Chapter - III
MEASURES OF CENTRAL TENDENCY

Meaning and Introduction

We have learnt in the previous chapter as to how large masses of primary data can be classified, tabulated and presented to enable the user to analyze the data. While this data is very useful for the statistician, the common man would look for a number that is representative of the entire data. In the words of Prof. R.A. Fisher, "the inherent inability of the human mind to grasp in its entirety a large body of numerical data compels us to seek relatively few constants that will adequately describe the data". Such numbers or values should summarize the characteristics of the entire data comprising the set of unequal values. This single numerical value is called an average. For example, if a person were studying the heights of students studying in various classes, it is possible to collect the heights of each and every student in each class. However, it is more meaningful to find out one number or value that can represent one class. We can consider the "average" height of each class as one such value.

An average is a representative figure, which is a 'gist', if not the substance of statistics. It is a single value around which other items of the distribution congregate. They are the values which lie between the two extreme observations, (i.e., the smallest and the largest observations), of the distribution and give us an idea about the concentration of the values in the central part of the distribution. Accordingly they are also sometimes referred to as the Measures of Central Tendency.

Definition

The terms "average" of "measures of central tendency" have been defined by various authors. Some of the definitions are given below.

According to **Simpson and Kafka** "a measure of Central Tendency is a typical value around which other figures congregate". It is a single value that represents a whole series. Its value always lies between the minimum and maximum values and is generally located in the centre or middle of the distribution.

In the words of **Ya Lun Chou**, "an average is a typical value in the sense that it is sometimes employed to represent all the individual values in a series of a variable."

"Averages are statistical constant which enable us to comprehend in a single effort the significance of the whole". - Bowley

"An average value is a single value within the range of the data that is used to represent all of the values in the series. Since an average is somewhere within the range of the data, it is sometimes called a measure of central Value". - **Croxton and Cowden**

Lawrence J. Kaplan states that "One of the most widely used set of summary figures is known as measure of location which is often referred to as averages, central tendency or central location. The purpose of computing an average value for a set of observations is to obtain a single value which is representative of all the items and which the mind can grasp simply and quickly. The single value is the point of location around which the individual items cluster".

Thus, from the above definitions, it is clear that an average is single value which represents a whole series.

Objectives / Functions / Uses of Averages:

Averages serve a very useful purpose in Statistics. Following are the objectives of Averaging data:

1. Represent the characteristics of the entire mass of data: An average reduces a complex mass of data into a single typical figure to enable the user to get a bird's eye view about the characteristics of the phenomenon under study. In the words of M.J. Mooney, "the purpose of an average is to represent a group of individual values in a simple and concise manner, so that the mind can have a quick understanding of the general size of the individuals in the group". The objective of an average is to enable the user to make inferences about the entire group by looking at the average. It should facilitate easy expression, reference and interpretation.

2. To facilitate comparison: An Average reduces a mass of complex data into one single figure. This makes it possible for the user to make comparisons between two or more sets of data and draw conclusions the characteristics of the separate sets of data. To illustrate, we can compare the average marks obtained by students from different colleges in a given examination and determine the best college. Similarly, average monthly sales of a product can help us understand if there is an impact of different seasons on the sales.

3. Describe Data: Average serve the objective of describing the main characteristics of the underlying data in a simple and brief manner.

4. To quantify relationships: One of the functions of an average is to help in establishing relationships between different groups in quantitative terms. For example, stating that income of an average American is more than that of an average Indian may not mean much. It is more meaningful if the respective incomes are expressed in terms of averages.

4. To reduce the impact of extreme values: Average help in setting off the impact of extreme values. For example, if a batsman has scored 300 and 10 runs respectively in two innings of a test match, the average score will read as 155. This is far more critical when number of observations is large.

5. Basis for further Analysis: Statistical analysis covers number of measures such as standard deviation, correlation, regression, etc to name a few. All these calculations are based on the "average". Thus, averages serve the purpose of facilitating further analysis.

6. To facilitate Sampling: We have briefly learnt about Sampling and various methods of sampling in the previous chapter. The concept of Sampling is to make inferences about a Universe (population) forming a sample as the basis. This objective is achieved only with the help of averages.

7. Facilitate Decision making: A business manager is interested in numbers such as sales, production, Occupancy (in case of hotel industry), capacity utilization (particularly in case of travel industry - Air, Train, bus), etc. All these numbers are critical for decision making and policy formulation. However, none of these numbers can be used as is. To facilitate decision making, all these numbers are stated in averages.

In short, Averages are very much useful (i) for describing the distribution in concise manner (ii) For comparative study of different distributions (iii) for computing various other statistical measures and lastly (iv) to take informed decisions.

Limitations of Averages

Averages are very useful Statistical tools, but they also suffer from certain limitations. The following are the limitations of Averages:

1. Misleading Conclusions: Average is a single numerical figure representing the characteristics of a given distribution. This number is vulnerable to errors in interpretation and can lead to misleading conclusions. For example, if the maximum and minimum temperatures of a particular city are 48^0 F and 2^0 F respectively, the average may still work out to 24^0 F. Looking at the average, the weather conditions might look very comfortable, but that is not really the case.

2. Choice of Average: There are different types of averages. Different types of averages are suitable for different objectives. The utility of average depends on a proper and judicious choice of the average. A wrong choice of the average might lead to erroneous results.

3. Incomplete Picture: An average does not provide the complete picture of a distribution. There may be number of distribution having the same average but differing widely in their structure and constitution. To form a complete idea about the distribution, the measures of central tendency are to be supplemented by some more measures such as dispersion, skewness and kurtosis.

4. Inadequate Representation of Data: In certain types of distribution like U-shaped or J-shaped distribution, an average, which is only a single point of concentration, does not adequately represent the centre series.

5. Absurd Conclusions: Sometimes, an average might throw up very absurd results that are not realistic. For example, the average size of a family might work out to a fractional number, which is not realistic.

Characteristics of a Good Average

The impact of some of the aforementioned limitations is not significant if the average being used has certain desirable characteristics. The following are the essentials of a good average:

1. It should be rigidly defined: An average should be rigidly defined so that there is no scope for confusion or manipulation. The average value will become very unstable and non representative of the base data if it is not well defined. It is ideal to use an average that is mathematically defined by way of formula.

2. It should be easy to calculate and simple to follow: Calculation of an average should be simple to understand. It should be easy to calculate, preferably without the help of calculators. If an average is too complex to understand and calculate, its use will be very limited. It should also be capable of expression in simple numerical terms without advanced mathematical intricacies.

3. It should be based on all observations in the series: An average will be truly representative of the whole mass of data if it is computed from all the observations.

4. It should not be affected much by an extreme values: An average will be representative of the data only if it can set off extreme values against each other. A few very small or very large observations should not unduly affect the value of a good average.

5. It should be capable of further statistical processing: An average should be capable of being used in calculation of other statistical measures such as standard deviation, correlation, etc.

6. It should be capable of further algebraic treatment: An average should lend itself readily to further algebraic treatment. For example, if averages of two or more sets of data are known, it should be possible to obtain the average of combined group.

7. It should posses sampling stability: An average should be least affected by sampling. If we take independent random samples of the same size from a given population and compute the average for each of these samples, the values so obtained from different samples should not be very different from one another.

8. It should be representative of the data: A good average should represent maximum characteristics of the underlying data.

Besides the above, a good average should ideally be nearest to most items of the series, should be capable of being ascertained through graphic methods and should be popular. Lastly, an average should be meaningful. For example, it is absurd to talk of an average of man's height and weight.

Types of Averages

Averages are of the following types:

A) Mathematical Averages:

I) Arithmetic Average or Mean
II) Geometric Mean
III) Harmonic Mean

B) Averages of Position:

I) Median
II) Mode

Of the above, Arithmetic Mean, Median and Mode are most widely used, followed by Geometric Mean and Harmonic Mean. Let us understand each of these Averages.

ARITHMETIC MEAN

Arithmetic Average or Mean of a series is the figure obtained by dividing the total values of the various items by their number. In other words it is the sum of the values divided by their number. Arithmetic means is the most widely used measure of central tendency.

Merits of Arithmetic Average

1. It is rigidly defined. Hence, different interpretations by different persons are not possible.
2. It is easy to understand and easy to calculate. In most of the series it is determinate and its value is definite.
3. It takes all values into consideration. Thus, it is more representative.
4. It can be subjected to further mathematical treatment. The properties of Arithmetic mean are separately explained elsewhere in the chapter.
5. It is used in the computation of various other statistical measures.
6. It is possible to calculate arithmetic average even if some of the details of the data are lacking. For example, it can be known even when only the number of items and their aggregate value are known, and details of various items are not available. Similarly, the aggregate value of items can be calculated if the number of items and the average are known.
7. Of all averages, arithmetic average is least affected by fluctuations of sampling. Thus, it is the most stable measure of central tendency.
8. It provides a good basis for comparison.
9. Arithmetic mean is impacted by every observation. It gives **weight to all items** in direct proportion to their size.

Demerits of Arithmetic Average

While Arithmetic mean satisfies most of the conditions of an ideal average, it suffers from certain drawbacks. Some of the demerits or limitations of Arithmetic Mean are listed below:

1. It cannot be determined by inspection.
2. It cannot be located graphically.
3. It cannot be used in the study of qualitative phenomena.
4. It can be significantly impacted by extreme values and may lead to erroneous conclusions. Abnormal items may considerably affect this average, particularly when the number of items is not large. Thus, it is desirable not to use arithmetic average when the distribution is unevenly spread.

5. It cannot be calculated if the distribution has open-ended classes i.e., "below 10" or "above 80".
6. It cannot be calculated even if a single observation is missing or lost or is illegible.
7. Arithmetic average can be a figure that does not exist in the series at all. Hence, it may be considered as a fictitious average.
8. It is not suitable in case of an extremely asymmetrical (skewed) Distribution. In fact, it is suitable only if the distribution is normally distributed.
9. It may result in misleading conclusions if the details of the data, from which it is computed, are not given. For example, if the profits of a company A in the last three years are Rs.20,000, Rs.25,000 and Rs.45,000 respectively and the profits of another company B in the last three years are Rs.40,000, Rs.35,000 and Rs.1,55,000 respectively, the average profits for the last 3 year period for both companies will work out to Rs. 30,000. This may result in the conclusion that the level of profitability is similar. This is misleading as the profit trends are very different.
10. Arithmetic mean has an upward bias. It gives greater importance to higher value of a series and lesser importance to lower values. It has an upward bias. One big observation among five items, four of which are small, will push up the average considerably. However, in a series of five items, if there are four big values and one small value, the average will not be pulled down very much.
11. It may sometimes lead to absurd conclusions. For example, the average size of a family might work out to a fractional number, which is not realistic.

Calculations of Arithmetic mean

Individual Observations

(a) Direct Method

$\overline{X} = \frac{\sum X}{N}$, where

$\overline{X}$ = Arithmetic Mean;

N = No. Of items

$\sum X$ = Total of the size of items

Steps: (1) Add all the values of the variable X and denote the total as $\sum X$

(2) Divide this total by number of items

(b) Short-Cut Method:

$\overline{X} = A + \frac{\sum d}{N}$, where

$\overline{X}$ = Arithmetic Mean;

A = Assumed Mean;

N = No. of Iteam

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

Steps: (1) Take an assumed mean

(2) Take the deviations of items from the assumed mean and denote the deviations as 'd'.

(3) Obtain the total of deviations and denote it as $\sum d$ and apply the formula.

Illustration 1

Calculate the arithmetic mean of the monthly incomes of the families in a certain locality in Hyderabad.

Family	A	B	C	D	E	F	G	H	I	J
Income	90	75	60	100	125	50	80	120	500	400

Solution: Calculation of Arithmetic Mean

(a) Direct Method

Family	Income Rs.
A	90
B	75
C	60
D	100
E	125
F	50
G	80
H	120
I	500
J	400
N=10	$\sum x = 1600$

$$\overline{X} = \frac{\Sigma X}{N} = \frac{1600}{10} = 160$$

(b) Short-Cut Method

Let us take 150 as assumed Mean. It is not necessary that assumed mean should be one of the items of the series. Any number can be taken as assumed mean

Calculation of Arithmetic Mean (Short-Cut Method)

Family	Income (Rs.)	d = (X-150)
A	90	-60
B	75	-75
C	60	-90
D	100	-50
E	125	-25
F	50	-100
G	80	-70
H	120	-30
I	500	+350
J	400	+250
N=10		$\sum d = +100$

$$X = A + \frac{\Sigma d}{N}, = 150 + \frac{100}{10} = \text{Rs.}160$$

Discrete Series:

Direct Method $\overline{X} = \frac{\Sigma fx}{N}$, where

$\sum fx$ = Total of the size multiplied by their respective frequency.

N = Total of frequency

Steps: (1) Multiply the size of items with its frequency and obtain the total. Denote is as $\sum fx$.

(2) Divide the total so obtained by the total of frequency.

Short-Cut Method

$\overline{X} = \frac{\Sigma fx}{N}$, where

A = Assumed Mean

$\sum fd$ = Total of the deviations from assumed mean multiplied with frequency.

N = Total of frequency.

Steps: (1) Take an assumed mean.

(2) Calculated the deviations of the size from the assumed mean.

(3) Multiply these deviations with respective frequencies and find out the total. Denote it as $\sum fd$.

Illustration 2 : From the following data of the marks obtained by 80 students of a class calculate the arithmetic mean.

Marks	10	20	30	40	50	60
No. of students	5	15	25	20	10	5

Solution : Calculation of Arithmetic Mean (Direct Method)

Marks (X)	No. of students	*fx*
10	5	50
20	15	300
30	25	750
40	20	800
50	10	500
60	5	300
	N=80	$\sum f$=2700

$\bar{X} = \frac{\Sigma fx}{N} = \frac{2700}{080} = 33.75$

Short-Cut Method: Let us take 40 as assumed mean

Calculation of Arithmetic mean (Short-Cut Method)

Marks X	No. of Students f	X-40 (d)	*fd*
10	5	-30	-150
20	15	-20	-300
30	25	-10	-250
40	20	0	0
50	10	+10	+100
60	5	+20	+100
	N=80		$\sum fd$=-500

$\bar{X} = A + \frac{\Sigma fd}{N}, = 40 + \frac{-500}{80} = 40\text{-}6.25 = 33.75$

Note: Such problems can also be worked out by taking common factor

Continuous Series:

Direct Method $\bar{X} = \frac{\Sigma f\cancel{x}m}{N}$, where

$\sum fm$ = Total of frequency multiplied with mid-values

N = Total of frequency

Steps: (1) Obtain the mid-point of each class and denote it as 'm'.

(2) Multiply the mid-points by respective frequencies, Obtain the total Denote it as $\sum fm$.

(3) Divide the total so obtained by the Total of frequency.

Short – Cut Method :

$$\bar{X} = A + \frac{\Sigma fd^2}{N} \times c$$

A = Assumed mean,

$\sum fd^1$ = Total of the deviations from assumed mean (after taking common factor) multiplied with frequency.

N = Total of frequency

C = Common factor

Steps: (1) Ascertain the mid-values of each class (m).

(2) Take an assumed average (A)

(3) Calculate the deviations from assumed average (mid-value minus assumed mean) (d)

(4) Ascertain the common factor and divide all deviations by the common factor. (d^1)

(5) Multiply the frequency with deviations (after taking common factor i.e.D^1) and obtain the total and devote is as $\sum fd^1$.

(6) Ascertain the total of frequency and apply the formula.

Step-deviation method: This method is not applicable to all the problems. This method can only be applied to the problem where any common factor can be taken in dx and we get $dx^1 = \frac{dx}{i}$ where i is the common factor. For above given example this method can be applied. Here we use the formula

$$\bar{X} = A + \frac{\sum fdx^1}{N} \times i$$

Where A is the Assumed Mean

F the frequency; I is the common factor

$N = \sum f$ and dx' is $= \frac{dx}{i}$

Illustration 3 : (Exclusive Class-intervals)

Calculate the arithmetic mean from the following table:

Monthly wages of domestic servants	No. of Servants	Monthly wages of domestic servants	No.of servants
0-10	1	50-60	35
10-20	4	60-70	10
20-30	10	70-80	7
30-40	22	80-90	1
40-50	30		

(C.A. Inter)

Solution: Calculation of Arithmetic Mean (Direct Method)

Monthly Wages	No. of Servants	Mid-Points	freq x midpoints
	f	m	fm
0-10	1	5	5
10-20	4	15	60
20-30	10	25	250
30-40	22	35	770
40-50	30	45	1350
50-60	35	55	1925
60-70	10	65	650
70-80	7	75	525
80-90	1	85	85
	N=120		**$\sum fm$=5620**

$$\bar{X} = \frac{\sum fm}{N} = \frac{5620}{120} = 46.83$$

Shrot-Cut Method:

Let us take 45 as assumed mean.

Calculation of Arithmetic Mean (Short-Cut Method)

Wages	Mid-Values	No. of Workers	Deviation Step from assumed mean (A=45)	deviation (d/10)	Freq.d'
(x)	m	f	d	d'	fd^1
0-10	5	1	-40	-4	-4
10-20	15	4	-30	-3	-12
20-30	25	10	-20	-2	-20
30-40	35	22	-10	-1	-22
40-50	45	30	0	0	+30
50-60	55	35	+10	+1	+35
60-70	65	10	+20	+2	+20
70-80	75	7	+30	+3	+21
80-90	85	1	+40	+4	+4
		N=120			$\sum fd^1$=22

$$\overline{X} = A + \left(\frac{\Sigma fd^1}{N} \times c\right) = 45 + \left(\frac{22}{120} x10\right) = 46.83$$

Illustration 4 : (Inclusive class-intervals)

Class-interval	50-59	40-49	30-39	20-29	10-19	0-9
Frequency	1	3	8	10	15	3

Solution:

Note: While Calculating mean it is not necessary to rearrange the series in ascending order.

Mid-Value = (Lower limit + Upper limit) / 2

Calculation of Arithmetic Mean

Class-Interval	Mid-Values	No. of Workers	Deviation Step from assumed mean (A=34.5)	deviation (d/10)	Freq x d'
(x)	m	f	d	d1	fd1
50-59	54.5	1	+20	+2	+2
40-49	44.5	3	+10	+1	+3
30-39	34.5	8	0	0	0
20-29	24.5	10	-10	-1	-10
10-19	14.5	15	-20	-2	-30
0-9	4.5	3	-30	-3	-9
		N=40			$\sum fd^1$=-44

$$\bar{X} = A + \left(\frac{\Sigma fd^1}{N} xc\right) = 34.5 + \left(\frac{-44}{40} x10\right) = 23.5$$

Illustration 5 : (Cumulative series – 'More than' type)

Given below is the distribution of 140 candidates obtaining marks X or higher in a certain examination (all marks are given in whole numbers).

X	10	20	30	40	50	60	70	80	90	100
C.F.	140	133	118	100	75	45	25	9	2	0

I.C.W.A.Inter

Solution: The above series has to be converted into simple exclusive series. The lowest class would be 10-20 and highest 90-100. The frequency of each class is found out by deducting the given cumulative frequency from the cumulative frequency of the previous class.

Calculation of Arithmetic mean

Marks obtained	Mid-Values	No. of Students	Deviation from assumed mean (A=55)	Step deviation (d/10)	Freq x d^1
X	m	*f*	d	d^1	fd^1
10-20	15	7	-40	-4	-28
20-30	25	15	-30	-3	-45
30-40	35	18	-20	-2	-36
40-50	45	25	-10	-1	-25
50-60	55	30	0	0	0
60-70	65	20	+10	+1	20
70-80	75	16	+20	+2	32
80-90	5	7	+30	+3	21
90-100	95	2	+40	+4	8
		N=140			Σfd^1=-53

$$\bar{X} = A + \left(\frac{\Sigma fd^2}{N} x\,c\right) = 55 + \left(\frac{-53}{140} x10\right) = 55 - 3.79 = 51.21 = 51 \text{ approx}$$

Illustration 6 : (Cumulative Series – 'Less than' type)

Calculate the average marks of the students from the following data.

Marks Below	10	20	30	40	50	60	70	80
No. of students	15	35	60	84	96	127	198	250

(B.Com Bangalore)

Solution:

This is to be converted into simple series. The lowest class is 0-10 and the highest 70-80. The frequency of the first class will be same i.e. 15. For the frequency with respect to other class intervals, the cumulative frequency of the previous class should be deducted from the cumulative frequency of that class. For example, the frequency for the next class 10-20 will be 35-15 = 20.

Calculation of Arithmetic Mean

Marks Obtained	**Mid-Values**	**No. of Students**	**Deviation from assumed mean (A=55)**	**Step deviation (d/10)**	**Freq x d^1**
X	m	f	d	d^1	fd^1
0-10	5	15	-30	-3	-45
10-20	15	20	-20	-2	-40
20-30	25	25	-10	-1	-25
30-40	35	24	0	-0	-0
40-50	45	12	+10	+1	+12
50-60	55	31	+20	+2	+62
60-70	65	71	+30	+3	213
70-80	75	52	+40	+4	+208
	N=250			**$\sum fd^1$=+385**	

$$\overline{X} = A + \left(\frac{\Sigma fd^1}{N} \times c\right) = 35 + \left(\frac{385}{250} x10\right) = 35 + 15.4$$
$$= 50.4 \text{ i.e} = 50 \text{ approx}$$

Illustration 7 : Preparation of frequency table and ascertainment of arithmetic mean.

The marks scored by 60 students in examination in statistics are given below. Form class-intervals of ten and calculate the arithmetic mean.

6	10	58	56	0	25	32	35	35	9
78	17	60	50	35	38	30	10	48	5
68	48	35	30	31	21	23	23	50	72
19	25	35	40	46	42	45	25	60	41
35	36	38	35	33	46	28	31	35	42
46	38	39	45	48	50	28	29	31	55

(B.Com Andhra, Mysore)

Solution:

Formation of frequency distribution Table.

Marks	**Tally Bars**	**Total**
0-10	1111	4
10-20	1111	4
20-30	~~1111~~ 1111	9
30-40	~~1111~~ ~~1111~~ ~~1111~~ ~~1111~~	20
40-50	~~1111~~ ~~1111~~ 11	12
50-60	~~1111~~ 1	6
60-70	111	3
70-80	11	2
		60

Calculation of Arithmetic Mean

Marks Obtained	Mid-Values	No. of Students	Deviation from assumed mean (A=35)	Step deviation (d/10)	Freq x d'
X	m	f	d	d'	fd'
0-10	5	4	-30	-3	-12
10-20	15	4	-20	-2	-8
20-30	25	9	-10	-1	-9
30-40	35	20	0	-0	0
40-50	45	12	+10	+1	+12
50-60	55	6	+20	+2	+12
60-70	65	3	+30	+3	+9
70-80	75	2	+40	+4	+8
		N=60			$\sum fd^1$= +12

$$\bar{X} = A + \left(\frac{\sum fd^1}{N} \times c\right) = 35 + \left(\frac{12}{60} X\ 10\right) = 35 + 2 = 37$$

Illustration 8 : (When size is given in minus value)

Calculate arithmetic mean from the following data:

Temperature 'C'	No. of days
-40 to -30	10
-30 to -20	28
-20 to 10	30
-10 to 0	42
0 to 10	65
10 to 20	180
20 to 30	10

(C.A. Inter)

Solution:

Calculation of Arithmetic Mean

Marks Obtained	Mid-Values	No. of Students	Deviation from assumed mean (A=35)	Step deviation (d/10)	Freq x d'
X	m	f	d	d'	fd'
-40 to -30	-35	10	-40	-4	-40
-30 to -20	-25	28	-30	-3	-84
-20 to -10	-15	30	-20	-2	-60
-10 to 0	-5	42	-10	-1	-42
0 to 10	+5	65	0	0	0
10 to 20	+15	180	+10	+1	+180
20 to 10	+25	10	+20	+2	+20
		N=365			$\sum fd^1$-26

$$\bar{X} = A + \left(\frac{\Sigma fd^1}{N} \times c\right) = 5 + \left(\frac{26}{36.5} \times 10\right) = 5 - 0.71 = 4.29$$

Hint: 'd' is ascertained as follows:
Deviations = Mid-Point minus assumed mean
For-35 d would be = -35 – (+5) = -35-5 = -40

Illustration 9 : (When a series has open-end class intervals)

Marks	No. of students
Below 10	8
10-20	12
20-30	15
30-40	25
40-50	15
Above 50	5

Solution: This is a case of open-end classes. A series is called open-end classes when lower limit of the first class and upper limit of the last class are not known. We have to make assumption about the unknown limits. This depends upon the class-intervals after the first class and before the last class. In the above case, the class-interval is uniform. In such a case, it would be proper to assume that the lower limit of the first class in zero and upper limit of the last class is 60.

Calculation of Arithmetic Mean

Marks Obtained	Mid-Values	No. of Students	Deviation from assumed mean (A=35)	Step deviation (d/10)	Freq x d'
X	m	f	d	d'	fd'
0-10	5	8	-30	-3	-24
10-20	15	12	-20	-2	-24
20-30	25	15	-10	-1	-15
30-40	35	25	0	0	0
40-50	45	15	+10	+1	+15
50-60	55	5	+20	+2	+10
		N=80			$\sum fd'$=-38

$$\bar{X} = A + \left(\frac{\Sigma fd^1}{N} \times c\right) = 35 + \left(\frac{-38}{80} \times 10\right) = 35 - 4.75 = 30.5 = 30 \text{ approax}$$

Illustration 10 : (When frequency is not given directly)
A market with 160 firms has the following distribution of average number of workers in various income groups: Find average salary paid in the whole market.

Income group	150-300	300-500	500-900	900-1500	1500-2000
No. of firms	50	35	25	30	20
Average No. of worker	12	25	35	15	10

Solution : In this problem frequencies are not given directly. The frequency (the number of workers in each income group) can be obtained by multiplying the number of firms with the average number of workers.

Calculation of Arithmetic Mean

Income Group X	Mid-Values m	No. of firms	Average No. of workers	frequency f	Freq x m fm
150-300	225	50	12	600	135000
300-500	400	35	25	875	350000
500-900	700	25	35	875	612500
900-1500	1200	30	15	450	540000
1500-2000	1750	20	10	200	350000
				N=3000	$\sum fm$=19875000

$\bar{X} = \frac{\sum fm}{N} = \frac{1987500}{3000} - \text{Rs.}\,662.5$

Illustration 11: When arithmetic mean is given and class-intervals have to be ascertained.

Find the class-intervals if the arithmetic mean of the following distribution is 22 and assumed mean is 25.

Step Deviation	-2	-1	0	+1	+2
Frequency	18	23	27	12	10

Solution:

Step Deviation	frequency	fd'
-2	18	-36
-1	23	-23
0	27	0
+`	12	+12
+2	10	+20
	N=90	$\sum fd^1$=-27

$\bar{X} = A + \left(\frac{\sum fd^1}{N} \text{ x c}\right) \Rightarrow 25 + \left(\frac{-25}{90} \text{ X C}\right) = 22, \Rightarrow -27$

C = 3 Thus, $\frac{3}{10}$ C =3=> 0.3 C=3; Thus, C= $\frac{3}{0.3}$ = 10

Assumed mean lies in the mid-value of that class with 'o' as step deviation. The lower and upper unit of this class are 25-10/2=20 and 25+10/2=30. Hence the class is 20-30. The other classes are 0-10, 10-20, 30-40 and 40=50

Illustration 12: When missing frequency is to be calculated.

Find the missing frequency from the following data:

Marks	Frequency
0-5	10
5-10	12
10-15	16
15-20	?
20-25	14
25-30	10
30-35	8

The average mark is 16.82 **B.Com Gujarat)**

Solution : Let the frequency of 15-20 class denoted as $f4$.

Marks Obtained	Mid-Values	No. of Students	Deviation from assumed mean (A=17.5)	Step deviation (d/5)	Freq x d'
X	m	f	d	d'	fd'
0-5	2.5	10	-15	-3	-30
5-10	7.5	12	-10	-2	-24
10-15	12.5	16	-5	-1	-16
15-20	17.5	$f4$	0	-1	0
20-25	22.5	14	+5	+1	+14
25-30	27.5	10	+10	+2	+20
30-35	32.5	8	+15	+3	+24
		$70+f4$			$\sum fd'=-12$

$\bar{X} = A + \left(\frac{\Sigma fd^1}{N} \times c\right) = 17.5 + \frac{-12}{70+f4} X\, 5 = 16.82, \frac{-60}{70+f4} = 16.82 - 17.5$

$\frac{-60}{70+f4} = -0.68, 0.68, (70 + f4) = 60, 47.6 + 0.68f4 = 60, 0.68f4 = 12.4f4$

$= \frac{12.4}{0.68} = 18$ approx.

Hence, missing frequency is 18.

Illustration 13 : When frequencies for the series is not given directly.

The number 3.2, 5.8, 7.9, and 4.5 have frequencies x, (x+2) (x-3) and (x+6) respectively. If the arithmetic mean is 4.876, find the value of x.

Solution:

Numbers	Frequency	
X	f	fx
3.2	x	3.2x
5.8	x+2	5.8 x + 11.6
7.9	x-3	7.9 x – 23.7
4.5	x+6	4.5 x + 27.0
	N=4x+5	$\sum fx$=21.4x + 14.9

$\bar{X} = \frac{\Sigma fx}{N}$, $4.876 = \frac{21.4x+14.9}{4x+5}$

$= 4.876\,(4x + 5) = 21.4x + 14.9$

$= 19.504x + 24.38 = 21.4x + 14.9$

$= 19.504x - 21.4x = 14.9 - 24.38$

$= -1.896x = -9.48$

$= 1.896x = 9.48, x = \frac{9.48}{1.896} = 5$

Illustration 14: When missing value is to be ascertained when mean and frequencies are given

From the following data of calculation arithmetic mean, find the missing item.

House Rent	**110**	**112**	**113**	**117**	**?**	**125**	**128**	**130**
No. f houses	25	17	13	15	14	8	6	2

Mean Rent = Rs.115.86

Solution:

House Rent X	No. of house *f*	*fx*
110	25	2750
112	17	1904
113	13	1469
117	15	1755
X	14	14x
125	8	1000
128	6	768
130	2	260
	N=100	**$\sum fx$=9906+14x**

$\bar{X} = \frac{\Sigma fx}{N}$, $115.86 = \frac{9906 + 14x}{100}$

11586=99014+14x, 14x = 11586-9906

14x=1680, x=120; Hence the missing item is 120

Illustration 15: When frequency distribution table is to be prepared.

The following are the monthly salaries in Rs. Of 20 employees of firm:

130	125	110	100	80	62	76	98	103	122
145	151	65	71	132	118	142	116	85	95

The firm gives bonuses of Rs.10, 15, 20, 25 and 30 for individuals in the respective salary groups; exceeding Rs.60 but not up to exceeding Rs.140 but not exceeding Rs.160. Fund the average bonus paid per employees.

(C.A.Inter)

Solution:

From the monthly salaries of the employees, we have to find the number f employees lying between the salary groups.

Salary Rs.	Tally Bars	Freq. *f*
61-80	~~1111~~	5
81-100	1111	4
101-120	111	4
121-140	1111	4
141-160	111	3

Calculation of average bonus paid per employee.

Bonus X	*freq.* *f*	*fx*
10	5	50
15	4	60
20	4	80
25	4	100
30	3	90
	N=20	**$\sum fx$=380**

$\bar{X} = \frac{\Sigma fx}{N}, = \frac{380}{20} = \text{Rs. } 19$

Illustration 16: (When unequal class-intervals are given)
The table given below shows the number of persons with different incomes in U.S.A. during the year 1929.

Income in thousands of dollars	**No. of Persons in lakhs**
0-1	13
1-2	90
2-3	81
3-5	117
5-10	66
10-25	27
25-50	6
50-100	2
100-1000	2

Calculate the average income per head.

Solution:

Calculation of Arithmetic Mean

Income (000 Dollars)	**Mid-Values**	**No.of persons (in Lakhs)**	**Deviation from assumed mean (A=7.5)**	**Feqxd'**
X	M	f	d	fd'
0-1	0.5	13	-7.0	-91.0
1-2	1.5	90	-6.0	-540.0
2-3	2.5	81	-5.0	-405.0
3.5	4.0	117	-3.5	-409.5
5-10	7.5	66	0	0
10-25	17.5	27	+10.0	+270.0
25-50	37.5	6	+30.0	+180.0
50-100	75.0	2	+67.5	+135.0
100-1000	550.0	2	+542.5	+1085.0
		N=404		$\sum fd$=224.5

$\overline{X} = \frac{\Sigma fx}{N} = 7.5 + \frac{224.5}{404} = 7.5 = +0.56 = 8.06$ thousand dollars

Illustration 17: (When mid points are given)
From the following data, calculate Arithmetic mean:

Mid-Point	5	15	25	35	45	55
Frequency	6	20	50	80	55	39

Solution:
Difference between mid-points = 15.5 = 10. Dividing the difference by 2, we get a figure of 5. Adding and subtracting 5 from each mid-point, we get the Class Intervals. The Arithmetic mean can be calculated as under.

Class Interval	**Mid point X**	**Frequency *f***	**Deviation from Assumed Mean D (d=X-30)**	***f*d***
0-10	5	6	-25	-150
10-20	15	20	-15	-300

20-30	25	50	-5	-250
30-40	35	80	+5	400
40-50	45	55	+15	825
50-60	55	39	+25	975
				1500

$\bar{X} = A + \frac{\Sigma fd}{N} = 30 + \frac{1500}{250} = 30 + 6 = 36$

Illustration 18 (When both Discrete and Continuous series are given)

From the following data, calculate Arithmetic mean:

Class Interval	2	4	6	8	10	11-16	17-22	22-32	32-40
Frequency	12	18	26	32	29	44	36	18	14

Solution:

Class	**Mid Point (x)** X	**Frequency (f)** F	**F *d**
2	2	12	24
4	4	18	72
6	6	26	156
8	8	32	256
10	5	29	145
11-16	13.5	44	594
17-22	19.5	36	702
22-32	27.5	18	495
32-40	35.5	14	497
		229	2941

$X = \frac{\Sigma fx}{N} = \frac{2941}{229} = 12.84$

Weighted Arithmetic Mean

Consider the performance of 2 batsmen. Batsman A scores 50 runs on a pitch that is tailor made for pace bowlers. In another match, Batsman B scores 50 on a wicket that is a Batsman's paradise. While the scores are equal, it is obvious that the performances are not the same. Similarly, if a student scores 70% marks in Accountancy and 90% in Sanskrit, while another students scores 90% marks in Accountancy and 70% in Sanskrit, the average marks scored by both students may be same, but their performances are different. In the first example, we will value the performance on the wicket tailor made for pace bowlers more than in Accountancy will be given greater importance compared to marks scored in Sanskrit. In other words, we are assigning greater weightage to these performances.

So far, while calculating Arithmetic mean, we have been giving equal weight to all observations. However, when different observations are given different weights, then we are calculating Weighted Arithmetic mean. To illustrate, in the above example, if the marks in Accountancy is considered twice as important as the marks in Sanskrit, we assign a weight of 2 and 1 respectively to the marks in by multiplying the individual value with its corresponding weight. In our example, the weighted score for Accountancy works out to 140 (70% *2) in case of first student and 180 (90%*2 in case of second student. Similary, the weighted score for Sanskrit works out to 90 (90% *1) in case of first student and 70 (70% *1) in case of second student. The next step is to calculate the sum of weighted scores. This works out to 230 (140+90) in case of first student and 250 (180+70) in case of second student. We then calculate the sum of weights, which is 3 (A weight of 2 for Accountancy and 1 for Sanskrit). The weighted Arithmetic mean is calculated by dividing the sum of weighted scores with sum of weights. In case of first student, the weighted Arithmetic mean is 230/3 = 76.67 whereas in case of second student, the weighted Arithmetic mean is 250/3 = 83.33. The weighted Arithmetic mean thus clearly brings out the relative importance of the two performances.

The assignment of weights is most critical in the calculation of weighted arithmetic mean. Weights could be either actual or arbitrary. If actual weights are available there is no problem in calculating the weighted mean. It however, weights are arbitrary, it becomes difficult to determine them. Different persons may assign different weights to various items. This can introduce a bit of subjectivity in calculation.

Weighted Arithmetic Mean is used when (i) importance of all items in a series is not equal. (ii) Classes of the same group contain widely varying frequencies. (iii) it is desired to calculate the average of the series from the average of its component parts. (iv) Ratios, percentages or rates are being averaged or (v) there is a change either in the proportion of values of items or in the proportion of their frequencies.

Calculation of Weighted Arithmetic Mean:

The formula for calculation of weighted arithmetic mean is as under.

Formula $\overline{X} = \dfrac{\Sigma WX}{\Sigma W}$

Where $\overline{X}_W$ = Weighted arithmetic mean.

ΣWX = Total of the variables multiplied by their respective weights

ΣW = Total of the weights.

Steps : (1) Multiply weights by the variables and obtain total. Denote it as ΣWX

(2) Divide the total by the sum of the weights i.e. ΣW

Illustration 19: An examination was held to decide about the award of scholarship in a college. The weights of various subjects were different. The marks obtained by 4 Candidates (out of 100 in each subject) are given below.

		Marks Obtained by			
Subject	**Weights**	**A**	**B**	**C**	**D**
Statistics	4	80	92	70	85
Accountancy	3	75	40	90	75
Economics	1	60	70	50	45
M. Law	2	45	50	60	65

Who should be awarded scholarship?

Solution:

In such a case weighted arithmetic mean should be calculated

Calculation of Weighted Arithmetic mean

	Weight	**A**		**B**		**C**		**D**	
Subject	W	X	WX	X	WX	X	WX	X	WX
Statistics	4	80	320	92	368	70	280	85	340
Accountancy	3	75	225	40	120	90	270	75	225
Economics	1	60	60	70	70	50	50	45	45
M. Law	2	45	90	50	100	60	120	65	130
	ΣW =10		**ΣWX =695**		**ΣWX =658**		**ΣWX =720**		**ΣWX =740**

$\overline{X}_W = \dfrac{\Sigma WX}{\Sigma W}$

$A = = \frac{695}{10} = 69.5, B = \frac{658}{10} = 65.8, C = \frac{720}{10} = 72.0, D = \frac{740}{10} = 74.0$

As the weighted arithmetic mean is highest in case of D, he should be awarded scholarship.

Note: Weighted mean should be calculated when the importance of the items in a series is not equal.

Combined Mean:

Consider a class with two sections. Section A consists of 40 students and Section B consists of 60 students. All the students appear for an examination and marks are awarded to each student. The students attempt to find the average marks scored by them in the examination. The students of section A collect marks awarded to each of the 40 students, add them up, divide the sum with 40 and arrive at an average score of 65. Similarly, students of section B collect marks awarded to each of the 60 students, add them up, divide the sum with 60 and arrive at an average score of 70. However, the lecturer of the class wishes to find out the average for the entire class. Should the lecturer attempt to collect the marks awarded to each of the 100 students, and them up and divide the sum with 100 or is there a shorter method of calculating the average for the entire class, if the averages for both sections and the number of students in each section is available? If we are calculating the mean of both sections ***combined***, by making use of the averages of the components (sections) of the total series (Class), we are using the concept of Combine4d Arithmetic mean.

If a series consists of two or more component series, the mean of the whole series can be expressed in terms of the mean of the component series. If, for example, a series relating to wages in a particular industry is divided in two parts - one relating to males and the other relating to females - and if we know the number of observations in each group and their respective means, we can find the combined mean of the two series as follows:

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

Where $\overline{X}_{12}$ is the combined mean of the two series, $\overline{X}_1$ and $\overline{X}_2$ the means of the two series respectively and N1 and N2 the number of observations in the two series.

In the example stated above, the combined mean of the entire class can be calculated as [(40*65)+(60*70)]/(40+60)=(2600+4200)/100=6800/100=68.

We will arrive at the same number even if we add up the score of all 100 students and divided the sum with 100.

Illustration 20: Means of two distributors of 100 and 150 items are 50 40 respectively. Find the combined Mean. ***(B.Com.Kakatiya)***

Solution: $\overline{X}_{12} = \frac{N1\overline{X} + N2\overline{X}2}{N1 + N2}$

$$= \frac{100x50 + 150x40}{100 + 150} = \frac{11000}{250} = 44$$

Illustration 21: There are three sections in B.Com. First year in a certain college. The number of students in each section and the average marks obtained by them in Statistics paper are as follows. Find the average marks obtained by the students of three sections taken together.

Section	Average marks	No. of Students
A	60	60
B	55	50
C	75	55

Solution:

Combined Mean:

$$\bar{X}_{123} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2 + N_3\bar{X}_3}{N_1+N_2+N_3}$$

$N_1 = 60, N_2 = 50, N_3 = 55, \bar{X}_1 = 60, \bar{X}_2 = 55, \bar{X}_3 = 75$

$$\bar{X}_{123} = \frac{(60x60)+(55x50)+(75x55)}{60+50+55} = \frac{3600+2750+4125}{165} \; 63.48$$

Properties of the Arithmetic Average

The arithmetic mean satisfies some mathematical properties, stated below:

1. The algebraic sum of the deviations of a given set of individual observations from the arithmetic mean is always zero. Symbolically. $\sum(X-\bar{X})=0$ In other words, the sum of the positive deviations from the arithmetic averages is equal to the sum of negative deviations. For this reason, arithmetic average is characterized as the centre of gravity.
2. The sum of squares of deviations of a set of observations is the minimum when deviations are taken from the arithmetic average. Symbolically. $\sum(X-A)^2$ is least when $A=\bar{X}$

   I.e., $\sum(X-X)^2$ is smaller than, $\sum$ (X-any value)^2.

   This is known as the property of 'least squares' in arithmetic mean.
3. If each of the values of a variate X is increased (or decreased) by a constant b, the arithmetic mean also increases (or decreases) by the same amount. Also, when the values of X are multiplied by a constant say k, the arithmetic mean is also multiplied by the same amount k.

To illustrate, if there are n terms, say X_1, X_2, X_3 .... Xn and the arithmetic mean is $\bar{X}$, if each of the terms is added with b, the new terms would be X_1+b, X_2+b X_3+b.... X_n+b. The sum of all terms will be $[(X_1+b)+(X_2+b)+(X_3+b)+....+(Xn+b)]$ which is equal to $[(X_1+X_2+X_3+....+Xn]+[b+b+b+....+b]$. Thus, the average will be

$= [(x_1+X_2+X_3+....+X2]+(b+b+b+.....+b)(N \text{ times})]/N$

$= \bar{X}+b$.

4. If a series consists of two or more component series the mean of the whole series can be determined as under:

$$X_{12} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1+N_2}$$

This is same as concept of combined mean.

Illustration 22: The man of 100 items was 50. Later on it was found that two items were misread as 18 and 6 instead of 81 and 66. Find the correct mean.

(B.Com Kakatiya)

Solution:

$\bar{X} = \frac{\sum X}{N}$ $\therefore \sum X = N\bar{X}$

$N = 100, \bar{X} = 50$ $\sum X = 100x50 = 5000$

But this is not correct $\sum X$

Correct $\sum X$ = Incorrect $\sum X$-Wrong items + Correct items

= 5000-18-6+81+66

= 5000-24+147=5123

Correct $\overline{X} = \frac{\text{Correct} \sum X}{N} = \frac{5123}{100} = 51.23$

Illustration 23: The man age of 100 labourers working in a factory running two shifts of 70 and 30 workers respectively is 121. The mean wage of 70 labourers working in the morning shift is Rs.130. Find the mean wage of workers working in the evening shift.

Solution:

$\overline{X}_{12} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1+N_2}$

$\overline{X} = 121, N_1 = 70, N_2 = 30, \overline{X}_1 = 130$

$= \frac{(70x130)+30\,\overline{X}_2}{100} = 121$

$= 9100+30\,\overline{X}_2 = 12100$

$= 30\,\overline{X}_2 = 12100\text{-}9100$

$30\,\overline{X}_2 = 3000$

$\overline{X}_2 = 100$

Hence the man wage of 30 workers in the evening shift is Rs.100.

Illustration 24 : The mean age of a combined group of men and women is 30 years. If the mean age of group of men is 32 and that of the group of women is 27, find out the percentage of men and women in the group. *(C.A.)*

Solution:

$\overline{X}\,12 = \frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1+N_2}$

$\overline{X}\,12 = 30, \overline{X}\,1 = 32, \overline{X}\,2 = 27$ N_1 and N_2 are not given

Let the percentage of Men be N_1. Percentage of women will be 100-N_1. So that $N_1+N_2 = 100$

$= \frac{(N_1x32)+(100-N_1)27}{100} = 30$

$= 32\,N_1 + 2700\text{-}27\,N_1 = 3000$

$= 5\,N_1 = 300$

$N_1 = 60$, 100-$N_1 = N_2$ 100=60 = 40

Hence the percentage of men and women in the group is 60 and 40 respectively.

Exercise 2 (a)

1. Compute Arithmetic mean of monthly incomes of 10 employees given below.

Sl.No.	1	2	3	4	5	6	7	8	9	10
Income (Rs)	8,000	10,000	9,000	6,500	8,500	12,500	20,000	14,500	13,000	9,500

2. Calculate the Arithmetic Mean for the following data

Family	A	B	C	D	E	F	G	H	I	J
Income (Rs)	60	80	10	75	150	140	75	45	100	90

3. Calculate Arithmetic mean from the following data

X	20	40	60	80	100	120	140
F	3	5	6	10	6	5	3

MEDIAN

Median, in simple terms, means middle. In Statistics, median represents the middle point in a series. Median divides the observations in two equal parts, in such a way that the number of observations smaller than median is equal to number of observations greater than it. However, in order to do this, the values, need to be arranged in an order, which can be Ascending or descending order. Thus, Median is the value of the middle item of a series arranged in an ascending or a descending order of magnitudes.

Thus, Median is the value of the middle item of a series arranged in an ascending or a descending order of magnitudes.

According to Prof. Horace Secrist, "Median of a series is the value of the item, actual or estimated, when a series is arranged in order of magnitude which divides the distribution into two parts". In the words of Prof. L.R. Corner, median is "that value of the variable which divides the group into two equal parts, one part comprising all values greater, and the other all values less then median".

Thus, computation of median essentially involves arranging the given data in order and then would divide the series into equal parts – one part containing 5 values that are less than the median value and the other containing 5 values that are more than the median value. If, however, there are even number of items in a series, there is no central item dividing the series in two equal parts. For example if there are 12 items in a series the median value would be between the values of 6th and 7th items. In such a case, Median will be the average of 6th and 7th items. Students may note that Median does not take into account the values of all the items in a series. For example, if the monthly incomes of five families (arranged in ascending order) are Rs.6,000, Rs.9,000, Rs.12,500, Rs.18,000 and Rs.20,000, the median value would be Rs.12,500. If, however, the monthly incomes of five families were Rs.3,000, Rs.8,000, Rs.12,500, Rs.15,000 and Rs.25,000, the median value would still be Rs.12,500, though the two series are different in their composition. Median is only a positional measure. It is the value of the middle item irrespective of all other values.

Merits of Median

1. It is rigidly defined.
2. It is easy to understand and easy to calculate.

3. It is not influenced much by items on the extremes.
4. It can be computed for distributions which have open end classes.
5. It can be determined by inspection.
6. It can be located graphically.
7. Median can be used while dealing with qualitative data where numerical measurements are not available, but where it is possible to rank the objects in some order.
8. Median can be calculated even if some of the extreme values are not available, provided we know the number of observations.
9. The median is centrally located. The absolute sum (disregarding positive and negative values) of the deviations of the individual values from the median is always the minimum.
10. The value of the median never falls into the open end interval and it can also be calculated without difficulty from grouped frequency distributions with classes of unequal width.

Demerits of Median

1. Median is not exact in many cases. For example, if number of observations is even, median is estimated by taking the average of the two middle terms. Similarly, when several values in the centre are of the same size, the median may be somewhat indeterminate.
2. Calculation of Median does not consider all the items of the series. Thus, it is not fully representative of the entire data.
3. Median is significantly impacted by fluctuations of sampling. Hence, it is less reliable.
4. The Median is not very suitable for further algebraic treatment. For example, this assumes that the observations in the median class are uniformly distributed which may not be the case.
5. Median of a frequency distribution is estimated based on simple interpolation. This assumes that the observations in the median class are uniformly distributed, which may not be the case.
6. It is unsuitable if it is desired to give greater importance to large or small values. Thus, the Median is most useful when we require a measure of location which is not affected by high or low value items.

Quartiles, Deciles and Percentiles:

There are other 'positional' values which divide a series in a number of parts. The most common positional values besides median are quartiles deciles and percentiles

The values which divide the given data in four equal parts are known as Quartiles. There will be three such values. The first quartile, which is also called the lower quartile (Q1), covers the first 25% of the series. Similarly, the second quartile covers first 50% items of the series and divides the series into two equal parts. It is same as Median. The third quartile which is also called the Upper quartile (Q3) covers the first 75% items of a series.

Deciles divide a series in 10 equal parts. There are nine deciles denoted by D1, D12, D3 etc. The first decile (D1) covers the first 10% of the series, the second decile (D2) covers the first 20% of the series, and so on. Percentiles divide the series in 100 equal parts There are 99 percentiles denoted by P1, P2, P3, etc.

Calculation of Median

Individual Observations:

Median (M) = Size of $\frac{N+1}{2}$ th time

Lower Quartile (Q1) = Size of $\frac{N+1}{4}$ th item

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

Steps: (1) Arrange the data in ascending order.

(2) Apply the formula.

Illustration: 25 (When odd number of values are given).

Obtain the value of median from the following data: 120, 170, 100, 110, 180, 220,160.

Solution: The data is arranged in ascending order

100, 110, 120, 160, 170, 180, 220.

There are 7 items. Thus, N=7

Calculation of Median

Median = Size of $\left(\frac{N+1}{2}\right)$th item = $\left(\frac{7+1}{2}\right)$ the item = 4^{th} item = 160

Illustration 16: (When even number of values are given)

Calculate Median, Quartiles, and 8^{th} Deciles from the following data the marks obtained by 10 students in an Examination.

27	31	19	35	23	16	40	46	17	37

Solution: The data has to be arranged in ascending order:

16	17	19	23	27	31	35	37	40	46

There are 10 items. Thus, =10

Median = Size of $\left(\frac{N+1}{2}\right)$ th items = $\left(\frac{10+1}{2}\right)$ th = 5.5^{th} item

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

$= \frac{27+31}{2} = \frac{58}{2} = 29$

Hence Median marks = 29

Q_1 = Size of $\left(\frac{N+1}{4}\right)^{th}$ item = Size of $\left(\frac{10+1}{4}\right)^{th}$ = 2.75^{th} item.

= Size of 2^{nd} item + 0.75 (Size of 3^{rd} item-Size of 2^{nd} item)

= 17+0.75(19-17) = 17+1.5 = 18.5 marks

Q_3 = Size of 3 $\left(\frac{N+1}{4}\right)^{th}$ item = Size of 3 $\left(\frac{10+1}{4}\right)^{th}$ item

= Size of 8.25^{th} item.

= Size of 8^{th} item + 0.25 (Size of 9^{th} item – Size of 8^{th} item)

= 37+0.25(40-37) = 37+.75 = 37.75

D_8 = Size of 8 $\left(\frac{N+1}{10}\right)^{th}$ item = Size of 8 $\left(\frac{11}{10}\right)^{th}$ item

= Size of 8.8^{th} item = Size of 8^{th} item + 0.8 (9^{th} itme-8^{th} item)

= 37+0.8 (4-37) = 37+2.4 = 39.4

Discrete Series:

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

Steps : (1) Arrange the data in ascending order.

(2) Find out cumulative frequencies.

(3) Apply the formula M=size of $\frac{N+1}{2}$

(4) Look at the cumulative frequency column and find that total which is either equal to $\frac{N+1}{2}$ or next higher than that and determine the value of the variable corresponding to this. The value of the corresponding variable is the value of Median.

Note: Similarly the other values such as. Q1, Q3 etc. can be ascertained.

Illustration 27:

Locate Median, Upper Quartile and lower Quartile from the following data.

Size of shoes	Frequency	Size of shoes	Frequency
4.0	10	6.5	15
4.5	18	7.0	10
5.0	22	7.5	8
5.5	25	8.0	7
6.0	40		

Solution: Calculation of Median and Quartiles

Size of shoes	**Frequency**	**Cumulative frequency**
4.0	10	10
4.5	18	28
5.0	22	50
5.5	25	75
6.0	40	115
6.5	15	130
7.0	10	140
7.5	8	148
8.0	7	155

Median = Size of $\left(\frac{N+1}{2}\right)^{th}$ item = Size of $\frac{155+1}{2}+1 = \frac{156}{2}$ th item.

= Size of 78^{th} item = 6.0

Hence, **Median = 6.0**

Lower Quartile (Q1) = Size of $\left(\frac{N+1}{2}\right)^{th}$ item = Size of $\left(\frac{155+1}{4}\right)^{th}$ item

= Size of 39^{th} item = 5.0 Hence Lower Quartile = 5.0

Upper Quartile (Q3)

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

= Size of 117^{th} item = 6.5. Hence **Upper Quartile = 6.5**

Continuous series:

Median=L1+= $\frac{L2-L1}{f1}(m-c)$ Where

L_1 and L_2	=	The Lower and Upper limit of the class in which median lies.
f_1	=	frequency of the median class
M	=	the middle number i.e. $\frac{N}{2}$
C	=	Cumulative frequency of the class preceding the median class.

It is also given in the following manner

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

Where

L_1	=	Lower limit of the median class
$\frac{N}{2}$	=	Middle number
$C.f$	=	Cumulative frequency of the class preceding the median class
f	=	frequency of the median class.
I	=	class interval of the median class

Steps: (1)Determine the particular class in which the median lies by using $\frac{N}{2}$ as the rank

(2) Apply the formula

Note : For lower Quartile and Upper Quartile the method to be applied is same. Q1 and Q3 classes can be ascertained by taking $\frac{N}{4}$ th and $\frac{3\,(N)th}{4}$ number.

Illustration 28: (Exclusive series)

Find the median and the quartiles from the following table.

Monthly Income	No. of Persons
Below 50	35
50-60	24
60-70	21
70-80	18
80-90	6
90 and above	3

(B.Com. Banaras)

Solution : Calculation of Median and Quartiles.

Income (Rs.)	f	c. f
Below 50	35	35
50-60	24	59
60-70	21	80
70-80	18	98
80-90	6	104
90 and above	3	107
	N=107	

Median = Size of $\frac{N}{2}$ th item = Size of $= \frac{107}{2}$ item.

Median lies in 50-60 Class.

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

$L_1 = 50; \frac{N}{2} = 53.5, f = 24, Cf = 35, i = 10$

$= 50 + \frac{53.5-35}{24} \times 10 = 50 + \frac{18.5}{24} \times 10 = 50{+}7.7 = 57.7$

Lower Quartile (Q1) = Size of $\frac{N}{4}$ th item.

$= \frac{107}{4} = 26.75$ th item.

Q_1 lies in the first class. This class is taken as 40-50

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

$L_1 = 40, \frac{N}{4} = 26.75, C.\,f = 0, f = 35, i = 10$

Note: As Q1 lies in the first class itself, cumulative frequency of the preceding group is taken as zero.

$40 + \frac{26.75-0}{35} \times 10 = 40 + 7.64 = 47.64$

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

$= \frac{3x107}{4}$ th item = 80.25 item

Q_3 lies in 70-80 class

$Q_1 = L_1 + \frac{\frac{3\,(N)}{4} - C.f}{f} x\, i$

$L_1 = 70, \frac{3N}{4} = 80.25, C.\, f_1 = 80, f = 18, i = 10$

$= 70 + \frac{80.25-80}{18} x\ 10 = 70+0.14 = 70.14$

Illustration 29: (Inclusive series)

The following table gives the distribution of males in an Indian Town. Find the media age and interpret.

Age Groups	Males	Age Groups	Males
0-9	2756	50-59	610
10-19	2124	60-69	245
20-29	1677	70-79	67
30-39	1481	80-89	6
40-49	1021	90-99	3

(B.Com. Madras)

Solution : Calculation of Median Age.

Age Groups	*f*	c. *f*
0-6	2756	2756
10-19	2124	4880
20-29	1677	6557
30-39	1481	8038
40-49	1021	9059
50-59	610	9669
60-69	245	9914
70-79	67	9981
80-89	6	9987
90-99	3	9990

Median = Size of $\frac{N}{2}$ th item, Size of $\frac{9990}{2}$ th item.

Size of 4995 th item

Median class = 20-29 i.e. 19.5-29.5

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

$L_1 = 19.5\ \frac{N}{2} = 4995, f = 1677, Cf = 4880, i = 10$

$= 19.5 + \frac{4995-4880}{1677} x\ 10 = 19.5 + \frac{115}{1677} x\ 10 = 19.5+0.7 = 20.2$ years

30-32	2	40-42	62
32-34	9	42-44	39
34-36	25	44-46	20
36-38	30	46-48	11
38-40	49	48-50	3

Calculate from the above data:

(i) The median and third quartile wages.

(ii) The number of wage earners receiving wages between Rs.37 and 45 per week.

Solution: Calculation of Median and third quartile wages.

Weekly wages	f	**C. f**
30-32	2	2
32-34	9	11
34-36	25	36
36-38	30	66
38-40	49	115
40-42	62	177
42-44	39	216
44-46	20	236
46-48	11	247
48-50	3	250
	N=250	

Median = Size of $\frac{N}{2}$ th item = Size of 250 = $\frac{125}{2}$ th item.

Median lies in the class 40-42

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

$L1 = 40$, $\frac{N}{2} = 125$, C. f.= 115, $f = 62$, i= 2

$= 40 + \frac{125-115}{62} \times 2 = 40+0.32$ = Rs. 40.32

Third Quartile = Size of $\left(\frac{3N}{4}\right)^{th}$ item = Size of $\left(\frac{3x250}{4}\right)^{th}$ item

Third Quartile lies in the wage group of 42-44

$Q3 = L1 + \frac{\frac{3N}{4} - C.f}{f} \times i$

$L1 = 42$, $\frac{3N}{4} = 187.5$, C.f. = 177, f = 39, i = 2

$= 422 + \frac{187.5-177}{39} \times 2 = 42+0.54$=Rs. 42.54

(ii) There are 30 persons in the wage group of 36-38. On the assumption that frequencies are equally distributed throughout the class, Rs.37 will be received by half of the number i.e.15. Similarly there are 20 persons in the wage group of 44-46. Half off it i.e. 10 persons are getting Rs.45 wages. Hence the total number of wages earners between 37 and 45 are 175 (15+49+62+39+10)

Illustration 31: (When missing frequency is to be ascertained using median formula)

An incomplete distribution is given below.

Variable	Frequency
0-10	10
10-20	20
20-30	?
30-40	40
40-50	?
50-60	25
60-70	15
	Total = 170

You are given that median value is 35. Using the median formula, fill up the missing frequencies. *(B.Com. Nagarjuna and C.A. – Adapted).*

Solution:

Let the missing frequency of the classes 20-30 and of 40-50 denoted as f_1, f_2 Total frequency = 170

Total frequency of the classes other than that of the missing frequencies = (10+20+40+25+15) = 110

Thus, $110+f_1+f_2 = 170 = f_1+f_2=170-110=60$

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

$L1 = 30, \frac{N}{2} = 85, f = 30+f_1. f=40, i=10$

$30 + \frac{85-(30+f1)}{40} \times 10 = 35 = 30 + \frac{85-30-f1}{4} = 35$

$\frac{55-f1}{4} = 35-30=5$. Thus, $55-f1=20$, $f1=55-20=35$

$f1+f2=60$, $35+f2=60$. Thus, $f2=60-35=25$

Hence missing frequencies are 35 and 25 respectively.

Illustration 32: (When cumulative frequencies are given and the class interval is not equal)

In a small branch of a bank in rural area following is the average deposit balance of current accounts during a month. Calculate median, 7th decile and 85th percentile.

	Deposit balance	No. of deposits
Less than	1,000	500
Less than	900	498
Less than	800	480
Less than	600	475
Less than	550	440
Less than	500	374
Less than	400	300

Less than	250	125
Less than	100	25

Solution : As cumulative frequencies are given they have to be converted into simple frequencies and in ascending order. It is not necessary to have equal class interval while calculating median etc.

Deposit balance	No. of deposits	C.F.
0-100	25	25
100-250	100	125
250-400	175	300
400-500	74	374
500-550	66	440
550-600	35	475
600-800	5	480
800-900	18	498
900-1000	2	500
	N=500	

Median = Size of (N/2)th item = Size of 500/2th item i.e. 250th item.

Median lies in the class 250-400

Median $= L1 + \frac{\frac{N}{2} - C.f}{f} x\, i$

$L1 = 250,\ \frac{N}{2} = 250,\ f = 175,\ C.f = 125,\ i = 150$

$= 250 + \frac{250-125}{175} x\ 150 = 250+107.14=357.14$

7^{th} Decile (D7) = Size of $7\left(\frac{N}{10}\right)$ th item = Size of $\frac{7x500}{10}$ th item.

= Size of 350^{th} item. Hence D7 lies in the class 400-500

D7 $= L1 + \frac{\frac{7N}{10} - C.f}{f} x\, i$

$L1 = 400,\ \frac{7N}{10} = 350,\ f = 74,\ C.f = 300,\ i = 150$

$D7 = 400 + \frac{350-300}{74} x\ 100 = 400+67.6=467.6$

85^{th} Percentile = Size of $85\left(\frac{N}{10}\right)^{th}$ item = Size $\frac{85x500}{100}$ th item.

= Size of 425 th item.

85^{th} Percentile lies in the class 500-550

P4 $= L1 + \frac{\frac{85N}{100} - C.f}{f} x\, i = 500 + \frac{425-374}{66} x\ 50$

$= 500 + \frac{51}{66} x\ 50 = 500 + 38.64 = 538.64$

Illustration 33: Find the median after making the necessary changes to make the Class intervals equal. Amend the following table and find the Median.

Class Intervals	5-10	10-25	25-45	45-55	55-65	65-85	85-95
Frequency	7	19	25	26	13	7	3

Solution: In this problem, there is a specific request to amend the Class intervals .This can be done as under:

The classes 5-10 and 10-25 can be combined into a single class of 5-25, with a frequency equal to 7+19 = 26. Similarly, the classes 45-55 and 55-65 can be combined into a single class of 45-65, with a frequency equal to 26+13=39. Lastly, the class 85-95, and a nonexistent class 95-105 can be combined into a class 85-105, with a frequency of 3.

X	*f*	C*f*
5-25	26	26
25-45	25	51
45-65	39	90
65-85	7	97
85-95	3	100
	N=100	

I = 20, N/2 = 50, L = 525; *f*= 25; C*f* = 26

Using Formula Median = 25+[(50-26)/25] *20

= 25 + (24*20)/25 = 25+19.20=44.20.

Locating Median Graphically

One of the advantages of Median is that it can be located graphically. This is done with the help of ogive curves.

One method of ascertaining Median graphically is to draw both the 'less than' ogive and 'more than' ogive on the same graph. If done so, they intersect at a point. A perpendicular from the point of intersection on the X-axis gives the value of median. This has been explained in the previous chapter.

Another method of locating the median graphically is explained below:

Step 1: Draw the less than ogive.

Step 2 : Calculate N/2 or (N+1)/2 where N=sum of frequencies. Mark the same on Y axis.

Step 3: Draw a horizontal straight line from the point marked on Y axis, parallel to X axis so that it meets the Less than ogive.

Step 4: From the point of intersection of the horizontal line and the ogive, drop a perpendicular to X axis.

Step 5 : The point of intersection of the perpendicular and the X axis is the median. It is important to note that there should be no gaps between consecutive classes. Hence, if any inclusive series is given, it should be first converted into an exclusive series.

Illustration 34: Draw a less than ogive of the following series and determine the median.

Class Interval	0-10	10-20	20-30	30-40	40-50	50-60
Frequency *f*	2	6	20	32	22	18

Solution :

Class Interval	0-10	10-20	20-30	30-40	40-50	50-60
Frequency *f*	2	6	20	32	22	18
Less than	2	8	28	60	82	100

Workers	4	6	10	20	12	6	2

MODE

While looking for a shirt is one of the Retail outlets, you will probably find a number of shirts available for different sizes. However, the number of shirts as well as the variety will be more for some sizes (size 40 for example) when compared to other sizes (size 44 for example). How does the Shirt manufacturer or the Store owner decide on which sizes to store more? This decision can be taken very easily by applying the concept of Mode.

Mode is the most common item of a series. It is derived from the French word 'la Mode' which signifies fashion. It is the value occurring most frequently in a set of observations and around which other items of the set cluster most densely. In other words, it is an actual value which has the highest concentration of items around it.

According to Croxton and Cowden, "The mode of distribution is the value at the point around which the items tend to be most heavily concentrated. It may be regarded as the most typical of a series of values".

In the words of A.M. Tuttle, "Mode is the value which has the greatest frequency density in its immediate neighbourhood".

Thus, Mode is a value around which there is highest concentration of values. It may not necessarily be the value which occurs the largest number of times in a series, as in some case the point of maximum concentration may be around some other value. In some cases, there may be more than one point of concentration of values. These are called bi-model or multi-modal series.

Merits of Mode

1. Mode is easy to understand and calculate.
2. It is not influenced much by items on the extremes.
3. It can be located even if the class-intervals are of unequal magnitudes, provided the modal class and the preceding and succeeding it are of the same magnitude.
4. It can be computed for distributions which have open end classes.
5. Mode is not an isolated value like the median. It is the term that occurs most in the series.
6. Mode is not a fictional value that is not found in the series.
7. It can be determined by inspection.
8. It can be located graphically
9. It has wide business application

Drawbacks of Mode

1. Calculation of Mode does not consider all the items of the series. Thus, it is not fully representative of the entire data.
2. It is not rigidly defined.
3. It is not capable of further mathematical treatment.
4. Mode is sometimes indeterminate. There may be 2 (Bi-modal) or more (Multimodal) values.
5. Mode is significantly impacted by fluctuations of sampling. Hence, it is less reliable.
6. Mode is considerably influenced by the choice of grouping. A change in the size of the class interval will change the value of the mode". It is a very unstable average and its true value is difficult to determine.

Calculation of Mode

Individual Observations

Steps: Count the number of times the various value repeat themselves and the value occurring the maximum number of times is the modal value.

Illustration 35: Calculate the mode from the following data of marks obtained by 10 students.

S.No.	1	2	3	4	5	6	7	8	9	10
Marks	15	32	29	17	32	32	25	29	20	35

Size of item	No. of times it occurs
15	1
17	1
20	1
23	1
20	1
32	3
35	1
Total	

As 32 occurs for maximum number of times i.e.3, hence the modal marks are 32.

Discrete Series:

Mode can be ascertained by just inspection. But in some cases, it is not possible to ascertain mode by inspection. In such cases we prepare grouping Table. A grouping Table has six columns. In the first column the maximum frequency is marked. In the second column frequencies are grouped in twos i.e. By adding frequencies of item number 1 and 2, 3 and 4 and so on. In the third column the first frequency is left and the remaining are grouped in twos i.e. By adding frequencies of item number 2 and 3, 4 and 5 so on. In column number four the frequencies are grouped in three's i.e. by adding the frequencies of item numbers 1, 2 and 3, 4, 5 and 6 and so on. In column number five the first frequency is left and the remaining frequencies are grouped in threes i.e. By adding the frequencies of item numbers 2, 3 and 4, 5, 6 and 7 and so on. In column number six the first two frequencies are left and then the remaining frequencies are grouped in three's i.e. By adding the frequencies of item number 3, 4 and 5, 6, 7 and 8 and on.

After preparing the grouping table, an analysis Table is prepared. While preparing this table column number are put on the left hand side various probable values of mode on the right hand side. The values against which frequencies are the highest are marked in the grouping table and entered in the analysis Table.

Illustration 36:

Height in inches	57	59	61	62	63	64	65	66	67	68
No. of persons	3	5	7	10	20	22	24	5	2	2

Solution: In the first column the maximum frequency of 65 is marked. In column 2, frequencies are grouped in twos. For example, we get the number 8 by adding the frequencies of item numbers 1 and 2, which is 3 and 5. In column 3, the first frequency is left and the remaining are grouped in two. For example, we get 12 by adding frequencies of item numbers 2 and 3, which is 5 and 7. In column 4, the frequencies are grouped in three. To illustrate, we get 15 by adding the frequencies of item numbers 1, 2 and 3, which is 3+5+7. In column 5, the first frequency is left and the remaining frequencies are taken groups in three.

Grouping Table

Height	1	2	3	4	5	6
57	3					
		8				
59	5			15		
			12			
61	7				22	
		17				
62	10					37
			30			
63	20			52		
		42				
64	22				66	
			46			

						51
65	24	29		31		
66	5		7		9	
67	2	4				
69	2					

To illustrate, we get 22 by adding the frequencies of item numbers 2, 3 and 4; which is 5+7+10. In column number six the first two frequencies are left and then the remaining frequencies are taken in groups of three. We get 37 by adding the frequencies of item number 3,4 and 5, which is 7+10+20. The values against which frequencies are the highest are marked for each column.

Analysis Table

Size of item containing maximum frequency

Col.No.	**62**	**63**	**64**	**65**	**66**
1				1	
2		1	1		
3			1	1	
4	1	1	1		
5		1	1	1	
6			1	1	1
Total	**1**	**3**	**5**	**4**	**1**

We then prepare an Analysis Table. The column numbers are put on the left hand side and various probable values of mode on the right hand side. The values against which frequencies are the highest are marked in the grouping table and entered in the analysis Table. For example, in the first column, the highest value is 24. This is highlighted in Bold in the grouping table. The value corresponding to 24 is 65. Thus, a score of 1 is written in the first row (representing column 1 of the Grouping table) under the column titled 65. Similarly, the highest value in case of second column is 42. This is highlighted in Bold in the grouping table. The values comprising 42 are 64 and 65. Thus, a score of 1 is written in the second row (representing column 2 of the Grouping table) under the columns titled 64 and 65. This exercise is done for all columns. The last step is to add the scores allocated to each column. The column getting the highest score will represent the modal value. In this example, the highest score of 5 is for value 64. Hence, mode is 64. In other words, the value 64 occurs for the maximum number of times i.e. 5 and hence, the modal height is 64.

Continuous Series:

In case of Continuous Series also, we prepare the Grouping Table and Analysis table. However, this is used to ascertain the Modal class. We then go one step ahead and obtain the value of Mode by Interpolation as is the case with Median. The following formula is used to calculate Mode (Z)

$$\text{Mode} : (Z) = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} * i$$

Where: L = is the lower limit of Modal Interval

f_1 is the Frequency corresponding to modal interval

f_0 f_2 are frequencies of classes preceding and succeeding the Modal Interval and i is the length of Modal Interval.

Steps for calculation of Mode

Ascertain Modal class by preparing grouping and analysis tables or by inspection.

Apply the above formula.

Alternatively, Mode can also be calculated with the help of the following formula, in case the above formula fails.

Mode : $(Z) = L + \frac{D1}{D1+D2} * i$

Where $D_1 = | f_1 - f_0 |$ and $D_2 = |f_1 = f_2|$

and D_2 = Here only the positive values are taken.

Note 1 : While calculating mode it is necessary that the class intervals are uniform. If they are unequal, they should be made equal on the assumption that frequencies are equally distributed throughout.

Note 2 : If two or more variables have the same highest frequency, then Mode is said to be ill-defined. In such cases, mode may be ascertained by the following formula based upon the relationship between Mean, Median and Mode. **Mode = 3 Median – 2 Mean.**

Illustration 37 : (Exclusive series)

The following table gives the length of life of 150 electric lamps.

Life (hours)	Frequency of Lamps
0-400	4
400-800	12
800-1200	40
1200-1600	41
1600-2000	27
2000-2400	13
2400-2800	9
2800-3200	4

Calculate Mode. *(B.Com Agra)*

Solution:

Grouping Table

	Frequency					
Life	**1**	**2**	**3**	**4**	**5**	**6**
0-400	4					
		16				
				56		
400-800	12					
			52			
					93	
800-1200	40					
		81				
						108
1200-1600	41					
			68			
				81		
1600-2000	27					
		40				
					49	
200-2400	13					

		22	26
2400-2800	9	13	
2800-3200	4		

Analysis Table

Col. No.	Size of group Containing maximum frequency				
	400-800	800-1200	1200-1600	1600-2000	2000-2400
1			1		
2			1	1	
3		1	1		
4			1	1	1
5	1	1	1		
6		1	1	1	
		3	6	3	1

Hence Modal group is 1200-1600

Mode = $L1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$

$L1 = 1200,\ f_1 = 41,\ f_0 = 40,\ f_2 = 27,\ i = 400$

$= 1200 + \frac{41-42}{82-40-27} \times 400 = 1200 + 26.67 = 1226.67$

Illustration 38: (When unequal class-intervals are given).

Variable	Frequency	Variable	Frequency
0-2	3	18-24	10
2-6	6	24-30	8
6-12	17	30-34	5
12-17	17	34-36	2
17-18	10		

(B.Com. Osmania)

Solution : As the class intervals are unequal, they have to be made equal by adjusting frequencies.

Size	Frequency
0-6	9
6-12	17
12-18	27
18-24	10
24-30	8
30-36	7

By inspection it is clear that modal class is 12-18

Mode = $L1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$

$L1 = 12,\ f_1 = 27,\ f_0 = 17,\ f_2 = 10,\ i = 6$

$= 12 + \frac{27-17}{54-17-10} \times 6 = 12 + 2.2 = 14.2$

Illustration 39: Find Mode from the following data.

X:	0-50	50-100	100-50	150-200	200-250	250-300	300-350
f:	4	13	37	40	43	22	18

Grouping Table

Solution:

X	1	2	3	4	5
0-50	4				
50-100	13	17		54	
100-150	37		50		90
150-200	40	77			
200-250	43		83		
250-300	22	65		105	
300-350	18		40		

Analysis Table

	0-50	50-100	100-150	150-200	200-250	250-300	300-350
1					1		
2			1	1			
3				1	1		
4				1	1	1	
5		1	1	1			
Total	0	1	2	4	3	1	0

Thus, the Modal Interval is 150-200.

alternative formula is applied:

Mode : (Z) = $L + \frac{D1}{D1+D2} * I$, where

L=150; $D_1 = |40 - 37|$ 3, $D_2 = |40 - 43| = 3$ i=50

Thus, Z=L + $L + \frac{D1}{D1+D2} * 50 = 150 + \frac{3}{3+3} * 50$

$= 150 = \frac{150}{6} = 150 + 25 = 175.$

Illustration 40: (When mode is ill-defined)

Size of item	Frequency
0-10	4
10-20	6
20-30	20
30-40	32
40-50	33
50-60	17
60-70	8
70-80	2

Grouping Table

Size	1	2	3	4	5	6
0-10	4					
		10				
10-20	6			30		
			26			

C-50

				58	
20-30	20				
		52			85
30-40	32				
			65		
				82	
40-50	33				
		50			
				58	
50-60	17				
			25		27
60-70	8				
		10			
70-80	2				

Analysis Table

Col.No.	Size of group Containing maximum frequency					
	10-20	20-30	30-40	40-50	50-60	60-70
1			1			
2		1	1			
3			1	1		
4			1	1	1	
5	1	1	1	1	1	1
6		1	1	1		
Total	1	3	5	5	2	1

This is a bi-modal series: Mode is to be ascertained by applying the formula.

Mode = 3 Median – 2 Mean.

Calculation of Mean and Median

Size	Midvalue	Frequency	Deviation from Assumed Mean	Step-Deviation	Freq x d'	Cumulative Frequency
X	m	f	d	d^1	fd^1	
0-10	5	4	-40	-4	-16	4
10-20	15	6	-30	-3	-18	10
20-30	25	20	-20	-2	-40	30
30-40	35	32	-10	-1	-32	62
40-50	45	33	0	0	0	95
50-60	55	17	+10	+1	17	112
60-70	65	8	+20	+2	16	120
70-80	75	2	+30	+3	6	122
		122			$\Sigma fd^1 = -67$	

$X = A + \frac{\Sigma fd1}{N} xC = 45 + \frac{-67}{122} x\ 10 = 45\text{-}5.5\text{-}\ 39.5$

Median = Size of $\frac{N}{2}$ th item. = Size of $\frac{122}{2}$ = 61[st] item.

Median Class = 30-40

Median = $L1 + \frac{\frac{N}{2} - c.f}{f} \text{ x i}$

$L1 = 30, \frac{N}{2} = 61, C.f = 30, f = 32, i = 10$

$30+\frac{61-30}{32} \times 10 = 30+9.7 = 39.7$

Mode = 3 Median-2 Mean = (3 x 39.7) – (39.5) = 119.1-79.0 = 40.1

Location of Mode Graphically:

As stated earlier, we can locate the mode of a frequency distribution graphically. The following steps need to be followed:

1. Draw a histogram of the given distribution.
2. Joint the top right corner of the highest rectangle i.e., modal rectangle by a straight line to the top right corner of the preceding rectangle. Similarly the top left corner of the highest rectangle is joined to the top corner of the rectangle on the right.
3. From the point of intersection of these two diagonal lines, draw a perpendicular on the horizontal axis.
4. The foot of the perpendicular indicates the mode.

Note: It is not necessary to draw the entire Histogram. A partial histogram showing the modal, pre-modal and post-modal classes is sufficient.

Illustration 41: Determine the mode of the following frequency distribution without using a formula.

Life of Electric Lamps (Hrs.)	No. of Lamps
100-1020	100
1020-1040	150
1040-1060	500
1060-1080	350
1080-1100	200

Solution:

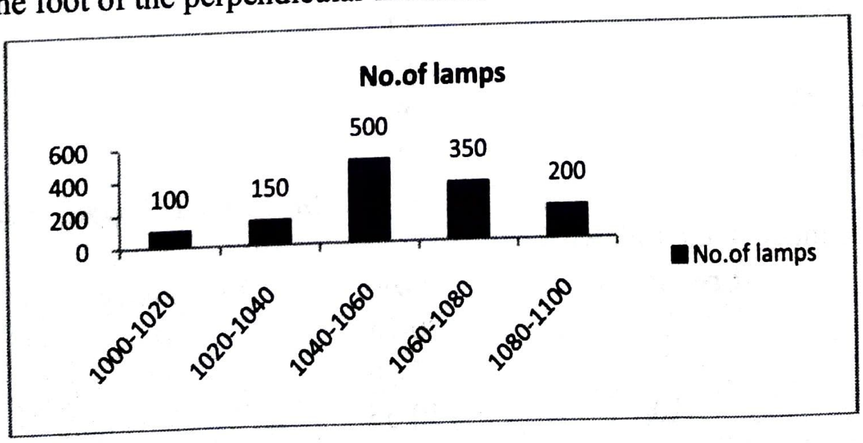
Step 1: Draw Histogram.

Step 2: The highest rectangle is the bar for the Class 1040-1060. This is the modal class. The preceding and succeeding rectangles are for classes 1020-1040 and 1060-1080 respectively. Draw a straight line joining the left top corner of the bar representing the class preceding the modal class and the right top corner of the bar representing the modal class.

Step 3: Draw a straight line joining the left top corner of the bar representing the modal class and the right top corner of the bar representing the class succeeding the modal class.

Step 4: From the point of intersection of the two lines, drop a perpendicular to the X axis.

Step 5: The foot of the perpendicular indicates the mode.



21. Draw a histogram for the following distribution and find the modal wage.

Wages in Rs.	10-15	15-20	20-25	25-30	30-35	35-40	40-45
No. of Wages Earners	60	140	110	150	120	100	90

22 Construct a Histogram from the following data find the modal value:

Marks	1-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41-45
Freq.	7	10	16	32	24	18	10	5	1

(B.Com Osmania, SVU) (Ans: 19.33)

GEOMETRIC MEAN

Geometric mean is the nth root of the product of n items of a series. If there are 2 numbers, say a and b, the Geometric mean of the two numbers is the square root of the product of the 2 numbers. Thus, Geometric mean would be lab. Similarly, if there are 3 numbers, Geometric mean of the three numbers would be the cube root of the product of the 3 numbers. Thus, Geometric mean would be $(abc)^{1/3}$. This concept can be applied to as many numbers as possible.

Merits of Geometric Mean

1. It is rigidly defined. Hence, different interpretations by different persons are not possible.
2. It takes all values into consideration. Thus, it is more representative.
3. It can be subjected to further mathematical treatment. The properties of Geometric mean have been separately explained.
4. it has a bias towards lower values.
5. It is not affected much by presence of extremely small or extremely large observations.
6. It is not much affected by the fluctuations of sampling.

Drawbacks of Geometric Mean

1. It is neither simple to understand nor easy to calculate.
2. It cannot be determined by inspection.
3. It cannot be located graphically.
4. It cannot be used in the study of qualitative phenomena.
5. It may be a fictitious value i.e. One that does not exist in the series.
6. It cannot be computed if any value in a series is zero or negative.
7. It brings out the property of the ratio of change and not of absolute difference as in the case of arithmetic mean.
8. It cannot be calculated even if a single observation is missing or lost
9. It cannot be calculated in case the distribution has open-ended classes i.e., "below 10" or "above 80".
10. The property of giving more weight to smaller items may in some cases prove to be a drawback of the geometric mean.

Calculation of Geometric Mean

Individual Observations:

$G.M. = \text{Anti log } \frac{\Sigma \log X}{}$

Steps: (1) Ascertain logarithms of variable X and denote the total as $\sum \log x$

(2) Divide $\sum \log X$ by N and take the antilog of the value so obtained.

Illustration 42:

Compute the geometric mean of the following: *(B.Com. Osmania)*

2000, 200, 20, 12, 8, 0.8

Solution :

Geometric Mean = Antilog $\frac{\sum \log X}{N}$

Size (X)	log X
200	3.3010
200	2.3010
20	1.3010
12	1.0792
8	0.9031
0.8	1.9031
$\sum$**Log X =**	**8.7884**

Given $\sum \log X = 8.7884$, N=6

G.M. = Antilog $\left\{\frac{8.7884}{6}\right\}$ = A.L (1.4647) = 29.16

Illustration 43: Calculate the Geometric mean of the following two series

(a)	(b)
2574	0.8974
475	0.0570
75	0.0081
5	0.5677
O.8	0.0002
0.08	0.0984
0.005	0.0854
0.0009	0.5672

(B.Com. Madras)

Solution:

Series A		Series B	
X	**log X**	**X**	**log X**
2574	3.4106	0.8974	$\bar{1}.9530$
475	2.6767	0.0570	$\bar{2}.7559$
75	1.8751	0.0081	$\bar{3}.9085$
5	0.6990	0.5677	$\bar{1}.7541$
0.8	$\bar{1}.9031$	0.0002	$\bar{4}.3010$
08	$\bar{2}.9031$	0984	$\bar{2}.9930$
005	$\bar{3}.6990$	0.0854	$\bar{2}.9315$
0009	$\bar{4}.9542$	0.5672	$\bar{1}.7539$
N=8	$\sum$log X=2.1208	N=8	$\sum$logx= $\overline{10}.3508$

Geometric Mean = Antilog $\left(\frac{\sum \log X}{N}\right)$

Series A = Antilog $\left(\frac{2.1209}{8}\right)$ = a.L (0.2651) = 1.841

Series B = A.L $\left(\frac{\overline{10}.3508}{8}\right)$ = A.L. $\left(\frac{\overline{16}+6.3508}{N8}\right)$ = A.L. (2.7939) = 06222

Discrete Series:

Geometric Mean = Antilog $\frac{\Sigma \log X}{N}$

Steps: (i) Find the logarithms of variable X

(ii) Multiply these logarithms with respective frequencies and obtain the total, denote it as $\sum f \log x$.

(iii) Divide $\sum f \log x$ by total frequency and find Antilog of the value.

Illustration 44: Find out Geometric mean from the following data:

X	10	20	30	40	50	60
F	15	18	22	16	12	7

Solution: Calculation of Geometric Mean.

X	f	logx	flogx
10	15	1.0000	15.0000
22	18	1.3010	23.4180
30	22	1.4771	32.4962
40	16	1.6021	25.6336
50	12	1.6990	20.3880
60	7	1.7782	12.4474
	90		129.3832

Geometric Mean = Antilog $\frac{\Sigma \log X}{N}$

$= \text{Al} \frac{129.3832}{90} = \text{A.L. } 1.4376 = 27.39$

Continuous Series: G.M. = Antilog $\frac{\Sigma \log X}{N}$

Steps: (1) Find the mid-points of the classes and ascertain their logarithms.

(2) Multiply these logarithms with their frequencies, obtain the total and denote it as $\sum f \log m$.

(3) Divide this total by the total of the frequency and find the antilog of the value so obtained.

Illustration 45: Compute the Geometric Mean of the following:

Weight (lbs.)	No. of Persons	Weight (lbs.)	No. of Persons
100-200	12	130-140	22
110-120	18	140-150	18
120-130	25	150-160	10

Solution:

Computation of Geometric Mean

Weight	f	Mid-point	log m	f logm
100-110	12	108	2.0212	24.2544
110-120	18	115	2.0607	37.0926
120-130	25	125	2.0969	52.4225
130-140	22	135	2.1303	46.8666
140-150	18	145	2.1614	38.9050
150-160	10	155	2.1903	21.9030
	N=105			**$\sum f\log m$=221.4443**

Geometric Mean = Antilog $\left(\frac{\Sigma \log m}{N}\right)$ = Antilog $\frac{221.4443}{105}$ = A.L. 2.10899 =128.5

Illustration 46: The price of a commodity increased by 5% from 1991 to 1992, 8% from 1992 to 1993 and 77% from 1993 to 1994. Find the average increase from 1992 to 1994.

X	Price at the end of the year Taking preceding years as 100	Log X
1992	105	2.0212
1993	108	2.0334
1994	177	2.2480
N=3		$\sum$logx=6.3026

Geometric Mean = A.L. $\frac{\sum \log X}{N}$ = A.L. $\frac{6.3206}{x}$ = A.L. 2.1009 = 126.2

Thus the average increase from 1992 to 1994 = 126.2-100 = 26.2%

Applications of Geometric Mean: Geometric Mean is especially suitable in averaging rates, percentages and rates of increase between two periods. It is most appropriate average to be used if more weightage is required to be given to smaller items. It is used in the construction of index numbers. Geometric mean is used to find average percentage increase in prices, sales, production etc.

Algebraic Properties of Geometric Mean

1. If each item of a series is replaced by the geometric mean of the series, the product of all items in the series remains unchanged. To illustrate, given 4 numbers 8, 64, 4 and 32, the geometric mean of the 4 numbers is (8*64*2*32)¼ = 16. The product of the 4 numbers = 8*64*4*33 = 65536. If all the 4 products are replaced with 16, then the product of the 4 numbers = 16*16*16*16=65536.

2. It is possible to calculate Combined Geometric mean for two or more series. If G1, G2.... Are the geometric means of the series of sizes n1, n2..., respectively, the combined geometric G of the combined n1+n2+ ...., is given by

Log G = $\frac{n_1 \log G_1 + n_2 \log G_2 + \cdots}{n_1 + n_2 + \cdots}$

Illustration 47: Three groups of observations contain 8, 7 and 5 observations. Their geometric means are 8.52, 10.12 and 7.75 respectively. Find the geometric mean of 20 observations in the single group formed by pooling the three groups.

Solution:

Here n1=8, n2 = 7 and n3 = 5,

G1=8.52, G2=10.12 and G3=7.75.

The combined geometric mean would be:

$$\text{Log G} = \frac{M_1 \log G_1 + n_2 \log G_2 + M_3 \log G_3}{n_1 + n_2 + n_3}$$

$$= \frac{8 \log 8.52 + 7 \log 10.12 + 5 \log 7.75}{8+7+5}$$

= [(8*0.9304)+(7*1.0052)+(5*0.8893)]/20 = $\frac{18.9261}{20}$ = 0.9463

Combined geometric mean, G=Antilog (0.9463) = **8.837**

3. Given a series with geometric mean "g", the product of ratios of the geometric mean to the numbers less than or equal to the geometric mean is always equal to the product of ratios of the geometric mean to the numbers greater than the geometric mean. To illustrate, given 3 numbers 4, 16 and 64 their Geometric mean is equal to (4*16*64)1/3 = 16. The numbers 4 and 16 are less than or equal to 16. Thus ratio of these terms to the geometric mean 4/16 and 16/16, which is equal to ¼ and 1 respectively. Product of these

two ratios = ¼ *1 = ¼. Similarly, 64 is a number greater than geometric mean = 16/64 = ¼.

In other words, in case of G.M., if the ratios of the geometric mean to the figures, which are equal to or less than, are multiplied together, their product would be equal to the product of the ratios of figures more than the geometric mean.

Illustration 48: Compared to the previous year, the overhead expenses went up by 32 percent in 1986, they increased by 40 percent in the next year and by 50 percent in the following year. Calculate the average rate of increase in overhead expenses over three years. Explain clearly the reason for your choice of overage.

Solution : If arithmetic mean is used:

Year	Percentage rise X
1986	32
1987	40
1988	50
Total	$\sum X = 122$

$\overline{X} = \frac{\sum X}{N} = \frac{122}{3} = 40.67$

The arithmetic mean calculated above does not fully represent the situation. Geometric mean is accepted as the most appropriate average.

Calculation of Geometric Mean

Percentage rise in Overhead expenses	Expenses at the end of the year assuming preceding year as 100 (X)	Logarithms of X
32	132	2.1206
40	140	2.1461
50	150	2.1761
Total		log X = 6.4828

Geometric mean, G.M = Antilog $\left(\frac{\sum \log X}{N}\right)$ = Antilog $\left(\frac{6.4828}{3}\right)$

= Antilog (2.1476) = 140.5

Average rate of increase in overhead expenses = 140.5-100=40.5 percent.

The choice of geometric average has been made because the rates are being averaged.

Weighted Geometric Mean

When different observations are given different weights, then we are calculating a Weighted Average. The concept of weighted Geometric mean is the same as weighted arithmetic mean. Weights are assigned to various observations and the following formula is applied:

Weighted Geometric mean (G.M1) = A.L. $\left(\frac{\sum W \log X}{\sum w}\right)$

Steps : 1 Calculate logarithms of the variable X. **2.** Multiply such values by their respective weights, obtain the total ($\sum W$, logx). **3.** Divide this total by the total of weight and find Antilog of the value so obtained.

Illustration 49: Find the Weighted Geometric Mean from the following data.

Group	Index Number	Weights
A	220	40
B	230	15
C	260	13
D	180	10
E	200	12
F	120	10

Solution:

Calculation of Weighted Geometric Mean

Group	Index No. X	Weights W	log X	W.X logx
A	220	40	2.3424	93.6960
B	230	15	2.3617	35.4225
C	260	13	2.4150	31.3950
D	180	10	2.2553	22.5530
E	200	12	2.3010	27.6120
F	120	10	2.0792	20.7920
		$\Sigma W=100$		$\Sigma w.\log X=231.4735$

Weighted G.M. = A.L. $\left(\frac{\Sigma w.logX}{\Sigma w}\right)$ = A.L. = A.L. $\frac{231.4735}{100}$ = **206.4**

HARMONIC MEAN

It is reciprocal of the arithmetic mean of the reciprocals of the individual observations.

Merits of Harmonic Mean

1. It is rigidly defined. Its value is always definite.
2. It is also based on all observations of the series. It cannot be calculated in the absence of even a single figure.
3. It is capable of further algebraic treatment.
4. It is not affected very much by fluctuation of sampling
5. It gives greater importance to small items and as such a single big item cannot push up its value.
6. It measures relative changes and is extremely useful in averaging certain types of ratios and rates.

Drawbacks of Harmonic Mean

1. It is not easy to understand or calculate.
2. It accords a very high weightage to small items.
3. It is not very useful for analysis of economic data.
4. It is usually a fictitious value that does not exist in a series.
5. It cannot be computed in case of zero or negative values.
6. It is not a good representative of a statistical series, unless the phenomenon is such where small items have to be given a very high weightage.

Calculation of Harmonic Mean

Individual Observations:

Harmonic Mean = $\frac{N}{\Sigma(\frac{1}{X})}$

Steps : (1) Calculate reciprocals of the various items and obtain the total

(2) Divide the number of items by the total of the reciprocals

Illustration 50 : Compute Harmonic mean for the following

20, 40, 50, 80, 100 *(B.Com. Osmania)*

Solution :

X	$\frac{1}{X}$
20	0.0500
40	0.0250
50	0.0200
80	0.0125
100	0.0100
N=5	$\Sigma\left(\frac{1}{X}\right)$=0.1175

Harmonic Mean = $\frac{N}{\Sigma(\frac{1}{X})} = \frac{5}{0.1175} = 42.5532$

Discrete Series: Harmonic Mean = $\frac{N}{\Sigma\left(fx\frac{1}{x}\right)}$

Steps: (1) Take the reciprocals of the various items of variable X

(2) Multiply the reciprocals by respective frequencies and obtain the total and denote it $\Sigma\left(f\,x\,\frac{1}{x}\right)$

(3) Apply the formula.

Illustration 51: Calculate the harmonic mean from the following data.

X	f	$\frac{1}{x}$	$f\,x\,\frac{1}{x}$
10	5	0.1000	0.5000
20	16	0.0500	0.8000
30	18	0.0333	0.5994
40	12	0.0250	0.3000
50	24	0.0200	0.4800
60	15	0.0167	0.2505
	N= 90		$\Sigma f\,x\frac{1}{2}$ = 2.9299

Harmonic Mean = $\frac{N}{\Sigma\left(fx\frac{1}{x}\right)} = \frac{90}{2.9299} = 30.7178$

Continuous Series:

The formula and the steps are same. Instead of taking reciprocals of the variables, we take the reciprocal of the mid-values. An easier method, would be to divide the frequencies with the respective mid-values.

Illustration 52: Calculate the harmonic mean from the following data.

Marks	0-10	10-20	20-30	30-40	40-50
Freq.	8	10	15	6	2

Solution:

Calculation of Harmonic Mean

Marks	f	mid-values	f/m
0-10	8	5	1.6000
10-20	10	15	0.6667
20-30	15	25	0.6000
30-40	6	35	0.1714
40-50	2	45	0.0444
	N=41		$\sum f/m$=3.0825

$$\text{Harmonic mean} = \frac{N}{\Sigma\frac{f}{m}} = \frac{41}{3.0825} = 13.30089$$

Note: Harmonic mean is useful for computing the average rate of increase in profits of a concern, average speed at which a journey has been performed or the average price at which an article has been sold.

Illustration 53: A person drives for 100 km at a speed of 30 kmph. He then makes the return journey at a speed of 20 kmph. What is the average speed per hour?

Solution: In such cases harmonic mean would give correct result.

$$\text{Harmonic Mean} = \frac{N}{\Sigma(1/x)}$$

$$= \frac{2}{\Sigma\left(\frac{1}{30}+\frac{1}{30}\right)} = \frac{2}{\Sigma(0.0333+0.0500)} = \frac{2}{0.0833} = 24 \text{ k. p. Hr.}$$

Weighted Harmonic Mean

The concept of weighted Harmonic mean is similar to the concept of Weighted Arithmetic and Geometric means. For example, if different distances are travelled at different speeds, the average speed can be computed using weighted H.M. Let us suppose that distances, w1, w2, ... Wn are travelled with speeds $X_1, X_2, \ldots, X_n$ per unit of time. If $t_1, t_2 \ldots\ldots, t_n$ are the respective times taken to cover the distances, then we have

$$T_1 = \frac{W1}{X1} \quad t2 = \frac{W1}{X2} \ldots, tn = \frac{Wn}{Xn}$$

Hence

$$\text{Average Speed} = \frac{\text{Total distance travelled}}{\text{Total time taken}}$$

$$= \frac{W1+W2+\cdots+Wn}{t1+t2+\cdots+tn} = \frac{W1+W2+\cdots+Wn}{\left(\frac{W1}{X1}+\frac{W2}{X2}+\cdots+\frac{Wn}{Xn}\right)} = \frac{\Sigma w}{\Sigma(w/X)}$$

which is the weighted harmonic mean of the speeds, the corresponding weights being the distances covered.

Illustration 54: A person walks 10 km. At 3 km. An hour, 6 km at 3 km. An hour and 4 km. AT 2 km. An hour. Find his average speed per hour.

Solution:

Computation of Weighted H.M.

Rate at which Travelled X	Distance travelled w	1/X	wx(1/X)
3	10	0.3333	3.333
3	6	0.3333	2.000
2	4	0.5000	2.000
Total	**20**		**7.333**

H.M. = 20/7.333=2.73

Hence the average speed was 2.73 k. Per hour.

Choice of a Suitable Average

We have learnt about a number of averages being used for statistical analysis of data. In order to arrive at the correct conclusions, it is important to choose the right measure or average. This choice should be based on various factors listed below:

1. **Objective:** The average should be chosen in accordance with the objective of investigation or inquiry. To determine the most fashionable or most frequently occurring item mode should be computed. Median should be the choice in case the object is to determine an average that would indicate its position or remaining in relation to all the values.
2. **Weight age to various values:** If all the values in a series are to be given equal importance, arithmetic mean will be a suitable average. If small items are to be given greater importance than the big ones, geometric mean is the best average. Harmonic mean is suitable where it is desired to give larger weight age to smaller items. Weighted arithmetic average is computed when it is desired to give due importance to different items of a series.
3. **Representative:** The average chosen should be such that it represents the basic characteristics of the data. For example, the geometric mean best represents the ratio or percentage changes while the arithmetic average represents best changes in absolute magnitudes.
4. **Nature and the form of distribution:** If the distribution is symmetrical, arithmetic mean, median or mode may be used almost interchangeably. In case the distribution tails off on either of the sides, mode or median would be preferable.
5. **Open Ended Classes:** In case of open-end classes, Mode and Median are preferable. Similarly, Median and Mode can be chosen if the data has extreme values.
6. **Unequal Class Intervals:** Median is most suitable average in case of a frequency distribution involving varying class intervals.
7. **Variation in data:** When the data is properly spread out and does not have huge variations, arithmetic average is most appropriate. This is the most frequently used average when we refer to various averages such as average cost of production, average price, etc.
8. **Unimodal /Multi modal distribution:** If the distribution uni-modal, Mode is suitable as a general purpose average. If the distribution is not unimodal, a mathematical average such as arithmetic or geometric mean should be considered.
9. **Type of Data:** When the data comprises rates, percentages or ratio instead of actual observations or quantities, geometric mean is most appropriate.

10. **Need for further Analysis:** If the average has to be used for further algebraic treatment, mathematical averages in general and arithmetic mean in particular is most suitable.
11. **Sampling Stability:** Mathematical averages, particularly the arithmetic mean, is most suitable in testing of hypothesis based on sampling techniques.
12. **Qualitative phenomenon:** median is appropriate in case of phenomena such as intelligence, honesty, etc that is not capable of precise quantitative studies.

Empirical Relationship between Mean, Median and Mode:

Mode touches the peak of the curve indicating maximum frequency. Median divides the area of the curve in two equal halves and Mean is the centre of gravity. The three values are inter-related. The following points list the relationship between the various averages.

- In a distribution, the relative positions of mean, median, and the mode depend upon the Skewness of the distribution. If the distribution is symmetrical, then Mean = Median = Mode.
- If the distribution is positively skewed (the longer tail of the distribution is towards the right), the Mode will be lesser than the Median, which in turn will be lower than the Arithmetic Mean. In case of a negatively skewed distribution, the Mode will be greater than the Median, which in turn will be greater than the Arithmetic Mean.
- In a given distribution, if all the observations are positive, Arithmetic Mean is greater than Geometric Mean, which in turn is greater than Harmonic Mean.
- With a distribution of moderate Skewness, median tends to be approximately 1/3rd as far away from the mean as from the mode.
- Mode=3 Median-2 Mean
- Mean-Mode=3 (Mean-Median). Thus, the difference between Mean and Mode is three times, the difference between mean and median. In other words median is closer to the mean than mode.
- Simplifying, we get Mean = ½ (3 Median – Mode)

Median = Mode+2/3 (Mean-Mode) = 1/3 (2 Mean-Mode)

These relationships in the values of mean, median and mode are of great value. In cases where mode is till-defined, or the series is bi-modal and mode cannot be calculated by using the formula, it can be ascertained by using these empirical relationships.

Comparison of various Measures of Central Tendencies

The arithmetic mean (or the mean), the geometric mean and the harmonic mean are rigidly defined. They are based upon all the observations and are suitable for further arithmetic treatment. The mean can be known even when the number of items and their aggregate values are given. The mean gives greater importance to bigger or smaller items of a given series. For example, the 'Chief Executive's salary would greatly affect the average salary of the employees in factory where majority consists of low paid workers. In a given series, it is difficult to locate the mean or the geometric mean or the harmonic mean by inspection, while the median and the mode can easily be located. Furthermore, in order to compute the mean, the geometric mean and the harmonic mean, every item in the series is required to be known while it is not so for the median or mode. Unlike median and in some cases the mode, the mean, geometric mean and the harmonic mean cannot be exactly located in a series. The median given the best result in a study of qualitative measurements such as intelligence, honesty, virtue, beauty, etc. Unlike mean, the median cannot be used to calculate the aggregate value of items if the number of items is known. Suppose the median of several series, whereas this is possible in case of mode, the extreme value of items have no effect on the mode provided they are not in the modal class. The mode can be easily determined from the graph.

The geometric mean is more useful when we want to find the rate of growth of population or the rate growth of industrial production in a country, etc. It is also very extensively used in the construction of index numbers. Like mean, the geometric mean is based on all the observations in a series and therefore, one cannot compute it if one of the observations is missing. Moreover, it is comparatively difficult to compute, and cannot be calculated if the series contains a negative value for an observation.

The harmonic mean is the most appropriate average when calculating the average speed of a vehicle (train) – when the speed is expressed in kilometers per hour, etc. Like mean and the geometric mean, the harmonic mean is also based on all the observations in a series and therefore, has the same shortcoming as the mean. Further, it is difficult to compute and it gives greater importance to small items.

Illustration 55: In a given distribution, Arithmetic Mean = 35, Median = 36, What is the Mode? ***(B.Com. Osmania)***

Solution:

Mode = 3 Median -2 Mean = 3 x 36 – 2 x 35 = 108–70 = 38

Illustration 56: In a moderately symmetrical distribution the value of median is 42.8 and the value of mode is 40. Find the mean.

Solution:

3 Median – 2 Mean = Mode

3 x 42.8 – 2 Mean = 40

128.4 – 2 Mean = 40

-2 Mean = 40-128.4

-2 Mean = -88.4

Mean = 44.2

Illustration 57: The mode and mean are 28 and 25 respectively. Find value of median. ***(B.Com. Kakatiya)***

Solution:

3 Median -2 Mean = Mode

3 Median -2 x 25 = 28

3 Median = 78

Median = 26

Exercise 4 (d)

1 Find the value of Geometric Mean:
12345, 1234.5, 123.4, 12.345, 1.2345 ***(B.Com. Kakatiya) (Ans 123.4)***

2. Compute the geometric mean of the data given below:
2200, 220, 22, 0.22, 0.022, 0.0022 ***(B.Com. Osmania) (Ans: 2.199)***

3. Compute the geometric mean of the following data:
6.5, 169.0, 11.0, 112.5, 14.2, 75.5, 35.5 ***(B.Com. Kakatiya) (Ans: 33.92)***

4. Calculate geometric mean of the following figures.
0.00400, 0.00567, 0.00900, 0.04561, 0.09221 ***(B.Com. Osmania) (Ans: 0.01537)***