4

Measures of Central Tendency

In the previous chapters, data collection and presentation of data were discussed. Even after the data have been classified and tabulated one often finds too much details for many uses that may be made of the information available. We, therefore, frequently need further analysis of the tabulated data. One of the powerful tools of analysis is to calculate a single *average value* that represents the entire mass of data. The word average is very commonly used in day-to-day conversation. For example, we often talk of average work, average income, average age of employees, etc. An 'average' thus is a single value which is considered as the most representative or typical value for a given set of data. Such a value is neither the smallest nor the largest value, but is a number whose value is somewhere in the middle of the group. For this reason an average is frequently referred to as a measure of central tendency or central value. Measures of central tendency show the tendency of some central value around which data tends to cluster.

OBJECTIVES OF AVERAGING

There are two main objectives of the study of averages :

(*i*) *To get one single value that describes the characteristics of the entire data.* Measures of central value, by condensing the mass of data in one single value, enable us to get an idea of the entire data. Thus one value can represent thousands, lakhs and even millions of values. For example, it is impossible to remember the individual incomes of millions of earning people of India and even if one could do it there is hardly any use. But if the average income is obtained, we get one single value that represents the entire population. Such a figure would throw light on the standard of living of an average Indian.

(*ii*) *To facilitate comparison.* Measures of central value, by reducing the mass of data in one single figure, enable comparisons to be made. Comparison can be made either at a point of time or over a period of time. For example, the figure of average sales for December may be compared with the sales figures of previous months or with the sales figure of another competitive firm.

CHARACTERISTICS OF A GOOD AVERAGE

Since an average is a single value representing a group of values, it is desirable that such a value satisfies the following properties :

(*i*) *It should be easy to understand.* Since statistical methods are designed to simplify complexity, it is desirable that an average be such that can be readily understood, its use is bound to be very limited.

(*ii*) *It should be simple to compute.* Not only an average should be easy to understand but also should be simple to compute so that it can be used widely. However, though ease of computation is desirable, it should not be sought at the expense of other advantages, *i.e.*, if in the interest of greater accuracy, use of a more difficult average is desirable one should prefer that.

(iii) *It should be based on all the observations.* The average should depend upon each and every observation so that if any of the observation is dropped average itself is altered.

(iv) *It should be rigidly defined.* An average should be properly defined so that it has one and only one interpretation. It should preferably be defined by an algebraic formula so that if different people compute the average from the same figures they all get the same answer (barring arithmetical mistakes).

(v) *It should be capable of further algebraic treatment.* We should prefer to have an average that could be used for further statistical computations. For example, if we are given separately the figures of average income and number of employees of two of more companies we should be able to compute the combined average.

(vi) *It should have sampling stability.* We should prefer to get a value which has what the statisticians call 'sampling stability'. This means that if we pick 10 different group of college students, and compute the average of each group, we should expect to get approximately the same values. It does not mean, however, that there can be no difference in the value of different samples. There may be some difference but those averages in which this difference, technically called 'sampling fluctuation,' is less are considered better than those in which this difference is more.

(vii) *It should not be unduly affected by the presence of extreme values.* Although each and every observation should influence the value of the average, none of the observations should influence it unduly. If one or two very small or very large observations unduly affect the average, *i.e.*, either increase its value or reduce its value, the average cannot be really typical of the entire set of data. In other words, extremes may distort the average and reduce its usefulness.

The following are the important measures of central tendency which are generally used in business :

A. Arithmetic mean, B. Median, C. Mode, D. Geometric mean, and E. Harmonic mean

A. ARITHMETIC MEAN

The most popular and widely used measure for representing the entire data by one value is what most laymen call an 'average' and what the statisticians call the arithmetic mean. Its value is obtained by adding together all the observations and by dividing this total by the number of observations.

Calculation of Arithmetic Mean—Ungrouped Data

For ungrouped data, arithmetic mean may be computed by applying any of the following methods :

(i) Direct method,

(ii) Short-cut method.

(*i*) Direct Method : The arithmetic mean, often simply referred to as mean, is the total of the values of a set of observations divided by their total number of observations. Thus, if $X_1, X_2, \dots\dots\dots, X_N$ represent the values of N items or observations, the arithmetic mean denoted by $\overline{X}$ is defined as :

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_N}{N} = \frac{\sum_{i=1}^{N} X_i}{N}$$

If the subscripts are dropped, the formula becomes :

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

It may be pointed out that in keeping with standard statistical practice, the symbol $\overline{X}$ will represent throughout this text the arithmetic mean of a set of observations.

Illustration 1. The monthly income (in rupees) of 10 employees working in a firm is as follows :

4487 4493 4502 4446 4475 4492 4572 4516 4468 4489

Find the average monthly income.

Solution. Let income be denoted by X.

ΣX = 4487+4493+4502+4446+4475+4492+4572+4516+4468+4489 = 44,940

$$\overline{X} = \frac{\Sigma X}{N} = \frac{44940}{10} = 4494$$

Hence the average monthly income is Rs. 4,494.

***(ii)* Short-cut Method.** The arithmetic mean* can also be calculated by taking deviations from any arbitrary point in which case the formula shall be :

$$\overline{X} = A + \frac{\Sigma d}{N}$$

where $d = (X - A)$

and A = Arbitrary point (or assumed mean).

It should be noted that any value can be taken as arbitrary point and the answer would be the same as obtained by the direct method.

Illustration 2. Calculate average monthly income by the short-cut method from data of Illustration 1 taking deviations from 4460 as the arbitrary point.

Solution. **CALCULATION OF AVERAGE INCOME**

X (Rs.)	$(X-4460)$ d
4487	+27
4493	+33
4502	+42
4446	−14
4475	+15
4492	+32
4572	+112
4516	+56
4468	+ 8
4489	+29
	Σd = +340

$$\overline{X} = A + \frac{\Sigma d}{N} = 4460 + \frac{340}{10} = 4460 + 34 = \text{Rs. } 4494.$$

One may find that the short-cut method takes more time as compared to direct method. However, this is true only for ungrouped data. In case of grouped data, considerable saving in time is possible by adopting the short-cut method.

Calculation of Arithmetic Mean—Grouped Data

For grouped data, arithmetic mean may be computed by applying any of the following methods :

(*i*) Direct method,

(*ii*) Short-cut method.

---

*This formula is derived as follows :

Let $d = X - A$ or $X = A + d$

Taking summation of both sides and dividing by N, we get

$$\frac{\Sigma X}{N} = \frac{\Sigma A}{N} + \frac{\Sigma d}{N} \quad \text{or} \quad \overline{X} = A + \frac{\Sigma d}{N}$$

The population mean is denoted by μ (μ is the Greek letter mu) and the sample mean by $\overline{X}$.

(i) Direct Method. When direct method is used

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

where X = mid-point of various classes.
f = the frequency of each class.
N = the total frequency.

Note. For computing mean in the case of grouped data the mid-points of the various classes are taken as representative of that particular class. The reason is that when the data are grouped, the exact frequency with which each value of the variable occurs in the distribution in unknown. We only know the limits within which a certain number of frequencies occur. For example, when we say that the number of persons within the income group 4000–4500 is 50 we cannot say as to how many persons out of 50 are getting 4001, 4002, 4003, etc. We, therefore, make an assumption while calculating arithmetic mean that the frequencies within each class are distributed uniformly or evenly over the range of the class-interval, *i.e.*, there will be as many observations below the mid-point as above it. Unless such an assumption is made, the value of mean cannot be computed.

Illustration 3. The following are the figures of profits earned by 1,400 companies during 2003-04.

Profits (Rs. Lakhs)	*No. of Companies*	*Profits (Rs. Lakhs)*	*No. of Companies*
200–400	500	1,000–1,200	100
400–600	300	1,200–1,400	80
600–800	280	1,400–1,600	20
800–1000	120		

Calculate the average profits for all the companies.

Solution. **CALCULATION OF AVERAGE PROFITS**

Profits (Rs. Lakhs)	*Mid-points X*	*No. of Companies f*	*fX*
200–400	300	500	1,50,000
400–600	500	300	1,50,000
600–800	700	280	1,96,000
800–1,000	900	120	1,08,000
1,000–1,200	1100	100	1,10,000
1,200–1,400	1300	80	1,04,000
1,400–1,600	1500	20	30,000
		N = 1,400	ΣfX = 8,48,000

$$\bar{X} = \frac{\sum fX}{N} = \frac{8,48,000}{1,400} = 605.71$$

Thus, the average profit is Rs. 605.71 lakhs.

---

* $$\bar{X} = \frac{\sum fX}{N} \qquad \text{(direct method)}$$

Now $d = \frac{X - A}{i}$, $\quad \therefore \; X = A + id$

Substituting the value of X in the direct method, we get

$$\bar{X} = \frac{\sum f(A+id)}{N} = \frac{\sum fA}{N} + \frac{\sum fd}{N} \times i$$

Hence $$\bar{X} = A + \frac{\sum fd}{N} \times i \qquad (\because \Sigma f = N)$$

(*ii*) **Short-cut Method***. When short-cut method is used, the following formula is applied :

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

where $$d = \frac{X - A}{i}$$

and i = size of the equal class interval.

Illustration 4. Calculate the average profit by the short-cut method from the data of Illustration 3.

Solution. **CALCULATION OF AVERAGE PROFITS**

Profits (Rs. Lakhs)	*Mid points X*	*f*	*(X–900)/200 d*	*fd*
200–400	300	500	–3	–1,500
400–600	500	300	– 2	–600
600–800	700	280	– 1	– 280
800–1000	900	120	0	0
1000–1200	1100	100	+ 1	+ 100
1200–1400	1300	80	+ 2	+ 160
1400–1600	1500	20	+ 3	+ 60
		N = 1,400		Σfd = – 2,060

$$\overline{X} = A + \frac{\Sigma fd}{N} \times i = 900 - \frac{2,060}{1,400} \times 200 = 900 - 294.29 = \text{Rs. } 605.71$$

Hence the average profit is Rs. 605.71 lakhs.

Correcting Incorrect Values

It sometimes happens that due to an oversight or mistake in copying certain wrong values are taken while calculating the mean. The problem is how to find out the correct mean. The process is very simple. From ΣX deduct wrong observations and add correct observations and then divide the correct ΣX by the number of observations and the result so obtained will give the value of the correct mean.

Illustration 5. The average weekly for a group of 25 persons working in a factory was calculated to be Rs. 378.40. It was later discovered that one figure was misread as 160 instead of the correct value Rs. 200. Calculate average wage.

Solution. $\Sigma X = N\overline{X} = 25 \times 378.4 = 9460$

Less: Incorrect figure 160

9300

Add: Correct figure 200

Total 9500

∴ *Correct* $\Sigma X = 9500$

Hence correct average $= \frac{9500}{25} = 380.$

(*b*) The mean of 200 observations was 50. Later on, it was discovered that two observations were wrongly read as 92 and 8 instead of 192 and 88. Find out the correct mean.

Solution.

Here $\overline{X} = 50$ and $N = 200$

$\Sigma X = 200 \times 50 = 10,000$

Less Incorrect observations 100

9,900

Add Correct observation correct total = $\frac{280}{10,180}$

Correct mean = $\frac{10,180}{200}$ = 50.9

Mathematical Properties of Arithmetic Mean

The important mathematical properties of arithmetic mean are :

1. The algebraic sum of the deviations of all the observations from arithmetic mean is always zero, *i.e.*, $\Sigma (X - \overline{X}) = 0$. This shall be clear from the following example :

X	$(X - \overline{X})$
10	− 20
20	− 10
30	0
40	+ 10
50	+ 20
$\Sigma X = 150$	$\Sigma (X - \overline{X}) = 0$

Here $\overline{X} = \frac{\Sigma X}{N} = \frac{150}{5} = 30$. When the sum of the deviations from the actual mean, *i.e.*, 30, is taken it comes out to be zero. It is because of this property that the mean is characterised as a *point of balance, i.e.,* the sum of the positive deviations from mean is equal to the sum of the negative deviations from mean.

2. The sum of the squared deviations of all the observations from arithmetic mean is minimum, that is, less than the squared deviations of all the observations from any other value than the mean. The following example would clarify the point :

X	$(X - \overline{X})$	$(X - \overline{X})^2$
2	− 2	4
3	− 1	1
4	0	0
5	+ 1	1
6	+ 2	4
$\Sigma X = 20$	$\Sigma (X - \overline{X}) = 0$	$\Sigma (X - \overline{X})^2 = 10$

The sum of the squared deviations is equal to 10 in the above case. If the deviations are taken from any other value, the sum of the squared deviations would be greater than 10. This is known as the least square property of the arithmetic mean and becomes the basis for defining the concept of standard deviation.

3. If we have the arithmetic mean and number of observations of two or more than two related groups, we can compute combined average of these groups by applying the following formula :

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

$\overline{X}_{12}$ = Combined mean of the two groups.

$\overline{X}_1$ = Arithmetic mean of the first group.

$\overline{X}_2$ = Arithmetic mean of the second group.

N_1 = Number of observations in the first group.

N_2 = Number of observations in the second group.

The following example will illustrate the application of the above formula :

Illustration 6(*a*). There are two branches of a company employing 100 and 80 employees respectively. If arithmetic means of the monthly salaries paid by two branches are Rs. 4570 and Rs. 6750 respectively, find the arithmetic mean of the salaries of the employees of the company as a whole.

Solution. We should compute the combined mean. The formula is

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

Given $N_1 = 100, \overline{X}_1 = 4570, N_2 = 80, \overline{X}_2 = 6750$

$$\therefore \quad \overline{X}_{12} = \frac{(100 \times 4570) + (80 \times 6750)}{100 + 80} = \frac{997000}{180} = 5538.89$$

If we have to find out the combined mean of three related groups, the above formula can be extended as follows :

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

Illustration 6. (*b*) The mean of marks in Statistics of 100 students of a class was 72. The mean of marks of boys was 75, while their number was 70. Find out the mean marks of girls in the class. (*MBA, Osmania Univ, 2006*)

Solution. We are given $N = 100$, $\overline{X}_{12} = 72$, $\overline{X}_1$, *i.e.*, mean marks of boys = 75, N_1 = number of boys = 70. We have to find out the mean marks of girls, *i.e.*, $\overline{X}_2$.

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

$$72 = \frac{70(75) + 30\overline{X}_2}{70 + 30}$$

$$7200 = 5250 + 30\,\overline{X}_2$$

$$\overline{X}_2 = \frac{1950}{30} = 65$$

Hence mean marks of girls in the class = 65.

Illustration 6. (*c*) The mean age of a combined group of men and women is 30 years. If mean age of the group of men is 32 and that of the group of women is 25, find out the percentage of men and women in the group.

Solution. Let N_1 represent percentage of men and N_2 percentage of women so that $N_1 + N_2 = 100$.

We are given $\overline{X}_{12} = 30$

$\overline{X}_1 = 32$ (mean age of group of men)

$\overline{X}_2 = 25$ (mean age of group of women)

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

$$30 = \frac{N_1(32) + N_2(25)}{100}$$

$$3000 = 32N_1 + (100 - N_1)25$$

$$32N_1 + 2500 - 25N_1 = 3000 \quad \text{or} \quad N_1 = 71.43$$

$$N_2 = 100 - 71.43 = 28.57$$

Hence the percentage of men and women is respectively 71.43 and 28.57.

Merits and Limitations of Arithmetic Mean

The arithmetic mean is the most popular average in practice. It is due to the fact that it possesses first six out of seven characteristics of a good average and no other average possesses such a large number of characteristics.

However, arithmetic mean is unduly affected by the presence of extreme values. Also in open-end frequency distribution, it is difficult to compute mean without making assumption regarding the size of the class-interval of the open-end classes. The arithmetic mean is usually neither the most commonly occurring value nor the middle value in a distribution and in extremely asymmetrical distribution, it is not a good measure of central tendency.

Weighted Arithmetic Mean

One of the limitations of the arithmetic mean discussed above is that it gives equal importance to all the observations. But there are cases where the relative importance of the different observations is not the same. When this is so, we compute weighted arithmetic mean. The terms 'weight' stands for the relative importance of the different observations. The formula for computing weighted arithmetic mean is :

$$\overline{X}_w = \frac{\Sigma WX}{\Sigma W}$$

where $\overline{X}_w$ represents the weighted arithmetic mean

X = The variable.

W = Weights attached to the variable X.

An important problem that arises while using weighted mean is regarding selection of weights. Weights may be either actual or arbitrary, *i.e.,* estimated. Needless to say, if actual-weights are available, nothing like this. However, in the absence of actual-weights, arbitrary or imaginary weights may be used. The use of arbitrary weights may lead to some error, but this is better than no weights at all. In practice, it is found that if weights are intelligently assigned keeping the phenomena in view, the error involved will be so small that it can be easily overlooked.

Weighted mean is specially useful in problems relating to the construction of index numbers and standardised birth and death rates.

Illustration 7. A contractor employs three types of workers—male, female and children. To a male worker he pays Rs. 200 per day, to a female worker Rs. 150 per day and to a child worker Rs. 100 per day. What is the average wage per day paid by the contractor?

Solution. The average wage is not the simple arithmetic mean, *i.e.,* $\frac{200+150+100}{3}$ = Rs. 150 per day. If we assume that the number of male, female and child workers is the same, this answer would be correct. For example, if we take 10 workers in each case then the average wage would be

$$\overline{X} = \frac{(10\times 200)+(10\times 150)+(10\times 100)}{10+10+10} = \frac{2000+1500+1000}{30} = \frac{4500}{30} = \text{Rs. }150$$

However, the number of male, female and child workers employed is generally different. If we know how many workers of each type are employed by the contractor in question, nothing like this. However, in the absence of this we take assumed weights. Let us assume that the number of male, female and child workers employed is 20, 15 and 5, respectively. The average wage would be the weighted mean calculated as follows :

Wage per day (Rs.) X	*No. of workers* W	WX
200	20	4000
150	15	2250
100	5	500
	$\Sigma W = 40$	$\Sigma WX = 6750$

$$\overline{X}_w = \frac{\Sigma WX}{\Sigma W} = \frac{6750}{40} = 168.75$$

Hence the average wage per day paid by the contractor is Rs. 168.75 to all types of workers.

B. MEDIAN

The median is the measure of central tendency which appears in the "middle" of an ordered sequence of values. That is, half of the observations in a set of data are lower than it and half of the observations are greater than it.

As distinct from the arithmetic mean which is calculated from the *value of every* observation in the series, the median is what is called a *positional* average. The term 'position' refers to the place of a value in a series. The place of the median in a series is such that an equal number of observations lie on either side of it. For example, if the income of five persons is Rs. 7000, 7200, 7500, 7600, 7800, then the median income would be Rs. 7500. Changing any or both of the first two values with any other numbers with value of 7500 or less and/or changing any of the last two values to any other values with values of 7500 and more, would not affect the value of the median which would remain 7500. In contrast, in case of arithmetic mean the change in value of single observation would cause the value of the mean to be changed. Median is thus the central value of the distribution or the value that divides the distribution into two equal parts. If there are even number of observations in a series, there is no actual value exactly in the middle of the series and as such the median is indeterminate. In this case, the median is arbitrarily taken to be halfway between the two middle observations. For example, if there are 10 observations in a series, the median position is 5.5, that is the median value is halfway between the value of the observations that are 5th and 6th in order of magnitude. Thus when N is odd, the median is in an actual value with the remainder of the series in two equal parts on either side of it. If N is even, then the median is a derived figure, *i.e.*, half the sum of two values.

Calculation of Median—Ungrouped Data

Arrange the data in ascending or descending order of magnitude. (Both arrangements would give the same answer.)

Apply the formula : Median = Size of $\frac{N+1}{2}$th observation.

Illustration 8. From the following data of wages of 7 workers, compute the median wage :

Wages (in Rs.)	4600	4650	4580	4690	4660	4606	4640

Solution :

CALCULATION OF MEDIAN

S. No.	*Wages arranged in ascending order*
1	4580
2	4600
3	4606
4	4640
5	4650
6	4660
7	4690

Median = Size of $\frac{N+1}{2}$th observation = $\frac{7+1}{2}$ = 4th observation.

Value of 4th observation is 4640. Hence median wages = Rs. 4640.

In the above illustration, the number of observations was odd and, therefore, it was possible to determine the value of 4th observation. When the number of observations is even, for example, if in the above case the number of observations are 8 the median would be the value of $\frac{8+1}{2}$ = 4.5th observation. For finding out the value of 4.5th observation, we shall take the average of 4th and 5th observations. Hence the median shall be

$$\frac{4640+4650}{2} = 4645$$

Calculation of Median—Grouped Data

Determine the particular class in which the value of median lies. Use $\frac{N}{2}$ to locate the median class and not $\frac{N+1}{2}$ because in the use of grouped data it is $N/2$ which divides the area of the curve into two equal parts.

Apply the following formula for determining the exact value of median :

$$\text{Median} = L + \frac{N/2 - p.c.f.}{f} \times i$$

L = Lower limit of median class, *i.e.,* the class in which the middle observation in the distribution lies.

$p.c.f.$ = Preceding cumulative frequency to the median class.

f = Frequency of the median class.

i = The class-interval of the median class.

Illustration 9. (*a*) 1,500 workers are working in an industrial establishment. Their age is classified as follows :

Age (yrs.)	*No. of workers*	*Age (yrs.)*	*No. of workers*
18–22	120	38–42	184
22–26	125	42–46	162
26–30	280	46–50	86
30–34	260	50–54	75
34–38	155	54–58	53

Calculate the median age.

Solution :

CALCULATION OF MEDIAN AGE

Age group	*f*	*c.f.*
18–22	120	120
22–26	125	245
26–30	280	525
30–34	260	785
34–38	155	940
38–42	184	1,124
42–46	162	1,286
46–50	86	1,372
50–54	75	1,447
54–58	53	1,500

Median = Size of $\frac{N}{2}$th observation = $\frac{1,500}{2}$ = 750th observation.

Hence median lies in the class 30–34.

$$\text{Median} = L + \frac{N/2 - p.c.f}{f} \times i = 30 + \frac{750-525}{260} \times 4 = 30+3.46 = 33.46$$

Hence the median age of the workers is 33.46 years.

(*b*) Calculate the median from the following data pertaining to the profits (in crore Rs.) of 125 companies :

Profits (Rs. crore)	*No. of companies*
less than 10	4
less than 20	16
less than 30	40
less than 40	76
less than 50	96
less than 60	112
less than 70	120
less than 80	125

(*MBA, MD Univ., 2000*)

Solution :

CALCULATION OF MEDIAN

Profits (Rs. Crore)	*No. of companies (f)*	*c.f.*
0 — 10	4	4
10 — 20	12	16
20 — 30	24	40
30 — 40	36	76
40 — 50	20	96
50 — 60	16	112
60 — 70	8	120
70 — 80	5	125

Med. = Size of $\frac{N}{2}$ th observation = $\frac{125}{2}$ = 62.5th observation.

Median lies in the class 30—40.

$$\text{Med.} = L + \frac{N/2 - p.c.f.}{f} \times i$$

$$L = 30, N/2 = 62.5, p.c.f. = 40, f = 36, i = 10$$

$$\therefore \quad \text{Med.} = 30 + \frac{62.5 - 40}{36} \times 10 = 30 + 6.25 = 36.25.$$

Hence 50% of the companies have profits upto Rs. 36.25 crores and the remaining 50% of the companies have profits more than Rs. 36.25 crores.

Merits and Limitations of Median

The median is superior to arithmetic mean in certain respects. For example, it is especially useful in case of open-end distribution and also it is not influenced by the presence of extreme values. In fact when extreme values are present in the data, the median is a more satisfactory measure of central tendency than the mean.

The sum of the deviations of observations from median (ignoring signs) is *minimum.* In other words, the absolute deviation of observations from the median is less than from any other value in the distribution. For example, the median of items 4, 6, 8, 10 and 12 is 8. The deviations from 8 ignoring signs are 4, 2, 0, 2, 4 and the total is 12. This total will be smaller than the one obtained if deviations are taken from any other value. Thus, if deviations are taken from 7, the deviations ignoring signs would be 3, 1, 1, 3, 5 and the total is 13. In an estimation situation, if one is interested in minimising the absolute amount of error and the sign of the error is not particularly important, then the median is preferable to arithmetic mean.

However, since median is a positional average, its value is not determined by each and every observation. Also median is not capable of algebraic treatment. For example, median cannot be used for determining the combined median of two or more groups. Also the median is less reliable average than the mean for estimation purposes since it is more affected by sampling fluctuations. Furthermore, the median tends to be rather unstable value if the number of observations is small.

Related Positional Measures or Quantities

Besides median, there are other measures which divide a series into a equal number of parts. Important amongst these are quartiles, deciles and percentiles. These quartiles, deciles and percentiles are all special cases of *quantities.* Quartiles are those values of the variate which divide the total frequency into four equal parts, deciles divide the total frequency in 10 equal parts and the percentiles divide the total frequency in 100 equal parts. Just as one point divides a series into two parts, three points would divide it into four parts, 9 points into 10 parts and 99 points into 100 parts consequently there are only 3 quartiles, 9 deciles and 99 percentiles for a series. The quartiles are denoted by symbol Q, deciles by D and percentiles by P. The subscripts 1, 2, 3, etc., beneath Q, D, and P would refer to the particular value that we want to compute. Thus Q_1 would denote first quartile Q_2 second quartile, Q_3 third quartile, D_1 first decile, D_8 eighth decile, P_1 first percentile, etc.

Graphically any set of these partition values serves to divide the area of the frequency curve or histogram into equal parts. If vertical lines are drawn at each quartile, for example, the area of the histogram will be divided by these lines into four equal parts. The 9 deciles divide the area of the histogram or frequency curve into 10 equal parts and the 99 percentiles divide the area into 100 equal parts.

In economics and business, quartiles are more widely used than deciles and percentiles. The quartiles are the points on the *X*-scale that divide the distribution into four equal parts. Obviously there are three quartiles, the second coinciding with the median. More precisely stated, the lower quartile, Q_1 is that point on the *X*-scale such that one-fourth of the total frequency is less than Q_1 and three-fourths is greater than Q_1. The upper quartile, Q_3, is that point on the X-scale such that three-fourths of the total frequency is below Q_3 and one-fourth is above it.

The deciles and percentiles are important in psychological and educational statistics concerning grades, rates, scores and ranks ; they are of use in economics and business in personnel department, productivity ratings and other situations.

Computation of Quartiles, Deciles, Percentiles, etc.

The procedure for computing quartiles, deciles, etc., is the same as for median.

For grouped data, the following formulae are used for quartiles, deciles and percentiles ;

$$Q_j = L + \frac{\frac{jN}{4} - p.c.f.}{f} \times i \qquad \text{for } j = 1, 2, 3$$

$$D_k = L + \frac{\frac{kN}{10} - p.c.f.}{f} \times i \qquad \text{for } k = 1, 2, \ldots, 9$$

$$P_1 = L + \frac{\frac{lN}{100} - p.c.f.}{f} \times i \qquad \text{for } l = 1, 2, \ldots, 99$$

where the symbols have their usual meanings and interpretation.

Illustration 10. The profits earned by 100 companies during 2009-10 are given below:

Profits (Rs. lakhs)	*No. of companies*	*Profits (Rs. lakhs)*	*No. of companies*
20 — 30	4	60 — 70	15
30 — 40	8	70 — 80	10
40 — 50	18	80 — 90	8
50 — 60	30	90 — 100	7

Calculate Q_1, median, D_4 and P_{80} and interpret the values. *(MBA, D.U., 2001)*

Solution. **CALCULATION OF Q_1, Q_2, D_4 AND P_{80}**

Profits (Rs. lakhs)	*f*	*c.f.*
20 – 30	4	4
30 – 40	8	12
40 – 50	18	30
50 – 60	30	60
60 – 70	15	75
70 – 80	10	85
80 – 90	8	93
90 – 100	7	100

Q_1 = Size of $N/4$th observation $= \frac{100}{4} = 25$th observation.

Hence Q_1 lies in the class 40 — 50.

$$Q_1 = L + \frac{N/4 - p.c.f.}{f} \times i = 40 + \frac{25-12}{18} \times 10 = 40 + 7.22 = 47.22$$

25 per cent of the companies earn an annual profit of Rs. 47.22 lakhs or less.

Median or Q_2 = Size of $\frac{2N}{4}$th observation = $\frac{200}{4}$ = 50th observation. Q_2 lies in the class 50—60.

$$Q_2 = L + \frac{2N/4 - p.c.f.}{f} \times i = 50 + \frac{50-30}{30} \times 10 = 50 + 6.67 = 56.67$$

50 per cent of the companies earn an annual profit of Rs. 56.67 lakhs or less.

D_4 = Size of $\frac{4N}{10}$th observation = 40th observation

D_4 lies in the class 50—60.

$$D_4 = L + \frac{4N/10 - p.c.f.}{f} \times i = 50 + \frac{40-30}{30} \times 10 = 50 + 3.33 = 53.33.$$

Thus 40 per cent of the companies earn an annual profit of Rs. 53.33 lakhs or less.

P_{80} = Size of $\frac{80N}{100}$th observation = $\frac{80 \times 100}{100}$ = 80th observation

P_{80} lies in the class 70—80.

$$P_{80} = L + \frac{80\,N/100 - p.c.f.}{f} \times i = 70 + \frac{80-75}{10} \times 10 = 70 + 5 = 75$$

This means that 80 per cent of the companies earn an annual profit of Rs. 75 lakhs or less and 20 per cent of the companies earn an annual profit of more than Rs. 75 lakhs.

Determination of Median, Quartiles, etc., Graphically

Median can be determined graphically by applying any of the following two methods :

1. Draw two ogives — one by 'less than' method and other by 'more than' method. From the point where both these curves intersect each other, draw a perpendicular on the X-axis. The point where this perpendicular touches the x-axis gives us the value of median.

2. Draw only one ogive by 'less than' method. Take the variable on the X-axis and frequency on the Y-axis. Determine the median value by the formula : median = Size of $\frac{N}{2}$th item. Locate this value on the Y-axis and from it draw a perpendicular on the cumulative frequency curve. From the point where it meets the ogive draw another perpendicular on the X-axis and the point where it meets the X-axis is the median.

The other partition values like quartiles, deciles and percentiles can also be determined graphically.

Illustration 11. Using the data of illustration 10, determine graphically the values of Q_1, Q_2, D_{40} and P_{80}.

Solution. Draw the ogive by the 'less than' method as shown in the graph.

To determine different quartiles, horizontal lines (broken) are drawn from the cumulative frequency values. For example, if we want to determine the value of median, a horizontal line can be drawn from the cumulative frequency value of 0.50 to the less than curve and then extending a vertical line to the horizontal axis. In a similar manner other values can be determined as shown in the graph. Therefore, $Q_1 = 47.22$, $Q_2 = 56.67$, $D_{40} = 53.33$ and $P_{80} = 75$. This may be noted down here that these graphical values are same as obtained by the formulae.

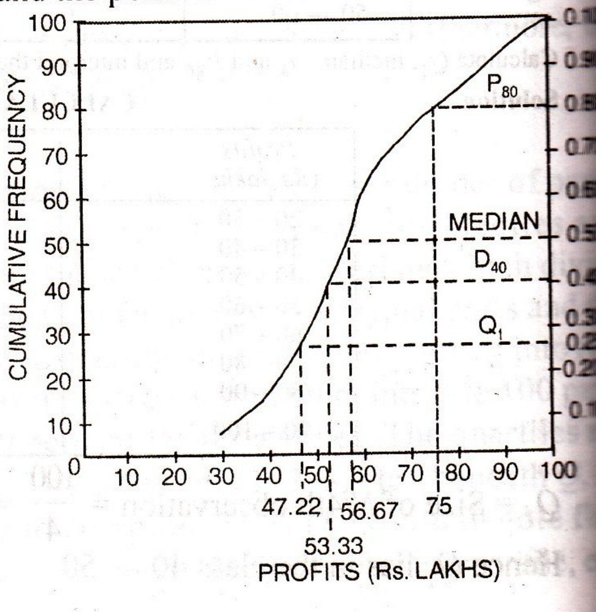


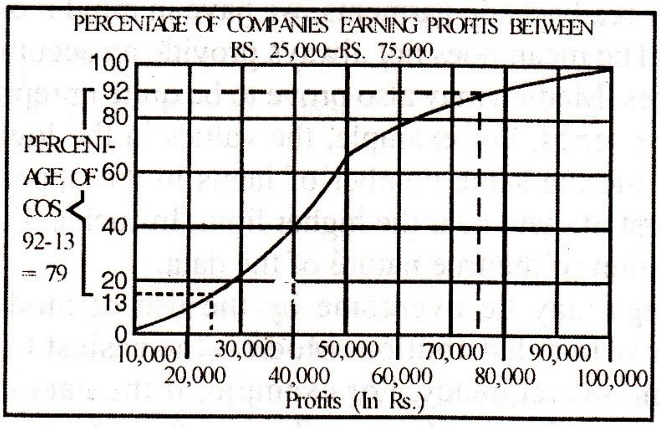
Illustration 12. You are given the net profits earned by some companies. Estimate graphically the percentage of companies getting profits between Rs. 25,000 and Rs. 75,000.

Profits (in Rs.)	*No. of companies*	*Profits (in Rs.)*	*No. of companies*
10,000—20,000	15	60,000—70,000	22
20,000—30,000	35	70,000—80,000	12
30,000—40,000	47	80,000—90,000	11
40,000—50,000	68	90,000—1,00,000	8
50,000—60,000	32		

Solution. Finding percentage from the given data :

Profits less than	*No. of companies*	*Percentage*
Rs. 20,000	15	6.0
" 30,000	50	20.0
" 40,000	97	38.8
" 50,000	165	66.0
" 60,000	197	78.8
" 70,000	219	87.6
" 80,000	231	92.4
" 90,000	242	96.8
" 1,00,000	250	100.0

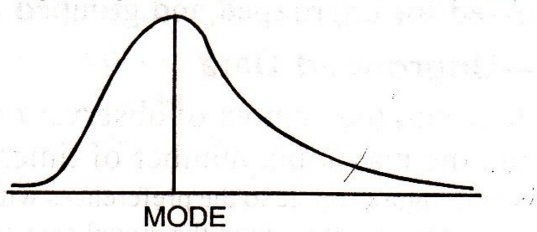
Plotting the data on the graph paper :



The graph shows clearly that the percentage of companies earning profits less than Rs. 75,000 is 92 and the percentage of companies earning profits less than Rs. 25,000 is 13. Thus the percentage of companies making profits between Rs. 25,000 and Rs. 75,000 is (92–13) = 79.

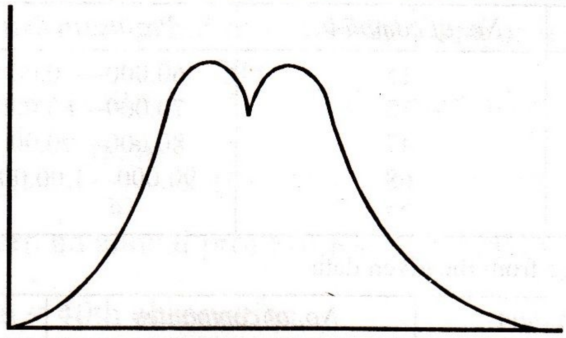
C. MODE

Mode is defined as that value which occurs the maximum number of times, *i.e.*, having the maximum frequency. For example, if we take the values of six different observations as 5, 8, 10, 8, 5, 8, mode will be 8 as it has occurred maximum number of times, *i.e.*, 3 times. Graphically, it is the value on the *X*-axis below the peak, or highest point, of the frequency curve as can be seen from the following diagram.



This interpretation of the statistical mode is analogous to that of the fashion mode. A person dressing with current styles is "in the mode". But a current fashion can be a poor description of what most persons are wearing because of the variety of styles worn by the general public. In statistics, the mode only tells us which single value occurs most often; it may, therefore, represent a majority of the total population.

It is possible that a distribution may be bimodal. This happens when there may be two or more values of equal or nearly equal occurrence as can be seen from the following diagram :



The presence of more than one mode has a special significance in statistical analysis, for it indicates potential trouble. It is usually dangerous to compare bimodal populations or to draw conclusions about them because they usually arise when there is some non-homogeneous factor present in the population.

If the collected data produce a bimodal distribution, the data themselves should be questioned. Quite often such a condition is caused by the taking of too small a sample; the difficulty can be remedied by increasing the sample size. In instances where a distribution is bimodal and nothing can be done to change it, the mode is obviously eliminated as a measure of central tendency.

There are many situations in which arithmetic mean and median fail to reveal the true characteristic of data. For example, when we talk of most common wage, most common income, most common height, most common size of shoe or ready-made garments, we have in mind mode and not the arithmetic mean or median discussed earlier. The mean does not always provide an accurate reflection of the data due to the presence of extreme values. Median may also prove to be quite unrepresentative of the data owing to an uneven distribution of the series. For example, the values in the lower half of a distribution range from, say, Rs. 10 to 100, while the same number of items in the upper half of the series range from Rs. 100 to Rs. 6,000 with most of them near the higher limit. In such a distribution the median value Rs. 100 will provide little indication of the true nature of the data.

Both these shortcomings may be overcome by the use of mode. Mode refers to that value which occurs most frequently in a distribution. Mode is the easiest to compute since it is the value corresponding to the maximum frequency. For example, if the data is :

Size of shoes	:	5	6	7	8	9	10	11
No. of persons	:	10	20	25	40	22	15	6

the modal size is '8' since it appears maximum number of times in the data.

Calculation of Mode

Determining the precise value of the mode of a frequency distribution is by no means an elementary calculation. Essentially it involves fitting mathematically some appropriate type of frequency curve to the grouped data and the determination of the value on the X-axis below the peak of the curve. However, there are several elementary methods of *estimating* the mode. These methods have been discussed for ungrouped and grouped data.

Calculation of Mode—Ungrouped Data

For determining mode count, the number of observations the various values repeat themselves, and the value which occurs the maximum number of times is the modal value.

Illustration 13. The following figures relate to the preferences with regard to size of screen (in inches) of T.V. sets of 30 persons selected at random from a locality. Find the modal size of the T.V. screen.

12	20	12	24	29
20	12	20	29	24
24	20	12	20	24
29	24	24	20	24
24	20	24	24	12
24	20	29	24	24

Solution. CALCULATION OF MODAL SIZE

Size in inches	*Tally*	*Frequency*
12	𝍸	5
20	𝍸 |||	8
24	𝍸 𝍸 |||	13
29						4
		Total 30

Since size 24 occurs the maximum number of times, therefore, the modal size of T.V. screen is 24 inches.

Calculation of Mode—Grouped Data

In the case of grouped data, the following formula is used for calculating mode :

$$Mo = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i \qquad ...(i)$$

where L = Lower limit of the modal class.

Δ_1 = The difference between the frequency of the modal class and the frequency of the pre-modal class, *i.e.*, preceding class.

Δ_2 = The difference between the frequency of the modal class and the frequency of the post-modal class, *i.e.*, succeeding class.

i = The size of the modal class.

Another form of this formula is :

$$Mo = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i \qquad ...(ii)$$

where L = Lower limit of the modal class.

f_1 = Frequency of the modal class.

f_0 = Frequency of the class preceding the modal class.

f_2 = Frequency of the class succeeding the modal class.

While applying the above formula for calculating mode, it is necessary to see that the class intervals are *uniform* throughout. If they are unequal, they should first be made equal on the assumption that the frequencies are equally distributed throughout the class, otherwise we will get misleading results.

A distribution having only one mode is called *unimodal.* If it contains more than one mode, it is called *bimodal* or *multimodal*. In the latter case, the value of mode cannot be determined by the above formula and hence mode is *ill-defined* when there is more than one value of mode.

Where mode is ill-defined, its value may be ascertained by the following approximate formula* based upon the relationship between mean, median and mode.

$$\text{Mode} = 3 \text{ Median} - 2 \text{ Mean} \qquad ...(iii)$$

Illustration 14. The following data relate to the sales of 100 companies :

Sales (Rs. lakhs)	*No. of companies*	*Sales (Rs. lakhs)*	*No. of companies*
Below 60	12	66 — 68	10
60—62	18	68—70	3
62—64	25	70—72	2
64—66	30		

Calculate the value of modal sales.

*See page 99.

Solution. Since the maximum frequency 30 is in the class 64—66, therefore, 64—66 is the modal class.

$$\text{Mode} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$

$$L = 64,\ \Delta_1 = (30 - 25) = 5,\ \Delta_2 = (30 - 10) = 20,\ i = 2$$

$$\text{Mode} = 64 + \frac{5}{5 + 20} \times 2 = 64 + \frac{10}{25} = 64.4$$

Hence modal sales are Rs. 64.4 lakhs.

Locating Mode Graphically

In a frequency distribution the value of mode can also be determined graphically. The steps in calculation are :

1. Draw a histogram of the given data.
2. Draw two lines diagonally on the inside of the modal class bar, starting from each upper corner of the bar to the upper corner of the adjacent bar.
3. Draw a perpendicular line from the intersection of the two diagonal lines to the X-axis (horizontal scale) which gives us modal value.

Illustration 15. The daily profits in rupees of 100 shops are given as follows :

Profits (in Rs. lakhs)	No. of shops	Profits (in Rs. lakhs)	No. of shops
0—100	12	300—400	20
100—200	18	400—500	17
200—300	27	500—600	6

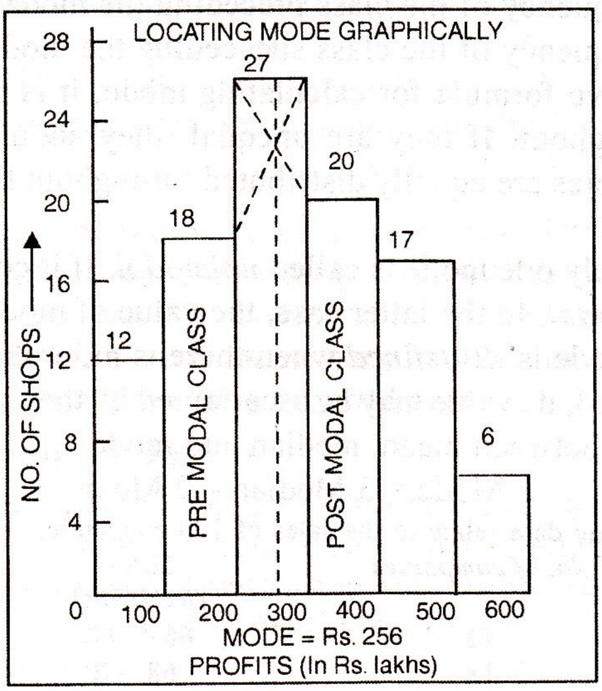
Draw the histogram and thence find the modal value. Check this value by direct calculation.

Solution.

Direct calculation :

Mode lies in the class 200—300.

$$\text{Mode} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i = 200 + \frac{9}{9 + 7} \times 100 = 256.25$$



From the above diagram, the modal value is also 256. Hence by both the methods we get the same value of mode.

Mode can also be determined from frequency polygon in which case perpendicular is drawn on the base from the apex of the polygon and the point where it meets the base gives the modal value.

However, graphic method of determining mode can be used only where there is one class containing the highest frequency. If two or more classes have the same highest frequency, mode cannot be determined graphically.

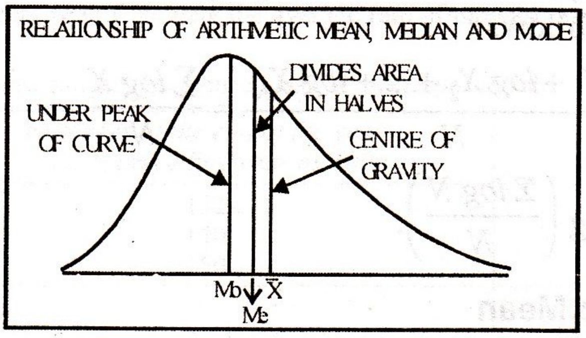
Merits and Limitations of Mode

Like median, the mode is not affected by extreme values and its value can be obtained in open-end distributions without ascertaining the class limits. Mode can be easily used to describe qualitative phenomenon. For example, when we want to compare the consumer preferences for different types of products, say, soap, toothpastes, etc., or different media of advertising, we should compare the modal preferences. In such distributions where there is an outstanding large frequency, mode happens to be meaningful as an average.

However, mode is not a rigidly defined measure as there are several formulae for calculating the mode, all of which usually give somewhat different answers. Also the value of mode cannot always be computed, such as, in case of bimodal distributions.

Relationship among Mean, Median and Mode

A distribution in which the values of mean, median and mode coincide is known as *symmetrical* distribution. Conversely stated, when the values of mean, median and mode are not equal, the distribution is known as *asymmetrical* or *skewed*. In moderately skewed or asymmetrical distributions, a very important relationship exists among mean, median and mode. In such distributions, the distance between the mean and the median is approximately one-thrid of the distance between the mean and mode as will be clear from the following diagram :



Karl Pearson has expressed this approximate relationship as follows :

$$\text{Mean} - \text{Median} = \frac{1}{3}\,(\text{Mean} - \text{Mode})$$

or

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

and

$$\text{Median} = \frac{2\ \text{Mean} + \text{Mode}}{3}$$

If we know any of the two values out of the three, we can compute the third from these relationships. The following example will illustrate this point :

Illustration 16. In a moderately asymmetrical distribution the Mode and Mean are 32.1 and 35.4 respectively. Calculate the Median.

Solution. Mode = 3 Median –2 Mean

Mode = 32.1, Mean = 35.4

Substituting the values

32.1 = 3 Median –2 (35.4) or 3 Med = 102.9 or Med. = 34.3.

2. GEOMETRIC MEAN

In business and economic problems, very often we are faced with questions pertaining to percentage rates of change over time. Neither the mean, the median nor mode is the appropriate average to use in these instances. For example, consider the following figures of sale of a company :

Year :	2007	2008	2009	2010
Sales (million tonnes) :	20.2	22.5	23.9	28.0

Suppose we want to find out the average percentage rate of change per year in sales. To answer this question we must specify what we mean by the 'Average percentage rate of change per year'. The most generally useful interpretation of this term is the constant percentage rate of change which if applied each year would take us from the first to the last figure. Hence in the above illustration we would be interested in that constant yearly percentage rate of change which would be required to move from 20.2 million tonnes of sales in 2007 to 28.0 million tonnes in 2010. None of the previously discussed averages provides the correct answer to this question. The correct answer can be obtained through the use of the geometric mean or, what amounts to the same thing, through the use of the familiar compound interest formula. In the discussion which follows, the geometric mean is defined, and the relationship between this average and compound interest calculations is indicated.

Geometric mean is defined as the *N*th root of the product of *N* observations of a given data. If there are two observations, we take the square root; if there are three observations, the cube root; and so on, symbolically.

$$G.M. = \sqrt[N]{X_1 \times X_2 \times X_3 \times ...... \times X_N}$$

where $X_1, X_2, X_3 ..., X_N$, refer to the various observations of the data.

When the number of observations is three or more the task of multiplying the number and of extracting the root becomes quite difficult. To simplify calculations logarithms are used. Geometric mean is then calculated as follows :

$$\log \text{G.M.} = \frac{\log X_1 + \log X_2 + ... + \log X_N}{N} = \frac{\Sigma \log X}{N}$$

$$\therefore \quad \text{G.M.} = \text{antilog}\left(\frac{\Sigma \log N}{N}\right).$$

Calculation of Geometric Mean

In ungrouped data, geometric mean is calculated with the help of the following formula :

$$\text{G.M.} = \text{A.L.}\left(\frac{\Sigma \log X}{N}\right)$$

In grouped data, for calculating geometric mean first we will find the midpoints and then apply the following formula :

$$\text{G.M.} = \text{A.L.}\left(\frac{\Sigma f \log X}{N}\right)$$

where X = midpoint.

Compound Interest Formula

The compound interest formula is expressed as follows :

$$P_n = P_0 (1 + r)^n$$

where P_n = amount accumulated at the end of n periods,

P_0 = Original principal,

r = Rate of interest expressed as a decimal, and

n = Number of compound periods.

It follows from the above formula that :

$$r = \sqrt[n]{\frac{P_n}{P_0}} - 1$$

If interest is compounded at different rates in each time period, and if these successive rates are denoted by $r_1, r_2, \ldots, r_n$, then the amount accumulated at the end of n periods with an original principal of P_0 is

$$P_n = P_0 (1 + r_1)(1 + r_2) \ldots\ldots (1 + r_n).$$

Applications of Geometric Mean

Geometric mean is specially useful in the following cases :

1. The geometric mean is used to find the average per cent increase in sales, production, population or other economic or business data. For example, from 2008 to 2010 prices increased by 5%, 10% and 18% respectively. The average annual increase is not 11% as given by the arithmetic average but 10.9% as obtained by the geometric mean. This average is also useful in measuring the growth of population, because population increases in geometric progression.
2. Geometric mean is theoretically considered to be the best average in the construction of index number.* It makes index numbers satisfy the time reversal test and gives equal weights to equal ratio of change.
3. It is an average which is most suitable when large weights have to be given to small values of observations and small weights to large values of observations, situations which we usually come across in social and economic fields.

The following examples illustrate the use of geometric mean.

Illustration 17. Compared to the previous year the overhead expenses went up by 32% in 2008; they increased by 4% in the next year and by 50% in the following year. Calculate the average rate of increase in the overhead expenses over the three years.

Solution. In average ratios and percentages, geometric mean is more appropriate. Applying geometric mean here :

% Rise	*Expenses at the end of the year taking preceding year as 100*	*log X*
32	132	2.1206
40	140	2.1461
50	150	2.1761
		$\Sigma \log X = 6.4428$

$$\text{G.M.} = \text{A.L.}\left(\frac{\Sigma \log X}{N}\right) = \text{A.L.}\left(\frac{6.4428}{3}\right) = \text{A.L. } 2.1476 = 140.5.$$

Average rate of increase in overhead expenses

$= 140.5 - 100 = 40.5\%.$

Illustration 18. The annual rates of growth of output of a factory in 5 years are 5.0, 7.5, 2.5, 5.0 and 10.0 respectively. What is the compound rate of growth of output per annum for the period ?

Solution.

CALCULATING COMPOUND RATE OF GROWTH

Annual rate of growth	*Output relatives at the end of the year*	*log X*
5.0	105.0	2.0212
7.5	107.5	2.0314
2.5	102.5	2.0107
5.0	105.0	2.0212
10.0	110.0	2.0414
		$\Sigma \log X = 10.1259$

$$\text{G.M.} = \text{A.L.}\left(\frac{\Sigma \log X}{N}\right) = \text{A.L.}\left(\frac{10.1259}{5}\right) = \text{A.L. } 2.0252 = 105.9.$$

The compound rate of growth of output per annum for the period is 105.9–100 = 5.9%.

*Please refer to chapter on Index Numbers.

Illustration 19. A piece of property was purchased for Rs. 2,00,000 and sold 10 years later for Rs. 23,26,000. What is the average annual rate of return on the original investment ?

Solution. 2,00,000 X^{10} = 3,26,000

$$X^{10} = \frac{3,26,000}{2,00,000} = 1.63$$

$$\text{Log } X = \frac{\text{Log } 1.63}{10} = \frac{0.2122}{10} = 0.0212$$

$$X = \text{A.L. } (0.0212) = 1.05 \text{ or } 105\%.$$

Hence the investment yielded a mean rate return of 105 – 100 = 5 per cent over the 10-year period.

Combined Geometric Mean

Just as we have talked of combined arithmetic mean, in a similar manner we can also talk of combined geometric mean. If the geometric mean of N observations is 6 and these N observations are divided into two sets first containing N_1 and second containing N_2 observations having G_1 and G_2 as the respective geometric means, then

$$\text{Log } G = \frac{N_1 \text{ Log } G_1 + N_2 \text{ Log } G_2}{N_1 + N_2}$$

Thus if the geometric mean of 5 observations is 20 and of another 10 observations is 35.28, the combined geometric mean shall be

$$\text{Log } G = \frac{5 \text{ Log } 20 + 10 \text{ Log } 35.28}{5+10} = \frac{(5 \times 1.3010) + (10 \times 1.5475)}{15}$$

$$= \frac{6.505 + 15.475}{15} = \frac{21.98}{15} = 1.465$$

$$\therefore \quad G = \text{A.L. } 1.465 = 29.17.$$

Illustration 20. 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 the 20 observations in the single group formed by pooling the three groups.

Solution.
$$\text{Log } G = \frac{N_1 \text{ Log } G_1 + N_2 \text{ Log } G_2 + N_3 \text{ Log } G_3}{N_1 + N_2 + N_3}$$

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

$$= \frac{(8 \times .9304) + (7 \times 1.0052) + (5 \times .8893)}{20}$$

$$= \frac{7.4432 + 7.0364 + 4.4465}{20} = \frac{18.9261}{20} = 0.9463$$

$$G = \text{A.L. } 0.9463 = 8.837.$$

Hence the combined geometric mean of the 20 observations taken together is 8.837.

Merits and Limitations of Geometric Mean

Geometric mean is highly useful in averaging ratios and percentages and in determining rates of increase and decrease. It is also capable of algebraic manipulation. For example, if the geometric mean of two or more series and their numbers of observations are known, a combined geometric mean can easily be calculated.

However, compared to arithmetic mean, this average is more difficult to compute and interpret. Also geometric mean cannot be computed when there are both negative and positive values in a series or more observations are having zero value.

E. HARMONIC MEAN

The harmonic mean is based on the reciprocal of the numbers averaged. It is defined as the reciprocal of the arithmetic mean of the reciprocal of the individual observation. Thus by definition

$$\text{H.M.} = \frac{N}{\left(\frac{1}{X_1}+\frac{1}{X_2}+\frac{1}{X_3}+\ldots+\frac{1}{X_N}\right)}$$

When the number of observations is large, the computation of harmonic mean in the above manner becomes tedious. To simplify calculations, we obtain reciprocals of the various observations and apply the following formulae :

For ungrouped data, $\text{H.M.} = \dfrac{N}{\Sigma\left(\frac{1}{X}\right)}$

For grouped data, $\text{H.M.} = \dfrac{N}{\Sigma\left(f \times \frac{1}{X}\right)}$ or $\dfrac{N}{\Sigma\left(\frac{f}{X}\right)}$ *

Illustration 21. (*i*) Calculate harmonic mean of numbers 10, 20 25, 40, 50. (*ii*) Calculate harmonic mean from the following frequency distribution :

X :	0—10	10—20	20—30	30—40	40—50
f :	8	15	20	4	3

Solution. (*i*)

CALCULATION OF HARMONIC MEAN

X	*1/X*
10	0.100
20	0.050
25	0.040
40	0.025
50	0.020
	$\Sigma\ 1/X = 0.235$

$$\text{H.M.} = \frac{N}{\Sigma\left(\frac{1}{X}\right)} = \frac{5}{0.235} = 21.28$$

(*ii*) **CALCULATION OF HARMONIC MEAN**

Variable	*X*	*f*	*f × 1/X*
0–10	5	8	1.600
10–20	15	15	1.000
20–30	25	20	0.800
30–40	35	4	0.114
40–50	45	3	0.067
		$N = 50$	$\Sigma\left(f \times \frac{1}{X}\right) = 3.581$

$$\text{H.M.} = \frac{N}{\Sigma\left(f \times \frac{1}{X}\right)} = \frac{50}{3.581} = 13.96.$$

*There is no need to first calculate $1/X$ and then multiply it by f. We can directly obtain f/X to simplify calculation.

Applications of Harmonic Mean

The harmonic mean is restricted in its field of applications.* It is useful for computing average rate of increase of profits or average speed at which a journey has been perform the average price at which an article has been sold. The rate usually indicates the rel between two different types of measuring units that can be expressed reciprocally. For exam if a man walked 20 km., in 5 hours, the rate of his walking speed can be expressed as follo

$$\frac{20 \text{ km.}}{5 \text{ hours.}} = 4 \text{ km. per hour}$$

where the unit of the first term is a km., and the unit of the second term is an *hour*. reciprocally,

$$\frac{5 \text{ hours}}{20 \text{ km.}} = \frac{1}{4} \text{ hour per km.}$$

where the unit of the *fitst term is an hour* and the unit of the *second term is a kilometre.*

Illustration 22. In a certain factory a unit of work is completed by A in 4 minutes, by B in 5 minutes, in 6 minutes, by D in 10 minutes and by E in 12 minutes.

(*a*) What is the average number of units of work completed per minute ?

(*b*) At this rate how many units will they complete in a six-hour day ?

Solution. (*a*) The average number of units per minute will be obtained by calculating the harmonic

CALCULATION OF HARMONIC MEAN

X	$1/X$
4	0.250
5	0.200
6	0.167
10	0.100
12	0.083
	$\Sigma\ 1/X = 0.8$

$$\text{H.M.} = \frac{N}{\Sigma\left(\frac{1}{X}\right)} = \frac{5}{0.8} = 6.25$$

Hence the average number of units completed per minute = 6.25.

$$\text{The average units per minute} = \frac{1}{6.25} = 0.16.$$

In 6 hours, *i.e.*, 360 minutes, total number of units produced will be 360 × 0.16 × 5 = 288 by all the five work combined.

Illustration 23. A toy factory has assigned a group of 4 workers to complete an order of 1,400 toys of a cert type. The productive rates of the four workers are given below :

Workers	*Productive rates*
A	4 minutes per toy
B	6 minutes per toy
C	10 minutes per toy
D	15 minutes per toy

Find the average minutes per toy by the group of workers.

Solution. If we assume that each of the four workers is assigned the same number of toys (constant value) to meet the orders, or $\frac{1,400}{4}$ = 350 toys per worker, the arithmetic mean would give the correct answer.

$$\bar{X} = \frac{4+6+10+15}{4} = \frac{35}{4} = 8\frac{3}{4} \text{ minutes per toy}$$

---

*The harmonic mean is a measure of central tendency for data expressed as rates, such as kms., per hour, tonnes per day, quantity per litre, miles per fortnight, etc.

Verification

Time required by A to complete 350 toys = 350 × 4 = 1,400 minutes
Time required by B to complete 350 toys = 350 × 6 = 2,100 minutes
Time required by C to complete 350 toys = 350 × 10 = 3,500 minutes
Time required by D to complete 350 toys = 350 × 15 = 5,250 minutes
12,250 minutes.

In 12,250 minutes 1,400 toys will be completed.

Hence, in completing one toy time taken will be

$$\frac{12{,}250}{1{,}400} = 8\frac{3}{4} \text{ minutes.}$$

However, if we assume that each worker works the same amount of time but produces different number of toys, harmonic mean would be more appropriate. This assumption is more true in practice (people working same amount of time but having different output).

$$\text{H.M.} \quad \frac{4}{\frac{1}{4}+\frac{1}{6}+\frac{1}{10}+\frac{1}{15}} = \frac{4 \times 60}{35} = 6\frac{6}{7} \text{ minutes per toy}$$

$$\text{Time required to complete 1,400 toys} = \frac{1{,}400 \times 48}{7} = 9{,}600 \text{ minutes}$$

Verification. Each worker works for $\frac{9{,}600}{4} = 2{,}400$ minutes

Toys produced by A in 2400 minutes $= \frac{2400}{4} = 600$

– " " B " " $= \frac{2400}{6} = 400$

– " " C " " $= \frac{2400}{10} = 240$

– " " D " " $= \frac{2400}{15} = 160$

Total = 1,400

Merit and Limitations of Harmonic Mean

The harmonic mean, like the arithmetic mean and geometric mean, is computed from all observations. It is useful in special cases for averaging rates.

However, harmonic mean cannot be computed when there are both positive and negative observations or one or more observations have zero value. It also gives largest weight to smallest observations and as such is not a good representation of a statistical series. In dealing with business problems, harmonic mean is rarely used.

Relationship among the Averages

In any distribution when the original observations differ in size, the values of A.M., G.M. and H.M. would also differ and will be in the following order :

$$\text{A.M.} \geq \text{G.M.} \geq \text{H.M.}$$

i.e., arithmetic mean is greater than geometric mean and geometric mean is greater than harmonic mean. The equality signs hold only if all the observations $X_1, X_2, \ldots\ldots X_N$ are identical.

Progressive Average

This average is also based upon the arithmetic mean. The important features of this average are :

(*i*) It is a cumulative average. In the calculation of this average all previous figures are added and no previous figure is left as is done in the case of moving average.* The progressive average for the first year would remain the same ; the progressive average for the second year is equal to $\frac{a+b}{2}$; for the third year $\frac{a+b+c}{3}$, for the fourth year $\frac{a+b+c+d}{4}$; and so on.

(*ii*) The average value can be obtained for all the years. The moving average, on the other hand, cannot be computed for all the years. The longer the period of moving average, the greater the number of years for which the moving average cannot be computed.

This average is generally used during the early years of the working of a business. For example the figures of sales, profits or production of each successive year may be compared with the respective figures for the entire previous period in order to find out how a business is growing.

The process of computing the progressive average shall be clear with the help of the following example :

Illustration 24. Calculate progressive average from the following data :

Year	*Sale of steel (in m. tonnes)*	*Year*	*Sale of steel (in m. tonnes)*
2004	12	2008	25
2005	14	2009	22
2006	15	2010	30
2007	18		

Solution.

CALCULATION OF PROGRESSIVE AVERAGE

Year	*Sales (in m. tonnes)*	*Progressive totals*	*Years included*	*Progressive average*
2004	12	12	1	12.00
2005	14	26	2	13.00
2006	15	41	3	13.67
2007	18	59	4	14.75
2008	25	84	5	16.80
2009	22	106	6	17.67
2010	30	136	7	19.43

The progressive average makes it clear that the above company is steadily progressing year after year.

Which Average to use ?

We have explained different methods of computing the various types of averages and also their distinctive features. At this point, the reader can question "which of these average should I use ?" Or "which of these is the best average to be used ?"

It must be clearly understood that no single average can be regarded as best for all purposes. The following two considerations should be kept in mind in the selection of an average :

1. The type of data available. Are they badly skewed (avoid the mean), gappy around the middle (avoid the median), or unequal in class-interval (avoid the mode) ?

2. The concept of the typical value required by the problem. Within the framework of descriptive statistics, the main requirement is to know what each average means and then select one that fulfils the purpose on hand. Is a composite average of all absolute or relative values

*For details please refer to chapter on Business Forecasting and Time Series Analysis.

arithmetic mean or geometric mean) ? Or, is a middle value needed (median), or the most common value (mode) ?

In some cases, it may even be advisable to work out more than one average and present them although, to be sure, this procedure creates an added burden for the reader as well as for the statistician. But the added burden is preferable to the use of a single average that may give an incomplete description. To use it alone is like looking through a keyhole : the part of the room you can see cannot give a full idea of the whole room.

Arithmetic Mean

In the following cases, arithmetic mean should not be used :

1. In highly skewed distributions.
2. In distributions with open-end intervals.
3. When the distribution is unevenly spread, concentration being small or large at irregular points.
4. When an average rate of growth or change over a period of time is required.
5. When the observations form a geometric progression, *i.e.*, 1, 2, 4, 8, 16, etc.
6. When averaging rates (*i.e.*, speed, fluctuations in the prices of articles, etc.).
7. When there are very large and very small values of observations arithmetic mean would be extremely misleading on account of undue influence of extreme values.

Leaving aside the above specific cases where either median, mode, geometric mean or harmonic mean is more appropriate, in other cases we should apply as a rule of thumb the arithmetic mean—the most popular and widely used average in practice.

Median. The median is generally the best average in open-end grouped distributions, especially where if plotted as a frequency curve one gets a J or reverse *J* curve, for example, in case of price distribution or income distribution. In such cases very high or very low values would cause the mean to be higher or lower than the most "common" values. In such instances, the median may be more representative to use in describing the mass of data.

Mode. Generally speaking, the significance of mode lies in the fact that it can be used to describe qualitative data. The mode can be used in problems involving the expression of preference where quantitative measurements are not possible. Thus the preferred type of package design among a number of alternative designs would be the modal design. If we want to compare consumer preferences for different kinds of products, or different kinds of advertising, we can compare the modal preferences expressed by different groups of people but we cannot calculate the median or mean. Mode is a particularly useful average for discrete series, *e.g.*, number of people wearing a given size of shoes, or number of children per household, etc. The mode is best suited where there is an outstandingly large frequency.

Geometric Mean. The geometric mean is typically used in averaging index numbers, rates of change, ratios and other sets of data expressed in percentage form. It is particularly important in Economics and Business Statistics in index number construction.

Harmonic Mean. Harmonic mean is useful in problems in which values of a variable are compared with a constant quantity of another variable, *i.e.*, rates, time, distance covered within certain time and quantities purchased or sold per unit, etc.

GENERAL LIMITATIONS OF AN AVERAGE

1. Since an average is a single value representing a group of values, it must be properly interpreted, otherwise, there is every possibility of jumping to wrong conclusion. This can be best illustrated with the help of a story. A person had to cross a river from one bank to another. He was not aware of the depth

of the river, so he enquired from another man who told him that the average depth of water is 160 cms. The man was 175 cms and he thought that he can very easily cross the river because all the time he would be above the water level. So he started. In the beginning the level of water was very shallow but as he reached the middle, the water was 500 cms deep and he lost his life. The man was drowned because he had a misconception that average depth means uniform depth throughout. But it is not so. An average represents a group of values and lies somewhere in between the two extreme values.

2. An average may give us a value that does not exist in the data. For example, the arithmetic mean of 100, 300, 250, 50, 100 is $\frac{800}{5} = 160$, a value that does not exist in the data.

3. At times an average may give absurd results. For example, if we are calculating average size of a family we may get a value 4.8 . But this is impossible as persons cannot be in fractions. However, we should remember that it is an average value representing the entire group of families.

4. Measures of central value fail to give us any idea about the formation of the series. Two or more series may have the same central value but may differ widely in composition. For example, observe the following two series :

Series A :	150	170	190	210	280
Series B :	300	500	20	78	102

In both series, average $\overline{X} = 200$.

5. We must remember that an average is a measure of central tendency. Hence unless the data show a clear-cut concentration of observations an average may not be meaningful at all. This evidently precludes the use of any average to typify a bimodal or a U-shaped or a J-shaped distribution.

MISCELLANEOUS ILLUSTRATIONS

Illustration 25. Calculate mean, median and mode for the following data pertaining to marks in Statistics out of 140 marks for 80 students in a class :—

Marks more than :	0	20	40	60	80	100	120
No. of Students :	80	76	50	28	18	9	3

(*MBA, K.U., 2002*)

Solution. **CALCULATION OF MEAN, MEDIAN AND MODE**

Marks	*No. of Students* f	*m.p.* X	*(X-70)/20* d	*fd*	*c.f.*
0—20	4	10	– 3	– 12	4
20—40	26	30	– 2	– 52	30
40—60	22	50	– 1	– 22	52
60—80	10	70	0	0	62
80—100	9	90	+ 1	+ 9	71
100—120	6	110	+ 2	+ 12	77
120—140	3	130	+ 3	+ 9	80
	$N = 80$			$\Sigma fd = -56$	

Mean : $\overline{X} = A + \frac{\Sigma fd}{N} \times i = 70 - \frac{56}{80} \times 20 = 70 - 14 = 56$

Median : Med = Size of $\frac{N}{2}$ th observation = $\frac{80}{2}$ = 40th observation Median lies in the class 40–60.

Median lies in the class 40—60

$$\text{Med} = L + \frac{N/2 - p.c.f}{f} \times i = 40 + \frac{40 - 30}{22} \times 20 = 40 + 9.09 = 49.09$$

Mode : Since the highest frequency is 26, mode lies in the class 20–40.

$$Mo = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i = 20 + \frac{22}{22+4} \times 20 = 20 + 16.92 = 36.92$$

Illustration 26. The following table gives the distribution of weekly wages of 600 workers of a factory :

Weekly wages (in Rs.)	*Frequency*	*Weekly wages (in Rs.)*	*Frequency*
Below 375	69	600—675	58
375—450	167	675—750	24
450—525	207	750—825	10
525—600	65		

(*a*) Draw an ogive for the above data and thence obtain the median value. Check it against calculated value.
(*b*) Obtain the limits of weekly wages of central 50 per cent of the workers. (*MBA, Delhi Univ., 1996*)

Solution.

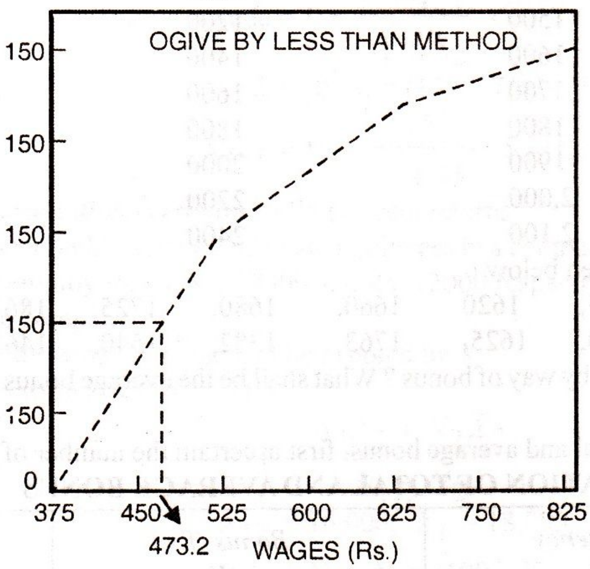
CALCULATION OF MEDIAN

Weekly wages (in Rs.)	*f*	*Cum freq.*
Less than 375	69	69
" " 450	167	236
" " 525	207	443
" " 600	65	508
" " 675	58	566
" " 750	24	590
" " 825	10	600

Med. = Size of $\frac{N}{2}$th observation = $\frac{600}{2}$ = 300th observation

Median lies in the class 450 – 525.

$$\text{Med} = L + \frac{N/2 - p.c.f.}{f} \times i = 450 + \frac{300 - 236}{207} \times 75 = 450 + 23.2 = \text{Rs. } 473.2$$



The median value as shown in the graph above is also Rs. 473.2.

(*b*) The limits of weekly wages of central 50 per cent of the workers shall be given by Q_1 and Q_3.

Q_1 = Size of $\frac{N}{4}$th observation = $\frac{600}{4}$ = 150th observation

Q_1 lies in the class 375—450.

$$Q_1 = L + \frac{N/4 - p.c.f.}{f} \times i$$

$$= 375 + \frac{150 - 69}{167} \times 75 = 375 + 36.38 = \text{Rs. } 411.38$$

Q_3 = Size of $\frac{3N}{4}$ th observation = $\frac{3 \times 600}{4}$ = 450th observation

Q_3 lies in the class 525–600.

$$Q_3 = L + \frac{3N/4 - p.c.f.}{f} \times i$$

$$= 525 + \frac{450 - 443}{65} \times 75 = 525 + 8.08 = \text{Rs. } 533.08.$$

Hence the limits of weekly wages of central 50 per cent workers are Rs. 411.38 and Rs. 533.08.

Illustration 27. In a factory, there are 100 skilled, 250 semi-skilled and 150 unskilled workers. It has been observed that on an average a unit length of a particular fabric is woven by a skilled worker in 3 hours, by semi-skilled workers in 4 hours and by an unskilled worker in 5 hours. After a training of 2 years, the semi-skilled workers are expected to become skilled and the unskilled workers to become semi-skilled. How much less time will be required after 2 years of training for weaving unit length of fabric by an average worker ?

Solution. Average time per worker before training is.

$$= \frac{(100 \times 3) + (250 \times 4) + (150 \times 5)}{100 + 250 + 150} = \frac{2050}{500} = 4.1 \text{ hours.}$$

Now after training the composition of workers is as follows :

Skilled-workers = 100 + 250 = 350

Semi-skilled workers = 150

Unskilled workers = Nil

Average time per worker after training is :

$$= \frac{(350 \times 3) + (150 \times 4)}{350 + 150} = \frac{1050 + 600}{500} = 3.3 \text{ hours.}$$

After 2 years 0.8 hour less would be required.

Note. An assumption has been made that there has been no turnover of workers.

Illustration 28. A Limited Company wants to pay bonus to workers. The bonus is to be paid as under :

Weekly wages (Rs.)	*Bonus (Rs.)*
1300 but not exceeding 1400	1000
1400 " " " 1500	1200
1500 " " " 1600	1400
1600 " " " 1700	1600
1700 " " " 1800	1800
1800 " " " 1900	2000
1900 " " " 2,000	2200
2,000 " " " 2,100	2400

Actual wages drawn by the workers is given below :

1325, 1378, 1420, 1620, 1455, 1620 1660, 1680, 1725, 1863, 1832, 1942, 1952, 1800, 2002, 2,028 2,100, 1610, 1625, 1763, 1382, 1540, 1463, 1578, 1723,

How much the company would need to pay by way of bonus ? What shall be the average bonus paid per member of the staff ?

(*MBA., Jodhpur Univ., 2005*)

Solution. For determining the figure of total and average bonus, first ascertain the number of persons in each wages group.

CALCULATION OF TOTAL AND AVERAGE BONUS

Weekly wages (Rs.)	*Frequency* f	*Bonus (Rs.)* X	*Total Bonus* fX
1300–1400	3	1000	3000
1400–1500	3	1200	3600
1500–1600	2	1400	2800
1600–1700	6	1600	9600
1700–1800	3	1800	5400
1800–1900	3	2000	6000
1900–2,000	2	2200	4400
2,000 and above	3	2400	7200
	$N = 25$		$\Sigma fX = 42,000$

The company would need Rs. 42,000 to pay bonus.

$$\text{Average bonus per workers} = \frac{42,000}{25} = \text{Rs. } 1680.$$

Illustration 29. In 500 small-scale industrial units, the return on investment ranged from 0 to 30 per cent; no units sustaining any loss. 5 per cent of the units had returns ranging from 0 per cent to (and including) 5 per cent, and 15 per cent of the units earned returns exceeding 5 per cent but not exceeding 10 per cent. The median rate of return was 15 per cent and the upper quartile 20 per cent. The uppermost layer of returns exceeding 25 per cent was earned by 50 units.

(*i*) Present the information in the form of a frequency table as follows :

Exceeding 0 per cent but not exceeding 5 per cent
" 5 " " " " 10 "
" 10 " " " " 15 "
and so on.

(*ii*) Find the rate of return around which there is maximum concentration of the units.

Solution. (*i*)

FREQUENCY TABLE

Rate of Return	*Industrial Units*
Exceeding 0 but not exceeding 5 per cent	$500 \times \frac{5}{100} = 25$
" 5 " " 10 "	$500 \times \frac{15}{100} = 75$
" 10 " " 15 "	$250 - 100 = 150$
" 15 " " 20 "	$375 - 250 = 125$
" 20 " " 25 "	$500 - 375 - 50 = 75$
" 25 " " 30 "	50
	Total = 500

(*ii*) For finding out the rate of return around which there is maximum concentration of the units, we will calculate mode. Mode lies in the class 10–15.

$$\text{Mode} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$

$$L = 10;\ \Delta_1 = (150 - 75) = 75;\ \Delta_2 = (150 - 125) = 25;\ i = 5$$

$$\text{Mode} = 10 + \frac{75}{75 + 25} \times 5 = 13.75.$$

Hence the maximum concentration is around 13.75 per cent returns.

Illustration 30. The mean monthly salary paid to all employees in a company is Rs. 16000. The mean monthly salaries paid to technical and non-technical employees are Rs. 18000 and Rs. 12000 respectively. Determine the percentage of technical and non-technical employees of the company.

Solution. Let percentage of technical personnel be denoted by X.

Non-technical employees would be $(100 - X)$%.

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

$$\bar{X}_{12} = 16,000,\ \bar{X}_1 = 18,000,\ \bar{X}_2 = 12,000$$

$$\text{Let } N_1 = X \therefore N_2 = (100 - X)$$

Substituting the values

$$16,000 = \frac{18000X + (100 - X)12000}{100}$$

$$1600000 = 18,000X + 1200000 - 12000$$

or

$$6000X = 400000$$

$$\bar{X} = \frac{400000}{6000} = 66.67.$$

Hence percentage of technical employees = 66.67 and percentage of non-technical employees = 100 – 66.67 = 33.33.

Illustration 31. A company invests Rs. 1 lakh at 10% annual rate of interest. What will be the total amount after 6 years if the principal is not withdrawn ?

Solution. Applying the compound interest formula :

$$P_n = P_0\left(1+\frac{r}{100}\right)^n$$

$$P_0 = 1{,}00{,}000, \quad r = 10, \quad n = 6$$

$$P_n = 1{,}00{,}000\left(1+\frac{10}{100}\right)^6.$$

Taking logarithms,

$$\log P_n = \log 1{,}00{,}000 + 6 \log\left(1+\frac{10}{100}\right)$$

$$= \log 1{,}00{,}000 + 6 \log 110 - 6 \log 100$$

$$= 5 + (6 \times 2.0414) - (6 \times 2) = 5.2484$$

$$P_n = AL\ 5.2484 = 1{,}77{,}200.$$

Thus the total amount after 6 years would be Rs. 1,77,200.

Illustration 32. In a certain factory a unit of work is completed by A in 10 minutes, by B in 15 minutes, by C in 12 minutes and by D in 20 minutes.

(*i*) What is the average number of units of work completed per minute ?

(*ii*) At this rate how many units will they complete in an 8-hour day ?

Solution. The average number of units of work per minute will be obtained by finding out the harmonic mean

$$H.M. = \frac{4}{\frac{1}{10}+\frac{1}{15}+\frac{1}{12}+\frac{1}{20}} = \frac{4 \times 120}{36} = \frac{40}{3}$$

Hence $\frac{40}{3}$ units per minute is the average rate.

The average units per minute is $\dfrac{1}{\frac{40}{3}} = \dfrac{3}{40}$

(*b*) In 8 hours, *i.e.*, 480 minutes, the total number of units produced will be $480 \times \frac{3}{40} \times 4 = 144$ by all the four workers combined.

Illustration 33. A machine was purchased for Rs. 10 lakh in 2006. Depreciation on the diminishing balance was charged @ 40% in the first year, 25% in the second year and 10% per annum during the next three years. What is the average depreciation charge during the whole period?

Solution. Since we are interested in finding out the average rate of depreciation, geometric mean will be the most appropriate average. The cost of machine can be ignored as it is immaterial in the rate calculation.

DETERMINING AVERAGE RATE OF DEPRECIATION

Year	*Diminishing value (for a value of Rs. 100)* X	*Log X*
2006	100 – 40 = 60	1.7782
2007	100 – 25 = 75	1.8751
2008	100 – 10 = 90	1.9542
2009	100 – 10 = 90	1.9542
2010	100 – 10 = 90	1.9542
		$\Sigma \log X = 9.5159$

$$\text{G.M.} = \text{AL}\left(\frac{\Sigma \log X}{N}\right) = \text{AL}\left(\frac{9.5159}{5}\right) = \text{AL } 1.9032 = 80.$$

The diminishing value being Rs. 80, the depreciation will be 100 – 80 = 20%.

Illustration 34. The following is the age distribution of 1,000 person working in a large industrial house :

Age group	*No. of persons*	*Age group*	*No. of persons*
20–25	30	45–50	105
25–30	160	50–55	70
30–35	210	55–60	60
35–40	180	60–65	40
40–45	145		

Due to continuous losses, it is desired to bring down the strength to 30% of the present number according to the following scheme :

(*i*) To retrench the first 15% from the lower group.

(*ii*) To absorb the next 45% in other branches.

(*iii*) To make 10% from the highest age group retire permanently, if necessary.

Calculate the age limits of the persons retained and those to be transferred to other departments. Also find the average age of those retained.

(*MBA,DU, 2005*)

Solution. The first 15% to be retrenched are from the lower group ; hence their total number comes to $\frac{15 \times 1,000}{100} = 150$. 30 belong to 20 – 25 age group and rest, *i.e.*, (150–30), *i.e.*, 120 belong to 25–30 age group. The next 45% are to be absorbed in other branches. They are $1000 \times \frac{45}{100} = 450$. They belong to the following age groups :

Age groups	*No. of persons*
25–30	40
30–35	210
35–40	180
40–45	20

Those to retire are 10% and belong to the highest age group. Their number comes to $1,000 \times \frac{10}{100} = 100$ and their age groups are :

Age groups	*No. of persons*
60–65	40
55–60	60

AVERAGE AGE OF THOSE RETAINED

Age groups	*m.p.* X	*No. of persons* f	$(X - 47.5)/5$ d	fd
40–45	42.5	125	–1	–125
45–50	47.5	105	0	0
50–55	52.5	70	+1	+70
		$N = 300$		$\Sigma fd = -55$

$$\bar{X} = A + \frac{\Sigma fd}{N} \times i = 47.5 - \frac{55}{300} \times 5 = 47.5 - 0.92 = 46.58.$$

Hence the average age of those retained is 47 years approximately.

Illustration 35. The median and mode of the following wage distribution are Rs. 33.5 and Rs. 34 respectively. However, three frequencies are missing. Determine their values.

Wages (in hundred Rs.)	0–10	10–20	20–30	30–40	40–50	50–60	60–70	Total
Frequencies	4	16	?	?	?	6	4	230

Solution. Let the missing frequencies be f_0, f_1 and f_2 corresponding to classes 20–30, 30–40 and 40–50 respectively. Since median and mode are 33.5 and 34, they lie in the class 30–40. The frequency of this class is f_1.

DETERMINING MISSING VALUES

Wages (in hundred Rs.)	*Frequency*	*Cum. frequency*
0–10	4	4
10–20	16	20
20–30	?	$20 + f_0$
30–40	?	$20 + f_0 + f_1$
40–50	?	$20 + f_0 + f_1 + f_2$
50–60	6	226
60–70	4	230
	$N = 230$	

$$f_0 + f_1 + f_2 = 230 - (4 + 16 + 6 + 4) = 200$$

$$f_2 = 200 - (f_0 + f_1) = 200 - f_0 - f_1$$

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

$$34 = 30 + \frac{f_1 - f_0}{2f_1 - f_0 - (200 - f_0 - f_1)} \times 10$$

$$\frac{4}{10} = \frac{f_1 - f_0}{2f_1 - f_0 - 200 + f_0 + f_1}$$

$$10f_1 - 10f_0 = 4(3f_1 - 200) \text{ or } 10f_1 - 10f_0 = 12f_1 - 800$$

$$-2f_1 - 10f_0 = -800 \text{ or } f_1 + 5f_0 = 400. \qquad \text{...(i)}$$

$$\text{Med.} = L + \frac{N/2 - \text{p.c.f.}}{f} \times i$$

$$33.5 = 30 + \frac{115 - 20 - f_0}{f_1} \times 10$$

$$3.5f_1 = 950 - 10f_0$$

$$7f_1 = 1{,}900 - 20f_0$$

$$7f_1 + 20f_0 = 1{,}900 \qquad \text{...(ii)}$$

From Eqns. (*i*) and (*ii*),

$$f_1 + 5f_0 = 400$$

$$7f_1 + 20f_0 = 1{,}900$$

Multiplying Eqn. (*i*) by 4,

$$4f_1 + 20f_0 = 1{,}600$$

$$7f_1 + 20f_0 = 1{,}900$$

$$-3f_1 = -300 \text{ or } f_1 = 100$$

Substituting the value of f_1 in Eqn. (*i*),

$$100 + 5f_0 = 400$$

$$5f_0 = 300 \text{ or } f_0 = 60$$

Since

$$f_0 + f_1 + f_2 = 200$$

$$f_2 = 200 - 100 - 60 = 40$$

Hence the missing frequencies are

$$f_0 = 60, f_1 = 100, f_2 = 40.$$

Illustration 36. Calculate the arithmetic mean and the median of the frequency distribution given below. Hence calculate the mode using the empirical relation between the three :

Height (in cms.)	*No. of students*	*Height (in cms.)*	*No. of students*
130–134	5	150–154	17
135–139	15	155–159	10
140–144	28	160–164	1
145–149	24		

Solution. **CALCULATION OF ARITHMETIC MEAN AND MEDIAN**

Height (in cms.)	*m.p.* X	*f*	*(X–147)/5* d	*fd*	*c.f.*
129.5–134.5	132	5	–3	–15	5
134.5–139.5	137	15	–2	–30	20
139.5–144.5	142	28	–1	–28	48
144.5–149.5	147	24	0	0	72
149.5–154.5	152	17	+1	+17	89
154.5–159.5	157	10	+2	+20	99
159.5–164.5	162	1	+3	+3	100
		$N = 100$		$\Sigma fd = -33$	

Mean : Mean $\bar{X} = A + \frac{\Sigma fd}{N} \times i = 147 - \frac{33}{100} \times 5 = 147 - 1.65 = 145.35$

Median : Median = Size of $\frac{N}{2} = \frac{100}{2} = 50$th observation.

Median lies in the class 144.5 – 149.5.

$$\text{Median} = L + \frac{N/2 - p.c.f.}{f} \times i = 144.5 + \frac{50 - 48}{24} \times 5 = 144.5 + .417 = 144.917.$$

Mode : Mode = 3 Median – 2 Mean = (3 × 144.917) – (2 × 145.35) = 144.051.

Illustration 37. Income of employees in an industrial concern are given below. The total income of the 10 employees in the class over Rs. 25000 is Rs. 3,00,000. Compute the mean income. Every employee belonging to the top 25% of the earners is required to pay 5% of his income to workers' relief fund. Estimate the contribution to this fund.

Income (Rs.)	*Employers*	*Income (Rs.)*	*Employers*
Below 5000*	90	15000–20000	80
5000–10000	150	20000–25000	70
10000–15000	100	25000 and over	10

Solution. **COMPUTATION OF MEAN**

Income (Rs.)	*m.p.(X)*	*f*	*fX*
0–5000	2500	90	2,25,000
5000–10,000	7500	150	11,25,000
10,000–15,000	12500	100	12,50,000
15,000–20,000	17500	80	14,00,000
20,000–25,000	22500	70	15,75,000
25,000 and over	30000 (given)	10	3,00,000
		$N = 500$	$\Sigma fX = 58,75,000$

$$\bar{X} = \frac{\Sigma fX}{N} = \frac{587500}{500} = \text{Rs. } 11750.$$

Number of employees belonging to the top 25% of the earners is $\frac{25}{100} \times 500 = 125$ and the distribution of these top earners as obvious from the above table is as follows :

Distribution of top 25% earners

* Since class intervals are equal, below 5000 should mean 0–5000

Income (Rs.)	Frequency
25000 and over	10
20000–25000	70
15000–20000	45

45 persons are to be taken in the last class, *i.e.*, 15000–20000 with the highest income level starting from 20000 and below. Under the assumption that the frequencies are equally distributed throughout the class, the calculation would be as follows :

80 persons have income in the range 15000–20000 = Rs. 5000

$\therefore$ 45 persons have income in the range = $\frac{5000}{80} \times 45 = 2812.5$ or 2812

Since we are interested in the top 45 earners in the income group 15000–20000, their salaries will range from (20000–2812) to 20000, *i.e.*, 17188 to 20000.

The distribution of top 125 persons is as follows :

Income (Rs.)	m.p. (X)	f	Total Income fX
25000 and over	—	10	300000 (given)
20000–25000	22500	70	1575000
17188–20000	18594	45	836730
		$N = 125$	$\Sigma fX = 2711730$

Hence the total income of the top 25% of earners is Rs. 27,11,730

5% Contribution to the fund = 0.05 × 2711730 = Rs. 135586.5

Illustration 38. Calculate the median and mode for the distribution of the weights of 150 students from the data given below :

Weight (in kg) :	30–40	40–50	50–60	60–70	70–80	80–90
Frequency :	18	37	45	27	15	8

Solution.

CALCULATION OF MEDIAN AND MODE

Weight (kg)	f	c.f.
30–40	18	18
40–50	37	55
50–60	45	100
60–70	27	127
70–80	15	142
80–90	8	150
	$N = 150$	

Median = Size of $\frac{N}{2}$ th observation = $\frac{150}{2}$ = 75th observation

Median lies in the class 50–60.

Median $= L + \frac{N/2 - p.c.f.}{f} \times i = 50 + \frac{75-55}{45} \times 10 = 50 + 4.444 = 54.44$

Mode : Since highest frequency is 45, mode lies in the class 50–60.

Mode $= L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i = 50 + \frac{8}{8+18} \times 10 = 50 + 3.08 = 53.08.$

Illustration 39. Following distribution gives the pattern of overtime work per week done by 100 employees of a company. Calculate median, first quartile, and 7th decile.

Overtime hours :	10–15	15–20	20–25	25–30	30–35	35–40
No. of employees :	11	20	35	20	8	6

(*MBA, Kurukshetra Univ., 200*

Solution.

CALCULATION OF MEDIAN, Q_1 AND D_7

Overtime (hrs.)	f	c.f.
10–15	11	11
15–20	20	31
20–25	35	66
25–30	20	86
30–35	8	94
35–40	6	100
	$N = 100$	

Med. = Size of $\frac{N}{2}$ th observation = $\frac{100}{2}$ = 50th observation

Median lies in the class 20–25.

$$\text{Med.} = L + \frac{N/2 - p.c.f.}{f} \times i = 20 + \frac{50-31}{35} \times 5 = 20 + 2.714 = 22.714$$

Q_1 = size of $\frac{N}{4}$ th observation = $\frac{100}{4}$ = 25th observation

Q_1 lies in the class 15–20.

$$Q_1 = L + \frac{N/4 - p.c.f.}{f} \times i = 15 + \frac{25-11}{20} \times 5 = 15 + 3.5 = 18.5$$

D_7 = Size of $\frac{7N}{10}$ th observation = $\frac{7 \times 100}{10}$ = 70th observation

D_7 lies in the class 25–30.

$$D_7 = L + \frac{7N/10 - p.c.f.}{f} \times 5 = 25 + \frac{70-66}{20} \times 5 = 25 + 1 = 26.$$

Illustration 40. In an examination of 675 candidates the examiner supplied the following information :

Marks obtained	*No. of candidates*	*Marks obtained*	*No. of candidates*
Less than 10%	7	Less than 50%	381
Less than 20%	39	Less than 60%	545
Less than 30%	95	Less than 70%	631
Less than 40%	201	Less than 80%	675

Calculate the mode and median of the percentage marks obtained. *(MBA, Rohilkhand Univ., 2002)*

Solution. **CALCULATION OF MEDIAN AND MODE**

Marks (in %)	*No. of candidates (f)*	*m.p. Xs*	*(X–45) / 10 d*	*fd*	*c.f.*
0–10	7	5	–4	–28	7
10–20	32	15	–3	–96	39
20–30	56	25	–2	–112	95
30–40	106	35	–1	–106	201
40–50	180	45	0	0	381
50–60	164	55	+1	+164	545
60–70	86	65	+2	+172	631
70–80	44	75	+3	+132	675

Median : Med. = Size of $\frac{N}{2}$ th observation = $\frac{675}{2}$ = 337.5th observation.

Median lies in the class 40–50.

$$\text{Med.} = L + \frac{N/2 - p.\,c.\,f.}{f} \times i$$

$L = 40$, $N/2 = 337.5$, $p.c.f. = 201$, $f = 180$, $i = 10$

$$\text{Med.} = 40 + \frac{337.5-201}{180} \times 10$$

$$= 40 + 7.58 = 47.58$$

Mode : By inspection mode lies in the class 40–50.

$$\text{Mode} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$

$$L = 40, \Delta_1 = 180 - 106 = 74, \Delta_2 = 180 - 164 = 16, i = 10.$$

$$\text{Mode} = 40 + \frac{74}{74+16} \times 10 = 40 + 8.22 = 48.22.$$

Illustration 41. The following is the average amount of dollars each major airline spends per passenger on food

American	7.41	Continental	2.77
United	7.24	US Air	2.68
Northwest	5.15	American west	2.00
TWA	5.09	Southwest	0.14
Delta	4.61		

What are the mean and median cost per passenger? Which would be the better figure to use for a new airline in developing its business plan? *(MBA, Delhi Univ., 2003)*

Solution. Calculation of $\bar{X}$ and median.

$$\bar{X} = \frac{7.41 + 7.24 + 5.15 + 5.09 + 4.61 + 2.77 + 2.68 + 2.00 + 0.14}{9} = \frac{37.09}{9}$$

$= \$ 4.12.$

Median. Arranging the given data in ascending order.

0.14 2.00 2.68 2.77 4.61 5.09 5.15 7.24 7.41

$$\text{Med} = \text{Size of } \frac{N+1}{2} \text{ th observation}$$

$$= \frac{9+1}{2} = 5\text{th observation}$$

Size of 5th observation is 4.61. Hence median = **$ 4.61.**

Median would be a better choice for new airlines in developing its business plan as median is not affected by extreme observations.

Illustration 42. Consider the following distribution :

X	0–10	10–20	20–30	30–40	40–50
f	12	18	20	25	23

Compute mean median and mode. *(MBA, G.G.S.I.P. Univ., 2009)*

Solution. **CALCULATION OF MEAN, MEDIAN AND MODE**

X	f	m.p. X	$(X–25) / 10$ d	fd	c.f.
0–10	12	5	–2	–24	12
10–20	18	15	–1	–18	30
20–30	20	25	0	0	50
30–40	25	35	+1	+25	75
40–50	23	45	+2	+46	98
	$N = 98$			$\Sigma fd = 29$	

Mean : Mean $\bar{X} = A + \frac{\Sigma fd}{N} \times i = 25 + \frac{29}{98} \times 10 = 25 + 2.96 = 27.96$

Median : Med. = Size of $\frac{N}{2}$ th item $= \frac{98}{2} = 49$th item

Median lies in the class 20–30.

$$\text{Med.} = L + \frac{N/2 - p.c.f.}{f} \times i$$

$$L = 20, N/2 = 49, p.c.f. = 30, f = 20, i = 10$$

$$= 20 + \frac{49 - 30}{20} \times 10 = 20 + 9.5 = 29.5$$

Mode : Mode lies in the class 30–40.

$$\text{Mode} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$

$$L = 30,\ \Delta_1 = (25 - 20) = 5,\ \Delta_2 = (25 - 23) = 2,\ i = 10.$$

$$\text{Mode} = 30 + \frac{5}{5+2} \times 10 = 30 + \frac{50}{7} = 30 + 7.14 = 37.14.$$

PROBLEMS

1. Answer following questions, each question carries **one** mark.
   (*i*) What is arithmetic mean? (*MBA, Madurai Kamaraj Univ., 2001*)
   (*ii*) What is meant by mode?
   (*iii*) What is the empirical formula linking mean, median and mode? (*MBA, Madurai Kamaraj Univ., 2004*)
   (*iv*) Give formula for geometric mean and harmonic mean in case of a continuous frequency distribution.
   (*v*) When mode is ill-defined?
   (*vi*) What are quartiles and percentiles?
   (*vii*) Is sum of deviations from arithmetic mean always zero?
   (*viii*) What is weighted mean?
   (*ix*) Is harmonic mean reciprocal of arithmetic mean?
   (*x*) Why arithmetic mean is most affected by extreme observations?
   (*xi*) When is mode useful over other averages?
2. Answer the following questions, each question carries **four** marks.
   (*i*) Define average. Write down the properties of an average. (*MA Eco., M.K. Univ., 2003*)
   (*ii*) What are the uses of geometric mean and harmonic mean?
   (*iii*) What is combined mean? Explain with the help of an example.
   (*iv*) Distinguish between median, quartiles, deciles and percentiles.
   (*v*) What is the arithmetic means of first n natural number, 1, 2 .... n ?
3. What are the various measures of central tendency? Why are they called measures of central tendency? (*MBA, UP. Tech. Univ., 2004*)
4. Give a brief description of different measures of central tendency. Why is arithmetic mean so popular?
5. (*a*) Is it necessarily true that being above average indicates that someone is superior? Explain.
   (*b*) What are quartiles of a distribution? Explain their uses.
6. (*a*) Define arithmetic mean and median and discuss their merits and demerits as measures of central tendency.
   (*b*) How would you account for the predominant choice of arithmetic mean as a measure of central tendency? Under what circumstances would it be appropriate to use mode, median, geometric mean and harmonic mean. (*MBA, KU, 2002; MBA, Delhi Univ., 2006*)
7. Comment on the following :
   (*i*) If first and third quartiles are 20 and 40 respectively the median will be 30.
   (*ii*) If daily wages paid to men and women employed in a factory are Rs. 100 and Rs. 90, the average wage per worker would be Rs. 90.
   (*iii*) A man claims that his average bank balance during the year is Rs. 3700. The bank, on the other hand, claims that he overdrew his account at least 10 times during the year and as such his claim is false.
   (*iv*) The increase in the price of commodity x is 20%. Then the price decreased 25% and again increased 15%. The resultant increase in the price is 10%.
   (*v*) The mode of a distribution cannot be less than the arithmetic mean.
   (*vi*) If Q_1, Q_2, Q_3, be respectively the lower quartile, the median and the upper quartile of a distribution, then $Q_2 - Q_1 = Q_3 - Q_2$.
   (*vii*) Arithmetic mean is the best measure of central tendency.
8. What is a statistical average? What are the desirable properties for an average to possess? Mention different types of averages and state why the arithmetic mean is the most commonly used amongst them.
9. (*a*) What are the essential requisites of a good measure of central tendency? Compare and contrast the commonly employed measure in terms of these requisites.
   (*b*) Prove that the arithmetic mean of two positive numbers a and b is at least as large as their geometric mean.
10. (*a*) What are the properties of a good average?
   (*b*) In each of the following cases, explain whether the description applies to mean, median or both :
   (*i*) Can be calculated from a frequency distribution with open-end classes?
   (*ii*) The values of all observations are taken into consideration in the calculation.
   (*iii*) The values of extreme observations do not influence the average.
11. (*a*) Under what circumstances would it be appropriate to use Arithmetic mean, Median or Mode? Discuss.

(b) Explain the properties of a good average. In the light of these properties which average do you think is the best and why? *(MBA, Jodhpur Univ., [illegible])*

11. (a) Give a brief note of the measures of central tendency together with their merits and demerits. Which is the best measure of central tendency and why? *(MBA, Osmania Univ., [illegible])*

(b) "Every average has its own peculiar characteristics. It is difficult to say which average is the best." Comment briefly. *(MBA, HPU, [illegible])*

12. For the following frequency table calculate mean, median and mode :

Weekly rent (in Rs.)	*No. of persons paying the rent*	*Weekly rent (in Rs.)*	*No. of persons paying the rent*
200–400	6	1,200–1,400	15
400–600	9	1,400–1,600	10
600–800	11	1,600–1,800	8
800–1,000	14	1,800–2,000	7
1,000–1,200	20		

[Mean = Med. = 1,100; Mode = 1,109.41]

13. Calculate the simple and weighted arithmetic mean price per bag of 20 kg. of coal purchased by an industry for the half year. Account for difference between the two.

Month	*Price per bag (Rs.)*	*Bag purchased*	*Month*	*Price per bag (Rs.)*	*Bag purchased*
Jan.	42.05	25	April	52.00	52
Feb.	51.25	30	May	44.25	10
Mar.	50.00	40	June	54.00	45

[$\overline{X} = 48.93$; $\overline{X}_w = 50.31$]

14. (a) Explain clearly the concepts of Geometric mean and Harmonic mean. Point out some of the business applications of these concepts.

(b) Calculate the geometric mean of the following price relatives :

Commodity	*Price Relative*	*Commodity*	*Price Relative*
Wheat	237	Sugar	124
Rice	198	Salt	107
Pulses	156	Oils	196

[G.M. = 163.4]

15. The following table gives the distribution of 100 accidents during seven days of the week of a given month. During the particular month there were 5 Mondays, Tuesdays and Wednesdays, and only four each of the other days. Calculate the number of accidents per day.

Days	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
No. of accidents	26	16	12	10	8	10	18

[14.13]

16. At harvesting time a farmer employed 10 men, 24 women and 16 boys to lift potatoes. A woman's work was three quarters as effective as that of a man, while a boy's work was only half. Find the daily wage bill if a man rate was Rs. 80 a day and the rates for the women and boys in proportion to their effectiveness. Calculate the average daily rate for all the workers.

17. The following table gives the daily wages (in rupees) in a certain commercial organisation :

Daily wages (Rs.)	*30–32*	*32–34*	*34–36*	*36–38*	*38–40*
No. of Workers	2	9	25	30	49
Daily wages (Rs.)	40–42	42–44	44–46	46–48	48–50
No. of Workers	62	39	20	11	3

Calculate from the above data :

(i) the median and the third quartile wages; and

(ii) the number of wage-earners receiving between Rs. 37 and Rs. 45.

[(i) Med. 40.32; $Q_3 = 42.54$, (ii) 175]

18. Six types of workers are employed in each of two workshops but at different rates of wages as follows :

Types of workers	*Workshop A Daily wages per worker*	*No. of workers*	*Wrokshop B Daily wages per worker*	*No. of workers*
Mechanic	92.50	2	93.00	11
Fitter	93.50	14	93.00	50
Electrician	94.00	20	94.25	8
Carpenter	93.00	7	93.50	10
Smith	93.00	6	93.50	10
Clerk	92.00	1	95.00	2

In which of the two workshops is the average rate of wages per worker higher and by how much?

19. (*a*) A motor car covered a distance of 50 km four times. The first time at 50 km p.h., the second at 20 km p.h., the third at 40 km p.h. and the fourth at 25 km p.h. Calculate the average speed and explain the choice of the average.
[29.63]

(*b*) A man gets three annual increases in salary. At the end of the first year he gets an increase of 4%, at the end of second year an increase of 6% on his salary as it was at the end of the first year, and at the end of the third year an increase of 9% on his salary as it was at the end of the second year. What is the average percentage increase?
[6.7]

(*c*) A machine is assumed to depreciate 44 per cent in value in the first year, 25% in the second year and 10% per annum for the next three years, each percentage being calculated on the diminishing value. What is the average percentage depreciation for the five years? (*MBA. Vikram Univ., 2001*)
[21.07%]

20. (*a*) Mr. *A* spends Rs. 1000 for apples costing Rs. 25 per kilogram and another Rs. 1000 for apples costing Rs. 20 per kilogram. What is the average price of apples per kilogram? (*MBA, Vikram Univ., 2005*)
[22.22]

(*b*) Three men take 12, 8, 6 hours respectively to husk an acre of corn. Determine the average number of hours to husk an acre.
[$\bar{X}$ = 8.67]

21. (*a*) If oranges for one rupee are bought at 10 paise each and for another rupee at 5 paise each, the average price would be $6\frac{2}{3}$ paise and not $7\frac{1}{2}$ paise. Explain and verify.

(*b*) You take a trip which entails travelling 900 kilometres by train at an average speed of 60 kilometres per hour, 3,000 kilometres by boat at an average of 25 kilometres per hour, 400 kilometres by plane at 350 kilometres per hour and finally 15 kilometres by taxi at 25 kilometres per hour. What is your average speed for the entire distance?
[31.6]

(*c*) A certain store made weekly profits of Rs. 5,000, Rs. 10,000 and Rs. 80,000 in 2008, 2009 and 2010 respectively. Determine the average rate of growth of this store's profits.
[266.6]

22. The following distribution represents the number of minutes spent by a group of teenagers in going to movies. What is the median?

Minutes/Week	*Number of teenagers*	*Minutes/Week*	*Number of teenagers*
0–99	27	400–499	58
100–199	32	500–599	38
200–299	65	600 and more	9
300–399	78		

[Med. = 333.6]

23. An investor buys Rs. 1,200 worth of shares in a company each month. During the first five months he bought the shares at a price of Rs. 10, Rs. 12, Rs. 15, Rs. 20 and Rs. 24 per share. After 5 months what is the average price paid for the shares by him?
[14.63]

24. The value of a machine decreases at a constant rate from the cost price of Rs. 10,000 to scrap value of Rs. 1,000 in ten years. Find the annual rate of decrease and the value of the machine at the end of one, two and three years.
[25.89]

25. An incomplete distribution is given below :

Variable :	10–20	20–30	30–40	40–50	50–60	60–70	70–80	Total
Frequency :	12	30	?	65	?	25	18	229

You are told that the median value is 46. Using the median formula, fill up the missing frequencies and calculate the arithmetic mean of the completed table.

[Freq. corresponding to 30 – 40 is 33.5;
and corresponding to 50 – 60 is 45.5 ; and $\bar{X}$ = 45.87]

26. The production of butter fat during 7 consecutive days was recorded for 300 cows. Calculate the average fat content.

Butter fat (lb) :	10–,	11—,	12—,	13—,	14—,	15—,	16—,	17—,
No. of cows :	8	25	50	75	60	48	22	12

[14 lb]

27. The number of telephone calls received in 293 successive one-minute intervals at an exchange are shown in the following table :

No. of calls :	0	1	2	3	4	5	6	7	8
Frequency :	10	28	35	45	65	52	32	12	14

Calculate mean and modal number of calls.
[3.89 ; 4]

8. Given below is the frequency distribution of the marks obtained by 90 students. Compute the arithmetic mean, median and mode :

Marks	No. of Students	Marks	No. of Students
20–29	2	60–69	18
30–39	12	70–79	10
40–49	15	80–89	9
50–59	20	90–99	4

(58.5, 57.5, 56.64)

29. The geometric mean of 10 observations on a certain variable was calculated to be 16.2. It was later discovered that one of the observations was wrongly recorded at 10.9 when in fact it was 21.9. Apply appropriate correction and calculate the correct geometric mean.
[17.08]

30. A recent college graduate was hired by a large manufacturing corporation and placed in their management training programme. As part of her training, she was assigned to five different departments of the corporation for various periods of time. At the end of the training period in each department, the supervisor graded her performance on a scale from zero to ten. At the end of the training programme, the training director computed an overall mean score based on the following consideration :

The marketing and production phases of her training were assumed to be of equal importance. Both of these were considered to be three times as important as the purchasing and financial phases. The accounting training was twice as important as the latter two.

If the supervisor's ratings were as follows, compute an appropriate mean score :

Department	*Score*
Finance	4
Marketing	7
Production	8
Purchasing	6
Accounting	9

31. Atul gets a pocket money allowance of Rs. 120 per month. Thinking that this was rather less, he asked his friends about their allowances and obtained the following data which includes his allowance also (Amounts in Rs.) :

112,	118,	110,	115,	125,	120,	120,	122,	115,	110,	110,	115,	113, 120,
118,	110,	115,	110,	117,	118,	115,	112,	115,	110,	115,	110,	112, 118,
120,	125,	118.										

He presented these data to his father and asked for an increase in his allowances as he was getting less than average amount. His father, a statistician, countered pointing out that Atul's allowance was actually more than the average amount. Reconcile these statements.

Would Atul or his friends getting less than the 'average' have no more cause for complaint if their allowances were increased to the 'average' amount? Give reasons for your answer.

32. From the following table showing the wage distribution in a certain factory, determine :
(*a*) the mean wage, (*b*) the median wage, (*c*) the modal wage,
(*d*) the wage limits for the middle 50% of the wage earners,
(*e*) The percentage of the workers who earned between Rs. 1750 and Rs. 2250,
(*f*) The percentage who earned more than Rs. 2500 per week, and
(*g*) The percentage who earned less than Rs. 2000 per week.

Weekly wage (Rs.)	*No. of workers*	*Weekly wage (Rs.)*	*No. of workers*
1200–1400	8	2200–2400	32
1400–1600	12	2400–2600	18
1600–1800	20	2600–2800	7
1800–2000	30	2800–3000	6
2000–2200	40	3000–3200	4

33. From the following distribution of travel time to work of a firm's employees, find the modal travel time :

Travel time (in minutes)	*Frequency*
Less than 80	218
Less than 70	215
Less than 60	195
Less than 50	156
Less than 40	85
Less than 30	50
Less than 20	18
Less than 10	2

[illegible] given the following incomplete frequency distribution. It is known that the total frequency is 1,000 and that the median is 413.11. Estimate by calculation the missing frequencies and find the value of the mode.

Sales (Rs. lakhs)	*No. of companies*	*Sales (Rs. lakhs)*	*No. of companies*
300–325	5	400–425	326
325–350	17	425–450	7
350–375	80	450–475	88
375–400	?	475–500	9

[127, 248; mode = 413.98]

35. (*a*) In a certain office a letter is typed by *A* in 4 minutes. The same letter is typed by *B*, *C* and *D* in 5, 6, 10 minutes respectively. What is the average time taken in completing one letter? How many letters do you expect to be typed in one day comprising 9 working hours?
[H.M. = 5.58 minutes per letter.
Letters typed in 8 hours (480 minutes) = 86]

(*b*) For an income distribution of a group of men, 20 p.c. of men have income below Rs. 3500, 35 p.c. below Rs. 7500, 60 p.c. below Rs. 17500 and 80 p.c. below Rs. 25000; the first and third quartile are Rs. 5500 and Rs. 20000. Put the above information in cumulative frequency distribution and find the median.

36. *The rate of a certain commodity in the first week of January.* 2008 *was 0.4 kg per rupee; it was 0.6 kg per rupee in the second week and 0.5 kg per rupee in the third week. Therefore, it is correct to say that the average price was 0.5 kg per rupee. Verify.*

37. Below given is the frequency distribution of weekly wages of 100 workers in a factory :

Weekly wages (Rs.)	*No. of workers*	*Weekly wages (Rs.)*	*No. of workers*
1120–1124	3	1145–1149	10
1125–1129	5	1150–1154	8
1130–1134	12	1155–1159	5
1135–1139	23	1160–1164	3
1140–1144	31		

Draw the ogive for the distribution and use it to determine the median wage of a worker and verify the result by the formula. How many workers earned weekly wages between Rs. 1132 and Rs. 1153?

38. Find the missing frequencies in the following distribution if *N* is 100 and median 30 :

Marks :	0–10	10–20	20–30	30–40	40–50	50–60
No. of students :	10	15	?	30	10	8

(*MBA, M.D. Univ., 1999*)

39. Draw an ogive for the following distribution. Read the median from the graph and verify your result by the mathematical formula. Also obtain the limits of income of central 50% of the employees.

Weekly Income (Rs.)	*No. of employees*	*Weekly Income (Rs.)*	*No. of employees*
Below 550	6	700–750	16
500–600	10	750–800	12
600–650	22	Above 800	15
650–700	30		

(*MBA, Delhi Univ., 1999*)

[Med. = 679.2; 626.7 to 747.7]

40. Following is the cumulative frequency distribution of preferred length of kitchen slabs obtained from the preference study on 50 housewives.

Length (in metres) more than	*Number of housewives*	*Length (in metres) more than*	*Number of housewives*
1.0	50	2.5	42
1.5	46	3.0	10
2.0	40	3.5	3

A manufacturer has to take decision on what length of slabs to manufacture. What length would you recommend and why?

41. Following are the data for marks obtained by students in a paper. The top 20% students will qualify for a prize. What is the lower limit of marks above which the student will get the prize?

Marks	*No. of students*	*Marks*	*No. of students*
0–10	5	50–60	10
10–20	7	60–70	4
20–30	8	70–80	4
30–40	10	80–90	2
40–50	10		

42. A factory pays workers on piece rate basis and also a bonus to each worker on the basis of individual output in each quarter. The rate of bonus payable is as follows :

Output (in units)	*Bonus (in rupees)*	*Output (in units)*	*Bonus (in rupees)*
70–74	400	90–94	700
75–79	450	95–99	800
80–84	500	100–104	1000
85–89	600		

The individual output of a batch of 50 workers is given below :

94	83	78	76	88	86	93	80	91	82
89	97	92	84	932	80	85	83	98	103
87	88	88	81	95	86	99	81	87	90
84	97	80	75	93	101	82	82	89	72
85	83	75	72	83	98	77	87	71	80

By suitable classification you are required to find :

(*i*) Average bonus per worker for the quarter.

(*ii*) Average output per worker.

[(*i*) 90.03 (*ii*) 86.1]

43. An individual purchases three qualities of ball-pens. The relevant data are given below :

Quality	*Price per ball-pen (Rs.)*	*Money spent (Rs.)*
A	10.00	500
B	10.50	300
C	20.00	200

Calculate an average price per ball-pen. (*MBA, Kurukshetra Univ., 2002*)

44. A number of particular articles have been classified according to their weights. After drying for two weeks the same articles have been weighted and similarly classified. It is known that the median weight in the first weighing it was 17.35. Some frequencies in the first weighing (*a* and *b*) and second weighing (*x* and *y*) are missing. It is known that $a = 1/3x$ and $b = 1/2y$. Find out the value of *a*, *b*, *x* and *y*.

	1st Weighing	*IInd Weighing*
0–5	*a*	*x*
5–10	*b*	*y*
10–15	11	40
15–20	52	50
20–25	75	30
25–30	22	28

[$a = 3, b = 6, x = 9, y = 12$]

45. Describe the method of constructing ogive. How would you determine median from it? Draw ogive and find median from the following data :

Marks :	0–15	15–30	30–45	45–60	60–75
No. of Students :	2	15	30	9	4

(*M. Com., AMU, 2001*)

Calculate the median and quartiles for the following data :

Class-Interval	*Frequency*	*Class-Interval*	*Frequency*
0–50	20	150-200	30
50-100	60	200-250	24
100-150	50	250-300	16

46. Calculate the mean and median for the following data :

Central wage (in Rs.) :	15	20	25	30	35	40	45
No. of wage earners :	3	25	19	16	4	5	6

(*MBA, Madurai Kamaraj Univ., 2007*)

*****