MEASURES OF CENTRAL TENDENCY AND DISPERSION
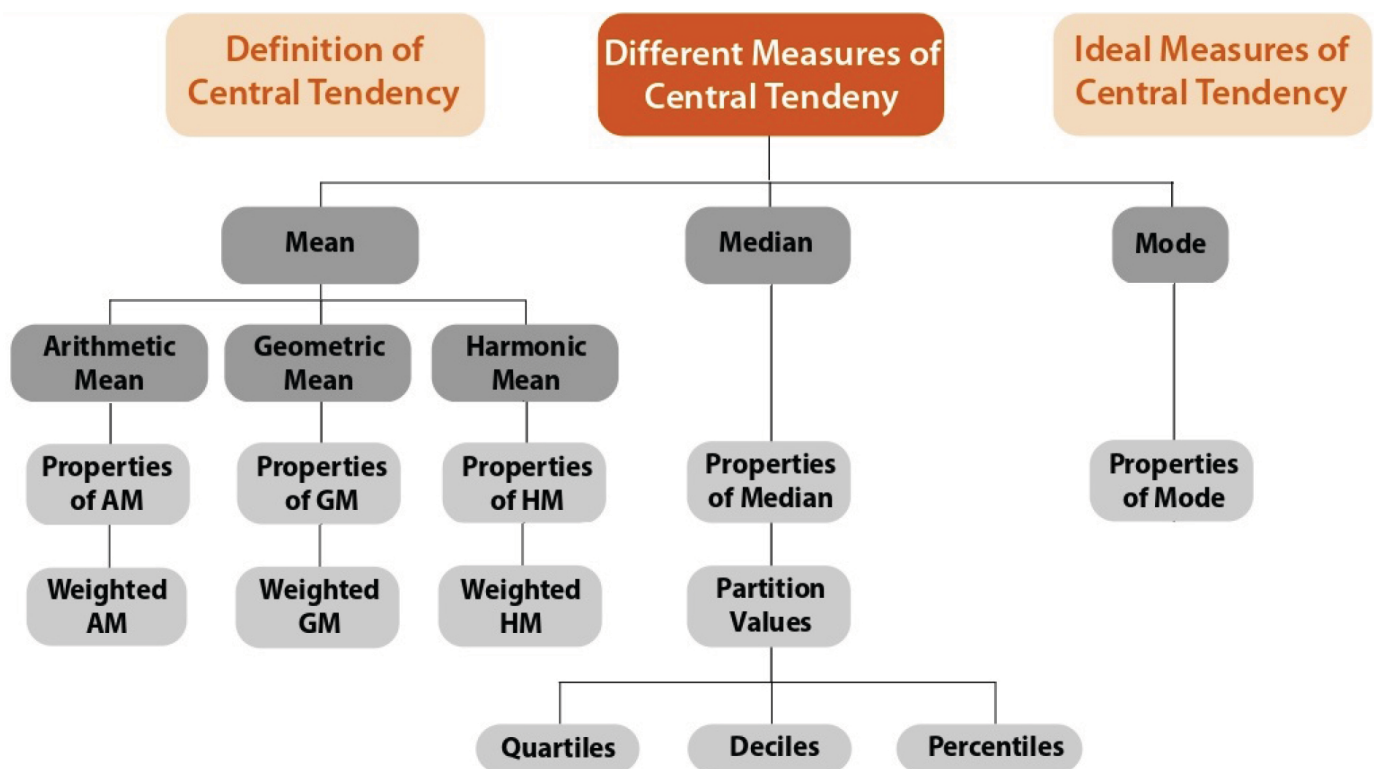
UNIT I: MEASURES OF CENTRAL TENDENCY

LEARNING OBJECTIVES

After reading this chapter, students will be able to understand:

◆ To understand different measures of central tendency, i.e. Arithmetic Mean, Median, Mode, Geometric Mean and Harmonic Mean, and computational techniques of these measures.

◆ To learn comparative advantages and disadvantages of these measures and therefore, which measures to use in which circumstance.

UNIT OVERVIEW 👉

```
Definition of          Different Measures of       Ideal Measures of
Central Tendency       Central Tendeny             Central Tendency
                              │
        ┌─────────────────────┼──────────────────────┐
      Mean                  Median                   Mode
        │                     │                       │
  ┌─────┼─────┐               │                       │
Arithmetic Geometric Harmonic │                       │
  Mean     Mean      Mean      │                       │
  │        │         │    Properties              Properties
Properties Properties Properties of Median          of Mode
of AM      of GM     of HM
  │        │         │    Partition
Weighted Weighted Weighted  Values
  AM       GM       HM        │
                    ┌─────────┼──────────┐
                Quartiles  Deciles   Percentiles
```

14.1.1 DEFINITION OF CENTRAL TENDENCY

In many a case, like the distributions of height, weight, marks, profit, wage and so on, it has been noted that starting with rather low frequency, the class frequency gradually increases till it reaches its maximum somewhere near the central part of the distribution and after which the class frequency steadily falls to its minimum value towards the end. Thus, central tendency may be defined as the tendency of a given set of observations to cluster around a single central or middle value and the single value that represents the given set of observations is described as a measure of central tendency or, location, or average. Hence, it is possible to condense a vast mass of data by a single representative value. The computation of a measure of central tendency plays a very important part in many a sphere. A company is recognized by its high average profit, an educational institution is judged on the basis of average marks obtained by its students and so on. Furthermore, the central tendency also facilitates us in providing a basis for comparison between different distribution. Following are the different measures of central tendency:

(i) Mean
  ◆ Arithmetic Mean (AM)
  ◆ Geometric Mean (GM)
  ◆ Harmonic Mean (HM)
(ii) Median (Me)
(iii) Mode (Mo)

14.1.2 CRITERIA FOR AN IDEAL MEASURE OF CENTRAL TENDENCY

Following are the criteria for an ideal measure of central tendency:

(i) It should be properly and unambiguously defined.

(ii) It should be easy to comprehend.

(iii) It should be simple to compute.

(iv) It should be based on all the observations.

(v) It should have certain desirable mathematical properties.

(vi) It should be least affected by the presence of extreme observations.

14.1.3 ARITHMETIC MEAN

For a given set of observations, the AM may be defined as the sum of all the observations divided by the number of observations. Thus, if a variable x assumes n values $x_1, x_2, x_3, \ldots\ldots..x_n$, then the AM of x, to be denoted by $\overline{x}$, is given by,

$$\overline{X} = \frac{x_1 + x_2 + x_3 + \ldots\ldots\ldots + x_n}{n}$$

$$= \frac{\sum\limits_{i=1}^{n} x_i}{n}$$

$$\overline{X} = \frac{\sum x_i}{n} \qquad\qquad ....(14.1.1)$$

In case of a simple frequency distribution relating to an attribute, we have

$$\overline{x} = \frac{f_1 x_1 + f_2 x_2 + f_3 x_3 + ... + f_n x_n}{f_1 + f_2 + f_3 + ... + f_n}$$

$$= \frac{\sum f_i x_i}{\sum f_i}$$

$$\overline{X} = \frac{\sum f_i x_i}{N} \qquad\qquad ....(14.1.2)$$

assuming the observation x_i occurs f_i times, i=1,2,3,……..n and N=$\sum f_i$.

In case of grouped frequency distribution also we may use formula (14.1.2) with x_i as the mid value of the i-th class interval, on the assumption that all the values belonging to the i-th class interval are equal to x_i.

However, in most cases, if the classification is uniform, we consider the following formula for the computation of AM from grouped frequency distribution:

$$\overline{x} = A + \frac{\sum f_i d_i}{N} \times C \qquad\qquad ....(14.1.3)$$

Where, $d_i = \frac{x_i - A}{C}$

        A = Assumed Mean

        C = Class Length

ILLUSTRATIONS:

Example 14.1.1: Following are the daily wages in Rupees of a sample of 9 workers: 58, 62, 48, 53, 70, 52, 60, 84, 75. Compute the mean wage.

Solution: Let x denote the daily wage in rupees.

Then as given, x_1=58, x_2=62, x_3= 48, x_4=53, x_5=70, x_6=52, x_7=60, x_8=84 and x_9=75.

Applying (14.1.1) the mean wage is given by,

$$\overline{x} = \frac{\sum_{i=1}^{9} x_i}{9}$$

$$= ₹ \frac{(58 + 62 + 48 + 53 + 70 + 52 + 60 + 84 + 75)}{9}$$

$$= ₹ \frac{562}{9}$$

$$= ₹ 62.44.$$

Example 14.1.2: Compute the mean weight of a group of BBA students of St. Xavier's College from the following data:

Weight in kgs.	44 – 48	49 – 53	54 – 58	59 – 63	64 – 68	69 – 73
No. of Students	3	4	5	7	9	8

Solution: Computation of mean weight of 36 BBA students

Table 14.1.1

Weight in kgs. (1)	No. of Student (f_i) (2)	Mid-Value (x_i) (3)	$f_i x_i$ (4) = (2) x (3)
44 – 48	3	46	138
49 – 53	4	51	204
54 – 58	5	56	280
59 – 63	7	61	427
64 – 68	9	66	594
69 – 73	8	71	568
Total	36	–	2211

Applying (14.1.2), we get the average weight as

$$\overline{x} = \frac{\sum f_i x_i}{N}$$

$$= \frac{2211}{36} \text{ kgs.}$$
$$= 61.42 \text{ kgs.}$$

Example 14.1.3: Find the AM for the following distribution:

Class Interval	350 – 369	370 – 389	390 – 409	410 – 429	430 – 449	450 – 469	470 – 489
Frequency	23	38	58	82	65	31	11

Solution: We apply formula (14.1.3) since the amount of computation involved in finding the AM is much more compared to **Example 14.1.2**. Any mid value can be taken as A. However, usually A is taken as the middle most mid-value for an odd number of class intervals and any one of the two middle most mid-values for an even number of class intervals. The class length is taken as C.

Table 14.1.2 Computation of AM

Class Interval	Frequency(f_i)	Mid-Value(x_i)	$d_i = \dfrac{x_i - A}{c}$ $= \dfrac{x_i - 419.50}{20}$	$f_i d_i$
(1)	(2)	(3)	(4)	(5) = (2)X(4)
350 – 369	23	359.50	– 3	– 69
370 – 389	38	379.50	– 2	– 76
390 – 409	58	399.50	– 1	– 58
410 – 429	82	419.50 (A)	0	0
430 – 449	65	439.50	1	65
450 – 469	31	459.50	2	62
470 – 489	11	479.50	3	33
Total	308	–	–	– 43

The required AM is given by

$$\overline{x} = A + \frac{\sum f_i d_i}{N} \times C$$

$$= 419.50 + \frac{(-43)}{308} \times 20$$

$$= 419.50 - 2.79$$

$$= 416.71$$

Example 14.1.4: Given that the mean height of a group of students is 67.45 inches. Find the missing frequencies for the following incomplete distribution of height of 100 students.

Height in inches	60 – 62	63 – 65	66 – 68	69 – 71	72 – 74
No. of Students	5	18	–	–	8

Solution: Let x denote the height and f_3 and f_4 as the two missing frequencies.

Table 14.1.3

Estimation of missing frequencies

CI	Frequency (f$_i$)	Mid - Value (x$_i$)	$d_i = \dfrac{x_i - A}{c}$ $\dfrac{x_i - 67}{3}$	f$_i$d$_i$
(1)	(2)	(3)	(4)	(5) = (2) x (4)
60-62	5	61	-2	-10
63 – 65	18	64	– 1	– 18
66 – 68	f$_3$	67 (A)	0	0
69 – 71	f$_4$	70	1	f$_4$
72 – 74	8	73	2	16
Total	31+ f$_3$+ f$_4$	–	–	– 12+f$_4$

As given, we have

$$31 + f_3 + f_4 = 100$$

$\Rightarrow$ $f_3 + f_4 = 69$ ………………………………..(1)

and $\bar{x} = 67.45$

$\Rightarrow$ $A + \dfrac{\sum f_i d_i}{N} \times C = 67.45$

$\Rightarrow$ $67 + \dfrac{(-12 + f_4)}{100} \times 3 = 67.45$

$\Rightarrow$ $(-12 + f_4) \times 3 = (67.45 - 67) \times 100$

$\Rightarrow$ $-12 + f_4 = 15$

$\Rightarrow$ $f_4 = 27$

On substituting 27 for f$_4$ in (1), we get

$f_3 + 27 = 69$, $\Rightarrow f_3 = 42$

Thus, the missing frequencies would be 42 and 27.

Properties of AM

(i)   **If all the observations assumed by a variable are constants, say k, then the AM is also k.** For example, if the height of every student in a group of 10 students is 170 cm, then the mean height is, of course, 170 cm.

(ii) the algebraic sum of deviations of a set of observations from their AM is zero

i.e. for unclassified data , $\sum(x_i - \overline{x}) = 0$

and for grouped frequency distribution, $\sum f_i(x_i - \overline{x}) = 0$ $\Bigg\}$ ....(14.1.4)

For example, if a variable x assumes five observations, say 58, 63, 37, 45, 29, then $\overline{x} = 46.4$. Hence, the deviations of the observations from the AM i.e. $(x_i - \overline{x})$ are 11.60, 16.60, –9.40, –1.40 and –17.40, then $\sum(x_i - \overline{x}) = 11.60 + 16.60 + (-9.40) + (-1.40) + (-17.40) = 0$.

(iii) AM is affected due to a change of origin and/or scale which implies that if the original variable x is changed to another variable y by effecting a change of origin, say a, and scale say b, of x i.e. y=a+bx, then the AM of y is given by $\overline{y} = a + b\overline{x}$.

For example, if it is known that two variables x and y are related by 2x+3y+7=0 and $\overline{x} = 15$, then the AM of y is given by $\overline{y} = \dfrac{-7 - 2\overline{x}}{3}$

$= \dfrac{-7 - 2 \times 15}{3} = \dfrac{-37}{3} = -12.33$.

(iv) If there are two groups containing n_1 and n_2 observations and $\overline{x}_1$ and $\overline{x}_2$ as the respective arithmetic means, then the combined AM is given by

$\overline{x} = \dfrac{n_1\overline{x}_1 + n_2\overline{x}_2}{n_1 + n_2}$ ……………………………(14.1.5)

This property could be extended to k>2 groups and we may write

$\overline{x} = \dfrac{\sum n_i\overline{x}_i}{\sum n_i}$ ……………………………….(14.1.6) i = 1, 2,........n.

Example 14.1.5: The mean salary for a group of 40 female workers is ₹5,200 per month and that for a group of 60 male workers is ₹6800 per month. What is the combined mean salary?

Solution: As given n_1 = 40, n_2 = 60, $\overline{x}_1$ = ₹5,200 and $\overline{x}_2$ = ₹6,800 hence, the combined mean salary per month is

$\overline{x} = \dfrac{n_1\overline{x}_1 + n_2\overline{x}_2}{n_1 + n_2}$

$= \dfrac{40 \times ₹5,200 + 60 \times ₹6,800}{40 + 60}$ = ₹6,160.

14.1.4 MEDIAN – PARTITION VALUES

As compared to AM, median is a positional average which means that the value of the median is dependent upon the position of the given set of observations for which the median is wanted. Median, for a given set of observations, may be defined as the middle-most value when the observations are arranged either in an ascending order or a descending order of magnitude.

As for example, if the marks of the 7 students are 72, 85, 56, 80, 65, 52 and 68, then in order to find the median mark, we arrange these observations in the following ascending order of magnitude: 52, 56, 65, 68, 72, 80, 85.

Since the 4th term i.e. 68 in this new arrangement is the middle most value, the median mark is 68 i.e. Median (Me) = 68.

As a second example, if the wages of 8 workers, expressed in rupees are

56, 82, 96, 120, 110, 82, 106, 100 then arranging the wages as before, in an ascending order of magnitude, we get ₹56, ₹ 82, ₹ 82, ₹ 96, ₹ 100, ₹ 106, ₹ 110, ₹ 120. Since there are two middle-most values, namely, ₹ 96, and ₹ 100 any value between ₹ 96 and ₹ 100 may be, theoretically, regarded as median wage. However, to bring uniqueness, we take the arithmetic mean of the two middle-most values, whenever the number of the observations is an even number. Thus, the median wage in this example, would be

$$M = \frac{₹\,96 + ₹\,100}{2} = ₹\,98$$

In case of a grouped frequency distribution, we find median from the cumulative frequency distribution of the variable under consideration. We may consider the following formula, which can be derived from the basic definition of median.

$$M = l_1 + \left(\frac{\frac{N}{2} - N_l}{N_u - N_l}\right) \times C \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(14.1.7)$$

Where,

l_1 = lower class boundary of the median class i.e. the class containing median.

N = total frequency.

N_l = less than cumulative frequency corresponding to l_1. (Pre median class)

N_u = less than cumulative frequency corresponding to l_2. (Post median class)

l_2 being the upper class boundary of the median class.

C = $l_2 - l_1$ = length of the median class.

Example 14.1.6: Compute the median for the distribution as given in **Example 14.1.3.**

Solution: First, we find the cumulative frequency distribution which is exhibited in **Table 14.1.4.**

Table 14.1.4

Computation of Median

Class boundary	Less than cumulative frequency
349.50	0
369.50	23
389.50	61
409.50 (l_1)	119 (N_l)
429.50 (l_2)	201(N_u)
449.50	266
469.50	297
489.50	308

We find, from the **Table 14.1.4**, $\frac{N}{2} = \frac{308}{2}$ = 154 lies between the two cumulative frequencies 119 and 201 i.e. 119 < 154 < 201 . Thus, we have N_l = 119, N_u = 201 l_1 = 409.50 and l_2 = 429.50. Hence C = 429.50 – 409.50 =20.

Substituting these values in (14.1.7), we get,

M $= 409.50 + \dfrac{154 - 119}{201 - 119} \times 20$

= 409.50+8.54

= 418.04.

Example 14.1.7: Find the missing frequency from the following data, given that the median mark is 23.

Mark	:	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
No. of students	:	5	8	?	6	3

Solution: Let us denote the missing frequency by f_3. Table 14.1.5 shows the relevant computation.

Table 14.1.5
(Estimation of missing frequency)

Mark	Less than cumulative frequency
0	0
10	5
$20(l_1)$	$13(N_l)$
$30(l_2)$	$13+f_3(N_u)$
40	$19+f_3$
50	$22+f_3$

Going through the mark column, we find that $20<23<30$. Hence $l_1=20$, $l_2=30$ and accordingly $N_l=13$, $N_u=13+f_3$. Also the total frequency i.e. N is $22+f_3$. Thus,

$$M = l_1 + \left(\frac{\frac{N}{2}-N_l}{N_u - N_l}\right) \times C$$

$$\Rightarrow \quad 23 = 20 + \frac{\left(\frac{22+f_3}{2}\right)-13}{(13+f_3)-13} \times 10$$

$$\Rightarrow \quad 3 = \frac{22+f_3-26}{f_3} \times 5$$

$$\Rightarrow \quad 3f_3 = 5f_3 - 20$$

$$\Rightarrow \quad 2f_3 = 20$$

$$\Rightarrow \quad f_3 = 10$$

So, the missing frequency is 10.

Properties of median

We cannot treat median mathematically, the way we can do with arithmetic mean. We consider below two important features of median.

(i) If x and y are two variables, to be related by y=a+bx for any two constants a and b, then the median of y is given by
$y_{me} = a + bx_{me}$

For example, if the relationship between x and y is given by $2x – 5y = 10$ and if x_{me} i.e. the median of x is known to be 16.
Then $2x – 5y = 10$

$\Rightarrow \qquad y = -2 + 0.40x$

$\Rightarrow \qquad y_{me} = -2 + 0.40\ x_{me}$

$\Rightarrow \qquad y_{me} = -2 + 0.40 \times 16$

$\Rightarrow \qquad y_{me} = 4.40.$

(ii)   For a set of observations, the sum of absolute deviations is minimum when the deviations are taken from the median. This property states that $\sum |x_i - A|$ is minimum if we choose A as the median.

PARTITION VALUES OR QUARTILES OR FRACTILES

These may be defined as values dividing a given set of observations into a number of equal parts. When we want to divide the given set of observations into two equal parts, we consider median. Similarly, quartiles are values dividing a given set of observations into four equal parts. So there are three quartiles – first quartile or lower quartile denoted by Q_1, second quartile or median to be denoted by Q_2 or Me and third quartile or upper quartile denoted by Q_3. First quartile is the value for which one fourth of the observations are less than or equal to Q_1 and the remaining three – fourths observations are more than or equal to Q_1. In a similar manner, we may define Q_2 and Q_3.

Deciles are the values dividing a given set of observation into ten equal parts. Thus, there are nine deciles to be denoted by $D_1, D_2, D_3,.....D_9$. D_1 is the value for which one-tenth of the given observations are less than or equal to D_1 and the remaining nine-tenth observations are greater than or equal to D_1 when the observations are arranged in an ascending order of magnitude.

Lastly, we talk about the percentiles or centiles that divide a given set of observations into 100 equal parts. The points of sub-divisions being $P_1, P_2,..........P_{99}$. P_1 is the value for which one hundredth of the observations are less than or equal to P_1 and the remaining ninety-nine hundredths observations are greater than or equal to P_1 once the observations are arranged in an ascending order of magnitude.

For unclassified data, the pth quartile is given by the $(n+1)p^{th}$ value, where n denotes the total number of observations. p = 1/4, 2/4, 3/4 for Q_1, Q_2 and Q_3 respectively. p=1/10, 2/10,............9/10. For $D_1, D_2,......,D_9$ respectively and lastly p=1/100, 2/100,....,99/100 for P_1, $P_2, P_3....P_{99}$ respectively.

In case of a grouped frequency distribution, we consider the following formula for the computation of quartiles.

$$Q = l_1 + \left(\frac{Np - N_l}{N_u - N_l}\right) \times C \quad \text{..} \quad (14.1.8)$$

The symbols, except p, have their usual interpretation which we have already discussed while computing median and just like the unclassified data, we assign different values to p depending on the quartile.

Another way to find quartiles for a grouped frequency distribution is to draw the ogive (less than type) for the given distribution. In order to find a particular quartile, we draw a line parallel to the horizontal axis through the point Np. We draw perpendicular from the point of intersection of this parallel line and the ogive. The x-value of this perpendicular line gives us the value of the quartile under discussion.

Example 14.1.8: Following are the wages of the labourers: ₹ 82, ₹ 56, ₹ 90, ₹ 50, ₹ 120, ₹ 75, ₹ 75, ₹ 80, ₹ 130, ₹ 65. Find Q_1, D_6 and P_{82}.

Solution: Arranging the wages in an ascending order, we get ₹ 50, ₹ 56, ₹ 65, ₹ 75, ₹ 75, ₹ 80, ₹ 82, ₹ 90, ₹ 120, ₹ 130.
Hence, we have

$$Q_1 = \frac{(n+1)}{4} \text{th value}$$

$$= \frac{(10+1)}{4} \text{th value}$$

$$= 2.75^{\text{th}} \text{ value}$$

$$= 2^{\text{nd}} \text{ value} + 0.75 \times \text{difference between the third and the } 2^{\text{nd}} \text{ values.}$$

$$= ₹ [56 + 0.75 \times (65 - 56)]$$

$$= ₹ 62.75$$

$$D_6 = (15 + 1) \times \frac{6}{10} \text{ th value}$$

$$= 6.60^{\text{th}} \text{ value}$$

$$= 6^{\text{th}} \text{ value} + 0.60 \times \text{difference between the } 7^{\text{th}} \text{ and the } 6^{\text{th}} \text{ values.}$$

$$= ₹ (80 + 0.60 \times 2)$$

$$= ₹ 81.20$$

$$P_{82} = (10+1) \times \frac{82}{100} \text{ th value}$$

$$= 9.02^{\text{th}} \text{ value}$$

$$= 9^{\text{th}} \text{ value} + 0.02 \times \text{difference between the } 10^{\text{th}} \text{ and the } 9^{\text{th}} \text{ values}$$

$$= ₹ (120 + 0.02 \times 10)$$

$$= ₹ 120.20$$

Next, let us consider one problem relating to the grouped frequency distribution.

Example 14.1.9: Following distribution relates to the distribution of monthly wages of 100 workers.

Wages in (₹) :	less than 500	500–699	700–899	900–1099	1100–1499	more than 1500
No. of workers :	5	23	29	27	10	6

Compute Q_3, D_7 and P_{23}.

Solution: This is a typical example of an open end unequal classification as we find the lower class limit of the first class interval and the upper class limit of the last class interval are not stated, and theoretically, they can assume any value between 0 and 500 and 1500 to any number respectively. The ideal measure of the central tendency in such a situation is median as the median or second quartile is based on the fifty percent central values. Denoting the first LCB and the last UCB by the L and U respectively, we construct the following cumulative frequency distribution:

Table 14.1.7
Computation of quartiles

Wages in rupees (CB)	No. of workers (less than cumulative frequency)
L	0
499.50	5
699.50	28
899.50	57
1099.50	84
1499.50	94
U	100

For Q_3, $\dfrac{3N}{4} = \dfrac{3 \times 100}{4} = 75$

since, 57<75 <84, we take N_l = 57, N_u=84, l_1=899.50, l_2=1099.50, c = $l_2 - l_1$ = 200 in the formula (11.8) for computing Q_3.

Therefore, $Q_3 = ₹ \left[899.50 + \dfrac{75-57}{84-57} \times 200 \right] = ₹ 1032.83$

Similarly, for D_7, $\dfrac{7N}{10} = \dfrac{7 \times 100}{10} = 70$ which also lies between 57 and 84.

Thus, $D_7 = ₹ \left[899.50 + \dfrac{70-57}{84-57} \times 200 \right] = ₹995.80$

Lastly for P_{23}, $\dfrac{23N}{100} = \dfrac{23}{100} \times 100 = 23$ and as 5 < 23 < 28, we have

$P_{23} = ₹ [499.50 + \dfrac{23-5}{28-5} \times 200]$

$= ₹ 656.02$

🕐 14.1.5 MODE

For a given set of observations, mode may be defined as the value that occurs the maximum number of times. Thus, mode is that value which has the maximum concentration of the observations around it. This can also be described as the most common value with which, even, a layman may be familiar with.

Thus, if the observations are 5, 3, 8, 9, 5 and 6, then Mode (Mo) = 5 as it occurs twice and all the other observations occur just once. The definition for mode also leaves scope for more than one mode. Thus sometimes we may come across a distribution having more than one mode. Such a distribution is known as a multi-modal distribution. Bi-modal distribution is one having two modes.

Furthermore, it also appears from the definition that mode is not always defined. As an example, if the marks of 5 students are 50, 60, 35, 40, 56, there is no modal mark as all the observations occur once i.e. the same number of times.

We may consider the following formula for computing mode from a grouped frequency distribution:

$$\text{Mode} = l_1 + \left(\frac{f_0 - f_{-1}}{2f_0 - f_{-1} - f_1} \right) \times c \qquad \text{...........................(14.1.9)}$$

where,

l_1 = LCB of the modal class.
       i.e. the class containing mode.
f_0 = frequency of the modal class
f_{-1} = frequency of the pre-modal class
f_1 = frequency of the post modal class
C = class length of the modal class

Example 14.1.10: Compute mode for the distribution as described in Example. 14.1.3

Solution: The frequency distribution is shown below:

Table 14.1.8
Computation of mode

Class Interval	Frequency
350 - 369	23
370 - 389	38
390 - 409	58 (f_{-1})
410 - 429	82 (f_0)
430 - 449	65 (f_1)
450 - 469	31
470 - 489	11

Going through the frequency column, we note that the highest frequency i.e. f_0 is 82. Hence, $f_{-1} = 58$ and $f_1 = 65$. Also the modal class i.e. the class against the highest frequency is 410 – 429.

Thus l_1 = LCB=409.50 and c=429.50 – 409.50 = 20

Hence, applying formulas (11.9), we get

$$Mo = 409.5 + \frac{82-58}{2\times82-58-65} \times 20$$

= 421.21 which belongs to the modal class. (410 – 429)

When it is difficult to compute mode from a grouped frequency distribution, we may consider the following empirical relationship between mean, median and mode:

Mean – Mode = 3(Mean – Median) …………………….(14.1.9A)

or Mode = 3 Median – 2 Mean

(14.1.9A) holds for a moderately skewed distribution. We also note that if y = a+bx, then

$y_{mo}=a+bx_{mo}$ ………………………………………(14.1.10)

Example 14.11: For a moderately skewed distribution of marks in statistics for a group of 200 students, the mean mark and median mark were found to be 55.60 and 52.40. What is the modal mark?

Solution: Since in this case, mean = 55.60 and median = 52.40, applying (11.9A), we get the modal mark as

Mode   = 3 × Median – 2 × Mean
        = 3 × 52.40 – 2 × 55.60
        = 46.

Example 14.1.12: If y = 2 + 1.50x and mode of x is 15, what is the mode of y?

Solution:
By virtue of (11.10), we have
       y_{mo} = 2 + 1.50 × 15
            = 24.50.

14.1.6 GEOMETRIC MEAN AND HARMONIC MEAN

For a given set of n positive observations, the geometric mean is defined as the n-th root of the product of the observations. Thus if a variable x assumes n values $x_1, x_2, x_3, \ldots\ldots, x_n$, all the values being positive, then the GM of x is given by

$G= (x_1 \times x_2 \times x_3 \ldots\ldots \times x_n)^{1/n}$ .... (14.1.11)

For a grouped frequency distribution, the GM is given by

$G= (x_1^{f_1} \times x_2^{f_2} \times x_3^{f_3} \ldots\ldots\ldots \times x_n^{f_n})^{1/N}$ .... (14.1.12)

Where N = $\sum f_i$

In connection with GM, we may note the following properties :

(i)   Logarithm of G for a set of observations is the AM of the logarithm of the observations; i.e.

$$\log G = \frac{1}{n}\sum \log x$$
................(14.1.13)

(ii)   if all the observations assumed by a variable are constants, say K > 0, then the GM of the observations is also K.

(iii)   GM of the product of two variables is the product of their GM's i.e. if $z = xy$, then

    GM of z = (GM of x) × (GM of y)   ....(14.1.14)

(iv)   GM of the ratio of two variables is the ratio of the GM's of the two variables i.e. if $z = x/y$ then

GM of $z = \dfrac{\text{GM of } x}{\text{GM of } y}$ ....(14.1.15)

Example 14.1.13: Find the GM of 3, 6 and 12.

Solution: As given $x_1=3$, $x_2=6$, $x_3=12$ and n=3.

Applying (14.1.11), we have G= $(3×6×12)^{1/3} = (6^3)^{1/3}=6$.

Example 14.1.14: Find the GM for the following distribution:

x :	2	4	8	16
f :	2	3	3	2

Solution: According to (14.1.12), the GM is given by

$$G = (x_1^{f_1} \times x_2^{f_2} \times x_3^{f_3} \times x_4^{f_4})^{1/N}$$

$$= (2^2 \times 4^3 \times 8^3 \times 16^2)^{1/10}$$

$$= (2)^{2.50}$$

$$= 4\sqrt{2}$$

$$= 5.66$$

Harmonic Mean

For a given set of non-zero observations, harmonic mean is defined as the reciprocal of the AM of the reciprocals of the observation. So, if a variable x assumes n non-zero values x_1, x_2, x_3,...............,x_n, then the HM of x is given by

$$H = \frac{n}{\sum(1/x_i)}$$
................(14.1.16)

For a grouped frequency distribution, we have

$$H = \frac{N}{\Sigma\left[\dfrac{f_i}{x_i}\right]}$$

..................(14.1.17)

Properties of HM

(i) If all the observations taken by a variable are constants, say k, then the HM of the observations is also k.

(ii) If there are two groups with n_1 and n_2 observations and H_1 and H_2 as respective HM's than the combined HM is given by

$$\frac{n_1 + n_2}{\dfrac{n_1}{H_1} + \dfrac{n_2}{H_2}}$$

..................(14.1.18)

Example 14.15: Find the HM for 4, 6 and 10.

Solution: Applying (14.1.16), we have

$$H = \frac{3}{\dfrac{1}{4} + \dfrac{1}{6} + \dfrac{1}{10}}$$

$$= \frac{3}{0.25 + 0.17 + 0.10}$$

$$= 5.77$$

Example 14.1.16: Find the HM for the following data:

x:	2	4	8	16
f:	2	3	3	2

Solution: Using (14.1.17), we get

$$H = \frac{10}{\dfrac{2}{2} + \dfrac{3}{4} + \dfrac{3}{8} + \dfrac{2}{16}}$$

$$= 4.44$$

Relation between AM, GM, and HM

For any set of positive observations, we have the following inequality:

$$AM \geq GM \geq HM$$ ………….. (14.1.19)

The equality sign occurs, as we have already seen, when all the observations are equal.

Example 14.1.17: compute AM, GM, and HM for the numbers 6, 8, 12, 36.

Solution: In accordance with the definition, we have

$$AM = \frac{6+8+12+36}{4} = 15.50$$

$$GM = (6 \times 8 \times 12 \times 36)^{1/4}$$
$$= (2^8 \times 3^4)^{1/4} = 12$$

$$HM = \frac{4}{\frac{1}{6}+\frac{1}{8}+\frac{1}{12}+\frac{1}{36}} = 9.93$$

The computed values of AM, GM, and HM establish (14.1.19).

Weighted average

When the observations under consideration have a hierarchical order of importance, we take recourse to computing weighted average, which could be either weighted AM or weighted GM or weighted HM.

$$\text{Weighted AM} = \frac{\sum w_i x_i}{\sum w_i}$$ ………….. (14.1.20)

$$\text{Weighted GM} = \text{Anti log} \left(\frac{\sum w_i \log x_i}{\sum w_i} \right)$$ ………….. (14.1.21)

$$\text{Weighted HM} = \frac{\sum w_i}{\sum \left(\frac{w_i}{x_i} \right)}$$ ………….. (14.1.22)

Example 14.1.18: Find the weighted AM and weighted HM of first n natural numbers, the weights being equal to the squares of the corresponding numbers.

Solution: As given,

x	1	2	3	….	n
w	1^2	2^2	3^2	….	n^2

$$\text{Weighted AM} = \frac{\sum w_i x_i}{\sum w_i}$$

$$= \frac{1 \times 1^2 + 2 \times 2^2 + 3 \times 3^2 + \ldots\ldots\ldots\ldots\ldots n \times n^2}{1^2 + 2^2 + 3^2 + \ldots\ldots\ldots\ldots\ldots + n^2}$$

$$= \frac{1^3 + 2^3 + 3^3 + \ldots\ldots\ldots + n^3}{1^2 + 2^2 + 3^2 + \ldots\ldots\ldots\ldots + n^2}$$

$$= \frac{\left[\dfrac{n(n+1)}{2}\right]^2}{\dfrac{n(n+1)(2n+1)}{6}} = \frac{3n(n+1)}{2(2n+1)}$$

$$\text{Weighted HM} = \frac{\sum w_i}{\sum \left(\dfrac{w_i}{x_i}\right)}$$

$$= \frac{1^2 + 2^2 + 3^2 + \ldots\ldots\ldots\ldots n^2}{\dfrac{1^2}{1} + \dfrac{2^2}{2} + \dfrac{3^2}{3} + \ldots\ldots\ldots \dfrac{n^2}{n}}$$

$$= \frac{1^2 + 2^2 + 3^2 + \ldots\ldots\ldots\ldots + n^2}{1 + 2 + 3 + \ldots\ldots\ldots + n}$$

$$= \frac{\dfrac{n(n+1)(2n+1)}{6}}{\dfrac{n(n+1)}{2}}$$

$$= \frac{2n+1}{3}$$

Although mode is the most popular measure of central tendency, there are cases when mode remains undefined. Unlike mean, it has no mathematical property. Mode is also affected by sampling fluctuations.

GM and HM, like AM, possess some mathematical properties. They are rigidly defined and based on all the observations. But they are difficult to comprehend and compute and, as such, have limited applications for the computation of average rates and ratios and such like things.

Example 14.1.19: Given two positive numbers a and b, prove that **AH=G²**. Does the result hold for any set of observations?

Solution: For two positive numbers a and b, we have,

$$A = \frac{a+b}{2}$$

$$G = \sqrt{ab}$$

And $$H = \cfrac{2}{\cfrac{1}{a} + \cfrac{1}{b}}$$

$$= \frac{2ab}{a+b}$$

Thus $$AH = \frac{a+b}{2} \times \frac{2ab}{a+b}$$

$$= ab = G^2$$

This result holds for only two positive observations and not for any set of observations.

Example 14.1.20: The AM and GM for two observations are 5 and 4 respectively. Find the two observations.

Solution: If a and b are two positive observations then as given

$$\frac{a+b}{2} = 5$$

$$\Rightarrow \quad a+b = 10 \quad \text{…………………………………..(1)}$$

and $\sqrt{ab} = 4$

$$\Rightarrow \quad ab = 16 \quad \text{………………………………..(2)}$$

$$\therefore (a-b)^2 = (a+b)^2 - 4ab$$

$$= 10^2 - 4 \times 16$$

$$= 36$$

$\Rightarrow$ $a - b = 6$ (ignoring the negative sign)………………………(3)

Adding (1) and (3) We get,

$\quad\quad 2a = 16$

$\Rightarrow\quad\quad a = 8$

From (1), we get $b = 10 - a = 2$

Thus, the two observations are 8 and 2.

Example 14.1.21: Find the mean and median from the following data:

Marks	:	less than 10	less than 20	less than 30
No. of Students	:	5	13	23

Marks	:	less than 40	less than 50
No. of Students	:	27	30

Also compute the mode using the approximate relationship between mean, median and mode.

Solution: What we are given in this problem is less than cumulative frequency distribution. We need to convert this cumulative frequency distribution to the corresponding frequency distribution and thereby compute the mean and median.

Table 14.1.19

Computation of Mean Marks for 30 students

Marks Class Interval (1)	No. of Students (f_i) (2)	Mid - Value (x_i) (3)	$f_i x_i$ (4)= (2)×(3)
0 – 10	5	5	25
10 – 20	13 – 5 = 8	15	120
20 – 30	23 – 13 = 10	25	250
30 – 40	27 – 23 = 4	35	140
40 – 50	30 – 27 = 3	45	135
Total	30	–	670

Hence the mean mark is given by

$$\overline{x} = \frac{\sum f_i x_i}{N}$$

$$= \frac{670}{30}$$

$$= 22.33$$

Table 14.1.10

Computation of Median Marks

Marks (Class Boundary)	No.of Students (Less than cumulative Frequency)
0	0
10	5
20	13
30	23
40	27
50	30

Since $\frac{N}{2} = \frac{30}{2} = 15$ lies between 13 and 23,

we have $l_1 = 20$, $N_l = 13$, $N_u = 23$

and $C = l_2 - l_1 = 30 - 20 = 10$

Thus,

Median $= 20 + \frac{15-13}{23-13} \times 10$

$= 22$

Since Mode = 3 Median – 2 Mean (approximately), we find that

Mode $= 3 \times 22 - 2 \times 22.33$

$= 21.34$

Example 14.1.22: Following are the salaries of 20 workers of a firm expressed in thousand rupees: 5, 17, 12, 23, 7, 15, 4, 18, 10, 6, 15, 9, 8, 13, 12, 2, 12, 3, 15, 14. The firm gave bonus amounting to ₹ 2,000, ₹ 3,000, ₹ 4,000, ₹ 5,000 and ₹ 6,000 to the workers belonging to the salary groups 1,000 – 5,000, 6,000 – 10,000 and so on and lastly 21,000 – 25,000. Find the average bonus paid per employee.

Solution: We first construct frequency distribution of salaries paid to the 20 employees. The average bonus paid per employee is given by $\dfrac{\sum f_i x_i}{N}$ Where x_i represents the amount of bonus paid to the ith salary group and f_i, the number of employees belonging to that group which would be obtained on the basis of frequency distribution of salaries.

<div align="center">

Table 14.1.11
Computation of Average bonus

</div>

Salary in thousand ₹ (Class Interval) (1)	Tally Mark (2)	No of workers (f_i) (3)	Bonus in Rupees x_i (4)	$f_i x_i$ (5) = (3) × (4)
1-5	||||	4	2000	8000
6-10	⊢⊣⊣	5	3000	15000
11-15	⊢⊣⊣ |||	8	4000	32000
16-20	||	2	5000	10000
21-25	|	1	6000	6000
TOTAL	–	20	–	71000

Hence, the average bonus paid per employee

$$= (₹)\dfrac{71000}{20}$$
$$= (₹)\,3550$$

SUMMARY

◆ The best measure of central tendency, usually, is the AM. It is rigidly defined, based on all the observations, easy to comprehend, simple to calculate and amenable to mathematical properties. However, AM has one drawback in the sense that it is very much affected by sampling fluctuations. In case of frequency distribution, mean cannot be advocated for open-end classification.

◆ Median is also rigidly defined and easy to comprehend and compute. But median is not based on all the observation and does not allow itself to mathematical treatment. However, median is not much affected by sampling fluctuation and it is the most appropriate measure of central tendency for an open-end classification.

◆ Mode is the most popular measure of central tendency, there are cases when mode remains undefined. Unlike mean, it has no mathematical property. Mode is also affected by sampling fluctuations.

◆ Relationship between Mean, Median and Mode

Mean – Mode = 3(Mean – Median)

Mode = 3 Median – 2 Mean

◆ Relation between AM, GM, and HM

$$AM \geq GM \geq HM$$

◆ GM and HM, like AM, possess some mathematical properties. They are rigidly defined and based on all the observations. But they are difficult to comprehend and compute and, as such, have limited applications for the computation of average rates and ratios and such like things.

EXERCISE — UNIT-I

Set A

Write down the correct answers. Each question carries 1 mark.

1. Measures of central tendency for a given set of observations measures
   (a) The scatterness of the observations
   (b) The central location of the observations
   (c) Both (a) and (b)
   (d) None of these.

2. While computing the AM from a grouped frequency distribution, we assume that
   (a) The classes are of equal length
   (b) The classes have equal frequency
   (c) All the values of a class are equal to the mid-value of that class
   (d) None of these.

3. Which of the following statements is wrong?
   (a) Mean is rigidly defined
   (b) Mean is not affected due to sampling fluctuations
   (c) Mean has some mathematical properties
   (d) All these

4. Which of the following statements is true?
   (a) Usually mean is the best measure of central tendency
   (b) Usually median is the best measure of central tendency
   (c) Usually mode is the best measure of central tendency
   (d) Normally GM is the best measure of central tendency

5. For open-end classification, which of the following is the best measure of central tendency?
   (a) AM   (b) GM   (c) Median   (d) Mode

6. The presence of extreme observations does not affect
   (a) AM   (b) Median   (c) Mode   (d) Any of these.

7. In case of an even number of observations which of the following is median?
   (a) Any of the two middle-most value

(b) The simple average of these two middle values

(c) The weighted average of these two middle values

(d) Any of these

8. The most commonly used measure of central tendency is

(a) AM   (b) Median   (c) Mode   (d) Both GM and HM.

9. Which one of the following is not uniquely defined?

(a) Mean   (b) Median   (c) Mode   (d) All of these measures

10. Which of the following measure of the central tendency is difficult to compute?

(a) Mean   (b) Median   (c) Mode   (d) GM

11. Which measure(s) of central tendency is(are) considered for finding the average rates?

(a) AM   (b) GM   (c) HM   (d) Both (b) and (c)

12. For a moderately skewed distribution, which of he following relationship holds?

(a) Mean – Mode = 3 (Mean – Median)   (b) Median – Mode = 3 (Mean – Median)

(c) Mean – Median = 3 (Mean – Mode)   (d) Mean – Median = 3 (Median – Mode)

13. Weighted averages are considered when

(a) The data are not classified

(b) The data are put in the form of grouped frequency distribution

(c) All the observations are not of equal importance

(d) Both (a) and (c).

14. Which of the following results hold for a set of distinct positive observations?

(a) $AM \geq GM \geq HM$ (b) $HM \geq GM \geq AM$

(c) $AM > GM > HM$ (d) $GM > AM > HM$

15. When a firm registers both profits and losses, which of the following measure of central tendency cannot be considered?

(a) AM   (b) GM   (c) Median   (d) Mode

16. Quartiles are the values dividing a given set of observations into

(a) Two equal parts   (b) Four equal parts (c) Five equal parts   (d) None of these

17. Quartiles can be determined graphically using

(a) Histogram   (b) Frequency Polygon (c) Ogive   (d) Pie chart.

18. Which of the following measure(s) possesses (possess) mathematical properties?

(a) AM   (b) GM   (c) HM   (d) All of these

19. Which of the following measure(s) satisfies (satisfy) a linear relationship between two variables?

    (a) Mean   (b) Median   (c) Mode   (d) All of these

20. Which of he following measures of central tendency is based on only fifty percent of the central values?

    (a) Mean   (b) Median   (c) Mode   (d) Both (a) and (b)

Set B

Write down the correct answers. Each question carries 2 marks.

1. If there are 3 observations 15, 20, 25 then the sum of deviation of the observations from their AM is

    (a) 0   (b) 5   (c) –5   (d) None of these.

2. What is the median for the following observations?

    5, 8, 6, 9, 11, 4.

    (a) 6   (b) 7   (c) 8   (d) None of these

3. What is the modal value for the numbers 5, 8, 6, 4, 10, 15, 18, 10?

    (a) 18   (b) 10   (c) 14   (d) None of these

4. What is the GM for the numbers 8, 24 and 40?

    (a) 24   (b) 12   (c) $8 \cdot \sqrt[3]{15}$ (d) 10

5. The harmonic mean for the numbers 2, 3, 5 is

    (a) 2.00   (b) 3.33   (c) 2.90   (d) $-\sqrt[3]{30}$.

6. If the AM and GM for two numbers are 6.50 and 6 respectively then the two numbers are

    (a) 6 and 7   (b) 9 and 4   (c) 10 and 3   (d) 8 and 5.

7. If the AM and HM for two numbers are 5 and 3.2 respectively then the GM will be

    (a) 16.00   (b) 4.10   (c) 4.05   (d) 4.00.

8. What is the value of the first quartile for observations 15, 18, 10, 20, 23, 28, 12, 16?

    (a) 17   (b) 16   (c) 12.75   (d) 12

9. The third decile for the numbers 15, 10, 20, 25, 18, 11, 9, 12 is

    (a) 13   (b) 10.70   (c) 11   (d) 11.50

10. If there are two groups containing 30 and 20 observations and having 50 and 60 as arithmetic means, then the combined arithmetic mean is

    (a) 55   (b) 56   (c) 54   (d) 52.

11. The average salary of a group of unskilled workers is ₹ 10,000 and that of a group of skilled workers is ₹ 15,000. If the combined salary is ₹ 12,000, then what is the percentage of skilled workers?

   (a) 40%   (b) 50%   (c) 60%   (d) none of these

12. If there are two groups with 75 and 65 as harmonic means and containing 15 and 13 observation then the combined HM is given by

   (a) 65   (b) 70.36   (c) 70   (d) 71.

13. What is the HM of 1,1/2, 1/3,……………1/n?

   (a) n   (b) 2n   (c) $\dfrac{2}{(n+1)}$ (d) $\dfrac{n(n+1)}{2}$

14. An aeroplane flies from A to B at the rate of 500 km/hour and comes back from B to A at the rate of 700 km/hour. The average speed of the aeroplane is

   (a) 600 km. per hour   (b) 583.33 km. per hour

   (c) $100 \sqrt{35}$ km. per hour   (d) 620 km. per hour.

15. If a variable assumes the values 1, 2, 3…5 with frequencies as 1, 2, 3…5, then what is the AM?

   (a) $\dfrac{11}{3}$ (b) 5   (c) 4   (d) 4.50

16. Two variables x and y are given by y= 2x – 3. If the median of x is 20, what is the median of y?

   (a) 20   (b) 40   (c) 37   (d) 35

17. If the relationship between two variables u and v are given by 2u + v + 7 = 0 and if the AM of u is 10, then the AM of v is

   (a) 17   (b) –17   (c) –27   (d) 27.

18. If x and y are related by x–y–10 = 0 and mode of x is known to be 23, then the mode of y is

   (a) 20   (b) 13   (c) 3   (d) 23.

19. If GM of x is 10 and GM of y is 15, then the GM of xy is

   (a) 150   (b) log 10 × log 15   (c) log 150   (d) None of these.

20. If the AM and GM for 10 observations are both 15, then the value of HM is

   (a) Less than 15   (b) More than 15

   (c) 15   (d) Can not be determined.

Set C

Write down the correct answers. Each question carries 5 marks.

1. What is the value of mean and median for the following data:

Marks:	5–14	15–24	25–34	35–44	45–54	55–64
No. of Students:	10	18	32	26	14	10

    (a) 30 and 28　　　　(b) 29 and 30　(c) 33.68 and 32.94　(d) 34.21 and 33.18

2. The mean and mode for the following frequency distribution

Class interval:	350–369	370–389	390–409	410–429	430–449	450–469
Frequency:	15	27	31	19	13	6

    are

    (a) 400 and 390　　　　(b) 400.58 and 390　(c) 400.58 and 394.50　(d) 400 and 394.

3. The median and modal profits for the following data

Profit in '000 ₹:	below 5	below 10	below 15	below 20	below 25	below 30
No. of firms:	10	25	45	55	62	65

    are

    (a) 11.60 and 11.50　　　　　(b) ₹ 11556 and ₹ 11267

    (c) ₹ 11875 and ₹ 11667　　　　(d) 11.50 and 11.67.

4. Following is an incomplete distribution having modal mark as 44

Marks:	0–20	20–40	40–60	60–80	80–100
No. of Students:	5	18	?	12	5

    What would be the mean marks?

    (a) 45　　　　　　(b) 46　　　　　　(c) 47　　　　　　(d) 48

5. The data relating to the daily wage of 20 workers are shown below:

    ₹ 50, ₹ 55, ₹ 60, ₹ 58, ₹ 59, ₹ 72, ₹ 65, ₹ 68, ₹ 53, ₹ 50, ₹ 67, ₹ 58,
    ₹ 63, ₹ 69, ₹ 74, ₹ 63, ₹ 61, ₹ 57, ₹ 62, ₹ 64.

    The employer pays bonus amounting to ₹ 100, ₹ 200, ₹ 300, ₹ 400 and ₹ 500 to the wage earners in the wage groups 50-55, 55-60 and so on, and lastly 70-75, during the festive month of October.

    What is the average bonus paid per wage earner?

    (a) ₹ 200　　　　　(b) ₹ 250　　　　　(c) ₹ 285　　　　　(d) ₹270

6. The third quartile and 65th percentile for the following data are

Profits in '000 ₹:	les than 10	10–19	20–29	30–39	40–49	50–59
No. of firms:	5	18	38	20	9	2

(a) ₹ 33,500 and ₹ 29,184
(b) ₹ 33,000 and ₹ 28,680
(c) ₹ 33,600 and ₹ 29,000
(d) ₹ 33,250 and ₹ 29,250.

7. For the following incomplete distribution of marks of 100 pupils, median mark is known to be 32.

Marks:	0–10	10–20	20–30	30–40	40–50	50–60
No. of Students:	10	–	25	30	–	10

What is the mean mark?

(a) 32   (b) 31   (c) 31.30   (d) 31.50

8. The mode of the following distribution is ₹ 66. What would be the median wage?

Daily wages (₹):	30–40	40–50	50–60	60–70	70–80	80–90
No of workers:	8	16	22	28	–	12

(a) ₹ 64.00   (b) ₹ 64.56   (c) ₹ 63.21   (d) ₹ 64.25

ANSWERS

Set A

1.	(b)	2.	(c)	3.	(b)	4.	(a)	5.	(c)	6.	(b)
7.	(b)	8.	(a)	9.	(c)	10.	(d)	11.	(d)	12.	(a)
13.	(b)	14.	(c)	15.	(b)	16.	(b)	17.	(c)	18.	(d)
19.	(d)	20.	(b)								

Set B

1.	(a)	2.	(b)	3.	(b)	4.	(c)	5.	(c)	6.	(b)
7.	(d)	8.	(c)	9.	(b)	10.	(c)	11.	(a)	12.	(c)
13.	(c)	14.	(b)	15.	(a)	16.	(c)	17.	(c)	18.	(b)
19.	(a)	20.	(c)								

Set C

1.	(c)	2.	(c)	3.	(c)	4.	(d)	5.	(c)	6.	(a)
7.	(c)	8.	(c)								