{2} - C\right)$

$= 0.73175 + \dfrac{0.0002}{25}\left(\dfrac{110}{2} - 45\right)$

$= 0.73175 + \dfrac{0.0002}{25}(55 - 45)$

$= 0.73175 + \dfrac{0.0002}{25}(10)$

$= 0.73175 + \dfrac{0.002}{25}$

$= 0.73175 + 0.00008 = 0.73199$

(iii)   Mode

Modal class

The class which has maximum frequency 0.73195–0.73215 is the modal class.

$$\text{Mode} = L + \frac{(f_m - f_1)}{(f_m - f_1) + (f_m - f_2)} \times h$$

Where

$L = 0.73195$, $f_m = 30$, $f_1 = 25$, $f_2 = 8$, $h = 0.0002$

$$\text{Mode} = 0.73195 + \frac{(30 - 25)}{(30 - 25) + (30 - 8)} \times 0.0002$$

$$= 0.73195 + \frac{5}{5 + 22} \times 0.0002$$

$$= 0.73195 + \frac{0.00100}{27}$$

$$= 0.73195 + 0.00004 = 0.73199$$

EXAMPLE 3.26

Find the median and mode from the following data.

X	1.8	1.6	1.4	1.2	1.0	0.8	0.6	0.4	0.2
f	6	9	15	24	33	20	13	7	5

SOLUTION

Writing the data in an ascending order.

X	f	C.f	C.B. $\left(x - \frac{h}{2}, x + \frac{h}{2}\right)$
0.2	5	5	0.1 – 0.3
0.4	7	12	0.3 – 0.5
0.6	13	25	0.5 – 0.7
0.8	20	45	0.7 – 0.9
1.0	33	78	0.9 – 1.1
1.2	24	102	1.1 – 1.3
1.4	15	117	1.3 – 1.5
1.6	9	126	1.5 – 1.7
1.8	6	132	1.7 – 1.9
	132		

(i)   Median

$$\text{Median} = \left(\frac{n}{2}\right) \text{th item}$$

$$= \frac{132}{2} \text{ th item}$$

$$= 66 \text{ th item}$$

66th item lies in the class 0.9–1.1. Therefore 0.9–1.1 is the median class.

$$\text{Median} = L + \frac{h}{f}\left(\frac{n}{2} - C\right)$$

Where

$L = 0.9, h = 0.2, f = 33, n = 132, c = 45$

$$\text{Median} = 0.9 + \frac{0.2}{33}\left(\frac{132}{2} - 45\right)$$

$$=.0.9 + \frac{0.2}{33}(66 - 45)$$

$$= 0.9 + \frac{0.2}{33}(21)$$

$$= 0.9 + \frac{4.2}{33}$$

$$= 0.9 + 0.127 = 1.03$$

(ii)   Mode

Modal Class

The class which has maximum frequency.

i.e.   0.9 – 1.1 is the modal class.

Mode is given by the relation.

$$\text{Mode} = L + \frac{(f_m - f_1)}{(f_m - f_1) + (f_m - f_2)} \times h$$

$$= 0.9 + \frac{(33 - 20)}{(33 - 20) + (33 - 20)} \times 0.2$$

$$= 0.9 + \frac{13}{13 + 9} \times 0.2$$

$$= 0.9 + \frac{13 \times 0.2}{22}$$

$$= 0.9 + \frac{2.6}{22}$$

$$= 0.9 + 0.118 = 1.02$$

3.5 GEOMETRIC MEAN

Let $x_1, x_2, ..., x_n$ be the n positive values, then the geometric mean may be defined as nth positive root of their product.

i.e. $G = \sqrt[n]{x_1 . x_2 .... x_n}$

$G = (x_1 . x_2 ...... x_n)^{1/n}$

When n is large, the calculation of the geometric mean becomes difficult and time consuming. For this reason we use the formula in terms of logarithims.

$$G = antilog\left[\frac{1}{n}(\Sigma \log x)\right]$$

If the data is organized into frequency distribution, then the above formula becomes

$$G = antilog\left[\frac{1}{\Sigma f}(\Sigma f \log x)\right]$$

The geometric mean is appropriate to average ratios and rate of change.

EXAMPLE 3.27

Find the geometric mean of the data given below.

10, 12, 18, 20, 25, 30, 35

SOLUTION

x	log x
10	1.0000
12	1.0792
18	1.2553
20	1.3010
25	1.3979
30	1.4771
35	1.5441
Σ	9.0546

$$G = \text{antilog} \frac{1}{n} (\Sigma \log x)$$

$$= \text{antilog} \frac{1}{7} (9.0546)$$

$$= \text{antilog} (1.2935) = 19.6569$$

EXAMPLE 3.28

Calculate geometric mean of the data given below.

Classes	1 – 3	4 – 6	7 – 9	10 – 12	13 – 15
Frequency	1	2	3	2	1

SOLUTION

Classes	F	x	log x	f log x
1 – 3	1	2	0.3010	0.3010
4 – 6	2	5	0.6990	1.3980
7 – 9	3	8	0.9031	2.7093
10 – 12	2	11	1.0414	2.0828
13 – 15	1	14	1.1461	1.1461
				7.6372

$$G = \text{antilog} \frac{1}{\Sigma f} (\Sigma f \log x)$$

$$= \text{antilog} \frac{1}{9} (7.6372)$$

$$= \text{antilog} (0.8486) = 7.0563$$

3.5.1 Advantages and disadvantages of Geometric Mean

Advantages

(i)   It is based on all the observations.

(ii)   It is rigorously defined.

(iii)   It gives equal weightage to all the observations.

(iv)   It is not much affected by sampling variability.

Disadvantages

(i)   In case of negative value, it can not be calculated.

(ii)   It vanishes if any observation is zero.

3.6 HARMONIC MEAN

Harmonic may be defined as the reciprocal of the arithmetic mean of the reciprocal of the values.

$$i.e. \quad H = \frac{n}{\Sigma\left(\frac{1}{x}\right)} \quad \text{Where } x \neq 0$$

If the data is arranged in the form of frequency distribution, then Harmonic mean is defined as

$$H = \frac{\Sigma f}{\Sigma f\left(\frac{1}{x}\right)} \quad \text{Or} \quad H = \frac{\Sigma f}{\Sigma\left(\frac{f}{x}\right)}$$

EXAMPLE 3.29

A motorcycle is running at the rate of 15 km/hour during the first 60 km at 20 km/hour, during second 60 km, 30 km/hour, during the 3rd 60 km. What would be the average speed?

SOLUTION

$$H = \frac{n}{\Sigma\left(\frac{1}{x}\right)}$$

$$= \frac{3}{\frac{1}{15} + \frac{1}{20} + \frac{1}{30}}$$

$$= \frac{3}{0.067 + 0.050 + 0.033}$$

$$= \frac{3}{0.1503} = 19.9557 \text{ km/hour}$$

EXAMPLE 3.30

Calculate Harmonic mean of the following frequency distribution.

Classes	0 - 4	4 - 8	8 - 12	12 - 16	16 - 20	20 - 24	24 - 28
Frequency	2	5	7	8	7	4	1

SOLUTION

Classes	f	x	f/x
0 – 4	2	2	1.0000
4 – 8	5	6	0.8383
8 – 12	7	10	0.7000
12 – 16	8	14	0.5714
16 – 20	7	18	0.3889
20 – 24	4	22	0.1816
24 – 28	1	26	0.0385
Σ	34	—	3.7137

$$\text{H.M.} = \frac{\Sigma f}{\Sigma\left(\frac{f}{x}\right)} = \frac{34}{3.7137} = 9.1553$$

3.6.1 Advantages and disadvantages of Harmonic Mean

Advantages

(i)   It is rigorously defined by mathematical formula.

(ii)   It is based on all the observations.

(iii)   It is not much affected by sampling stability.

Disadvantages

(i)   It is not easily understandable.

(ii)   It can not calculate if any observation is zero.

(iii)   It gives high weightage to small values.

SUMMARY

The formulae and the methods of computing mean, median, mode, geometric mean and harmonic mean are summarized below:

APPLICATION	FORMULA
Arithmetic Mean:	
Ungrouped data	
Direct method	$\overline{X} = \dfrac{\Sigma X}{n}$
Short cut method	$\overline{X} = A + \dfrac{\Sigma D}{n}$
Grouped Data	$\overline{X} = \dfrac{\Sigma fx}{\Sigma f}$
Short cut method	$\overline{X} = A + \dfrac{\Sigma fD}{\Sigma f}$, $D = X - A$
Step deviation method	$\overline{X} = A + \dfrac{\Sigma fu}{\Sigma f} \times h$, $u = \dfrac{X - A}{h}$
Weighted mean	$\overline{X}_w = \dfrac{\Sigma WX}{\Sigma W}$
Combined mean	$\overline{X}_c = \dfrac{\Sigma n_i \overline{X}_i}{\Sigma n_i}$
Median	
Ungrouped data	Median $= \dfrac{n+1}{2}$ th value
Grouped data	Median $= 1 + \dfrac{h}{f}\left(\dfrac{n}{2}. C\right)$
Graphically	From Ogive

Mode:	
Ungrouped data	Value which occurs maximum no. of times
Grouped data	$\text{Mode} = 1 + \dfrac{f_m - f_1}{(f_m - f_1) + (f_m - f_2)} \times h$
Graphically	From histogram
Empirical relation	Mode = 3 Median − 2 Mean
Geometric Mean	
Ungrouped data	$G = \text{antilog} \dfrac{1}{n} (\Sigma \log x)$
Grouped data	$G = \text{antilog} \dfrac{1}{\Sigma f} (\Sigma f \log x)$
Harmonic Mean:	
Ungrouped data	$H = \dfrac{n}{\Sigma \left(\dfrac{1}{x}\right)}$
Grouped data	$H = \dfrac{\Sigma f}{\Sigma (f/x)}$

EXERCISES

3.1   Define statistical average. What are the qualities of good average?

3.2   (a) Define and explain arithmetic mean.

(b) Calculate arithmetic mean for the following values.

(i)   1, 2, 3, 4, 5

(ii)   3, 5, 7, 5, 7, 9

(iii) 15, 18, 7, 12, 13, 20, 23, 18

(iv)   48, 34, 69, 25, 22, 34, 56, 68

(c) The weights of 10 students are given below. Find Mean.

40.2, 50.3, 43.1, 50.0, 52.8, 60.0, 57.2, 48.5, 49.5, 52.0

3.3   (a) From the following data. Calculate arithmetic mean.

x	0	1	2	3	4	5
f	40	20	50	30	20.	10.

(b) Given below the frequency distribution of number of classrooms in different colleges. Find arithmetic mean.

No. of Rooms x	25	28	35	40	45
No. of Colleges f	3	7	10	3	2

3.4   (a) The marks obtained by 100 students are given below. Calculate average marks.

Marks	10–25	25–40	40–55	55–70	70–85	85–100
Freq.	6	20	44	26	3	1

(b) The following frequency distribution shows the daily income of 100 house holds in a locality.

Income	35–39	40–44	45–49	50–54	55–59	60–64	65–69
Freq.	13	15	28	17	12	10	5

Calculate arithmetic mean.

(c) The mean age of a group of 100 persons was found to be 35. Later it was discovered that age 50 was misread as 25. Find the corrected mean.

3.5   The height of college students measured to nearest inches is given below.

Height (inches)	60 – 62	63 – 65	66 – 68	69 – 71	72 – 74
No. of Student	05	18	42	27	08

Calculate arithmetic mean.

3.6   (a)   The marks in English of Ist year class are given below.

Marks	15–19	20–24	25–29	30–34	35–39
f	9	18	35	17	5

Calculate arithmetic mean by using.

(i)   Shortcut method

(ii)   Step deviation method

(b)   The following data have been obtained from a frequency distribution of continuous variable X after making substitution.

$$U = \frac{X - 130}{5}$$

U	-3	-2	-1	0	1	2	3
f	5	12	26	32	13	8	4

(i)   Find arithmetic mean by direct method.

(ii)   Find arithmetic mean by step deviation method.

3.7   (a)   Define the weighted arithmetic mean.

(b)   Calculate the weighted mean for the following items.

Items	Expenditure	Weights
Food	290	7.5
Rent	54	2.0
Clothing	98	1.5
Fuel & Light	75	1.0
Misc.	75	0.5

(c)   A student's final marks in Statistics, Computer science, mathematics and English are 75, 62, 85 and 57. Find the weighted arithmetic mean if weights of 3, 4, 2 and 1 respectively are attached to the subjects.

3.8 **(a)** A distribution consists of four components with frequencies 15, 12, 16, 21 having their means 16.5, 20.3, 21.6 and 26.2. Find the mean of the combined distribution.

(b) The average marks obtained by the students of three sections in statistics class are given below.

Section	No. of Students	Average Marks
A	50	75
B	75	62
C	60	68

Find average marks of the whole class.

(c) A variable Y is determined from a variable X by the equation $Y = 10 - 4X$.

Find Y when, X = -3, -2, -1, 0, 1, 2, 3, 4, 5 and show that $\overline{y} = 10 - 4\overline{x}$.

3.9 **(a)** Deviations from X = 10.5 of 10 items are given below.

−1.3, 2.0, 2.9, 7.5, −4.6, −3.4, 8.2, 9.3, −7.4, 5.6

Calculate the arithmetic mean.

(b) Find the mean from the following observations and show that

$$\Sigma(X - \overline{X}) = 0$$

6.5, 2.33, 7.4, 7.25, 6.50, 9.7, 8.35, 2.6, 2.43

(c) From the following frequency distribution show that sum of deviations of values from their mean is zero.

i.e. $\Sigma f(X - \overline{X}) = 0$

Classes	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
F	5	12	15	25	8	3	2

3.10 **(a)** Define median. Given a simple example of median.

(b) Calculate median from the following data.

(i) Heights of 9 students in inches are

60, 58, 54, 53, 55, 63, 51, 52, 57

(ii) No. of Road side accidents in 10 cities are

10, 6, 12, 8, 7, 15, 13, 5, 7, 14

(iii) Wages of 10 workers are

88, 70, 72, 125, 115, 95, 81, 90, 95, 90

3.11 (a) Calculate median for the following data.

No. of children	0	1	2	3	4	5	6	7
Frequency	2	3	4	6	5	4	3	3

(b) Given below is the frequency distribution of family members in 50 families. Find the median as the average family size.

No. of persons	2	3	4	5	6
No. of families	5	8	12	20	5

3.12 Compute the median from the following data.

Groups	10 - 19	20 - 29	30 - 39	40 - 49	50 - 59	60 - 69
F	15	17	24	31	21	12

3.13 Daily wages of factory workers are given below.

Daily wages	40-60	60-80	80-100	100-120
No. of Workers	13	23	101	182
Daily wages	120-140	140-160	160-180	
No. of Workers	105	19	7	

Calculate the median.

3.14 (a) Give the advantages and Disadvantages of median.

(b) Pocket money of the student during convocation is given below. Locate the median graphically.

Rs.	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30
No. of Students	12	9	18	7	4

Also check your result by actual calculations.

3.15 (a) Define mode. Explain it with example.

(b) Find the mode for the following data.

(i) 7, 4, 9, 10, 12, 15, 7, 9, 7

(ii) 70, 75, 60, 62, 68, 70, 60

(iii) 10, 8, 15, 12, 6, 9, 7, 16

3.16 (a) Calculate the mode from the following data.

Size of shoes	3	4	5	6	7	8	9	10
Freq.	12	18	20	13	18	27	19	9

(b) Calculate the mode from the following frequency distribution showing the number of children per house.

No. of Children	0	1	2	3	4	5	6	7	8	9	10
No. of houses	4	10	12	18	25	20	16	8	4	2	1

(c) Calculate median and mode from the following frequency distribution.

No. of Children	0	1	2	3	4	5	6	7
f		1	3	3	6	5	4	2	1

3.17 (a) The age distribution of employees in a factory is given below. Find the model age of the employees.

Age of Employee (Years)	15–20	20–25	25–30	30–35	35–40
Frequency.	5	23	58	104	141
Age of Employee (Years)	40–45	45–50	50–55	55–60	
Freq.	98	43	19	6	

(b) The following table shows the distribution of maximum loads in short tons supported by certain cables produced by a company.

Maximum Load	f
9.3 – 9.7	2
9.8 – 10.2	5
10.3 – 10.7	12
10.8 – 11.2	17
11.3 – 11.7	14
11.8 – 12.2	6
12.3 – 12.7	3
12.8 – 13.2	1

Determine mean, median and mode.

3.18   (a)   The weight of 40 male students at a college are given in the following frequency distribution.

Weight	118-126	127-135	136-144	145-153	154-162	163-171	172-180
Freq.	3	5	9	12	5	4	2

Calculate the mean, median and mode.

(b)   Compute Mean, Median and Mode from the following data.

x	2.4	2.0	1.6	1.2	0.8	0.4
f	6	9	15	30	12	8

3.19   (a)   The daily profits in rupees of 120 shops are given below.

Profit per shop	0-100	100-200	200-300	300-400	400-500	500-600
No. of Shops	15	18	25	40	15	7

Draw histogram. Calculate the mode graphically and verify the result by actual calculations.

(b)   From the following data find the missing frequency when mean is 15.38.

x	10	12	14	16	18	20
f	3	7	?	20	8	5

3.20   (a)   Discuss the empirical relation between mean, median and mode.

(b)   For a certain frequency distribution the value of mean was 11 and the median was 12. Find the approximate value of the mode.

(c)   Find the value of mode by using empirical relation between averages for the following data.

Marks	10–19	20–29	30–39	40–49	50–59
No. of Students	5	25	40	20	10

3.21   Which average will be suitable to compare.

(i)   Height of students.

(ii)   Size of shoes

(iii)   Intelligence of students.

(iv)   Number of Petals of a flower.

(v)   Average income of different people.

(vi)   Average size of ready-made garments.

(vii)   Weight of students.

(viii)   Marks obtained by the students.

3.22 Find the mean and Median of the following data.

Height (inches)	45 - 50	50 - 55	55 - 60	60 - 65	65 - 70	70 - 75
No. of Persons	2	7	12	18	13	3

3.23 Compute Mean and Median.

Monthly income (Rs.)	No. of Families	Monthly income (Rs.)	No. of families
110 – 119	2	160 – 169	18
120 – 129	4	170 – 179	13
130 – 139	17	180 – 189	6
140 – 149	28	190 – 199	5
150 – 159	25	200 – 209	2

3.24 (a) Calculate Q_1, Q_3, D_7, Median, P_{20} and P_{60} from the following data.

Classes	1-10	11-20	21-30	31-40	41-50	51-60	61-70
F.	2	4	7	8	6	3	2

(b) Calculate Q_1, Q_3, D_5 and P_{80} from the following data.

20, 35, 18, 27, 35, 40, 48, 33, 42, 35, 28

(c) Find the median, the quartiles 8th docile and 65th percentile for the distribution of examination marks given below:

Marks	30-39	40-49	50-59	60-69	70-79	80-89	90-99
No. of students	8	87	190	304	211	85	20

3.25 (a) Define geometric mean.

(b) What do you meant by Harmonic mean?

3.26 (a) Find the geometric mean of the following data.

    (i)   3, 5, 6, 6, 7, 10, 12

    (ii)   15, 18, 17, 20, 25, 23

    (iii) 75, 62, 76, 78, 59, 67

    (iv) 110, 115, 108, 112, 120, 128, 130

    (v)   45, 32, 37, 46, 39, 36, 41, 48, 36

(b) Find geometric mean from the following frequency distribution.

x	2	3	4	5	6
f	5	7	8	3	2

3.27 **(a)** Given the following frequency distribution of weights, calculate geometric mean.

Weight (grams)	65-84	85-104	105-124	125-144	145-164	165-184	185-204
f	9	10	17	10	5	4	5

(b) Give the following frequency distribution.

Classes	15 – 19	20 – 24	25 – 29	30 – 34	35 – 39
f	15	17	25	18	12

Find the geometric mean.

3.28 Calculate the geometric mean from the following frequency distribution.

Classes	100 - 200	200 - 300	300 - 400	400 - 500	500 - 600
f	15	18	30	20	17

3.29 Aslam gets a rise of 15% in salary at the end of his first year of service and further 25% and 30% at the end of the second and third year respectively. The rise in each case being calculated on his salary at the beginning of the year. To what annual percentage increase in this equivalent?

(b) A man traveling 100 kilometer has 5 stages at equal interval. The speed of the man in the various stages was observed to 10, 16, 20, 14, 15 kilometer per hour. Find the average speed at which the man travels.

3.30 **(a)** Define Harmonic mean

(b) Calculate Harmonic mean from the following data.

    (i)   1, 2, 3, 4, 5

    (ii)   12, 18, 16, 20, 25, 30

    (iii)   75, 60, 65, 85, 60, 50

(c) A man traveling 100 kilometer has 5 stages at equal interval. The speed of the man in the various stages was observed to 10, 16, 20, 14, 15 kilometer per hour. Find the average speed at which the man travels.

3.31 The reciprocal of 11 values of x are given below.

0.0500, 0.0454, 0.0400, 0.0333, 0.0285, 0.0232, 0.0213, 0.0200, 0.0182, 0.0151, 0.0143

Calculate harmonic mean and arithmetic mean.

3.32 **(a)** Find the harmonic mean from the following data.

Weight (gm)	65 – 84	85 – 104	105 – 124	125 – 144
f	9	10	17	10
Weight (gm)	145 – 164	165 – 184	185 – 204	
f	5	4	5	

(b) Compute Harmonic mean from the following data.

Hourly Wages	40-50	50-60	60-70	70-80	80-90
Frequency	4	8	16	8	4

3.33 For the data given below calculate

(i) Arithmetic mean

(ii) Harmonic Mean

(iii) Median

Marks	30 – 40	40 – 50	50 – 60	60 – 70
frequency	1	3	12	15
Marks	70 – 80	80 – 90	90 – 100	
frequency	14	11	5	

3.34 Calculate Mean, Median and Geometric mean from the following data

Daily wages (Rs)	No. of Workers
Below 05	7
Below 10	52
Below 15	107
Below 20	170
Below 25	215
Below 30	259
Below 35	285
Below 40	300

3.35　CHOOSE THE CORRECT ANSWER.

(i)　The sum of the deviations taken from mean is

　　(a)　Always equal to zero

　　(b)　Sometimes equal to zero

　　(c)　Never equal to zero

　　(d)　None of these

(ii)　The median of 12, 5, 6, 8 and 4 is

　　(a)　7　　　　　　　　　　　　(b)　6

　　(c)　8　　　　　　　　　　　　(d)　12

(iii)　The mode of 10, 8, 6, 5, 6, 7 is

　　(a)　10　　　　　　　　　　　(b)　7

　　(c)　6　　　　　　　　　　　　(d)　5

(iv)　The mode of "STATISTICS" is

　　(a)　S　　　　　　　　　　　　(b)　S and T

　　(c)　C　　　　　　　　　　　　(d)　none of these

(v)　Most suitable average for qualitative data is

　　(a)　Mean　　　　　　　　　　(b)　Median

　　(c)　Mode　　　　　　　　　　(d)　None of these

(vi)　In a symmetrical distribution, mean, median and mode are

　　(a)　Equal　　　　　　　　　　(b)　Unequal

　　(c)　Sometimes equal　　　　　(d)　None of these

(vii)　For any two values, the mean and median are always

　　(a)　Equal　　　　　　　　　　(b)　Unequal

　　(c)　ill define　　　　　　　　(d)　None of these

(viii)　The arithmetic mean of 5 values is 10, then sum of these values is

　　(a)　5　　　　　　　　　　　　(b)　10

　　(c)　15　　　　　　　　　　　(d)　50

(ix)　If $\Sigma(x - 20) = 0$, then mean will be

　　(a)　10　　　　　　　　　　　(b)　15

　　(c)　20　　　　　　　　　　　(d)　0

(x) The most frequent value in the set of data if it exists is called

    (a)   Mean   (b)   Median

    (c)   Mode   (d)   None of these

(xi) When the observations are not of equal importance then we use

    (a)   Simple mean   (b)   Weighted mean

    (c)   Combined mean   (d)   Median

(xii) For graphic representation of median, we have to draw

    (a)   Histogram   (b)   Historigram

    (c)   Ogive   (d)   Pie chart

(xiii) For graphic representation of mode, we have to draw

    (a)   Histogram   (b)   Historigram

    (c)   Frequency polygon   (d)   Ogive

(xiv) If the data contains an extreme value, the suitable average is

    (a)   Mean   (b)   Median

    (c)   Mode   (d)   Weight mean

(xv) A distribution which has one mode is called

    (a)   Unimodal   (b)   Bimodal

    (c)   Multimodal   (d)   None of these

(xvi) A distribution which has two modes is called

    (a)   Uni-model   (b)   Bi-model

    (c)   Multi-model   (d)   None of these

(xvii) Sum of squared deviation of the observations from mean is

    (a)   zero   (b)   1

    (c)   Minimum   (d)   None of these

(xviii) In case of skewed distribution, mean median and mode are

    (a)   equal   (b)   sometimes equal

    (c)   not equal   (d)   none of these

(xix) The mean is based on:

    (a)   All values   (b)   Small values

    (c)   Extreme values   (d)   Non of these

(xx)   Arithmetic mean of the positive numbers a and b is:

    (a) $\frac{ab}{2}$

    (b) $\frac{2a}{b}$

    (c) $\frac{a+b}{2}$

    (d) $\frac{2}{a+b}$

(xxi) If $\bar{x} = 10$ and $y = 2x + 5$, then $\bar{y}$ is equal to:

    (a)   20

    (b)   25

    (c)   30

    (d)   50

(xxii) Data must be arranged before calculating:

    (a)   Mode

    (b)   Median

    (c)   Mean

    (d)   None of these

(xxiii) The sum of deviation from mean is:

    (a)   Minimum

    (b)   Negative

    (c)   Zero

    (d)   None of these

(xxiv) Change of origin and scale is used for calculation of the:

    (a)   Mean

    (b)   Median

    (c)   Mode

    (d)   None of these

(xxv) Mode for the word **PROFESSOR** is

    (a)   R

    (b)   S

    (c)   O

    (d)   R, S and O

(xxvi) The mean, median and mode for constant "a" is

    (a)   0

    (b) $\frac{a}{2}$

    (c)   a

    (d) a^2

Answers 3.35

(i)	(a)	(ii)	(b)	(iii)	(c)	(iv)	(b)	(v)	(c)
(vi)	(a)	(vii)	(a)	(viii)	(d)	(ix)	(c)	(x)	(c)
(xi)	(b)	(xii)	(c)	(xiii)	(a)	(xiv)	(b)	(xv)	(a)
(xvi)	(b)	(xvii)	(c)	(xviii)	(c)	(xix)	(a)	(xx)	(c)
(xxi)	(b)	(xxii)	(b)	(xxiii)	(c)	(xxiv)	(a)	(xxv)	(d)
(xxvi)	(c)								

SHORT QUESTION ANSWERS

1. Define average.

Ans. An average is a single value, which represents the set of data as a whole.

2. What are the qualities of good average.

Ans. 1. An average should be rigidly defined.
2. It should be easy to understand.
3. It should be easy to calculate.
4. It should be based on all the observations.
5. It should be unaffected by extreme observation.

3. What are the commonly used average?

Ans. The commonly used averages are
1. Arithmetic mean 2. Median 3. Mode
4. Geometric mean 5. Harmonic mean

4. Define arithmetic mean.

Ans. It is defined as sum of all the observations divided by the number of observation. It is denote by $\bar{X}$.

i.e. $\bar{X} = \dfrac{\Sigma x}{n}$

5. Write down at least two properties of mean.

Ans. 1. The sum of deviations of the observations from their mean is zero.

$\Sigma(X - \bar{X}) = 0$

$\Sigma f(X - \bar{X}) = 0$

2. The sum of squared deviations of observations from mean is minimum.

3. Combined mean is

$\bar{X}_c = \dfrac{\Sigma n_i \bar{X}_i}{\Sigma n_i}$

6. Define median.

Ans. Median is the middle most value of a set of data, when arranged in order of magnitude.

7. Give the merits of the median.

Ans. (a) It is easy to calculate and understand.
(b) It is not affected by extreme values.
(c) It can be computed even in open end classes.
(d) It is possible to locate graphically.

8.   **Define Mode.**

Ans. Mode is the value which occurs maximum number of times in the set of data.

9.   **Define weighted mean.**

Ans. When the observations are not of equal importance we assign weights to their relative importance.

i.e. $\bar{X}_w = \dfrac{\Sigma WX}{\Sigma W}$

10.   **What is the empirical relation between mean, median and mode?**

Ans. In symmetrical distribution mean, median and mode are equal.

In a moderately skewed distribution.

Mode = 3 Median − 2 Mean

11.   **For a certain frequency distribution, the value of mean is 15 and median is 20. What will be value of mode?**

Ans. We know that

Mode = 3 median − 2 mean

= 3 (20) − 2(15)

= 60 − 30 = 30

12.   **Define Geometric mean**

Ans. Let $x_1, x_2, \ldots x_n$ be n positive values, then geometric mean may be defined as nth positive root of their product
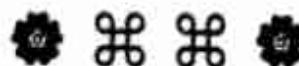
i.e. $G = \sqrt[n]{x_1 \cdot x_2 \ldots x_n}$

$G = (x_1 \cdot x_2 \ldots x_n)^{1/n}$

13.   **Define Harmonic mean**

Ans. Harmonic mean may be defined as the reciprocal of the arithmetic mean of the reciprocal of the values.

i.e. $H = \dfrac{n}{\Sigma\left(\dfrac{1}{x}\right)}$

❀ ⌘ ⌘ ❀

✦✦✦✦✦✦✦✦ CHAPTER # 4 ✦✦✦✦✦✦✦✦✦

MEASURE OF DISPERSION

4.1 INTRODUCTION

The measure of central tendency does not tell us any thing about the spread of the data, because any two sets of data may have the same central tendency with vast difference magnitude of their variability. Consider two types of data;

(a)   10, 12, 11, 14, 13

(b)   2, 10, 18, 27, 3

These two data have same mean 12, but differ in their variations. There is more variation in data (b) as compared to data (a). This illustrates the fact that measure of central tendency is not sufficient. We there for need some additional information concerning with how the data are dispersed about the average. This is done by measuring the dispersion. By dispersion we means the degree to which numerical data tend to spread about an average value.

There are two types of measures of dispersion.

(i)   Absolute dispersion

(ii)   Relative dispersion

(i)   Absolute Measure of Dispersion

An absolute dispersion is one that measure the dispersion in term of some units. For example, if the units of data are rupees, kilograms, centimeter etc. the units of the measure of dispersion will also be rupees, kilograms, centimeter etc.

(ii)   Relative Measure of Dispersion

It is expressed in the form of ratio, coefficient and it is independent of the units of measurements. It is useful for comparison of data of different nature.

MEASURES OF DISPERSION

The main measure of dispersion are the followings.

(i)   The Range

(ii)   The semi Interquartile Range or the Quartile Deviation

(iii)   The Mean Deviation

(iv)   The variance and the standard deviation

4.2   THE RANGE

It is defined as the difference between the largest and the smallest observations in a set of data.

$$\text{Range} = R = X_m - X_o$$

Where X_m = The largest observation

X_o = The smallest observation

The range is very simple measure of variability and only takes into account two most extreme observations.

Its relative measure known as the co-efficient of dispersion.

$$\text{Co-efficient of Range} = \frac{X_m - X_o}{X_m + X_o}$$

EXAMPLE 4.1

Calculate Range and Co-efficient of Range from the following data.

15, 20, 18, 16, 30, 42, 12, 25

SOLUTION

$$X_m = 42, \ X_o = 12$$

$$R = X_m - X_o$$

$$= 42 - 12$$

$$= 30$$

$$\text{Coefficient of Range} = \frac{X_m - X_o}{X_m + X_o}$$

$$= \frac{42 - 12}{42 + 12} = \frac{30}{54} = 0.56$$

EXAMPLE 4.2

Calculate Range and Coefficient of Range from the following data.

Classes	1 – 3	4 – 6	7 – 9	10 – 12	13 – 15
Freq.	1	2	3	2	1

SOLUTION

R = Mid value of the highest class — Mid value of the largest class

Classes	f	x
1 – 3	1	2
4 – 6	2	5
7 – 9	3	8
10 – 12	2	11
13 – 15	1	14

$X_m = 14$, $X_o = 2$

$$R = X_m - X_o$$
$$= 14 - 2$$
$$= 12$$

$$\text{Co-efficient of Range} = \frac{X_m - X_o}{X_m + X_o}$$

$$= \frac{14 - 2}{14 + 2} = \frac{12}{16} = 0.75$$

4.3 THE SEMI INTER QUARTILE RANGE OR QUARTILE DEVIATION

The semi-inter quartile range is defined as half of the difference between the third and the first quartiles.

$$Q.D. = \frac{Q_3 - Q_1}{2}$$

Where Q_1 and Q_3 are the first and the third quartiles of the data.

Its relative measure called the coefficient of quartile deviation is defined by the relation.

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

EXAMPLE 4.3

Given below are the marks obtained by 9 students.

45, 32, 36, 37, 46, 39, 36, 48, 41

Find the quartile deviation and coefficient of quartile deviation.

| SOLUTION |

Arranging the data in an order

32, 36, 36, 37, 39, 41, 45, 46, 48

n = 9

$$Q_1 = \left(\frac{n+1}{4}\right) \text{ the value}$$

$$= \left(\frac{9+1}{4}\right) \text{ th value}$$

$$= 2.5 \text{th value}$$

$$= 2\text{nd value} + 0.5 (3\text{rd value} - 2\text{nd value})$$

$$= 36 + 0.5(36 - 36)$$

$$= 36 + 0.5(0) = 36 \text{ marks}$$

$$Q_3 = 3\left(\frac{n+1}{4}\right) \text{ th value}$$

$$= 3\left(\frac{9+1}{4}\right) \text{ th value}$$

$$= \frac{30}{4} \text{ th value}$$

$$= 7.5 \text{th value}$$

$$= 7\text{th value} + 0.5(8\text{th value} - 7\text{the value})$$

$$= 45 + 0.5(46 - 45)$$

$$= 45 + 0.5(1) = 45.5 \text{ marks}$$

$$Q.D. = \frac{Q_3 - Q_1}{2}$$

$$= \frac{45.5 - 36}{2} = \frac{9.5}{2} = 4.75$$

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

$$= \frac{45.5 - 36}{45.5 + 36} = \frac{9.5}{81.5} = 0.117$$

| EXAMPLE 4.4 |

Compute the quartile deviation and coefficient of quartile deviation from the following data

Groups	10 – 19	20 – 29	30 – 39	40 – 49	50 – 59	60 – 69
f	15	17	24	31	21	12

SOLUTION

First of all we calculate Q_1 and Q_3

Groups	f	C.B.	C.f
10 - 19	15	9.5 - 19.5	15
20 - 29	17	19.5 - 29.5	32
30 - 39	24	29.5 - 39.5	56
40 - 49	31	39.5 - 49.5	87
50 - 59	21	49.5 - 59.5	108
60 - 69	12	59.5 - 69.5	120

$$Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - C\right) \quad \text{......... (i)}$$

$$Q_1 = \frac{n}{4} \text{ th value}$$

$$= \frac{120}{4} \text{ th value}$$

$$= \text{30th value}$$

30th value lies in the class 19.5 - 29.5

Therefore

$l = 19.5, \ h = 10, \ f = 17, \ n = 20, \ C = 15$

Putting these values in (i), we have

$$Q_1 = 19.5 + \frac{10}{17}\left(\frac{120}{4} - 15\right)$$

$$= 19.5 + \frac{10}{17}(30 - 15)$$

$$= 19.5 + \frac{10}{17}(15)$$

$$= 19.5 + \frac{150}{17}$$

$$= 19.5 + 8.82 = 28.32$$

$$Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - C\right) \quad \text{......... (ii)}$$

$$Q_3 = \frac{3n}{4} \text{ th value}$$

$$= \frac{3(120)}{4} \text{ th value}$$

$$= \text{90th value}$$

So 90th value lies in the group $49.5 - 59.5$
Therefore
$l = 49.5$, $h = 10$, $f = 21$, $c = 87$
Putting the values in (ii), we have

$$Q_3 = 49.5 + \frac{10}{21}\left(\frac{3 \times 120}{4} - 87\right)$$

$$= 49.5 + \frac{10}{21}(90 - 87)$$

$$= 49.5 + \frac{10}{21}(3)$$

$$= 49.5 + 1.43 = 50.93$$

$$Q.D. = \frac{Q_3 - Q_1}{2}$$

$$= \frac{50.93 - 28.32}{2} = \frac{22.61}{2} = 11.31$$

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

$$= \frac{50.93 - 28.32}{50.93 + 28.32} = \frac{22.61}{79.25} = 0.285$$

4.4 THE MEAN DEVIATION OR AVERAGE DEVIATION

Mean deviation is defined as the mean of the absolute deviations measured either from mean, median or mode. By absolute deviations we mean that all the deviations are positive.

$$M.D. = \frac{\Sigma \mid X - \bar{X} \mid}{n} \quad \text{(M.D. from Mean)}$$

$$M.D. = \frac{\Sigma \mid X - \text{Median} \mid}{n} \quad \text{(M.D. from Median)}$$

$$M.D. = \frac{\Sigma \mid X - \text{Mode} \mid}{n} \quad \text{(M.D. from Mode)}$$

(In case of Group Data)

(i)   Mean deviation from Mean $= \dfrac{\Sigma f \mid X - \bar{X} \mid}{\Sigma f}$

(ii)   Mean deviation from Median $= \dfrac{\Sigma f \mid X - \text{Median} \mid}{\Sigma f}$

(iii) Mean deviation from Mode $= \dfrac{\Sigma f \mid X - \text{Mode} \mid}{\Sigma f}$

Co-efficient of M.D. (Mean) $= \dfrac{\text{Mean deviation from Mean}}{\text{Mean}}$

Co-efficient of M.D. (Median) $= \dfrac{\text{M.D. from Median}}{\text{Median}}$

Co-efficient of M.D. (Mode) $= \dfrac{\text{M.D. from Mode}}{\text{Mode}}$

EXAMPLE 4.5

Calculate mean deviation from the mean and the median for the values
30, 36, 32, 33, 35, 39, 36.5, 35 and 34

SOLUTION

Mean $= \dfrac{\Sigma x}{n} = \dfrac{310.5}{9} = 34.5$

To find the median, we first arrange the values
30, 32, 33, 34, 35, 35, 36, 36.5 and 39

$\tilde{X} = 35$

X	$|X - \bar{X}|$	$|X - \tilde{X}|$
30	$|30 - 34.5| = 4.5$	5.0
36	1.5	1.0
32	2.5	3.0
33	1.5	2.0
35	0.5	0
39	4.5	4.0
36.5	2.0	1.5
35	0.5	0
34	0.5	1.0
310.5	$\Sigma |X - \bar{X}| = 18.0$	17.50

M.D. (From Mean) $= \dfrac{\Sigma |X - \bar{X}|}{n}$

$= \dfrac{18}{9} = 2.0$

$$\text{M.D. (From Median)} = \frac{\Sigma |X - \tilde{X}|}{n}$$

$$= \frac{17.5}{9} = 1.94$$

EXAMPLE 4.6

Compute mean deviation from Mean. Also calculate the coefficient of mean deviation

Weights	f	Weight	f
50 – 53	23	65 – 68	66
53 – 56	24	68 – 71	49
56 – 59	39	71 – 74	38
59 – 62	46	74 – 77	21
62 – 65	54	77 – 80	12

SOLUTION

Weight	F	x	$u = \dfrac{x-a}{h}$	fu	$|x - \bar{x}|$	$f|x - \bar{x}|$
50 – 53	23	51.5	−5	−115	13.08	300.84
53 – 56	24	54.5	−4	−96	10.08	241.92
56 – 59	39	57.5	−3	−117	7.08	276.12
59 – 62	46	60.5	−2	−92	4.08	187.68
62 – 65	54	63.5	−1	−54	1.08	58.32
65 – 68	66	66.5	0	0	1.92	126.72
68 – 71	49	69.5	1	49	4.92	241.08
71 – 74	38	72.5	2	76	7.92	300.96
74 – 77	21	75.5	3	63	10.92	229.32
77 – 80	12	78.5	4	48	13.92	167.04
—	372	—	—	−238	—	2130.00

$$\bar{x} = a + \frac{\Sigma fu}{\Sigma f} \times h$$

Let $\quad u = \frac{x - a}{n}$, $a = 66.5$ and $h = 3$

$$\bar{x} = 66.5 + \frac{(-238)}{372} \times 3$$

$$= 66.5 - 1.92 = 64.58$$

$$\text{M.D.} = \frac{\Sigma f \, |x - \bar{x}|}{\Sigma f}$$

$$= \frac{2130}{372} = 5.726$$

$$\text{Coefficient of M.D.} = \frac{\text{M.D.}}{\text{Mean}}$$

$$= \frac{5.726}{64.58} = 0.089$$

4.5 THE VARIANCE AND STANDARD DEVIATION

The Variance

The variance is defined as the mean of the squares of deviation of all the observations from their mean.

The concept of variance was introduced in 1918 by R.A. Fisher. Due to its importance variance is commonly used measure of dispersion. The symbolic definition for variance is

$$S^2 = \frac{\Sigma(x - \bar{x})^2}{n}$$

Or $\quad S^2 = \frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2$

In case of frequency distribution variance may be defined as

$$S^2 = \frac{\Sigma f(x - \bar{x})^2}{\Sigma f}$$

Or $\quad S^2 = \frac{\Sigma fx^2}{\Sigma f} - \left(\frac{\Sigma fx}{\Sigma f}\right)^2$

Standard Deviation

The positive square root of the variance is called standard deviation. Symbolically

$$S = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n}}$$

Or $\quad S = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$

In case of frequency distribution

$$S = \sqrt{\frac{\Sigma f(x - \bar{x})^2}{\Sigma f}}$$

Or $\quad S = \sqrt{\frac{\Sigma fx^2}{\Sigma f} - \left(\frac{\Sigma fx}{\Sigma f}\right)^2}$

The standard deviation is expressed in the same units as the observations themselves.

EXAMPLE 4.7

Calculate the variance and standard deviation from the following marks obtained by 9 students.

45, 32, 37, 46, 39, 36, 41, 48, 36

SOLUTION

X	X^2
45	2025
32	1024
37	1369
46	2116
39	1521
36	1296
41	1681
48	2304
36	1296
360	14632

$$S^2 = \frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2$$

$$= \frac{14632}{9} - \left(\frac{360}{9}\right)^2$$

$$= 1625.78 - 1600$$

$$= 25.78 \ (\text{Mark})^2$$

and $S = \sqrt{\dfrac{\Sigma x^2}{n} - \left(\dfrac{\Sigma x}{n}\right)^2}$

$$= 5.08 \ \text{Marks}$$

EXAMPLE 4.8

Determine the variance and standard deviation from the following data.

Groups	20 – 24	25 – 29	30 – 34	35 – 39	40 – 44	45 – 49	50 – 54
f	1	4	8	11	15	9	2

SOLUTION

Classes	F	x	fx	fx²
20 – 24	1	22	22	484
25 – 29	4	27	108	2916
30 – 34	8	32	256	8192
35 – 39	11	37	407	15059
40 – 44	15	42	630	26460
45 – 49	9	47	423	19881
50 – 54	2	52	104	5408
Total	50	—	1950	78400

$S^2 = \dfrac{\Sigma fx^2}{\Sigma f} - \left(\dfrac{\Sigma fx}{\Sigma f}\right)^2$

$= \dfrac{78400}{50} - \left(\dfrac{1950}{50}\right)^2$

$= 1568 - 1521$

$S^2 = 47$

Standard Deviation

$S = \sqrt{\dfrac{\Sigma fx^2}{\Sigma f} - \left(\dfrac{\Sigma fx}{\Sigma f}\right)^2}$

$= \sqrt{47} = 6.856$

Change of Origin and Scale (Short cut method for calculating variance and standard deviation)

Let $\quad u = \dfrac{x - a}{h}$

Then $\quad S^2 = h^2 \left[\dfrac{\Sigma fu^2}{\Sigma f} - \left(\dfrac{\Sigma fu}{\Sigma f} \right)^2 \right]$

Where $\quad\quad\quad h = $ class interval

Standard Deviation $= S = h \sqrt{\dfrac{\Sigma fu^2}{\Sigma f} - \left(\dfrac{\Sigma fu}{\Sigma f} \right)^2}$

EXAMPLE 4.9

Calculate variance and standard deviation by using short cut method to the following data.

Weight	28 – 31	32 – 35	36 – 39	40 – 43	44 – 47
f	1	14	56	172	245
Weight	48 – 51	52 – 55	56 – 59	60 – 63	64 – 67
f	263	156	67	23	3

SOLUTION

The necessary calculations are given below.

Weight (Kg)	f	x	$u = \dfrac{x - 49.5}{4}$	fu	fu^2
28 - 31	1	29.5	-5	-5	25
32 – 35	14	33.5	-4	-56	224
36 – 39	56	37.5	-3	-168	504
40 – 43	172	41.5	-2	-344	688
44 – 47	245	45.5	-1	-245	245
48 – 51	263	49.5	0	0	0
52 – 55	156	53.5	1	156	156
56 – 59	67	57.6	2	134	268
60 – 63	23	61.5	3	69	207
64 – 67	3	65.5	4	12	48
Σ	1000	—	—	-447	2365

$$S = h \times \sqrt{\frac{\Sigma fu^2}{\Sigma f} - \left(\frac{\Sigma fu}{\Sigma f}\right)^2}$$

$$= 4 \times \sqrt{\frac{2365}{1000} - \left(\frac{-447}{1000}\right)^2}$$

$$= 4\sqrt{2.365 - 0.1998}$$

$$= 4\sqrt{2.1652}$$

$$= 4 \times 1.47 = 5.88 \text{ kg}$$

Variance $(S)^2 = (5.88)^2$

$$= 34.5744 \ (\text{kg})^2$$

4.5.1 Co-efficient of Variation

The most important of all the relative measures of dispersion is co-efficient of variation. It is used to compare the variations in two or more set of data or distributions that are measured in different units. For example, one may be measured in hours and the other in rupees. The group which has lower value of co-efficient of variation is comparatively more consistent. The coefficient of variation is defined as

$$C.V. = \frac{S}{\bar{x}} \times 100$$

EXAMPLE 4.10

Goals scored by two teams A and B in a hockey season were as follows.

No. of goals x	Number of matches	
	A	B
0	24	17
1	9	9
2	8	6
3	5	5
4	4	3

By calculating the co-efficient of variation in each case, find which team may be considered more consistent.

SOLUTION

No. of Goals	Team A			Team B		
x	f	fx	fx²	f	fx	fx²
0	24	0	0	17	0	0
1	9	9	9	9	9	9
2	8	16	32	6	12	24
3	5	15	45	5	15	45
4	4	16	64	3	12	48
Total	50	56	150	40	48	126

Team A

$$\text{Mean} = \frac{\Sigma fx}{\Sigma f} = \frac{56}{50} = 1.12$$

$$S = \sqrt{\frac{\Sigma fx^2}{\Sigma f} - \left(\frac{\Sigma fx}{\Sigma f}\right)^2}$$

$$= \sqrt{\frac{150}{50} - \left(\frac{56}{50}\right)^2}$$

$$= \sqrt{3 - 1.2544}$$

$$= \sqrt{1.7456} = 1.321$$

$$\text{C.V.} = \frac{S}{\bar{x}} \times 100$$

$$= \frac{1.321}{1.12} \times 100 = 117.95\,\%$$

Team B

$$\text{Mean} = \frac{\Sigma fx}{\Sigma f} = \frac{48}{40} = 1.20$$

$$S = \sqrt{\frac{\Sigma fx^2}{\Sigma f} - \left(\frac{\Sigma fx}{\Sigma f}\right)^2}$$

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

$$= \sqrt{1.71} = 1.308$$

$$\text{C.V.} = \frac{S}{\bar{x}} \times 100$$

$$= \frac{1.308}{1.20} \times 100 = 109.0\%$$

We see that the coefficient of variation for the team B is smaller than that for the team A.

Hence team B is more consistent than team A.

4.5.2 Properties of Variance and Standard Deviation

The variance and standard deviation have the following properties.

(i)   The variance and standard deviation of a constant is zero. If "a" is a constant, then

$$var(a) = 0$$
$$S.D.(a) = 0$$

(ii)   The variance and standard deviation are independent of origin.

$$var(x + a) = var(x)$$
$$var(x - a) = var(x)$$

and $S.D.(x \pm a) = S.D.(x)$

(iii)   When all the values are multiplied with a constant, the variance of the values is multiplied by square of the constant, and their standard deviation is multiplied by the constant.

i.e. $var(ax) = a^2 \, var(x)$

$$var\left(\frac{x}{a}\right) = \frac{1}{a^2} var(x)$$

and $S.D.(ax) = a.S.D.(x)$

$$S.D.\left(\frac{x}{a}\right) = \frac{1}{a} . \, S.D. \, (x)$$

(iv)   The variance of the sum or difference of two independent variables is equal to the sum of their respective variances.

If x and y are two independent variables, then

$$var(x + y) = var(x) + var(y)$$
$$var(x - y) = var(x) + var(y)$$

and $S.D. \, (x + y) = S.D. \, (x) + S.D. \, (y)$

$$S.D. \, (x - y) = S.D. \, (x) + S.D. \, (y)$$

(v)   If sets of data constants of $n_1, n_2, ..., n_k$ values having corresponding means $\bar{x}_1, \bar{x}_2, ..., \bar{x}_k$ and variances $S_1^2, S_2^2, ... S_k^2$, then the combined variance of a set data is given by

$$S_c^2 = \frac{\Sigma n_i [S_i^2 + (\bar{x}_i - \bar{x}_c)^2]}{\Sigma n_i}$$

Where $\bar{x}_c = \dfrac{\Sigma n_i \bar{x}_i}{\Sigma n_i}$

EXAMPLE 4.11

A distribution consists of 3 components with frequency 100, 120 and 150 having means 5.5, 15.8 and 10.5 and standard deviations 2.4, 4.2 and 3.7 respectively. Find the coefficient of variation for the combined distribution.

SOLUTION

$$n_1 = 100, \; n_2 = 120, \; n_3 = 150$$

$$\bar{x}_1 = 5.5, \; \bar{x}_2 = 15.8, \; \bar{x}_3 = 10.5$$

$$S_1 = 2.4, \; S_2 = 4.2, \; S_3 = 3.7$$

Combined mean $= \bar{X}_c = \dfrac{\Sigma n_i \bar{x}_i}{\Sigma n_i}$

$$\bar{X}_c = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + x_3\bar{x}_3}{n_1 + n_2 + n_3}$$

$$= \frac{100(5.5) + 120(15.8) + 150(10.5)}{100 + 120 + 150}$$

$$= \frac{550 + 1896 + 1575}{370}$$

$$= \frac{4021}{370} = 10.87$$

$$S_c^2 = \frac{\Sigma n_i [S_i^2 + (\bar{x}_1 - \bar{X}_c)^2]}{\Sigma n_i}$$

$$S_c^2 = \frac{n_1[S_1^2 + (\bar{x}_1 - \bar{X}_c)^2] + n_2[S_2^2 + (\bar{x}_2 - \bar{X}_c)^2] + n_3[S_3^2 + (\bar{x}_3 - \bar{X}_c)^2]}{n_1 + n_2 + n_3}$$

$$= \frac{100[5.76 + (5.5 - 10.87)^2] + 120[17.64 + (15.8 - 10.87)^2] + 150[13.69 + (10.5 - 10.87)^2]}{100 + 120 + 150}$$

$$= \frac{100[5.76 + 28.84] + 120[17.64 + 24.30] + 150[13.69 + 0.137]}{370}$$

$$= \frac{100(34.6) + 120(41.94) + 150(13.827)}{370}$$

$$= \frac{3460 + 5032.8 + 2074.05}{370}$$

$$= \frac{10566.85}{370} = 28.56$$

$$S_c = \sqrt{28.56} = 5.34$$

$$\text{Combined C.V. } = \frac{S_c}{\overline{x}_c} \times 100$$

$$= \frac{5.34}{10.87} \times 100 = 49.13\%$$

4.6 MOMENTS

The measure of location alongwith measure of variability are useful to describe a data set but fail to tell anything about the shape of the distribution. For this purpose we need to define certain other measures. Important measures about the shape of the distribution is called moments. These measures are discussed as skewness and Kurtosis.

Moments are defined as the arithmetic means of the powers of the deviations of the observations from any value. These are denoted by u_1, u_2, u_3, u_4 for population data and m_1, m_2, m_3 and m_4 etc for sample data.

4.6.1 Moments about Mean

The rth sample moments about mean is denoted by m_r is given by

$$m_r = \frac{\Sigma(x - \overline{x})^r}{n} \text{ , (for ungroup data)}$$

$$m_r = \frac{\Sigma f(x - \overline{x})^r}{\Sigma f} \text{ , (for grouped data)}$$

where r = 1, 2, 3, 4,.........

m_1 is always zero while m_2 is called variance

$$m_1 = \frac{\Sigma(x - \overline{x})^1}{n} = 0$$

$$m_2 = \frac{\Sigma(x - \overline{x})^2}{n} = \text{variance}$$

$$m_3 = \frac{\Sigma(x - \overline{x})^4}{n} \text{ and } m_4 = \frac{\Sigma(x - \overline{x})^4}{n}$$

EXAMPLE 4.12

Calculate first four moments about mean for the following data given below.

4, 7, 5, 9, 8, 3, 6

SOLUTION

x	$x - \bar{x}$	$(x - \bar{x})^2$	$(x - \bar{x})^3$	$(x - \bar{x})^4$
4	-2	4	-8	16
7	1	1	1	1
5	-1	1	-1	1
9	3	9	27	81
8	2	4	8	16
3	-3	9	-27	81
6	0	0	0	0
42	0	28	0	196

First of all we calculate mean as

$$\bar{x} = \frac{\Sigma x}{n} = \frac{42}{7} = 6$$

$$m_r = \frac{\Sigma(x - \bar{x})^r}{n} \; ; r = 1, 2, 3, 4$$

$$m_1 = \frac{\Sigma(x - \bar{x})}{n} = \frac{0}{7} = 0$$

$$m_2 = \frac{\Sigma(x - \bar{x})^2}{n} = \frac{28}{7} = 4$$

$$m_3 = \frac{\Sigma(x - \bar{x})^3}{n} = \frac{0}{7} = 0$$

$$m_4 = \frac{\Sigma(x - \bar{x})^4}{n} = \frac{196}{7} = 28$$

EXAMPLE 4.13

Calculate first four moments about mean from the following data

SOLUTION

Classes	f	x	fx	$x - \bar{x}$	$f(x - \bar{x})$	$f(x - \bar{x})^2$	$f(x - \bar{x})^3$	$f(x - \bar{x})^4$
20 - 24	1	22	22	−17	−17	289	−4913	83521
25 - 29	4	27	108	−12	−48	576	−6912	82944
30 - 34	8	32	256	−7	−56	392	−2744	19208
35 - 39	11	37	407	−2	−22	44	−88	176
40 - 44	15	42	630	−3	45	135	405	1215
45 - 49	9	47	423	8	72	576	4608	36864
50 - 54	2	52	104	13	26	338	4394	57122
Total	50	−	1950	−	0	2350	−5250	281050

$$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{1950}{50} = 39$$

$$m_r = \frac{\Sigma f(x - \bar{x})^r}{\Sigma f} \; ; r = 1, 2, 3, \text{ and } 4$$

$$m_1 = \frac{\Sigma f(x - \bar{x})^1}{\Sigma f} = 0$$

$$m_2 = \frac{\Sigma f(x - \bar{x})^2}{\Sigma f} = \frac{2350}{50} = 47$$

$$m_3 = \frac{\Sigma f(x - \bar{x})^3}{\Sigma f} = \frac{-5250}{50} = -105$$

$$m_4 = \frac{\Sigma f(x - \bar{x})^4}{\Sigma f} = \frac{281050}{50} = 5621$$

4.6.2 Moments about an Arbitrary Value

The rth sample moment about any arbitrary origin a is denoted by m'_r is defined as

$$m'_r = \frac{\Sigma (x - a)^r}{n} = \frac{\Sigma D^r}{n}$$

where $D = x - a$

$r = 1, 2, 3, 4, \ldots\ldots$

So

$$m'_1 = \frac{\Sigma(x-a)^1}{n} = \frac{\Sigma D}{n}$$

$$m'_2 = \frac{\Sigma(x-a)^2}{n} = \frac{\Sigma D^2}{n}$$

$$m'_3 = \frac{\Sigma(x-a)^3}{n} = \frac{\Sigma D^3}{n}$$

$$m'_r = \frac{\Sigma(x-a)^4}{n} = \frac{\Sigma D^4}{n}$$

The rth sample moments for grouped data about any arbitrary value a is denoted by m'_r is defined as

$$m'_r = \frac{\Sigma f(x-a)^r}{\Sigma f} = \frac{\Sigma f D^r}{\Sigma f}, r = 1, 2, ,3 ,4$$

$$m'_1 = \frac{\Sigma f(x-a)^1}{\Sigma f} = \frac{\Sigma f D}{\Sigma f}$$

$$m'_2 = \frac{\Sigma f(x-a)^2}{\Sigma f} = \frac{\Sigma f D^2}{\Sigma f}$$

Similar

$$m'_3 = \frac{\Sigma f(x-a)^3}{\Sigma f} = \frac{\Sigma f D^3}{\Sigma f} \text{ and}$$

$$m'_4 = \frac{\Sigma f(x-a)^4}{\Sigma f} = \frac{\Sigma f D^4}{\Sigma f}$$

4.6.3 Relationship of Moments between mean and about Arbitrary Origin

$$m_1 = 0$$
$$m_2 = m'_2 - (m'_1)^2$$
$$m_3 = m'_3 - 3m'_2 m'_1 + 2(m'_1)^3$$
$$m_4 = m'_4 - 4m'_3 m'_1 + 6m'_2(m'_1)^2 - 3(m'_1)^4$$

EXAMPLE 4.14

(a)   Calculate first four moments about origin and convert them into moment about mean.

(b)   Also calculate fist four moments about $x = 5$ and convert them into moments about mean from the following data: 4, 7, 5, 9, 8, 3, 6.

SOLUTION

(a)   Moments about origin

x	x^2	x^3	x^4
4	16	64	256
7	49	343	2401
5	25	125	625
9	81	729	6561
8	64	512	4096
3	9	27	81
6	36	216	1296
42	280	2016	15,316

$$m'_r = \frac{\sum x^r}{n} \; ; r = 1, 2, 3, 4$$

$$m'_1 = \frac{\sum x}{n} = \frac{42}{7} = 6$$

$$m'_2 = \frac{\sum x^2}{n} = \frac{280}{7} = 40$$

$$m'_3 = \frac{\sum x^3}{n} = \frac{2016}{7} = 288$$

$$m'_4 = \frac{\sum x^4}{n} = \frac{15316}{7} = 2188$$

Converting into moments about mean, we have

$$m_1 = 0$$
$$m_2 = m'_2 - (m'_1)^2 = 40 - (6)^2 = 40 - 36 = 4$$
$$m_3 = m'_3 - 3m'_2 m'_1 + 2(m'_1)^3$$
$$= 288 - 3(40 (6) + 2(6)^3$$
$$= 288 - 720 + 432$$
$$m_3 = 0 \text{ and}$$
$$m_4 = m'_4 - 4m'_3 m'_1 + 6m'_2 (m'_1)^2 - 3(m'_1)^4$$
$$= 2188 - 4(288) (6) + 6(40) (6)^2 - 3(6)^4$$
$$= 2188 - 6912 + 8640 - 3888$$
$$m_4 = 28$$

(b)   Moments about $x = 5$

x	x – 5	$(x – 5)^2$	$(x – 5)^3$	$(x – 5)^4$
4	–1	1	–1	1
7	2	4	8	16
5	0	0	0	0
9	4	16	64	256
8	3	9	27	81
3	–2	4	–8	16
6	1	1	1	1
42	7	35	91	371

$$m'_r = \frac{\Sigma (x – a)^r}{n} = \frac{\Sigma (x – 5)^r}{n} \; ; r = 1, 2, 3, 4$$

$$m'_1 = \frac{\Sigma (x – 5)}{n} = \frac{7}{7} = 1$$

$$m'_2 = \frac{\Sigma (x – 5)^2}{n} = \frac{35}{7} = 5$$

$$m'_3 = \frac{\Sigma (x – 5)^3}{n} = \frac{91}{7} = 13$$

$$m'_4 = \frac{\Sigma (x – 5)^4}{n} = \frac{371}{7} = 53$$

Converting into moments about mean, we have

$m_1 = 0$

$m_2 = m'_2 – (m'_1) = 5 – (1)^2 = 4$

$m_3 = m'_3 – 3m'_2 m'_1 + 2(m'_1)^3$
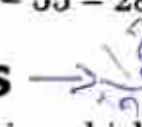
$13 – 15 + 2 = 0$

$m_4 = m'_4 – 4m'_3 m'_1 + 6m'_2 (m'_1)^2 – 3(m'_1)^4$

$= 53 – 4 (13) (1) + 6 (5) (1)^2 – 3(1)4$

$= 53 – 52 + 30 – 3$

$m_4 = 83 – 55 = 28$

4.7 SKEWNESS

A distribution in which the values equidistant from the mean have equal frequencies is called symmetrical. A distribution is called skewed if it is not symmetrical. A skewed distribution has a curve with a longer tail on any direction. If the right tail is longer than the left tail, the distribution is said to have positive

skewness. If the left tail is longer than the right tail, it is said to have negatively skewed. In positive skewness, the mean is greater than the median and the median is greater than the mode.

    i.e.   Mean > Median > Mode

And in the negatively skewed distribution

        Mode > Median > Mean

Mean > Median > Mode
Positive skewness

Mode   Median   Mean

Mean < Median < Mode
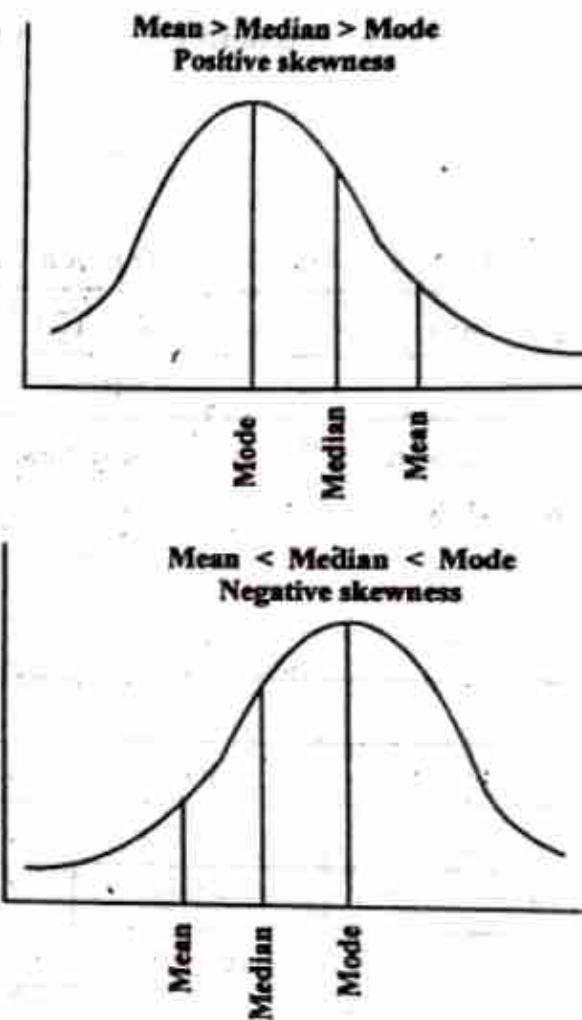**Negative skewness**

Mean   Median   Mode

4.7.1 Measure of Skewness

Karl Pearson introduced a coefficient of skewness denoted by S_k and defined by

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

We know that mode is sometimes ill-defined and it is difficult to locate by simple methods. In such cases the median is calculated and the formula becomes,

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

This coefficient usually varies between −3(negative skewness) and +3(positive skewness)

Another measure of skewness was suggested by Bowley. The Bowley's coefficient of skewness is

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

It values lies between −1 and +1.

Skewness may also be defined by moment ratio given as

$$b_1 = \frac{m_3}{\sqrt{m_2{}^3}}$$

EXAMPLE 4.15

The weight of 38 students at a college are given below.

Weight	118 – 126	127 – 135	136 – 144	145 – 153	154 – 162	163–171
f	3	5	9	12	5	4

(i)   Calculate Karl Pearson coefficient of skewness.

(ii)   Calculate Bowley's coefficient of skewness.

SOLUTION

(i)   Karl Pearson coefficient

Weight	F	x	fx	fx²	C.B	C.f
118 – 126	3	122	366	44652	117.5 – 126.5	3
127 – 135	5	131	655	85805	126.5 – 135.5	8
136 – 144	9	14	1260	176400	135.5 – 144.5	17
145 – 153	12	149	1788	266412	144.5 – 153.5	29
154 – 162	5	158	790	124820	153.5 – 162.5	34
163 – 171	4	167	668	111556	162.5 – 171.5	38
Σ	38	—	5527	809645		

$$\text{Mean} = \bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{5527}{38} = 145.45$$

$$S = \sqrt{\frac{\Sigma fx^2}{\Sigma f} - \left(\frac{\Sigma fx}{\Sigma f}\right)^2}$$

$$= \sqrt{\frac{809645}{38} - \left(\frac{5527}{38}\right)^2}$$

$$= \sqrt{21306.45 - 21154.94}$$

$$= \sqrt{151.51} = 12.31$$

$$\text{Mode} = 1 + \frac{(f_m - f_1)}{(f_m - f_1) + (f_m - f_2)} \times h$$

$$= 144.5 + \frac{(12 - 9)}{(12 - 9) + (12 - 9)} \times 9$$

$$= 144.5 + \frac{3}{3 + 7} \times 9$$

$$= 144.5 + \frac{27}{10}$$

$$= 144.5 + 2.7 = 147.2$$

$$S_k = \frac{\text{Mean} - \text{Mode}}{S}$$

$$= \frac{145.45 - 147.2}{12.31}$$

$$= \frac{-1.75}{12.31} = -0.142$$

(ii)   **Bowlay's Co-efficient**

$$Q_1 = 1 + \frac{h}{f}\left(\frac{n}{4} - C\right)$$

$$Q_1 = \frac{n}{4} \text{ th value}$$

$$= \frac{38}{4} \text{ th value} = 9.5 \text{ th value}$$

9.5th value lie in the group 135.5 – 144.5

$$Q_1 = 135.5 + \frac{9}{9}\left(\frac{38}{4} - 8\right)$$

$$= 135.5 + (1.5)$$

$$= 137.0$$

$$Q_3 = 1 + \frac{h}{f}\left(\frac{3n}{4} - C\right)$$

$$Q_3 = \frac{3n}{4} \text{th value}$$

$$= \frac{3 \times 38}{4} \text{ th value}$$

$$= 28.5 \text{th value}$$

Therefore

$$Q_3 = 144.5 + \frac{9}{12}\left(\frac{3 \times 38}{4} - 17\right)$$

$$= 144.5 + \frac{9}{12}(11.5)$$

$$= 144.5 + 8.625$$

$$= 153.13$$

$$\text{Median} = 1 + \frac{h}{f}\left(\frac{n}{2} - C\right)$$

$$\text{Median} = \frac{n}{2}\text{th value}$$

$$= \frac{38}{2}\text{th value} = 19\text{th value}$$

19th value lie in the group 144.5 − 153.5, therefore

$$\text{Median} = 144.5 + \frac{9}{12}\left(\frac{38}{2} - 17\right)$$

$$= 146.0$$

Bowley's coefficient of skewness

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

$$= \frac{153.13 + 137.0 - 2(146.0)}{153.13 - 137.0}$$

$$= \frac{290.13 - 292}{16.13}$$

$$= -0.116$$

Empirical Relation between Measure of Dispersion

$$\text{Mean deviation} = \frac{4}{5}\,(\text{Standard Deviation})$$

$$\text{Semi-Inter Quartile Range} = \frac{2}{3}\,(\text{Standard Deviation})$$

$$\text{Semi-Inter Quartile Range} = \frac{5}{6}\,(\text{Mean Deviation})$$

SUMMARY

The formula and methods of computing the different types of dispersion and skewness are summarized below:

APPLICATION	FORMULA
Range	$R = $ Maximum value $-$ Min value i.e. $R = X_m - X_o$
Quartile Deviation	$Q.D. = \dfrac{Q_3 - Q_1}{2}$ Where $Q_1 = 1 + \dfrac{h}{f}\left(\dfrac{n}{4}\text{-}C\right)$ $Q_3 = 1 + \dfrac{h}{f}\left(\dfrac{3n}{4} - C\right)$
Coefficient of Q.D.	$\dfrac{Q_3 - Q_1}{Q_3 + Q_1}$
Mean deviation	
Ungroupd data	(i) $M.D = \dfrac{\Sigma\lvert X - \text{Mean}\rvert}{n}$ (ii) $M.D = \dfrac{\Sigma\lvert X - \text{Median}\rvert}{n}$ (iii) $M.D = \dfrac{\Sigma\lvert X - \text{Mode}\rvert}{n}$
Grouped data	(i) $M.D. = \dfrac{\Sigma f\lvert X - \text{Mean}\rvert}{\Sigma f}$ (ii) $M.D. = \dfrac{\Sigma f\lvert X - \text{Median}\rvert}{\Sigma f}$ (iii) $M.D. = \dfrac{\Sigma f\lvert X - \text{Mode}\rvert}{\Sigma f}$
Standard deviation	
Ungrouped data	$S = \sqrt{\dfrac{\Sigma(X - \overline{X})^2}{n}}$ Or $S = \sqrt{\dfrac{\Sigma X^2}{n} - \left(\dfrac{\Sigma X}{n}\right)^2}$
Short cut method	$S = \sqrt{\dfrac{\Sigma d^2}{n} - \left(\dfrac{\Sigma d}{n}\right)^2}$ Where $d = X - a$

Grouped Data	
Direct method	$S = \sqrt{\dfrac{\Sigma f (X - \overline{X})^2}{\Sigma f}}$ Or $S = \sqrt{\dfrac{\Sigma fx^2}{\Sigma f} - \left(\dfrac{\Sigma fx}{\Sigma f}\right)^2}$
Step deviation method	$S = \sqrt{\dfrac{\Sigma fu^2}{\Sigma f} - \left(\dfrac{\Sigma fu}{\Sigma f}\right)^2} \times h$
Coefficient of variation	$C.V = \dfrac{S}{\overline{X}} \times 100$
Variance	Square of standard deviation
Measure of Skewness	
Person's formula	(i) $Sk = \dfrac{\text{Mean} - \text{Mode}}{\text{S.D.}}$ (ii) $= \dfrac{3(\text{mean} - \text{median})}{\text{S.D.}}$

EXERCISES

4.1 What is meant by dispersion? Discuss the various measures of dispersion.

4.2 What is range and how is calculated?

4.3 Find the range and coefficient of range from the following data.

(i) 15, 12, 18, 16, 11, 19, 25, 13, 17, 21

(ii) 105, 103, 110, 108, 106, 115, 110, 109, 102

4.4 (a) Find range from the following frequency distribution.

Classes	70 – 74	75 – 79	80 – 84	85 – 89	90 – 94
f	2	5	12	8	7

(b) Find range and its coefficient from the following data.

Daily wages	120	150	170	200	250	300
f	15	20	27	23	15	10

4.5 (a) Define semi-interquartile range or quartile deviation.

(b) Find the quartile deviation from the following data. Also calculate the coefficient of quartile deviation.

(i) 15, 12, 18, 16, 11, 19, 25, 13, 17, 21

(ii) 105, 103, 110, 108, 106, 115, 110, 109, 102

4.6 Calculate the quartile deviation from the following data.

Mark	30 – 39	40 – 49	50 – 59	60 – 69	70 – 79
f	8	87	190	86	20

4.7 Find the quartile deviation and its coefficient from the following data.

Classes	41 – 50	51 – 60	61 – 70	71 – 80	81 – 90	91– 100
f	30	36	43	104	73	14

4.8 (a) Define mean deviation and its coefficient.

(b) Calculate mean deviation from the following data.

9, 2, 6, 12, 8, 13, 23, 16, 6, 5

(c) Calculate mean deviation from median from the following data.

9, 2, 6, 12, 18, 13, 23, 16, 6, 5

4.9   Calculate mean deviation from mean and its coefficient from the data given below.

Income	35 – 39	40 – 44	45 – 49	50 – 54	55 – 59	60 – 64	65 – 69
F	13	15	28	17	12	10	5

4.10   Calculate mean deviation from median.

Classes	86 – 90	91 – 95	96 – 100	101– 105	106– 110	111– 115
f	6	4	10	6	3	1

4.11   Calculate mean deviation from mode from the data given below. Also calculate its coefficient.

Groups	1 – 3	3 – 5	5 – 7	7 – 9	9 – 11	11 – 13	13 – 15	15 – 17
F	6	53	83	56	21	26	4	4

4.12   **(a)**   Define variance and standard deviation. Describe their properties.

(b)   Calculate variance and standard deviation from the following data.

  (i)   1, 2, 3, 4, 5

  (ii)   3, 5, 7, 13, 15, 17, 23, 27

  (iii)   10, 8, 7, 9, 5, 12, 8, 6, 8, 2

4.13   Calculate the variance and standard deviation of the data given below. Also calculate the coefficient of variation.

x	63	64	65	66	67	68	69	70	71
f	4	6	10	20	30	13	12	3	2

(b)   Calculate variance and standard deviation from the following frequency distribution.

x, fx, fx^2

Classes	1-3	4-6	7-9	10-12	13-15
f	1	2	3	2	1

4.14   Calculate the variance and standard deviation for weight distribution of 120 students at the Govt. College, Lahore.

Weight	45 – 49	50 – 54	55 – 59	60 – 64	65 – 69	70 – 74
f	1	4	17	28	25	18
Weight	75 – 79	80 – 84	85 – 89	90 – 94	95 – 99	
f	13	6	5	2	1	

4.15 Determine the variance, standard deviation and coefficient of variation of the following data.

Classes	20 – 24	25 – 29	30 – 34	35 – 39	40 – 44	45 – 49	50 – 54
f	1	4	8	11	15	9	2

4.16 The following frequency distribution shows the marks of students.

Marks	30 – 39	40 – 40	50 – 59	60 – 69	70 – 79	80 – 89
f	10	14	26	20	18	12

Using the transformation $u = \dfrac{x - 54.5}{10}$.

Calculate the variance and coefficient of variation.

4.17 It is often stated that in frequency distribution there exists the approximate relation $\dfrac{\text{Mean deviation}}{\text{Standard deviation}} = 0.8$. Test this statement in the following distribution.

Weight (grams)	65 – 84	85 – 104	105 – 124	125 – 144
f	9	10	17	10
Weight (grams)	145 – 164	165 – 184	185 – 204	
f	5	4	5	

4.18 The breaking strength of 20 test pieces of a certain alloy is given as under.

95, 103, 97, 130, 76, 73, 78, 95, 89, 68, 82, 79, 69, 67, 83, 108, 94, 87, 93, 117

Calculate the average breaking strength of the alloy and standard deviation. Calculate the percentage of observations lying within the limit : mean ±S, mean ± 2S, mean ± 3S, where S stands for standard deviation.

4.19 What do you understand by variance? The wages of 1000 employees ranges from Rs. 4.50 to Rs. 19.50. They are grouped in 15 classes with a common class interval of Re. 1, and the class frequencies from the lowest class to the highest class are 6, 17, 35, 48, 65, 90, 131, 173, 155, 117, 75, 52, 21, 9 and 6. Find the mean wage and its standard deviation.

4.20 (a) Find the arithmetic mean and the standard deviation of the observations, 40, 40, 50, 60, 70, 70, 80, 80, 90. Also find the mean and standard deviation after increasing the observation by

(i) 10 units

    (ii)   10 percent

(b) What will be the standard deviation and the variance in each of the following cases.

    (i)   2x   (ii)   x + 2

    (iii)   2x + 4

    If var(x) = 25.

4.21 Compute the mean wages and co-efficient of variation for the employees working in two factors are given below.

Wages	Factories	
	Factory A	Factory B
30 – 35	12	4
35 – 40	18	10
40 – 45	29	31
45 – 50	32	67
50 – 55	16	35
55 – 60	8	15

4.22 Compare the variability of expenditure in two towns as given below.

Expenditure (Rupees)	No. of Families	
	Town A	Town B
21 – 30	3	2
31 – 40	61	14
41 – 50	132	20
51 – 60	153	27
61 – 70	140	28
71 – 80	51	7
81 – 90	2	2

4.23 The following are the marks of two students in different subjects.

A	75	60	72	75	56	82	50
B	90	35	95	88	19	60	76

Who is better student? Who is more consistent student?

standard deviation is 8 on the same test. Find the mean and standard deviation of the combined group of 100 children.

(b)   Coefficient of variation of two series are 75% and 90% and their standard deviations are 15 and 18 respectively. Find their mean.

4.25   Given the following data. Calculate the variance and standard deviation by step deviation method.

Classes	20 – 24	25 – 29	30 – 34	35 – 39
f	1	4	8	11
Classes	40 – 44	45 – 49	50 – 59	
f	15	9	2	

4.26   (a)   What is meant by skewness? Distinguish between positive and negative skewness.

(b)   What can you say of skewness in each of the following distributions?

   (i)   Mean = 19, Mode = 52
   (ii) $Q_1 = 136$, Median = 160, $Q_3 = 184$
   (iii)   Mean = 78, Median = 61

4.27   The heights of 100 college students measured to nearest inch are given below.

Height	60-62	63-65	66-68	69-71	72-74
f	5	18	42	27	8

Calculate Coefficient of skewness by Karl Person method.

4.28   Find the coefficient of skewness by Bowley's formula from the following frequency distribution and interpret the result.

Age Group	0 – 10	10 – 20	20 – 30	30 – 40
f	18	16	15	13
Age Group	40 – 50	50 – 60	60 – 70	70 – 80
f	10	5	2	1

4.29   Calculate

   (i)   Bowley's coefficient of skewness
   (ii)   Karl Pearson's coefficient of skewness from the following data.

Classes	20 – 24	25 – 29	30 – 34	35 – 39	40 – 44	45 – 49	50 – 54
Cumulative frequency	22	50	268	495	730	946	1000

4.30 The daily income of employees range from Rs. 0 to Rs. 18. They are grouped in intervals of Rs. 2 and class frequencies from lowest to the highest class are 5, 39, 69, 41, 29, 22, 16, 7, 5. Find coefficient of skewness.

4.31

Classes	2.5 – 2.9	3.0 – 3.4	3.5 – 3.9	4.0 – 4.4	4.5 – 4.9
f	18	22	36	48	42
Classes	5.0 – 5.4	5.5 – 5.9	6.0 – 6.4	6.5 – 6.9	7.0 – 7.4
f	26	14	10	03	01

Required
(i)   Standard Deviation
(ii)   Variance
(iii)   Pearsonian measure of skewness

4.32 (a)   What do you meant by moments?
      (b)   In a certain distribution, the first and second moments about value 2 are −1 and 16 respectively. find the mean and variance of the distribution.
      (c)   Find b_1 and b_2, given that first four raw moments are:
            $m'_1 = -2$, $m'_2 = 15$, $m'_3 = -25$, and $m'_4 = 80$

4.33 (a)   Given the following information. $\Sigma f = 290$ $\Sigma fx = 2610$, $\Sigma fx^2 = 23780$, $\Sigma fx^3 = 219530$ and $\Sigma fx^4 = 2056100$. Calculate first four moments about mean.
      (b)   First three moments of distribution about $x = 4$ are 1,4 and 10 respectively. Is the distribution symmetrical or skewed.
      (c)   Lower and upper quartiles of a distribution are 142.36 and 167.73 respectively. While median is 153.50. Find the coefficient of Skewness.

4.34 (a)   Calculate first four moments about mean. Also calculate Skewness and Kurtosis for the following data: 1, 2, 3, 4, 5.
      (b)   Calculate first four moments about mean.

Class	f
1 – 3	1
4 – 6	2
7 – 9	3
10 – 12	2
13 – 15	1

4.35 Calculate the first four moments about the mean for the following data. Also calculate Skewness and Kurtosis.

x	1	2	3	4	5	6	7	8	9
f	1	6	13	25	30	22	9	5	2

4.36 SELECT THE CORRECT ANSWER:

(i)   Which one of the following is not a measure of dispersion?

    (a)   Range   (b)   Standard deviation

    (c)   Second quartile   (d)   Variance

(ii)   The standard deviation is

    (a)   Square of the variances   (b)   Half of the Variance

    (c)   Two times standard deviation (d)   Square root of the Variance

(iii)   The Coefficient of variation is measure in

    (a)   The same unit as the means and Standard deviation

    (b)   Percent

    (c)   Square unit   (d)   None of these

(iv)   If the tail of a frequency distribution is to the right, the Coefficient of Skewness is

    (a)   Zero   (b)   Positive

    (c)   Negative   (d)   None

(v)   Symmetrical distribution will always have skewness equal to.

    (a)   Zero   (b)   Negative

    (c)   Positive   (d)   None

(vi)   If a distribution has zero variance, then which of the following is true?

    (a)   All of the observations are negative

    (b)   All the observations are positive

    (c)   All the observations are equal

    (d)   None of these

(vii)   For a normal distribution, the measure of kurtosis equal to

    (a)   Zero   (b)   3

    (c)   Negative   (d)   Positive

(viii) If the original units are measure in kg, the variance is

    (a)   Also measured in Kg   (b)   Measure in Kg squared

    (c)   Measure in half Kg   (d)   None of the these

(ix) If standard deviation of frequency distribution is 10, means is 40 and mode is also 40, then coefficient of Skewness is:

(a) Zero

(b) Positive

(c) Negative

(d) None of these

ANSWERS 4.36

(i)	(c)	(ii)	(d)	(iii)	(b)	(iv)	(b)	(v)	(a)
(vi)	(c)	(vii)	(b)	(viii)	(b)	(ix)	(a)		

SHORT QUESTION AND ANSWERS

1. **Define dispersion.**

Ans: By dispersion we mean the degree to which numerical data tends to spread about an average value.

2. **What is the difference between absolute dispersion and relative dispersion?**

Ans: **Absolute Dispersion:**

An absolute dispersion is one that measure the dispersion in term of same units. e.g. if the units of data are rupees, kilograms etc, the units of the measure of dispersion will also be rupees, kilograms etc.

Relative Dispersion:

It is expressed in the form of ratio and coefficients. It is independent of the units of measurements.

3. **What are the main measures of dispersion?** Type

Ans: The main measures of dispersion are the followings:

(i) The Range

(ii) The semi inter quartile range or quartile deviation.

(iii) The mean deviation

(iv) The variance and standard deviation

4. **Define Range**

Ans: Range is defined as the difference between the largest and the smallest observation in a set of data

$$\text{Range} = R = X_m - X_o$$

$$\text{Coefficient of Range} = \frac{X_m - X_o}{X_m + X_o}$$

5. **What do you meant by Quartile deviation?**

Ans: Quartile deviation is defined as half of the difference between the third and the first quartile. *i.e.*

$$Q.D = \frac{Q_3 - Q_1}{2}$$

6. **Define mean deviation.**

Ans: Mean deviation is defined as the mean of the absolute deviations measured either from mean, median or mode. By absolute deviation we mean that all the deviations are positive.

$$M.D. = \frac{\Sigma|X - Mean|}{n}$$

or

$$M.D. = \frac{\Sigma|X - Median|}{n}$$

or

$$M.D. = \frac{\Sigma|X - Mode|}{n}$$

7. **Define variance.**

Ans: The variance is defined as the mean of the squares of deviations of all the observations from their mean.

or

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

8. **Define standard deviation.**

Ans: The positive square root of the variance is called standard deviation symbolically.

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

9. **What do you know about coefficient of variation?**

Ans: It is used to compare the variations in two or more than two sets of data. The group which has lower value of coefficient of variation is comparatively more consistent. The coefficient of variation is defined as:

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

10. **What are the properties of variance.**

Ans: The variance has the following properties:

(i)   The variance of a constant is zero, var (a) = 0.

(ii)   var $(x \pm a)$ = var (X)

(iii)   var $(ax) = a^2$ var$^2(x)$

(iv)   var $(x \pm y)$ = var (x) + var (y)

11.   Define symmetrical distribution.

Ans: A distribution in which the values equidistant from the mean have equal frequency is called symmetrical distribution.

12.   What do you meant by skewness?

Ans: A distribution is called skewed if it is not symmetrical.

13.   Define positive skewness.

Ans: If the right tail is longer than the left tail, the distribution is said to have positive skewness. In case of positively skewness:

$$\text{Mean} > \text{Median} > \text{Mode}$$

14.   Define negative skewness.

Ans: If the left tail is longer than right tail, it is called negative skewness. In case of negative skewness.

$$\text{Mode} > \text{Median} > \text{Mean}$$

15:    What is the measure of Skewness?

Ans: (i) Karl Person Coefficient of Skewness:

$$S_k = \frac{\text{Mean} - \text{Mode}}{S}$$

(ii) Bowley coefficient of Skewness:

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

16.   Define moments.

Ans: Moments are defined s the arithmetic means of the powers of the deviations of observations from any value.

17.   What do you mean by kurtosis?

Ans: It is the degree of peakness or flatness of unimodel frequency curve.

18.   Define moment ratio.

Ans: The ratio between the moment about mean are called moment ratio. i.e.

$$b_1 = \frac{m_3^2}{m_2^3} \text{ and } b_2 = \frac{m_4}{m_2^2}$$

❀ ⌘ ⌘ ❀