CHAPTER

14 Measures of Central Tendency and Dispersion

Unit-1: Measures of Central Tendency

WHAT IS CENTRAL TENDENCY?

Let's take an example, if you talk about the weight or height of students in your class, what would you observe?

As we can see the majority of data is tending to cluster towards the center which we call the central tendency and the value which represents the center is called as measure of central tendency.

In the simplest cases, the measure of central tendency is an average of a set of measurements.

That's why, when our marks in exams come – we see average marks of students or when we go to work – we see the average salary of employees.

Central Tendency refers to the tendency of data to cluster around a central or typical value. It represents the value that best represents the center of a distribution. In other words, it gives us an idea of where most of the data points are concentrated.

LET'S SEE SOME DIFFERENT MEASURES OF CENTRAL TENDENCY

- **Arithmetic Mean (AM)**: It is also known as the average, which is the sum of all the values divided by the total number of values.
- **Median (Me):** It is the middle value in a sorted dataset. It divides the dataset into two equal halves, with half of the values below and half above the median.
- **Mode (Mo):** It is the value that occurs most frequently in a dataset. It represents the value that appears with the highest frequency.
- **Geometric Mean (GM):** It is the nth root of the product of n values. It is often used for calculating growth rates or when dealing with exponential data.
- **Harmonic Mean (HM):** It is the reciprocal of the arithmetic mean of the reciprocals of a set of values. It is useful when dealing with rates, ratios, or averages of rates.

  Now, let's understand all the above measures of central tendency in detail.

ARITHMETIC MEAN

Arithmetic Mean (Mean) = $\frac{\text{Sum of observations}}{\text{Number of observations}}$

Let's variable x has n values: $x_1, x_2, x_3, x_4, x_5, \ldots\ldots\ldots x_n$, then Arithmetic Mean is denoted by $\bar{x}$, which is equal to $\frac{x_1 + x_2 + \ldots.. + x_n}{n}$

Example 1. The following figures give the marks of 10 students in a class test:
Marks obtained: 12, 8, 17, 13, 15, 9, 18, 11, 6, 1. Find the arithmetic mean.

(a) 8 (b) 9 (c) 10 (d) 11

Sol. (d) To find Arithmetic Mean,

As we know, Arithmetic Mean is given by the formula,

$$A.M = \frac{\text{Sum of all observations}}{\text{Number of observations}}$$

Given Observations: 12 8 17 13 15 9 18 11 6 1

Here, number of observations = 10

$$\text{Arithmetic Mean (A.M)} = \frac{12+8+17+13+15+9+18+11+6+1}{10} = \frac{110}{10} = 11$$

Hence, the correct option is (d).

Example 2. Following are the Monthly salary in multiple of 1000 INR of a sample of 10 employees: 41, 32, 45, 65, 67, 78, 90, 75, 33, 44. Compute the mean wage (in 1000 INR).

(a) 40 (b) 45 (c) 57 (d) 59

Sol. (c) As per the definition,

$$A.M = \frac{x_1 + x_2 + \ldots.. + x_n}{n} \text{ or } A.M = \frac{\text{Sum of all observations}}{\text{Number of observations}}$$

Given Observations: 41, 32, 45, 65, 67, 78, 90, 75, 33, 44

Here, number of observations = 10

According to the formula, Arithmetic mean (A.M). is given by

$$A.M = \frac{x_1 + x_2 + \ldots.. + x_n}{n}$$

$$\bar{x} = \frac{41+32+45+65+67+78+90+75+33+44}{10}$$

$$\bar{x} = \frac{570}{10}$$

$$\bar{x} = 57$$

Hence, the correct option is (c).

ANOTHER METHODS TO FIND MEAN

1. **Direct method:** If $x_1, x_2, x_3, \ldots.. x_n$ be n observations with respective frequencies $f_1, f_2, f_3, \ldots.. f_n$, then

$$\text{Mean} = \frac{f_1x_1 + f_2x_2 + \ldots\ldots + f_nx_n}{f_1 + f_2 + \ldots\ldots + f_n} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

2. **Step-deviation method:** When the data values are quite large, the step-deviation method is used to find the mean.

$$\text{Mean} = a + \frac{\sum_{i=1}^{n} f_i u_i}{\sum_{i=1}^{n} f_i} \times h$$

where, a = assumed mean

x_i = class-marks = $\dfrac{\text{upper limit class + Lower limit class}}{2}$

$d_i = x_i - a$ = deviation from i^{th} class

$u_i = \dfrac{x_i - a}{h}$ and h = length of class

Example 3. In a test of 10 marks, 100 students get marks as below:

Marks obtained	Number of Students
1	2
2	3
3	5
4	10
5	30
6	10
7	18
8	12
9	6
10	4

What is the average marks obtained?

(a) 4.89 (b) 5.89 (c) 5.79 (d) 6.99

Sol. (b) Let Marks obtained by students be x_i and frequency be represented as f_i

Let the assumed mean, a = 5

Using Shortcut method,

Marks obtained (x_i)	Number of students (f_i)	$x_i \times f_i$
1	2	2
2	3	6
3	5	15
4	10	40
5	30	150
6	10	60
7	18	126
8	12	96
9	6	54
10	4	40
	$\Sigma f_i = 100$	$\Sigma = 589$

Now, Arithmetic mean will be given as,

$$\text{A.M} = \frac{589}{100} = 5.89$$

Hence, the correct option is (b).

Example 4. The following data give the daily earnings (in ₹) of 20 workers in a factory:

Daily earnings (in ₹):	100	140	170	200	250
No. of workers:	5	2	6	4	3

Calculate the average daily earnings using:

(I) Direct Method

(II) Step-deviation Method.

(a) 167.5 (b) 165.85 (c) 510.70 (d) 166.5

Sol. (I) As per the definition direct method is given as,

$$A.M = \frac{x_1 + x_2 + \ \ldots\ + x_n}{n} \text{ or } A.M = \frac{\text{Sum of all observations}}{\text{Number of observations}}$$

Number of observations = 20

Observations = 100, 100, 100, 100, 100, 140, 140, 170, 170, 170, 170, 170, 170, 200, 200, 200, 250, 250, 250

Now, Arithmetic Mean (A.M) will be,

$$A.M = \frac{5\times100+2\times140+6\times170+4\times200+3\times250}{20} = \frac{3350}{20} = 167.5$$

(II) As per the definition, step-deviation method is given as,

$$\text{Mean} = a + \frac{\sum_{i=1}^{n} f_i u_i}{\sum_{i=1}^{n} f_i} \times h$$

Daily Earnings (x_i)	Number of workers (f_i)	$x_i - a$	$u_i = \frac{x_i - a}{h}$ where, h = 10	$f_i \times u_i$
100	5	100 − 170 = −70	−7	−35
140	2	140 − 170 = −30	−3	−6
170	6	170 − 170 = 0	0	0
200	4	200 − 170 = 30	3	12
250	3	250 − 170 = 80	8	24
	$\Sigma f_i = 20$			$\Sigma f_i \times u_i = -5$

$$\text{Mean} = a + \frac{\sum_{i=1}^{n} f_i u_i}{\sum_{i=1}^{n} f_i} \times h = 170 + \frac{(-5)}{20} \times 10 = 170 - 2.5 = 167.5$$

Hence, the correct option is (a).

Example 5. Compute the mean weight of your friends' group:

Weight in kg	41 − 45	46 − 50	51 − 55	56 − 60	61 − 65	66 − 70
No. of friends:	4	3	9	8	7	2

(a) 55 (b) 54.2727 (c) 56.7272 (d) 59.2882

Sol. (b) Represent the given data in tabular form,

Weight in kg	Number of friends (f_i)	Mid Points (x_i)	$f_i \times x_i$
41 – 45	4	43	4 × 43 = 129
46 – 50	3	48	3 × 48 = 144
51 – 55	9	53	9 × 53 = 477
56 – 60	8	58	8 × 58 = 464
61 – 65	7	63	7 × 63 = 441
66 – 70	2	68	2 × 68 = 136
	$\Sigma f_i = 33$		$\Sigma f_i x_i = 1791$

Now, Mean will be given as, $\bar{x}=\dfrac{\Sigma f_i x_i}{\Sigma f_i}$

$\bar{x}=\dfrac{1791}{33} = 54.2727$

Hence, the correct option is (b).

Example 6. Given below is the distribution of marks obtained by 140 students in an examination:

Marks:	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99
No. of students:	7	15	18	25	30	20	16	7	2

(a) 67.5 (b) 56.85 (c) 51.070 (d) 50.714

Sol. (d) Represent the given data in tabular form,

Marks	Number of Students (f_i)	Mid Points (x_i)	$f_i \times x_i$
10 - 19	7	14.5	7 × 14.5 = 101.5
20 - 29	15	24.5	15 × 24.5 = 367.5
30 - 39	18	34.5	18 × 34.5 = 621
40 - 49	25	44.5	25 × 44.5 = 1112.5
50 - 59	30	54.5	30 × 54.5 = 1635
60 - 69	20	64.5	20 × 64.5 = 1290
70 - 79	16	74.5	16 × 74.5 = 1192
80 - 89	7	84.5	7 × 84.5 = 591.5
90 - 99	2	94.5	2 × 94.5 = 189
	$\Sigma f_i = 140$		$\Sigma f_i x_i = 7100$

Now, Mean will be given as, $\bar{x} = \dfrac{\Sigma f_i x_i}{\Sigma f_i}$

$\bar{x}=\dfrac{7100}{140}$

$\bar{x} = 50.714$

Hence, the correct option is (d).

Example 7. Calculate mean from the following data:

Marks	No. of students
0 – 10	4
10 – 20	16
20 – 30	40
30 – 40	76
40 – 50	96
50 – 60	112
60 – 70	120
70 – 80	125

(a) 50.20 (b) 53.60 (c) 51.70 (d) 66.5

Sol. (b) According to the given data, we get the following:

Marks	Number of students (f_i)	Mid Points (x_i)	$f_i \times x_i$
0–10	4	5	4 × 5 = 20
10 – 20	16	15	16 × 15 = 240
20 – 30	40	25	40 × 25 = 1000
30 – 40	76	35	76 × 35 = 2660
40 – 50	96	45	96 × 45 = 4320
50 – 60	112	55	112 × 55 = 6160
60 – 70	120	65	120 × 65 = 7800
70 – 80	125	75	125 × 75 = 9375
	$\Sigma f_i = 589$		$\Sigma f_i x_i = 31575$

Now, Mean will be given as,

$$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i}$$

$$\bar{x} = \frac{31575}{589}$$

$\bar{x} = 53.607 \approx 53.60$

Hence, the correct option is (b).

Example 8. The mean of the following frequency distribution is 50 but the frequencies f_1 and f_2 in classes 20 – 40 and 60 – 80 are missing. Find the missing frequencies.

Class:	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100	Total
Frequency:	17	f_1	32	f_2	19	120

(a) 28, 24 (b) 26, 26 (c) 22, 30 (d) 32, 20

Sol. (a) Given: Mean $(\bar{x})$ = 50

$\Sigma f_i = 120$

Represent the data in tabular form,

Classes	Frequency (f_i)	Mid Points (x_i)	$f_i \times x_i$
0 – 20	17	10	$17 \times 10 = 170$
20 – 40	f_1	30	$f_1 \times 30 = 30 f_1$
40 – 60	32	50	$32 \times 50 = 1600$
60 – 80	f_2	70	$f_2 \times 70 = 70 f_2$
80 – 100	19	90	$19 \times 90 = 1710$
	$\Sigma f_1 = 68 + f_1 + f_2$		$\Sigma f_i d_i = 3480 + 30f_1 + 70f_2$

According to the question, $\Sigma f_i = 120$

$\Rightarrow 120 = 68 + f_1 + f_2$

$\Rightarrow f_1 + f_2 = 52$ ...(i)

Now, Mean will be given as, $\bar{x} = \dfrac{\Sigma f_i x_i}{\Sigma f_i}$

$$50 = \frac{3480 + 30f_1 + 70f_2}{120}$$

$6000 = 3480 + 30f_1 + 70f_2$

On simplifying it, we get

$3f_1 + 7f_2 = 252$ ...(ii)

On solving, multiply equation (i) by 3 and subtract (i) and (ii), we get

$3f_1 + 3f_2 = 156$

$3f_1 + 7f_2 = 252$

$-\quad -\qquad -$

$-4f_2 = -96$

$$f_2 = \frac{96}{4}$$

$f_2 = 24$

Put $f_2 = 24$ in (i), we get

$f_1 + 24 = 52$

$f_1 = 52 - 24$

$f_1 = 28$

So, $f_1 = 28$, $f_2 = 24$

Hence, the correct option is (a).

PROPERTIES OF ARITHMETIC MEAN

1. If all the observation of the dataset are constants, let 'a', then the arithmetic mean will also be equal to 'a'.

   E.g.: If the weight of each student of group of 6 students of grade 10 is 60 kg, then the mean weight will also be 60 kg.

2. The algebraic sum of deviations of the given set of observations from their arithmetic mean is zero.

   That is, $(x - \bar{x}) = 0$ or for a frequency distribution $\Sigma f(x - \bar{x}) = 0$

   E.g.: For the given set of observations 2, 3, 4, 5, 6 then, $\bar{x} = \frac{2+3+4+5+6}{5} = 4$

   Thus, $\Sigma(x - \bar{x}) = (2-4) + (3-4) + (4-4) + (5-4) + (6-4)$

   $= -2 - 1 + 0 + 1 + 2 = 0$

   i.e., $\Sigma(x - \bar{x}) = 0$

3. If each observation of a dataset is increased or decreased by a constant 'k', then the arithmetic mean of new data will also gets increased or decreased by the same constant 'k'.

   E.g.: The arithmetic mean of 5 observations: 5, 10, 15, 20 and 25 is 15. If each observations is increased by 3, then the new mean will also increase by 3 i.e., new arithmetic mean = 15 + 3 = 18

4. AM is affected with the change of origin and scale i.e., if the variable 'x' is changed to a new variable 'y' by a change of origin 'a' and scale 'b'

   $\Rightarrow y = a + bx$, then

   $AM_y = a + b \times AM_x$ or $\bar{y} = a + b\bar{x}$

   E.g.: If x and y are related by the equation $3x + y = 6$ and $\bar{x} = 1$ then

   $\bar{y} = 6 - 3\bar{x}$

   $\Rightarrow \bar{y} = 6 - 3(1) \Rightarrow \bar{y} = 3$

5. **Combined mean:** If the arithmetic means of two or more sets of data are known, then the arithmetic mean of the combined data can be obtained.

   If n_1 and n_2 are the number of observations and $\bar{x}_1$ and $\bar{x}_2$ are their respective means of two sets of data then the combined mean with observations is given by $\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$

Example 9. The sum of deviations of a certain number of observations measured from 4 is 72 and the sum of observations of the same value from 7 is –3. Find the number of observations and their mean.

(a) 24, 7.5 (b) 24, 6.77 (c) 25, 6.88 (d) 25, 6.5

Sol. (c) Given: Sum of deviations from 4 = 72

Sum of observations of the same value from 7 = –3

To find: Number of observations and Mean

Let n be the number of observations,

According to the question,

$(x_1 - 4) + (x_2 - 4) + ... + (x_n - 4) = 72$

and, $(x_1 - 7) + (x_2 - 7) + ... + (x_n - 7) = 3$

So, $(x_1 + x_2 + ... + x_n) - 4n = 72$

$\Rightarrow (x_1 + x_2 + ... + x_n) = 72 + 4n$

$\Rightarrow \Sigma x_1 = 72 + 4n$ ...(i)

Also, $(x_1 - 7) + (x_2 - 7) + ... + (x_n - 7) = -3$

$\Rightarrow (x_1 + x_2 + ... + x_n) - 7n = -3$

$\Rightarrow (x_1 + x_2 + ... + x_n) = -3 + 7n$

$\Rightarrow \Sigma x_i = 7n - 3$ ...(ii)

On equating (i) and (ii), we get

$\Rightarrow 72 + 4n = 7n - 3$

$\Rightarrow 7n - 4n = 72 + 3$

$\Rightarrow 3n = 75$

$\Rightarrow n = 25$

Now, Mean is given as,

$$\text{Mean} = \Sigma \frac{x_i}{n}$$

Put n = 25 in order to find mean

$$= \frac{7n-3}{n} = \frac{7(25)-3}{25} = \frac{172}{25} = 6.88$$

Hence, the correct option is (c).

Example 10. The arithmetic mean of a set of 5 observations 5, 10, 15, 20 and 25 is 15 However, if each item is increased by 3, then the arithmetic mean will be

(a) 18 (b) 18.5 (c) 19 (d) 19.5

Sol. (a) **Detailed method:** According to the question,

Old Observations: 5, 10, 15, 20, 25

Mean = 1.5

If each observation is increased by 3, then

New observations: 8, 13, 18, 23, 28

Mean is given by the formula, $\text{Mean} = \frac{\Sigma x_i}{n}$

$$\Rightarrow \text{Mean} = \frac{8+13+18+23+28}{5}$$

$$\Rightarrow \text{Mean} = \frac{90}{5}$$

$\Rightarrow$ Mean = 18

Short-cut method: Since, the arithmetic mean of the given set of observations is 15. Now, each observation is increased by 3, thus the mean will also increase by 3.

Therefore, new mean = 15 + 3 = 18

Hence, the correct option is (a).

Example 11. The arithmetic mean of a set of 5 observations: 2, 4, 6, 8, 10 is 6. If each item is multiplied by 2, then the new arithmetic mean will be?

(a) 18 (b) 12 (c) 15 (d) 20

Sol. (b) **Detailed method:** According to the question,

Old Observations: 2, 4, 6, 8, 10

Mean = 6

If each observation is multiplied by 2, then

New observations: 4, 8, 12, 16, 20

Mean is given by the formula, Mean $= \frac{\sum x_i}{n}$

$\Rightarrow$ Mean $= \frac{4 + 8 + 12 + 16 + 20}{5}$

$\Rightarrow$ Mean $= \frac{60}{5}$

$\Rightarrow$ Mean = 12

Short-cut method: Since, the arithmetic mean of the given set of observations is 6.

Now, each observation gets multiplied by 2, thus the mean will also be multiplied by 2.

Therefore, new mean = 6 × 2 = 12

Hence, the correct option is (b).

Example 12. If it is known that 2 variables x and y are related by equation 2x + 3y = 5 and x = 1, then y is

(a) 1 (b) 0 (c) 5 (d) 2

Sol. (a) Given equation: 2x + 3y = 5; x = 1

Put x = 1 in the above equation, we get

2(1) + 3y = 5

$\Rightarrow$ 2 + 3y = 5

$\Rightarrow$ 3y = 3 $\Rightarrow$ y = 3

Hence, the correct option is (a).

Example 13. The mean salary for a group of 40 female workers is ₹52000 per month and that for a group of 50 male workers is ₹68000 per month. What is the combined mean salary?

(a) ₹65000 (b) ₹60888.88 (c) ₹51070 (d) ₹16650

Sol. (b) Given,

Number of Females (n_1) = 40

Mean salary of Females ($\bar{x}_1$) = ₹52000

Number of Males (n_2) = 50

Mean Salary of Males ($\bar{x}_2$) = ₹68000

We know, $\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$

$\therefore$ Combined Salary $\bar{x} = \dfrac{(40)(52000)+(50)(68000)}{40+50} = \dfrac{(2080000)+(3400000)}{90}$

$= 60{,}888.88$

Hence, the correct option is (b).

Example 14. A distribution consists of three components with total frequencies of 200, 250 and 300 having means 25, 10 and 15 respectively. Find the mean of the combined distribution.

(a) 4 (b) 8 (c) 12 (d) 16

Sol. (d) Tabulate the data,

Let the frequencies be f_1 and the mean be x_i

f_i	x_i	$f_i \times x_i$
200	25	$200 \times 25 = 5000$
250	10	$250 \times 10 = 2500$
300	15	$300 \times 15 = 4500$
$\Sigma f_i = 750$		$\Sigma f_i \times x_i = 12000$

Now, Mean of Combined distribution will be given as,

$= \dfrac{\Sigma f_i x_i}{\Sigma f_i} = \dfrac{12000}{750} = 16$

Hence, the correct option is (d).

Example 15. The mean weight of 150 students (boys and girls) in a class is 60 kg. The mean weight of boy students is 70 kg and that of girl students is 55 kg. The number of boys and girls respectively in the class are

(a) 50, 100 (b) 100, 50 (c) 120, 30 (d) 30, 120

Sol. (a) Let the number of boys be denoted as x and the number of girls be denoted as y.

As we know, Mean is given by the formula,

$\text{Mean} = \dfrac{\text{Sum of all observations}}{\text{Number of observations}}$

Thus, Sum of all observations = Mean × Number of observations

Total number of students = x + y

$\therefore x + y = 150$

$\Rightarrow x = 150 - y$ ...(i)

Now, Sum of weight of boys = '70x' kg and Sum of weight of girls = '55y' kg

Total weight of students (including boys and girls) = 150 × 60 = 9000 kg

$\Rightarrow 70x + 55y = 9000$ ...(ii)

Now, on solving (i) and (ii), we get

$70(150 - y) + 55y = 9000$

$\Rightarrow 10500 - 70y + 55y = 9000$

$\Rightarrow 15y = 1500$

$\Rightarrow y = 100$

Put $y = 100$ in (i), we get

$\Rightarrow x = 150 - 100$

$\Rightarrow x = 50$

So, number of Boys = 50 and number of girls = 100

Hence, the correct option is (a).

Example 16. Fifty students took up a test. The result of those who passed the test is given below:

Marks:	4	5	6	7	8	9
Number of Students:	8	10	9	6	4	3

If the average for all 50 students was 5.16 marks, find the average of those who failed.

(a) 2.1 (b) 3.9 (c) 1.1 (d) 3.3

Sol. (a) According to the question,

Mean marks of 50 students = 5.16

Tabulate the data,

Let the marks be f_i and the number of students be x_i

f_i	x_i	$f_i \times x_i$
4	8	$4 \times 8 = 32$
5	10	$5 \times 10 = 50$
6	9	$6 \times 9 = 54$
7	6	$7 \times 6 = 42$
8	4	$8 \times 4 = 32$
9	3	$9 \times 3 = 27$

As per the question,

Let the average mark of failed student be y.

Let the number of failed students be F.

Now, number of failed students will be given by,

$\Rightarrow 50 - (8 + 10 + 9 + 6 + 4 + 3) = F$

$\Rightarrow F = 50 - 40$

$\Rightarrow F = 10$

So, According to the question,

$$\Rightarrow \frac{(32 + 50 + 54 + 42 + 32 + 27 + 10y)}{50} = 5.16$$

$\Rightarrow 237 + 10y = 258$

$\Rightarrow 10y = 21$

$\Rightarrow y = 2.1$

Hence, the correct option is (a).

CORRECTING INCORRECT MEAN

Example 17. The mean marks of 100 students were found to be 40. Later on, it was discovered that a score of 53 was misread as 83. Find the correct mean corresponding to the correct score.

(a) 67.5 (b) 39.7 (c) 51.70 (d) 66.5

Sol. (b) According to the question,

Initial mean (Incorrect mean) = 40

Number of observations (n) = 100

Wrong observation (x_w) = 83 and Right observation (x_r) = 53

Now, Incorrect mean, $\bar{x}_{in} = \frac{\sum x_w}{n}$

$\Rightarrow 40 = \frac{\sum x_w}{100}$

$\Rightarrow \sum x_w = 4000$ ...(i)

Now, $\sum x_r = \sum x_w - x_w + x_r = 4000 - 83 + 53 = 3970$

Thus, Correct mean, $\bar{x}_c = \frac{\sum x_r}{n} = \frac{3970}{100} = 39.7$

Hence, the correct option is (b).

> **Note:** We can use the direct formula for the above:
>
> i.e., $\bar{x}_c = \bar{x}_{in} + \frac{x_c - x_{in}}{n}$,
>
> where, $\bar{x}_c$ represents the correct mean, $\bar{x}_{in}$ is the incorrect mean, x_c is the correct observation and x_{in} is the incorrect observation.

Example 18. The mean salary paid to 1000 employees of an establishment was found to be ₹180.40. Later on, after disbursement of salary, it was discovered that the salary of the two employees was wrongly entered as ₹297 and ₹165. Their correct salaries were ₹197 and ₹185. Find the correct mean salary.

(a) 168.5 (b) 179.85 (c) 510.70 (d) 180.32

Sol. (d) According to the question,

Initial mean (Incorrect mean) = 180.40

Number of employees (n) = 1000

Wrong observations (x_w) = 297, 165 and Right observations (x_r) = 197, 185

Now, Incorrect mean, $\bar{x}_{in} = \frac{\sum x_w}{n}$

$\Rightarrow 180.40 = \frac{\sum x_w}{1000}$

$\Rightarrow \sum x_w = 180400$ ...(i)

Now, $\sum x_r = \sum x_w - x_w + x_r = 180400 - 297 - 165 + 197 + 185 = 180230$

Thus, Correct mean, $\bar{x}_c = \frac{\sum x_r}{n} = \frac{180320}{1000} = 180.320$

Hence, the correct option is (d).

WEIGHTED ARITHMETIC MEAN

Let $w_1, w_2, \ldots\ldots w_n$ be the weights attached to n observations $x_1, x_2, \ldots x_n$ respectively then the weighted arithmetic mean, denoted by $\bar{x}w$ is given by

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

Example 19. A candidate obtained the following percentage of marks in different subjects in the Half-yearly Examination: English 46%, Statistics 67%, Cost Accountancy 72%, Economics 58%, Income Tax 53%. It is agreed to give double weights to marks in English and Statistics as compared to other subjects. What is the simple and weighted arithmetic mean?

(a) 56.4, 59.42 (b) 58.42, 59.2 (c) 59.2, 58.42 (d) None of these

Sol. (a) Let N be the number of subjects i.e., 5.

Tabulate the data and solve,

According to the question, it is agreed to give double weights to marks in English and Statistics as compared to other subjects

So, weights of English and Statistics = 2 and weights of other subjects = 1

Subject	Marks (x)	Weight (w)	w × x
English	46	2	92
Statistics	67	2	134
Cost Accountancy	72	1	72
Economics	58	1	58
Income Tax	53	1	53
	$\Sigma x = 296$	$\Sigma w = 7$	$\Sigma wx = 409$

So, Simple Mean will be given as

$$\bar{x} = \frac{\sum x}{N} = \frac{296}{5} = 59.2$$

Now, weighted arithmetic mean will be given by the formula,

$$\overline{X_w} = \frac{\sum wx}{\sum w} = \frac{409}{7} = 58.42$$

Hence, the correct option is (c).

Example 20. 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) One of these

Sol. (b) We know that,

Measures of central tendency for a given set of observations measures **the central location of the observations.**

Measures of central tendency is a mathematical concept that measures the central most value of a given series which can represent the whole set of the series. In the given series, the central location of the observation is measured through different tools in order to ascertain the most efficient value that can represent the series.

Hence, the correct option is (b).

Example 21. 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.

Sol. (c) While computing the AM from a grouped frequency distribution, we assume that all the values of a class are equal to the mid-value of that class.

Hence, the correct option is (c).

Example 22. Which of the following statements is wrong?

(a) Mean is rigidly defined

(b) Mean is not affected due a sampling fluctuations

(c) Mean has some mathematical properties

(d) All of these

Sol. (b) Mean is not affected due to sampling fluctuations is a wrong statement.

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.

Hence, the correct option is (b).

Example 23. If there are 3 observations 15, 20, 15 then the sum of deviation of the observations from their AM is

(a) 0 (b) 5 (c) –5 (d) None of these

Sol. (a) Given observations: 15, 20, 25

Firstly, we will find AM of the given data

$$AM = \frac{15 + 20 + 25}{3}$$

$$AM = \frac{60}{3}$$

AM = 20

Thus, the deviation of the observations from their AM is

15 – 20, 20 – 20, 25 – 20 i.e. –5, 0, 5

Therefore, the sum of deviation = –5 + 0 + 5 = 0

Hence, the correct option is (a).

Example 24. 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

Sol. (c) Given: $\bar{x}_1 = 50$, $n_1 = 30$, $\bar{x}_2 = 60$ and $n_2 = 20$

The combined arithmetic mean is

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} = \frac{30(50) + 20(60)}{30 + 20}$$

$$= \frac{1500 + 1200}{50} = \frac{2700}{50} = 54$$

Hence, the correct option is (c).

Example 25. 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

Sol. (a) Let the percentage of skilled worker be x, so unskilled worker will be (100 – x).

Therefore $n_1 = x, n_2 = x, \bar{x}_i = 12000$

$\bar{x}_1 = 15000$, $\bar{x}_2 = 10000$

$$\Rightarrow \bar{x}_i = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

$$\Rightarrow 12000 = \frac{x(15000) + (100 - x)(10000)}{\{x + 100 - x\}}$$

$\Rightarrow 1200000 = 15000x + 1000000 - 10000x$

$\Rightarrow 200000 = 5000x$

$\Rightarrow x = 40\%$

Hence, the correct option is (a).

Example 26. 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 **(ICAI)**

(a) 17 (b) –17 (c) –27 (d) 27

Sol. (c) Given,

Relationship between u and v is given by

$2u + v + 7 = 0$

AM of u = 10

$\Rightarrow v = -2u - 7$

Put u = 10 in the above equation to get AM of v, i.e.

$\bar{v} = -2(10) - 7$

$= -20 - 7 = -27$

Therefore, AM of v is –27.

Hence, the correct option is (c).

PRACTICE QUESTIONS (PART A)

1. The weights of a group of individuals are recorded in kilograms: 65, 70, 75, 80, 85, 90, 95. Compute the mean weight.

(a) 75 kg (b) 80 kg (c) 85 kg (d) 90 kg

2. Following are the daily wages in thousands of a sample of workers: ₹58, ₹62, ₹48, ₹53, ₹70, ₹52, ₹60, ₹84, ₹75. Compute the mean wage.

(a) ₹60,000 (b) ₹61,740 (c) ₹62,440 (d) None of these

3. Find the mean salary of 60 workers in a factory from the following table.

Salary in ₹	Number of Workers
3000	16
4000	12
5000	10
6000	8
7000	6
8000	4
9000	3
10000	1

(a) ₹5065.50 (b) ₹6000 (c) ₹5083.33 (d) None of the above

4. If the mean of the following distribution is 6, then the value of P is

X:	2	4	6	10	P + 5
f:	3	2	3	1	2

(a) 7 (b) 5 (c) 11 (d) 8

5. The mean of 6, 4, 1, 5, 6, 10 and 3 is 5. If each number is added with 2, then the new mean is ______.

(a) 7 (b) 5 (c) 6 (d) 10

6. The mean of a set of numbers is 8. If each number is multiplied by 3, then the new mean is ______.

(a) 24 (b) 8 (c) 3 (d) 12

7. What is the value of mean for the following data: **(ICAI)**

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

(a) 30 (b) 29 (c) 33.68 (d) 34.21

8. Calculate the arithmetic mean from the following frequency distribution:

Marks	0-10	10-20	20-30	30-40	40-50	50-60
No. of students	10	9	25	30	16	10

(a) 37 (b) 38.5 (c) 31.3 (d) 30

9. Given below is the distribution of marks obtained by 140 students in an examination:

Marks	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99
No. of students	7	15	18	25	30	20	16	7	2

Find the mean of the distribution.

(a) 52.83 (b) 55.2 (c) 63.5 (d) None of these

10. Find the combined mean from the following data:

	Series X	Series Y
Arithmetic mean	12	20
Number of items	80	60

(a) 10.54 (b) 15.43 (c) 3.50 (d) None of these

11. The mean of 20 observations is 85, but it was later found that two of the observations were wrongly read as 75 and 70 instead of 57 and 60. Find the correct mean.

(a) 80.7 (b) 75.5 (c) 63.5 (d) 83.6

12. Calculate weighted mean from the following data:

Value	10	12	15	18	20
Weight	2	5	12	4	7

(a) 18.74 (b) 17.55 (c) 13.50 (d) 15.73

Answer Key

1. (b) 2. (c) 3. (c) 4. (a) 5. (a) 6. (a) 7. (c) 8. (c) 9. (a) 10. (b) 11. (d) 12. (d)

CRITERIA FOR AN IDEAL MEASURE OF CENTRAL TENDENCY

- It should be properly and unambiguously defined
- It should be easy to comprehend.
- It should be simple to compute.
- It should be based on all the observations
- It should have certain desirable mathematical properties.
- It should be least affected by the presence of extreme observations

MERITS AND DEMERITS OF ARITHMETIC MEAN

Merits

Arithmetic mean possesses the following merits

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

3. It is based on all the observations.
4. It is suitable for further mathematical treatment.
5. Of all the averages, arithmetic mean is affected least by fluctuations of sampling.

Demerits

Arithmetic mean has the following drawbacks:

1. It is very much affected by extreme values.
2. In a distribution with open-end classes the value of mean cannot be computed without making assumptions regarding the size of the class' interval of the open-end classes.
3. It can neither be determined by inspection nor can it be located graphically.
4. It cannot be computed for qualitative data such as honesty, beauty, intelligence etc.
5. It may lead to wrong conclusions if the details of the data from which it is obtained are not available.

MEDIAN

- Median, for a given set of observations, when arranged in an ascending order or a descending order of magnitude. It may be defined as the middle-most value.
- As distinct from the arithmetic mean, which is based on all the items of the distribution, the median is what is called a positional average.
- The position of the median in a distribution is such that the number of observations above it is equal to the number of observations below it.

CALCULATION OF MEDIAN: INDIVIDUAL OBSERVATIONS

For ungrouped data consisting of n observations, the calculation of median involves the following steps:

Step 1. Arrange the given set of observations in an ascending or descending order of magnitude.

Step 2. The median is given by

(a) When n is odd, the value is given by: $\left(\frac{n+1}{2}\right)$th observation

(b) When n is even, the value is given by: $\dfrac{\frac{n}{2}\text{th observation} + \left(\frac{n}{2}+1\right)\text{th observation}}{2}$

Example 27. What is the median for the following observations?

10, 2, 7, 9, 13

(a) 5 (b) 9 (c) 6 (d) 11

Sol. (b) Given data: 10, 2, 7, 9, 13

Arranging the data in increasing order, we get 2, 7, 9, 10, 13

Clearly, the number of observations is 5 i.e. odd

$\therefore$ Median $= \left(\frac{n+1}{2}\right)$th term

$= \left(\frac{5+1}{2}\right)$

= 3rd term = 9

Therefore, the median is 9.

Hence, the correct option is (b) i.e., 9.

Example 28. The following number of goals were scored by a team in a series of 10 matches: 2, 3, 4, 5, 0, 1, 3, 3, 4, 3. Find the median of these scores.

(a) 5 (b) 6 (c) 3 (d) None of these

Sol. (c) Arranging the given data in ascending order: 0, 1, 2, 3, 3, 3, 3, 4, 4, 5

Here, number of observations (n) = 10, which is even

Thus, median $= \frac{\frac{n}{2}\text{th} + \left(\frac{n}{2}+1\right)\text{th term}}{2}$

$= \frac{\frac{10}{2}\text{th} + \left(\frac{10}{2}+1\right)\text{th term}}{2}$

$= \frac{\text{5th term} + \text{6th term}}{2}$

$= \frac{3+3}{2} = \frac{6}{2} = 3$

Therefore, the median of given observations is 3.

Hence, the correct option is (c) i.e. 3.

Example 29. The median of the observations 42, 72, 35, 92, 67, 85, 72, 81, 51, 56 is

(a) 69.5 (b) 72 (c) 64 (d) 61.5 **(Dec 2022)**

Sol. (a) Given observations: 42, 72, 35, 92, 67, 85, 72, 81, 51, 56

Arranging the observations in ascending order, we get

35, 42, 51, 56, 67, 72, 72, 81, 85, 92

Here, the number of observations (n) = 10, which is even.

Thus, Median = Average of two middle terms

$\Rightarrow$ Median $= \frac{67+72}{2} = \frac{139}{2} = 69.5$

Therefore, the median of given observations is 69.5.

Example 30. The median of $x, \frac{x}{2}, \frac{x}{3}, \frac{x}{5}$ is 10. Find x, where x > 0.

(a) 24 (b) 32 (c) 8 (d) 16

Sol. (a) Given: Median = 10

Clearly, the given observations $x, \frac{x}{2}, \frac{x}{3}, \frac{x}{5}$ are even in number and in decreasing order, thus

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

$$\Rightarrow \text{Median} = \frac{\text{2nd term} + \text{3rd term}}{2}$$

$$\Rightarrow 10 = \frac{1}{2}\left(\frac{x}{2} + \frac{x}{3}\right)$$

$$\Rightarrow 20 = \frac{5x}{6}$$

$\Rightarrow 5x = 120$

$\Rightarrow x = 24$

Therefore, the value of x is 24.

Hence, the correct option is (a) i.e. 24.

CALCULATION OF MEDIAN: DISCRETE SERIES

In the case of discrete series, where the variable takes the values X_1, X_2, ... X_n with respective frequencies $f_1, f_2, \ldots f_n$ with $\Sigma f = N$, median is the size of $\left(\frac{N+1}{2}\right)$th observation.

In this case, the calculation of median involves the following steps:

Step 1: Prepare the 'less than' cumulative frequency (c. f) distribution.

Step 2: Find $\frac{N+1}{2}$

Step 3: See the c.f. just greater than or equal to $\frac{N+1}{2}$

Step 4: The value of the variable corresponding to the c.f. obtained in Step 3 gives the required median.

Example 31. For the distribution:

X	1	2	3	4	5	6
f	6	9	10	14	12	8

The value of median is

(a) 3.5 (b) 3 (c) 4 (d) 5

Sol. (c) Given data,

X	1	2	3	4	5	6
f	6	9	10	14	12	8
c.f.	6	15	25	39	51	59

Here, N = 59

We know that, median will be the observation having cumulative frequency just equal or greater than $\left(\frac{N+1}{2}\right)$th term.

Thus, median = $\left(\frac{59+1}{2}\right)$th term = 30^{th} term

Thus, the cumulative frequency just greater than 30 is 39 which corresponds to 4.
Therefore, the median is 4.
Hence, the correct option is (c) i.e. 4..

Example 32. Calculate median from the following data

X	10	20	30	40	50	60	70
Y	1	5	12	20	19	9	4

(a) 50 (b) 70
(c) 40 (d) Cannot be determined

Sol (c) Here,

X	f	Less than c.f.
10	1	1
20	5	6
30	12	18
40	20	38
50	19	57
60	9	66
70	4	70
	$N = \Sigma f = 70$	

Thus, $\frac{N+1}{2} = \frac{71}{2} = 35.5$

Therefore, c.f just greater than 35.5 is 38 which corresponds to the value of 40.
i.e., median = 40
Hence, the correct option is (c).

CALCULATION OF MEDIAN – CONTINUOUS SERIES

In the case of continuous series, median is the size of $\frac{N}{2}$th observation, where $N = \sum f$ is the total frequency. The calculation of median in this case involves the following steps:

Step 1: Prepare the 'less than' cumulative frequency (c.f.) distribution.

Step 2: Find $\frac{N}{2}$

Step 3: See the c.f. just greater than or equal to $\frac{N}{2}$

Step 4: Find the class corresponding to the c.f. obtained in Step 3. This is called the median class.

Step 5: Apply the following interpolation formula for calculating the median:

$$\text{Median} = l + \frac{\frac{N}{2} - c.f}{f} \times h$$

where, l = lower limit of the median class, f = frequency of the median class, c.f. = cumulative frequency of the class preceding the median class, and h = size or width of the median class.

Example 33. The marks obtained by 100 students in a certain examination are given below:

Marks:	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
No. of Students:	10	9	25	30	16	10

Calculate the median marks.

(a) 32 (b) 50 (c) 30 (d) None

Sol. (a)

Marks	Number of students (f)	Cumulative frequency (c.f)
0 – 10	10	10
10 – 20	9	19
20 – 30	25	44
30 – 40	30	74
40 – 50	16	90
50 – 60	10	100
	N = 100	

Thus, c.f $> \frac{N}{2}$ i.e., c.f. > 50 is 74

Therefore, median class = 30 – 40

$\Rightarrow l = 30, f = 20, c.f = 44$ and $h = 10$

$$\text{Now, Median} = l + \frac{\frac{N}{2} - c.f}{f} \times h$$

$$= 30 + \frac{50 - 44}{30} \times 10 = 32$$

Hence, the correct option is (a).

PROPERTIES OF MEDIAN

1. 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 $y_{me} = a + bx_{me}$.
2. For a set of observations, the sum of absolute deviations is minimum when the deviations are taken from the median. This property states that $\Sigma|x - A|$ is minimum if we choose A as the median.

Example 34. Consider two variables, x and y, related by the equation $y = 3x + 5$. If the median of x is 10, what is the median of y?

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

Sol. (d) Given: $y = 3x + 5$ and the median of x is 10.

To find median of y, put $x = 10$ in the given equation

$\Rightarrow y = 3 \times 10 + 5$

$\Rightarrow y = 30 + 5$

$\Rightarrow y = 35$

Thus, the median of y is 35.

Hence, the correct option is (d) i.e. 35.

Example 35. 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

Sol. (c) Given: $y = 2x - 3$ and the median of x is 20.

To find median of y, put $x = 20$ in the given equation

$\Rightarrow y = 2 \times 20 - 3$

$\Rightarrow y = 40 - 3$

$\Rightarrow y = 37$

Thus, the median of y is 37.

Hence, the correct option is (c).

Example 36. 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

Sol. (b) In case of an even number of observations then the simple average of these two middle values is the median of the even number of observations.

E.g.: 2, 3, 4, 7, 8, 10

Median of the even number of given observations $= \frac{4+7}{2} = \frac{11}{2} = 5.5$

Hence, the correct option is (b).

Example 37. What is the median for the following observations?

5, 8, 6, 9, 11, 4

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

Sol. (b) Given data: 5, 8, 6, 9, 11, 4

Arranging the data in increasing order, we get

4, 5, 6, 8, 9, 11

Clearly, the number of observations is 6 i.e. even

$$\therefore \text{ Median } = \frac{\frac{n}{2}\text{th} + \left(\frac{n}{2}+1\right)\text{th term}}{2}$$

$$= \frac{\frac{6}{2}\text{th} + \left(\frac{6}{2}+1\right)\text{th term}}{2} = \frac{3\text{rd term} + 4\text{th term}}{2}$$

$$= \frac{6+8}{2} = \frac{14}{2} = 7$$

Therefore, the median of given observations is 7.

Hence, the correct option is (b).

PRACTICE QUESTIONS (PART B)

1. Find the median for the following data: 19 22 17 20 12

(a) 12 (b) 20 (c) 22 (d) 19

2. Find the median for the following data: 58 49 64 70 91 34

(a) 58 (b) 70 (c) 61 (d) None of these

3. The marks obtained by 9 students in a test are 25, 20, 15, 45, 18, 7, 10, 38, 12. Find the median marks.

(a) 20 (b) 18 (c) 10 (d) None of these

4. What is the value of median for the following data?

Marks	5-14	15-24	25-34	35-44	45-54	55-64
Number of students	10	18	32	26	14	10

(a) 28 (b) 30 (c) 32.94 (d) 33.18

5. Two variables x and y are given by $y - 5x - 5 = 0$. If the median of x is 20, what is the median of y?

(a) 100 (b) 105 (c) 1105 (d) None of these

6. Find the median of the following data:

x	10	5	7	11	8
f	15	20	15	18	12

(a) 8 (b) 9 (c) 10 (d) 11

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

(a) 10 (b) 16 (c) 5 (d) 18

8. The median 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

(a) 11.60 (b) 11556 (c) 11875 (d) 11.50

9. Find the median of the following data:

Mark	0–10	10–20	20–30	30–40	40–50	50–60
No. of students	4	6	20	10	7	3

(a) 15.6 (b) 25.8 (c) 11.5 (d) 27.5

10. Given below is the distribution of marks obtained by 140 students in an examination:

Marks	10–19	20–29	30–39	40–49	50–59	60–69	70–79	80–89	90–99
No. of students	7	15	18	25	30	20	16	7	2

Find the median of the distribution.

(a) 55.60 (b) 65.75 (c) 51.17 (d) 61.50

Answer Key

1. (d) 2. (c) 3. (b) 4. (c) 5. (b) 6. (a) 7. (a) 8. (c) 9. (d) 10. (c)

Merits of Median

Median possesses the following merits

1. It is rigidly defined.
2. It is easy to calculate and simple to understand.
3. It can be computed while dealing with a distribution with open end classes
4. Being a- positional average, it is not much affected by extreme observations.
5. it is the most appropriate average to be used while dealing with qualitative data.
6. It can sometimes be located by inspection and can also be determined graphically.

Demerits

Median has the following limitations:

1. Median, being a positional average, is not based on each and every item of the distribution.
2. It is not suitable for further mathematical treatment. For example, it is not possible to find the combined median of two or more groups.

3. It cannot be determined exactly for an ungrouped data consisting of an even number of observations. It is determined approximately as the mid-point of two middle 1 observations.
4. In comparison to arithmetic mean, it is much affected by sampling fluctuations.
5. For calculating the median, it is necessary to arrange the data in order of magnitude.

PARTITION VALUE OR QUARTILES OR DECILES OR PERCENTILES

Partition values, such as quartiles, deciles, and percentiles, are statistical measures that divide a given set of observations into equal parts.

- **Quartiles**: Quartiles divide the data into four equal parts, each containing 25% of the observations. The three quartiles are denoted as Q_1, Q_2 (which is the median), and Q_3. Q_1 represents the value below which 25% of the data falls, Q_2 is the median (50^{th} percentile), and Q_3 represents the value below which 75% of the data falls.
- **Deciles:** Deciles divide the data into ten equal parts, each containing 10% of the observations. Deciles are often used to analyze income distributions or rank data. The first decile (D_1) represents the value below which 10% of the data falls, the second decile (D_2) represents the value below which 20% of the data falls, and so on until the ninth decile (D_9).
- **Percentiles:** The value which divides the data into hundred equal parts, each containing 1% of the observations. There are percentiles which are . Percentiles are used to understand the relative position of a particular value in a dataset.

  For example, the 75^{th} percentile (P_{75}) represents the value below which 75% of the data falls, and the 90^{th} percentile (P_{9C}) represents the value below which 90% of the data falls. Thus, is the median.

CALCULATION OF QUARTILES - INDIVIDUAL OBSERVATIONS

For ungrouped data consisting of n observations (not necessarily all distinct), the calculation of k^{th} quartile Q_k (k = 1, 2, 3,) involves the following steps:

Step 1: Arrange the given data in an ascending order of magnitude.

Step 2: The value of k^{th} quartile Q_k is given by $\frac{k(n+1)}{4}$th observation.

$$\text{Thus, } Q_1 = \frac{(n+1)}{4}\text{th observation}$$

$$Q_2 = \frac{2(n+1)}{4}\text{th observation}$$

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

Example 38. Consider the following set of observations: 19, 12, 27, 14, 21, 18, 9, 15. What is the value of the first quartile?

(a) 11.5 (b) 12 (c) 12.5 (d) 13

Sol. (c) Given observations: 19, 12, 27, 14, 21, 18, 9, 15

Arranging the observations in increasing order, we get 9, 12, 14, 15, 18, 19, 21, 27

We know that,

First quartile,

$$Q_1 = \left(\frac{n+1}{4}\right)^{th} \text{value} = \left(\frac{8+1}{4}\right)^{th} \text{value} = \left(\frac{9}{4}\right)^{th} \text{value} = (2.25)^{th} \text{ value}$$

$= 2^{nd}$ value + 0.25 × Difference between 3^{rd} and 2^{nd} values

$= 12 + 0.25(14-12) = 12 + 0.25(2) = 12 + 0.50 = 12.25$

Therefore, the value of the first quartile is 12.50.

Hence, the correct option is (c) i.e. 12.50.

CALCULATION OF DECILES - INDIVIDUAL OBSERVATIONS

For ungrouped data consisting of n observations (not necessarily all distinct), the calculation of k^{th} decile $D_k(k = 1, 2, \ldots, 9)$ involves the following steps:

Step 1: Arrange the given data in an ascending order of magnitude.

Step 2: The value of k^{th} decile D_k is given by $\frac{k(n+1)}{10}$th observation.

Thus, $D_1 = \frac{(n+1)}{10}$ th observation

$D_2 = \frac{2(n+1)}{10}$th observation

$D_3 = \frac{3(n+1)}{10}$th observation and so on

CALCULATION OF PERCENTILES - INDIVIDUAL OBSERVATIONS

For ungrouped data consisting of n observations (not necessarily all distinct), the calculation of k^{th} percentile $P_k(k = 1, 2, \ldots, 99)$ involves the following steps:

Step 1: Arrange the given data in an ascending order of magnitude.

Step 2: The value of k^{th} percentile P_k is given by $\frac{k(n+1)}{100}$th observation.

Thus, $P_1 = \frac{(n+1)}{100}$ th observation

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

$P_3 = \frac{3(n+1)}{100}$th observation and so on

Example 39. Consider the set of numbers: 14, 8, 19, 22, 17, 9, 13, 16. What is the value of the third decile?

(a) 11.8 (b) 13.4 (c) 13.5 (d) 14

Sol. (a) Given observations: 14, 8, 19, 22, 17, 9, 13, 16, Arranging the observations in ascending order, we get 8, 9, 13, 14, 16, 17, 19, 22

We know that,

Third quartile, $D_3 = \frac{3}{10} \times (n + 1)^{th}$ value

$= \frac{3}{10} \times (8 + 1)^{th}$ value

$= \frac{3}{10} \times 9^{th}$ value $= 2.7^{th}$ value

$= 2^{nd}$ value + 0.7 × Difference between 3^{rd} and 2^{nd} values

$= 9 + 0.70(13 - 9) = 9 + 2.8 = 11.8$

Hence, the correct option is (a) i.e. 11.8.

Example 40. Consider the following set of observations: 19, 12, 27, 14, 21, 18, 9, 15. What is the value of the first quartile?

(a) 11.5 (b) 12 (c) 12.5 (d) 13

Sol. (c) Given observations: 19, 12, 27, 14, 21, 18, 9, 15

Arranging the observations in increasing order, we get

9, 12, 14, 15, 18, 19, 21, 27

We know that,

First quartile $Q_1 = \left(\frac{n+1}{4}\right)^{th}$ value $= \left(\frac{8+1}{4}\right)^{th}$ value $= \left(\frac{9}{4}\right)^{th}$ value

$= (2.25)^{th}$ value

= 2nd value + 0.25 × Difference between 3rd and 2nd values

$= 12 + 0.25(14 - 12)$

$= 12 + 0.25(2) = 12 + 0.50 = 12.50$

Therefore, the value of the first quartile is 12.50.

Hence, the correct option is (c).

Example 41. The marks obtained by 9 students in a test are 25, 20, 15, 45, 18, 7, 10, 38 and 12. Find the value of P_{70}.

(a) 25 (b) 7 (c) 10 (d) None of these

Sol. (a) Arranging the given data in ascending order, we get

7, 10, 12, 15, 18, 20, 25, 38, 45

Thus, $P_{70} = \frac{70(n + 1)}{100} = \frac{70(9 + 1)}{100} = \frac{70(10)}{100}$ = 7th observation = 25

Therefore, the value of P_{70} is 25.

Hence, the correct option is (a).

CALCULATION OF QUARTILES - DISCRETE SERIES

In case of discrete frequency distribution where the variable X takes the values $X_1, X_2, \dots X_n$ with respective frequencies $f_1, f_2, \dots f_n$ with $\Sigma f = N$, the calculation of each quartile $Q_k(k = 1, 2, 3)$ involves the following steps:

Step 1: Prepare the 'less than' cumulative frequency distribution.

Step 2: Find $\frac{k(N+1)}{4}$

Step 3: See the c.f. just greater than or equal to $\frac{k(N+1)}{4}$

Step 4: The value of X corresponding to the c.f. obtained in Step 3 gives the required value of Q_k

CALCULATION OF DECILES – DISCRETE SERIES

In case of discrete frequency distribution where the variable X takes the values $X_1, X_2, \ldots X_n$ with respective frequencies $f_1, f_2, \ldots f_n$ with $\sum f = N$, the calculation of each decile D_k (k = 1, 2, 3, ... 9) involves the following steps:

Step 1: Prepare the 'less than' cumulative frequency distribution.

Step 2: Find $\frac{k(N+1)}{10}$

Step 3: See the c.f. just greater than or equal to $\frac{k(N+1)}{10}$

Step 4: The value of X corresponding to the c.f. obtained in Step 3 gives the required value of D_k.

CALCULATION OF PERCENTILES – DISCRETE SERIES

In case of discrete frequency distribution where the variable X takes the values $X_1, X_2, \ldots X_n$ with respective frequencies $f_1, f_2, \ldots f_n$ with $\sum f = N$, the calculation of each percentile P_k (k = 1, 2, 3, ... 99) involves the following steps:

Step 1: Prepare the 'less than' cumulative frequency distribution.

Step 2: Find $\frac{k(N+1)}{100}$

Step 3: See the c.f. just greater than or equal to $\frac{k(N+1)}{100}$

Step 4: The value of X corresponding to the c.f. obtained in Step 3 gives the required value of P_k.

Example 42. Calculate the value of Q_1 and P_{65} from the following data:

X	10	5	7	11	8
f	15	20	15	18	12

(a) 7, 10 (b) 35, 52.65 (c) 10, 7 (d) None of these

Sol. (a) Given data:

X	f	c.f. (less than)
5	20	20
7	15	35

X	f	c.f. (less than)
8	12	47
10	15	62
11	18	80
	N = ∑f = 80	

For Q_1,

We have, $\frac{N+1}{4} = 20.25$

Thus, c.f. just greater than 20.25 is 35 which corresponds to X = 7.

Therefore, $Q_1 = 7$

For P_{65},

We have, $\frac{65(N \quad 1)}{4} \quad 52.65$

Thus, c.f. just greater than 52.65 is 62 which corresponds to X = 10.

Therefore, $P_{65} = 10$

Hence, the correct option is (a).

CALCULATION OF QUARTILES – CONTINUOUS SERIES

In case of continuous frequency distribution, the calculation of Q_k (k = 1, 2, 3) involves the following steps:

Step 1: Prepare the 'less than' cumulative frequency distribution.

Step 2: Find $\frac{KN}{4}$, where N = ∑f is the total frequency.

Step 3: See the c.f. just greater than or equal to $\frac{KN}{4}$

Step 4: Find the class, the class corresponding to c.f. obtained in Step 3.

Step 5: The value of Q_k is then obtained by using the following interpolation formula:

$$Q_k = l + \frac{\frac{kN}{4} - c}{f} \times h$$

where l = lower limit of Q_k class

C = c.f. of the class preceding the Q_k class

f = frequency of the Q_k class, and h = size or width of Q_k class.

CALCULATION OF DECILES – CONTINUOUS SERIES

In case of continuous frequency distribution, the calculation of D_k (k = 1, 2, ..., 9) involves the following steps:

Step 1: Prepare the 'less than' cumulative frequency distribution.

Step 2: Find $\frac{KN}{4}$, where $N = \Sigma f$ is the total frequency.

Step 3: See the c.f. just greater than or equal to $\frac{KN}{10}$

Step 4: Find the D_k class, the class corresponding to c.f. obtained in Step 3.

Step 5: The value of Q_k is then obtained by using the following interpolation formula:

$$Q_k = l + \frac{\frac{kN}{10} - c}{f} \times h$$

where l = lower limit of D_k class

C = c.f. of the class preceding the D_k class

f = frequency of the D_k class, and h = size or width of D_k class.

CALCULATION OF PERCENTILES – CONTINUOUS SERIES

In case of continuous frequency distribution, the calculation of P_k (k = 1, 2, ..., 99) involves the following steps:

Step 1: Prepare the 'less than' cumulative frequency distribution.

Step 2: Find $\frac{kN}{100}$, where $N = \Sigma f$ is the total frequency.

Step 3: See the c.f. just greater than or equal to $\frac{KN}{100}$

Step 4: Find the P_k class, the class corresponding to c.f. obtained in Step 3.

Step 5: The value of is then obtained by using the following interpolation formula:

$$P_k = l + \frac{\frac{kN}{100} - c}{f} \times h$$

where l = lower limit of P_k class

C = c.f. of the class preceding the P_k class

f = frequency of the P_k class, and h = size or width of P_k class.

Example 43. The third quartile and 65th percentile for the following data are:

Profit in '000₹	Less 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

Sol. (a) According to the given data,

Class Interval	Class- Boundaries	Frequency	Cumulative frequency
Less than 10	Less than 9.5	5	5

Class Interval	Class- Boundaries	Frequency	Cumulative frequency
10-19	9.5-19.5	18	23
20-29	19.5-29.5	38	61
30-39	29.5-39.5	20	81
40-49	39.5-49.5	9	90
50-59	49.5-59.5	2	92
		$n = \sum f_i = 9$	

Third quartile Q_3 = Value of $\left(\frac{3n}{4}\right)^{th}$ observation = $\left(\frac{3 \times 92}{4}\right)^{th}$ = 69^{th} observation

69th observation lies in the class interval 29.5 - 39.5, so it is a quartile class.

Here, l = 29.5, cf = 61, f = 20, c = 39.5-29.5 = 10

We know that,

$$Q_3 = l + \frac{\frac{3n}{4} - cf}{f} \times c = 29.5 + \frac{69 - 61}{20} \times 10 = 33.5$$

$\Rightarrow Q_3 = 33.5 \times 1000$ ($\because$ Profits is given in thousands)

= 33500

Thus, third quartile is ₹33,500.

Now, 65th percentile P_{65} = Value of $\left(\frac{65n}{100}\right)^{th}$ observation

$$= \left(\frac{65 \times 92}{100}\right)^{th} = \text{Value of } 59.8\text{th observation}$$

Since, 59.8th observation lies in a class 19.5 - 29.5. So, it is 65th percentile class.

Here, l = 19.5, cf = 23, f = 38, c = 10

We know that,

$$P_{65} = l + \frac{\frac{65n}{100} - cf}{f} \times c = 19.5 + \frac{59.8 - 23}{38} \times 10 = 29.184$$

$P_{65} = 29.184 \times 1000$ ($\because$ Profits is given in thousands)

P_{65} = ₹29184

Hence, the correct option is (a) i.e., ₹33,500 and ₹29,184.

PRACTICE QUESTIONS (PART C)

1. 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
2. The third decile for the numbers 15, 10; 20, 25, 19, 11, 9, 12 is
   (a) 13 (b) 10.70 (c) 11 (d) 11.50

3. Consider the set of numbers: 14, 8, 19, 22, 17, 9, 13, 16. What is the value of the third decile?

(a) 11.8 (b) 13.4 (c) 13.5 (d) 14

4. The fourth decile for the numbers 12, 15, 18, 20, 22, 25, 28, 30 is

(a) 12.50 (b) 19.20 (c) 28.0 (d) 30.0

5. Following are the wages of the laborer's:

₹82, ₹56, ₹90, ₹50, ₹120, ₹75, ₹75, ₹80, ₹130, ₹65

Find D_6 and P_{82}.

(a) 16.2 and 100.2 (b) 81.2 and 120.2

(c) 46.2 and 100.2 (d) None of these

6. 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

7. The marks obtained by 9 students in a test are 25, 20, 15, 45, 18, 7, 10, 38 and 12. Find the values of Q_3 and D_2.

(a) 31.5 and 10 (b) 11 and 10

(c) 10 and 7 (d) None of these

8. From the following data, calculate the median and the first and third quartile wages.

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

(a) 37.7667 and 42.5385 (b) 30.5660 and 24.5385

(c) 37.7667 and 32.7380 (d) None of these

9. Find the 45th and 57th percentiles for the following data on marks obtained by 100 students:

Marks	20-25	25-30	30-35	35-40	40-45	45-50
No. of students	10	20	20	15	15	20

(a) 30 and 37.34 (b) 33.35 and 37.34

(c) 40.5 and 44.6 (d) None of these

10. Calculate the values of Q_3, D_4 from the following data:

X	10	5	7	11	8
f	15	20	15	18	12

Answer Key

1. (c) **2.** (b) **3.** (a) **4.** (b) **5.** (a) **6.** (b) **7.** (a) **8.** (a) **9.** (b)
10. $Q_3 = 10$, $D_4 = 7$

MODE

Mode is a measure of central tendency that represents the value or values in a dataset that occur most frequently. It is the data point with the highest frequency or the data points with equal highest frequencies.

E.g.: In dataset 1, 2, 3, 4, 4, 5, 2, 3, 4, the mode is 4 because it appears three times, which is more than any other value in the dataset.

Bimodal Distribution: If there are two modes in a distribution it is called Bimodal Distribution.

E.g.: 1, 2, 2, 2, 4, 5, 8, 6, 6, 6 then the modes are 2 and 6, as both values appear equally most frequently.

Multi-Modal Distribution

When we have multiple modes in a distribution it is called Multi- Modal Distribution.

E.g.: Dataset 1, 2, 2, 4, 4, 6, 8, 6 has three modes: 2, 4, and 6.

No Mode: When all observations have the same frequencies then the distribution has no mode.

E.g.: 1, 2, 3, 4, 5 has no mode as each value appears only once and has the same frequency.

Calculation of Mode: Discrete Series

In discrete frequency distribution, mode can be determined just by inspection. It is the value of the variable corresponding to the maximum frequency

E.g.: Consider the following distribution:

X	1	2	3	4	5	6	7
f	2	3	12	5	2	6	3

Here, the highest frequency is 12 which corresponds to X = 3.

Therefore, Mode for the given distribution is 3.

Note: While determining mode by inspection in the case of discrete frequency distribution, an error of judgement is possible when the difference between the greatest frequency and the frequency preceding it or succeeding it is very small and the values are heavily concentrated on either side.

Example 44. 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

Sol. (b) Given data: 5, 8, 6, 4, 10, 15, 18, 10

We know that,

For a given set of observations, mode may be defined as the value that occurs the maximum number of times.

Here, occurs maximum number of times (2 times)).

Therefore, the modal value is 10.

Hence, the correct option is (b) i.e. 10.

Example 45. The heights (in centimeters) of a group of students in a class were recorded. The data set is as follows: 141, 150, 154, 152, 162, 162, 142, 155, 160, 167, 147, 152, 165. What is the mode(s) of the heights in the data set?

(a) 142
(b) 165
(c) 152 and 160
(d) No mode

Sol. (c) Given data: 141, 150, 154, 152, 162, 162, 142, 155, 160, 167, 147, 152, 165

We know that,

For a given set of observations, mode may be defined as the value that occurs the maximum number of times.

Here, we can see that the heights 152 and 160 both appear two times, which is more frequent than any other height in the data set.

Therefore, the modes of the heights in the data set are 152 and 160.

Hence, the correct option is (c).

Example 46. What is the mode for the following set of numbers? 7, 9, 12, 15, 18, 21, 24, 27

(a) 7
(b) 27
(c) 15, 18
(d) No mode

Sol. (d) We know that,

The mode represents the value(s) that occur most frequently in a set of numbers. In the given set, each number appears only once, and there are no repeated values. Therefore, there is no value that occurs more frequently than others, resulting in no mode for this set of numbers.

Hence, the correct option is (d).

Example 47. Modal group is:

Height in cms:	61 - 63	63 - 65	65 - 67	67 - 69	69 - 71
No.of students:	15	118	142	127	18

(a) 65 - 67
(b) 69 - 71
(c) 63 - 65
(d) none

Sol. (a) We know that,

For a given set of observations, mode may be defined as the value that occurs the maximum number of times.

Clearly, the highest frequency 142 is which lies in the class 65 - 67.

Therefore, the modal class is 65 – 67.

Hence, the correct option is (a) i.e., 65 – 67.

CALCULATION OF MODE - CONTINUOUS FREQUENCY DISTRIBUTION

The first step is to find the modal class, i.e., the class corresponding to the maximum frequency. The value of mode is then obtained by applying the following interpolation formula:

$$\text{Mode} = l + \frac{f_0 - f_{-1}}{2f_0 - f_{-1} - f_1} \times c$$

where, l = LCB of the modal class,

f_0 = frequency of modal class,

f_{-1} = frequency of pre-modal class,

f_1 = frequency of post-modal class,

c = class-length of the modal class

Example 48. The 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

(a) 390 (b) 390.50 (c) 394.50 (d) 394

Sol. (c) According to the question,

Class	Frequency f	Class-boundaries
350-369	15	349.5-369.5
370-389	27	369.5-389.5
390-409	31	389.5-409.5
410-429	19	409.5-429.5
430-449	13	429.5-449.5
450-469	6	449.5-469.5

Clearly, the maximum frequency is 31.

∴ The modal class is 389.5-409.5.

Here,

l = lower frequency point of modal class = 389.5

f_1 = frequency of the modal class = 31

f_0 = frequency of the preceding class = 27

f_2 = frequency of the succeeding class = 19

c = class length of the modal class = 20

$$\text{Mode} = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times c = 389.5 + \left(\frac{31 - 27}{2(31) - 27 - 19}\right) \times 20 = 394.50$$

Hence, the correct option is (c).

Example 49. Following is an incomplete distribution having modal mark as 44.

Marks:	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100
No. of students:	8	18	?	12	5

What would be the mean marks? **(ICAI)**

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

Sol. (d) Let us assume that ? = x

Given, mode is 44.

⇒ Modal class is 40 – 60.

Thus, $l = 40$, $f_1 = x$, $f_0 = 18$, $f_2 = 12$ and $c = 20$

We know that,

$$\text{Mode } Z = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times c$$

$$\Rightarrow 44 = 40 + \frac{x - 18}{2x - 18 - 12} \times 20$$

$$\Rightarrow 44 - 40 + \frac{x - 18}{2x - 30} \times 20$$

$$\Rightarrow \frac{1}{5} = \frac{x - 18}{2x - 30}$$

On cross multiplication, we get

$2x - 30 = 5(x - 18)$

$\Rightarrow 2x - 30 = 5x - 90$

$\Rightarrow 3x = 60$

$\Rightarrow x = \frac{60}{3}$

$\Rightarrow x = 20$

⇒ ? = 20

Now, to find mean:

Class	Frequency (f)	Mid value (x)	$d = \frac{x - A}{h} = \frac{x - 50}{20}$ $A = 50, h = 20$	f × d
0 – 20	5	10	–2	–10
20 – 40	18	30	–1	–18
40 – 60	20	50 = A	0	0
60 – 80	12	70	1	12
80 – 100	5	90	2	10
	n = 60			$\sum f \cdot d = -6$

$$\text{Mean } \bar{x} = A + \frac{\sum f \cdot d}{n} \times h = 50 - \frac{6}{60} \times 20 = 48$$

Therefore, the mean marks is 48.

Hence, the correct option is (d) i.e. 48.

It may be remarked that the above formula for computing mode is based on the following assumptions:

1. The frequency distribution must be continuous with exclusive type classes without any gaps. If the data are not given in the form of continuous classes, it must first be converted into continuous classes before applying the above formula.
2. The class intervals must be uniform throughout, i.e., the size of all the class intervals must be the same. If they are unequal, they should first be made equal on the assumption that frequencies are uniformly distributed over all the classes.

PROPERTIES OF MODE

- If $y = a + bx$ then $y_{mode} = a + bx_{mode}$

EMPIRICAL RELATIONSHIP OF MEAN, MEDIAN AND MODE

For a grouped frequency distribution, we may consider the following empirical relationship between mean, median and mode:

Mean – Mode = 3(Mean – Median)

⇒ Mean – Mode = 3 Mean – 3 Median

⇒ 3 Median = 3 Mean – Mean + Mode

⇒ 3 Median = 2 Mean + Mode

Example 50. If mean $(\bar{X})$ is 10 and mode (Z) is 7, then find out the value of median (M).

(a) 9 (b) 17 (c) 3 (d) 4.33 **(Dec 2022)**

Sol (a) Given: Mean (M) = 10, Mode (Z) = 7

To find: Median (M)

We know that,

3 Median = 2 Mean + Mode

⇒ 3M = 2(10) + 7

⇒ 3M = 27

⇒ M = 9

Therefore, the value of median (M) is 9.

Example 51. 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?

(a) 46 (b) 16 (c) 51 (d) 66

Sol. (a) Given: Mean mark = 55.60 and Median mark = 52.40

As we know,

Mode = 3 × Median – 2 Mean = 3(52.40) – 2(55.60) = 46

Hence, the correct option is (a).

Example 52. If $y = 2 + 1.50x$ and mode of x is 15, what is the mode of y?

(a) 7.5 (b) 24.5 (c) 10.7 (d) 17.5

Sol. (b) According to the question,

As we know, $y_{mo} = a + bx_{mo}$

Now, Put the values and compute,

$y_{mo} = 2 + 1.50(15) = 2 + 22.5 = 24.5$

Hence, the correct option is (b).

MERITS AND DEMERITS OF MODE

Merits

Mode possesses the following merits:

1. It is simple to understand and easy to calculate.
2. In some cases it can be located merely by inspection.
3. It can be determined graphically from a histogram.
4. It is not at all affected by extreme observations and can be calculated even if extreme values are not known.
5. It can be conveniently determined for distribution with open end classes.

Demerits

Mode has the following drawbacks:

1. It is not rigidly defined
2. It is not based on all the observations.
3. It is not suitable for further mathematical treatment.
4. As compared to mean, mode is affected to a greater extent by the fluctuations of sampling.
5. The value of mode cannot always be determined. In some cases, we may have a bi-modal distribution.

Example 53. What is the modal value for the numbers 5, 8, 6, 4, 10, 15, 18, 10?

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

Sol. (b) Given data: 5, 8, 6, 4. 10, 15, 18, 10

We know that,

For a given set of observations, mode may be defined as the value that occurs the maximum number of times.

Here, 10 occurs maximum number of times (2 times).

Therefore, the modal value is 10.

Hence, the correct option is (b).

Example 54. 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 **(ICAI)**

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

Sol. (b) Given: $x - y - 10 = 0$

To find the mode of y, put $x = 23$ in the given equation.

$\Rightarrow 23 - y - 10 = 10$

$\Rightarrow 13 - y = 0$

$\Rightarrow y = 13$

Therefore, mode of y is 13.

Hence, the correct option is (b) i.e. 13.

Example 55. Which of the following measure(s) satisfies (satisfy) a linear relationship between two variables?

(a) Mean (b) Median (c) Mode (d) None of these

Sol. (d) **GM:** Measure of the central tendency is difficult to compute.

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.

GM: For a given set of n positive observations, the geometric mean is defined as the n-th root of the product of the observations.

Median: 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.

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.

Hence, the correct answer is option (d).

Example 56. Which of the 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)

Sol. (b) **Median:** of central tendency is based on only fifty percent of the central values.

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.

Median: 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.

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.

Hence, the correct option is (b).

PRACTICE QUESTIONS (PART D)

1. The mode of the numbers 7, 7, 7, 9, 10, 11, 11, 11, 12 is
   (a) 11 (b) 12 (c) 7 (d) 7 & 11
2. Mode is

Variable:	2	3	4	5	6	7
No. of men:	5	6	8	13	7	4

   (a) 6 (b) 4 (c) 5 (d) None of these
3. Calculate mode from the following data:

Height in inches:	56	58	59	60	61	62	63	64	66	68
No. of persons:	3	7	6	9	20	22	24	5	3	1

   (a) 68 (b) 20 (c) 63 (d) 56
4. Compute mode for the distribution as described:

Weight in kgs.	40–45	46–50	51–55	56–60	61–65	66–70
No. of friends:	4	3	9	8	7	2

   (a) 67.55 (b) 54.29 (c) 70.65 (d) 66.50
5. Given below is the distribution of weights of a group of 60 students in a class:

Weights (in kg):	30–34	35–39	40–44	45–49	50–54	55–59	60–64
No. of Students:	3	5	12	18	14	6	2

   Find the mode of the distribution.
   (a) 47.5 (b) 65.85 (c) 10.70 (d) 46.59

Answer Key

1. (d) 2. (c) 3. (c) 4. (c) 5. (a)

GEOMETRIC MEAN

The geometric mean, usually abbreviated as G.M., of a set of n observations $x_1, x_2, \ldots\ldots, x_n$ is the nth root of their product. That is,

$$GM = (x_1 x_2 x_3 \ldots\ldots\ldots\ldots x_n)^{\frac{1}{n}}$$

Note: It may be noted that if there are two observations, G.M. can be computed by taking the square root; if there are three observations, G.M. can be computed by taking the cube root of their product and so on. **E.g.:**,

G.M. of two numbers 4 and 9 is $\sqrt{4 \times 9} = 2 \times 3 = 6$

G.M. of three numbers 1, 4 and 128 is $\sqrt[3]{1 \times 4 \times 128} = 2^3 = 8$

Using Calculator tricks we can find Log and antilog for any number:

To find Logarithm	To find Anti-logarithm
1. Type number and press root for 15 times	1. Multiply number by 0.000070274
2. Subtract 1 after step 1	2. Add 1 after step 1
3. Divide number by 0.000070274	3. Press multiply & =“ button for 15 times

Example 57. What is the geometric mean (GM) for the numbers 2, 8 and 32?

(a) 4 (b) 6 (c) 8 (d) 12

Sol (c) Given: $x_1 = 2, x_2 = 8, x_3 = 32$ and $n = 3$

We know that,

$$GM = (x_1 x_2 x_3 \ldots\ldots\ldots x_n)^{\frac{1}{n}}$$

$$= (2 \times 8 \times 32)^{\frac{1}{3}}$$

$$= (2 \times 2 \times 4 \times 4 \times 4 \times 2)^{\frac{1}{3}}$$

$$= 2 \times 4$$

$$= 8$$

Therefore, the required geometric mean is 8.

Hence, the correct option is (c).

Example 58. What is the GM for the numbers 8, 24 and 40?

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

Sol. (c) Given: $x_1 = 8, x_2 = 24, x_3 = 40$ and $n = 3$

We know that,

$$GM = (x_1 x_2 x_3 \ldots\ldots\ldots x_n)^{\frac{1}{n}}$$

$$= (8 \times 24 \times 40)^{\frac{1}{3}}$$

$$= 8(3 \times 5)^{\frac{1}{3}}$$

$$= 8(15)^{\frac{1}{3}}$$

$$= 8\sqrt[3]{15}$$

Hence, the correct option is (c).

Example 59. The geometric mean of three numbers is 12 and two of the numbers are 4 and 16. What is the third number?

(a) 12 (b) 32 (c) 27 (d) 48

Sol. (c) Let the third number be x.

We know that, G.M. = $\sqrt[n]{a_1 \times a_2 \times a_3 \times .... \times a_n}$

Since, Geometric mean of 4, 16 and x = 12

$\Rightarrow$ G.M. = $\sqrt[3]{4 \times 16 \times x}$

$\Rightarrow 12 = \sqrt[3]{64 \times x}$

On cubing both the sides,

$\Rightarrow (12)^3 = 64 \times x$

$\Rightarrow x = \frac{(12)^3}{64}$

$\Rightarrow x = 27$

Hence, option (c) i.e. 27 is the correct answer.

CALCULATION OF GEOMETRIC MEAN – DISCRETE SERIES

The geometric mean of a set of n observations $X_1, X_2, ... X_n$ with their respective frequencies $f_1, f_2, ... f_n$ is given by:

$$G.M. = \left(X_1^{f_1} \times X_2^{f_2} \times ... X_n^{f_n}\right)^{\frac{1}{n}}$$

where N = Σf is the total frequency.

Taking logarithm of both sides, we obtain

$$\text{Log (G.M.)} = \frac{1}{N} \log (X_1^{f1} \times X_2^{f2} \times ... X_n^{fn})^{1/N} = \frac{1}{N} [\log (X_1^{f1} \times X_2^{f2} \times ... X_n^{fn})$$

$$= \frac{1}{N} [f_1 \log X_1 + f_2 \log X_2 +....... + f_n \log X_n] = \frac{1}{N}\Sigma f \log X$$

$$G.M. = AL\left[\frac{1}{N}\Sigma f \log X\right]$$

CALCULATION OF GEOMETRIC MEAN – CONTINUOUS SERIES

In the case of grouped or continuous frequency distribution, G.M. is given by

$$G.M. = AL\left[\frac{1}{N}\Sigma f \log X\right]$$

where $X_1, X_2, ..., X_n$ are the class marks (or mid-values) of a set of grouped data with corresponding class frequencies $f_1, f_2, ..., f_n$

Example 60. Find the geometric mean from the following data:

Diameter (mm)	130	135	140	145	143	148	149	150
No. of Screws	3	4	6	6	3	5	2	2

(a) 142 mm (b) 165 mm (c) 110 mm (d) 149 mm

Sol. (a) According to the given data,

Diameter (mm) (X)	f	log X	f log X
130	3	2.1139	6.3417
135	4	2.1303	8.5212
140	6	2.1461	12.8766
145	6	2.1614	12.9684
143	3	2.1553	6.4659
148	5	2.1703	10.8515
149	2	2.1732	4.3464
150	2	2.1761	4.3522
	$N = \sum f = 31$		$\sum f \log X = 66.7239$

$$G.M. = AL\left[\frac{1}{N}\sum f\log X\right] = AL\left[\frac{66.7239}{31}\right] = AL(2.1524) = 142.0 \text{ mm}$$

PROPERTIES OF GEOMETRIC MEAN

1. Logarithm of G for a set of observations is the AM of the logarithm of the observations i.e., $\log G = \frac{1}{r}\Sigma \log X$
2. If all the observations assumed by a variable are constants, (k), then the GM of the observations is also k.
3. 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
4. GM of the ratio of two variables is the ratio of the GM's of two variables i.e., if $z = \frac{x}{y}$ then $\text{GM of } z = \frac{\text{GM of } x}{\text{GM of } y}$

USES OF GEOMETRIC MEAN

1. Geometric mean is used primarily to average data for which the ratio of consecutive terms remains approximately constant. This occurs, for example, with such data as rates of change, ratios, percent increase in sales, population sizes over consecutive time periods and the like.
2. It is the most appropriate average to be used in the construction of index numbers.
3. It is the most suitable average to be used when it is desired to give more weightage to smaller items and vice-versa.

MERITS AND DEMERITS OF GEOMETRIC MEAN

Merits

1. It is rigidly defined.
2. It is based on all the observations.
3. It is suitable for further mathematical treatment. For example, if the geometric means of two or more sets of data are known, then the geometric mean of the combined data can also be obtained. If G_1 and G_2 are the geometric means of two sets of data with number of observations n_1 and n_2 respectively, then the geometric mean G of the combined data with $n_1 + n_2$ observations is given by:

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

4. It gives less weight to large items and more to small ones than does the arithmetic mean.
5. Unlike AM, geometric mean is affected to a lesser extent by extreme observations.
6. It is not affected much by fluctuations of sampling.

Demerits

Geometric mean has the following drawbacks.

1. It is difficult to understand.
2. It is not easy to calculate for a non-mathematical person.
3. If any of the observations is zero, the geometric mean becomes zero and if any one of the observations is negative, the value of GM cannot be calculated.

Example 61. Which of the following measures of the central tendency is difficult to compute?

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

Sol. (d) **GM:** measure of the central tendency is difficult to compute.

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.

GM: For a given set of n positive observations, the geometric mean is defined as the nth root of the product of the observations.

Median: 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.

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.

Hence, the correct answer is option (d).

Example 62. If GM of x is 10 and GM of y is 15, then the GM of xy is

(a) 150 (b) $\log 10 \times \log 15$ (c) $\log 150$ (d) None of these

Sol. (a) Given: GM of x = 10 and GM of y = 15

We know that,

GM of the product of two variables is the product of their GM's.

Thus, GM of xy = GM of x × GM of y = 10 × 15 = 150

Therefore, GM of xy is 150.

Hence, the correct option is (a).

PRACTICE QUESTIONS (PART E)

1. Find the geometric mean of 2, 4, 8, 12, 16 and 24.

(a) 29.49 (b) 5.86 (c) 8.16 (d) 6.56

2. Find the geometric mean of 3, 6, 24 and 48.

(a) 9 (b) 12 (c) 16 (d) None of these

3. Find the GM for the following distribution:

x	2	4	8	16
f	2	3	3	2

(a) 2 (b) $4\sqrt{2}$ (c) 5 (d) $6\sqrt{3}$

4. The rates of returns from three different shares are 100%, 200% and 400% respectively, the average rate of return will be _______.

(a) 350% (b) 233.33% (c) 200% (d) 300%

5. Find the geometric mean for the following distribution:

Marks	0-10	10-20	20-30	30-40	40-50	50-60
No. of students	10	9	25	30	16	10

(a) 21.57 (b) 25.5 (c) 26.73 (d) None of these

Answer Key

1. (c) **2.** (b) **3.** (b) **4.** (c) **5.** (c)

HARMONIC MEAN

The harmonic mean, usually abbreviated as H.M., is defined as the reciprocal of the arithmetic mean of the reciprocals of the given set of observations.

Symbolically,

$$HM = \frac{n}{\Sigma \frac{1}{X}}$$

CALCULATION OF HARMONIC MEAN: INDIVIDUAL OBSERVATIONS

The harmonic mean of a set of n observations $X_1, X_2, \ldots\ldots, X_n$ (not necessarily all distinct) is given by

$$HM = \frac{n}{\frac{1}{X_1}+\frac{1}{X_2}+\ldots\ldots\ldots+\frac{1}{X_n}} = \frac{n}{\Sigma\frac{1}{X}}$$

Example 63. Find the harmonic mean of 5 numbers 4, 5, 6, 10 and 12.

(a) 2.5 (b) 6.25 (c) 1.50 (d) None of these.

Sol. (b) By definition of H.M.

$$\frac{n}{\sum\frac{1}{x}} = \frac{5}{\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\frac{1}{10}+\frac{1}{12}} = \frac{5}{\frac{15+12+10+6+5}{60}} = \frac{5\times 60}{48} = \frac{25}{4} = 6.25$$

Therefore, the required harmonic mean of given numbers is 6.25.

Hence, the correct option is (b).

Example 64. The harmonic mean for the numbers 2, 3, 5 is

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

Sol. (c) Given: $x_1 = 2, x_2 = 3, x_3 = 5$ and $n = 3$

We know that,

$$HM = \frac{n}{\sum\frac{1}{x_1}} = \frac{3}{\frac{1}{2}+\frac{1}{3}+\frac{1}{5}} = \frac{3\times 30}{15+10+6} = \frac{90}{31} = 2.90$$

Therefore, the harmonic mean of the given numbers is 2.90.

Hence, the correct option is (c).

Example 65. What is the HM of $1, \frac{1}{2}, \frac{1}{3}, \ldots\ldots\ldots \frac{1}{n}$?

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

Sol. (c) We know that,

H.M of $x_1, x_2, x_3, \ldots\ldots\ldots x_n$ (none of them is zero) is given by:

$$H.M = \frac{n}{\sum_{i=1}^{n}\left(\frac{1}{x_1}\right)}$$

Thus, H.M of $1, \frac{1}{2}, \frac{1}{3}, \ldots\ldots\ldots \frac{1}{n} = \frac{n}{1+2+3+\ldots\ldots\ldots n} = \frac{n}{\frac{n(n+1)}{2}} = \frac{2}{(n+1)}$

Hence, the correct option is (c).

CALCULATION OF HARMONIC MEAN: DISCRETE SERIES

The harmonic mean of a set of n observations $X_1, X_2, \ldots\ldots, X_n$ with their respective frequencies $f_1, f_2, \ldots\ldots\ldots, f_n$ is given by:

$$HM = \frac{n}{\Sigma\left[\frac{f}{X}\right]}$$

CALCULATION OF HARMONIC MEAN: CONTINUOUS SERIES

In the case of continuous series, we take $X_1, X_2, \ldots X_n$ are the class marks (or mid-values) of a set of grouped data with corresponding class frequencies $f_1, f_2, \ldots f_n$, then harmonic mean is given by:

$$HM = \frac{n}{\Sigma\left[\frac{f}{X}\right]}$$

Example 66. Find the harmonic mean for the following distribution.

X	10	15	20	25	30
f	2	5	4	9	5

(a) 10.15 (b) 15.41 (c) 19.73 (d) None of these.

Sol. (c) According to the given data,

X	f	$\frac{f}{X}$
10	2	0.2
15	5	0.34
20	4	0.2
25	9	0.36
30	5	0.167
	$N = \sum f = 25$	$\Sigma\frac{f}{X} = 1.267$

Thus, H.M $\frac{\sum f}{\Sigma\frac{f}{X}} = \frac{25}{1.267} = 19.73$

Hence, the correct option is (c).

USES OF HARMONIC MEAN

The harmonic mean is used

1. In averaging speeds when equal distances are covered with varying speed

2. In finding the average cost of some commodity when several different purchases are made by investing the same amount of money each time.

PRACTICE QUESTIONS (PART F)

1. 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) 100 km/hour (b) 583.33 km/hour

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

2. A fire engine rushes to a place of fire accident with a speed of 110 kmph and after the completion of operation returned to the base at a speed of 35 kmph. The average speed per hour in per direction is obtained as _______ of those speeds. **(Dec 2020)**

(a) Speed average of (b) HM of

(c) GM of (d) Half of HM of

Answer Key

1. (b) 2. (b)

WEIGHTED HARMONIC MEAN

Let $w_1, w_2, \ldots, w_n$ be the weights attached to n observations $X_1, X_2, \ldots\ldots X_n$ respectively. Then the Weighted Harmonic Mean, denoted by H.M. $_{w,}$ is defined as

$$H{\cdot}M_{.w} = \frac{\sum w_i}{\sum\left(\frac{w_i}{x_i}\right)}$$

USE OF WEIGHTED HARMONIC MEAN

- In actual practice, the weighted harmonic mean is most frequently used in averaging speeds when different distances are covered with varying speeds.
- In finding the average cost of some commodity when several different purchases are made by putting in a different amount of money each time.

PROPERTIES

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

If there are two groups with n_1 and n_2 observations and H_1 and H_2 as respective HM's then Combined HM is given by:

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

Example 67. The harmonic mean for 40 observations of group 1 data is 520 and for 50 observations of group 2 data is 680. What is the combined harmonic mean?

(a) 450.60 (b) 598.20 (c) 800.00 (d) None of these

Sol. (b) According to the given data,

$n_1 = 40$, $n_2 = 50$, $H_1 = 520$ and $H_2 = 680$

Thus, the combined HM

$$= \frac{40+50}{\frac{40}{520}+\frac{50}{680}} = \frac{90}{\frac{1}{13}+\frac{5}{68}} = 598.20$$

Therefore, the combined harmonic mean is 598.20.

Hence, the correct option is (b).

Example 68. If there are two groups with 75 and 65 as harmonic means and containing 15 and 13 observations, then the combined HM is given by

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

Sol. (c) We know that,

If there are two groups with n_1 and n_2 observations and H_1 and H_2 as respective HM's, then the combined HM is given by

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

Here, $n_1 = 15$, $n_2 = 13$, $H_1 = 75$ and $H_2 = 65$

Thus, the required combined harmonic mean

$$= \frac{15+13}{\frac{15}{75}+\frac{13}{65}} = \frac{28}{\frac{1}{5}+\frac{1}{5}} = \frac{28}{\frac{2}{5}} = 28 \times \frac{5}{2} = 14 \times 5 = 70$$

Therefore, the combined harmonic mean is 70.

Hence, the correct option is (c).

MERITS AND DEMERITS OF HARMONIC MEAN

Merits

Harmonic mean possess the following merits.

1. It is rigidly defined.
2. It is based on all the observations.
3. Since the reciprocals of the values of the observations are involved, It is not as much affected by one or two big observations.
4. It is not affected very much by fluctuations of sampling.

Demerits

Harmonic mean has the following drawbacks.

1. It is not easy to understand.
2. It is difficult to calculate.
3. Its value cannot be computed if any one of the observations is zero.
4. It gives the largest weight to the smallest item. This is generally not a desirable feature and as such this average is not very useful for the analysis of economic data.

Example 69. Which of the following measure(s) possesses (possess) mathematical properties?

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

Sol. (d) AM: For a given set of observations, the AM may be defined as the sum of all the observations divided by the number of observations.

GM: For a given set of n positive observations, the geometric mean is defined as the n-th root of the product of the observations.

HM: For a given set of non-zero observations, the harmonic mean is defined as the reciprocal of the AM of the reciprocals of the observation.

Hence, the correct option is (d).

RELATIONSHIP BETWEEN AM, GM AND HM

1. For any set of data, the values of A.M., G.M., and H.M. are connected by the following relation: AM $\geq$ GM $\geq$ HM
   - If all the values in the dataset are equal then AM = GM = HM
   - If all the values in the dataset are distinct then AM > GM > HM
2. For two numbers, we have $(G.M.)^2 = A.M. \times H.M.$

Example 70. The A.M. and G.M. of two numbers are 15 and 12 respectively. Find the H.M. of the numbers.

(a) 2 (b) 15 (c) 20 (d) 9.6

Sol. (d) We know that,

$AM \times HM = GM^2$

$\Rightarrow 15 \times HM = (12)^2$

$\Rightarrow HM = \frac{144}{15}$

$\Rightarrow HM = 9.6$

Therefore, the value of HM is 9.6.

Hence, the correct option is (d).

Example 71. 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) Cannot be determined.

Sol. (c) Given: AM = GM = 15

We know that,

$AM \times HM = GM^2$

$\Rightarrow 15\ HM = (15)^2$

$\Rightarrow HM = \frac{(15)^2}{15}$

$\Rightarrow HM = 15$

Therefore, the value of HM is 15.

Hence, the correct option is (c).

Example 72. 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

Sol. (b) If a and b are two positive observations such that $a > b$, then

According to the question

$\frac{a+b}{2} = 6.5$ ($\because$ AM = 6.5)

$a + b = 13$ ...(i)

and $\sqrt{ab} = 6$ ($\because$ GM = 6)

$ab = 36$ ...(ii)

We know, $(a - b)^2 = (a + b)^2 - 4ab$

$(a - b)^2 = (13)^2 - 4 \times 36$

$(a - b)^2 = 169 - 144$

$(a - b)^2 = 25$

$a - b = 5$ ...(iii)

Adding (i) and (iii), we get

$2a = 18$

$a = 9$

From (i), we get

$b = 1.3 - 9 = 4$

Therefore, the two numbers are 9 and 4.

Hence, the correct option is (b).

Example 73. 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

Sol. (d) Given: AM = 5 and GM = 3.2

We know that,

$AM \times HM = GM^2$

$\Rightarrow GM2 = 5 \times 3.2$

$\Rightarrow GM^2 = 1.6$

$\Rightarrow$ GM = 4

Therefore, GM will be 4.

Hence, the correct option is (d).

Example 74. 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

Sol. (c) AM > GM > HM true for a set of distinct positive observations.

AM: For a given set of observations, the AM may be defined as the sum of all the observations divided by the number of observations.

GM: For a given set of n positive observations, the geometric mean is defined as the n-th root of the product of the observations.

HM: For a given set of non-zero observations, the harmonic mean is defined as the reciprocal of the AM of the reciprocals of the observation.

Hence, the correct option is (c).

Example 75. For a moderately skewed distribution, which of the following relationships 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)

Sol. (a) For a moderately skewed distribution, following relationship holds

Mean – Mode = (Mean – Mediam)

Hence, the correct option is (a).

PRACTICE QUESTIONS (PART G)

1. Find the harmonic mean of the following numbers: 2, 3, 6, 8, 10

(a) 4.08 (b) 5.18 (c) 9.04 (4) None of these

2. Two values yielded an arithmetic mean of 24 and a harmonic mean of 6. The geometric mean of these values is ______.

(a) 8 (b) 12 (c) 14 (d) 16

3. The average of 5 quantities is 6 and the average of 3 is 8. What is the average of the remaining two?

(a) 4 (b) 5 (c) 3 (d) 3.5

4. From the record on sizes of shoes sold in a shop, one can compute the following to determine the most preferred shoe size.

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

5. In a moderately skewed distribution, the mode and median are 20 and 24 respectively. The value of mean will be ______.

(a) 21 (b) 26 (c) 30 (d) None

6. The mean of 'n' observation is 'X'. If 'K' is added to each observation, then the new mean is _______.

(a) X (b) XK (c) X – K (d) X + K

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. Mean of 7, 9, 12, x, 4, 11 & 5 is 9. Find the missing observation.
   (a) 13 (b) 15 (c) 12 (d) None of these
9. The median of following numbers, which are given in ascending order is 25. Find the value of if data is 11, 13, 15, 19, (x + 2), (x + 4), 30, 35, 39, 46
   (a) 22 (b) 20 (c) 15 (d) 30
10. For a distribution Mean, Median and Mode are 23, 24 and 25. 5 respectively, then it is most likely _____ skewed distribution.
   (a) Positively (b) Symmetrical (c) Asymptotically (d) Negatively
11. There are n numbers. When 50 is subtracted from each of these number the sum of the numbers so obtained is –10. When 46 is subtracted from each of the original n numbers, then the sum of numbers so obtained is 70. What is the mean of the original n numbers?
   (a) 56.8 (b) 25.7 (c) 49.5 (d) 53.8
12. Calculate the mode for the following data:

Monthly wages (₹)	200-250	250-300	300-350	350-400	400-450	450-500	500-550	550-600
No. of workers	4	6	20	12	33	17	8	2

   (a) 259.50 (b) 350.78 (c) 400.42 (d) 428.38
13. The mean of 100 items was found to be 40. Later on it was discovered that two items of 45, 35 were wrongly taken as 35 and 25. Find the correct mean.
   (a) 40.2 (b) 43.5 (c) 40 (d) None of these
14. An incomplete distribution is given below:

Class Interval	0-10	10-20	20-30	30-40	40-50	50-60	Total
Total Frequency	10	___	30	___	25	20	125

   The missing frequencies if the median value is 33 are
   (a) 10 and 15 (b) 15 and 20 (c) 15 and 25 (d) 20 and 25
15. A survey was conducted by a group of students as a part of their environmental awareness program, in which they collected the following data regarding the number of plants in 200 houses in a locality. Find the mean number of plants per house.

Number of plants	0-2	2-4	4-6	6-8	8-10	10-12	12-14
Number of houses	1	2	1	5	6	2	3

   (a) 162 (b) 20 (b) 15.5 (d) 8.1

Answer Key

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

Unit-2: Measure of Dispersion

WHAT IS DISPERSION?

The central tendency measures we've explored so far provide valuable insights into data, but they alone cannot offer a comprehensive description. Regardless of which measure we use, it doesn't reveal the distribution pattern of the data. We may encounter situations where different sets of observations share the same mean but exhibit substantial variations in their measurements around this average..

E.g.: consider the following three sets of observations, each containing 9 items:

Mean										Total	Mean
Set-A	20	20	20	20	20	20	20	20	20	180	20
Set-B	16	17	18	19	20	21	22	23	24	180	20
Set-C	12	14	16	18	20	22	24	26	28	180	20

All the three sets have the same mean i.e. 20 yet they are quite different.

Therefore, we can say that we need some more measures in addition to the central tendency to describe the data completely.

Definition of Dispersion: Dispersion in statistics is a way of describing how spread out a set of data is. It may be defined as the amount of deviation of the observations.

MEASURES OF DISPERSION

The degree to which the numerical data tends to deviate from the average value is called the variation or dispersion of the data.

The measure of dispersion can be classified as:

1. Absolute measures of dispersion
2. Relative measures of dispersion

ABSOLUTE MEASURES OF DISPERSION.	RELATIVE MEASURES OF DISPERSION
❑ Absolute measures are dependent on the unit of the variable under consideration. ❑ Easy to comprehend and compute. **Different Measures ways:** ❍ Range ❍ Mean Deviation ❍ Standard Deviation ❍ Quartile Deviation	❑ Relative measures of dispersion are unit free. ❑ For comparing two or more distributions, relative measures of dispersion are considered. **Different Measures ways:** ❍ Coefficient of Range. ❍ Coefficient of Mean Deviation ❍ Coefficient of Variation ❍ Coefficient of Quartile Deviation.

Example 1. Dispersion measures **(ICAI)**

(a) The scatterness of a set of observations

(b) The concentration of a set of observations

(c) Both (a) and (b)

(d) Neither (a) and (b)

Sol. (a) We know that,

Dispersion measures scatterness of a set of observations .

Hence, the option (a) is correct.

Example 2. When it comes to comparing two or more distributions, we consider **(ICAI)**

(a) Absolute measures of dispersion (b) Relative measures of dispersion

(c) Both (a) and (b) (d) Either (a) or (b)

Sol. (c) For comparing two or more distributions we consider both Absolute measure of dispersion and Relative measure of dispersion.

Hence, the option (c) is correct.

RANGE

DEFINITION

The range of a set of data is defined as the difference between the largest and the smallest value in the set.

Range = Largest value - Smallest value

For a grouped frequency distribution, it is the difference between upper limit of the highest class and lower limit of the smaller class.

Range = Upper class boundary (U.C.B) – Lower class boundary (L.C.B)

Example 3. Following are the wages of 10 workers expressed in INR. 45, 72, 78, 90, 65, 20, 90, 65, 50, 70. Find the range

(a) 60 (b) 59 (c) 63.63 (d) None of the above

Sol. (d) Given: The wages of 10 workers in Rs. are 45, 72, 78, 90, 65, 20, 90, 65, 50, 70

Maximum wage (H) = 90

Minimum wage (L)= 20

So, the range = H – L = 90 – 90 – 20 = 70

Hence, the correct answer is option (d).

Example 4. The following data represents the heights (in centimeters) of a group of students in a class:

Height (cm)	Frequency
100 - 120	5
120 - 140	8
140 - 160	12

Height (cm)	Frequency
160 – 180	10
180 – 200	6

What is the range of heights for the given grouped frequency data?

(a) 100 cm (b) 120 cm (c) 160 cm (d) 200 cm

Sol. (a) We know that,

Range = Upper class boundary (U.C.B) – Lower class boundary (L.C.B)

According to the data given,

Minimum height (L.C.B) = 100 cm

Maximum height (U.C.B) = 200 cm

Therefore, the range of height = 200 cm − 100 cm = 100 cm

Hence, the correct option is (a) i.e., 100 cm.

COEFFICIENT OF RANGE

The range is an absolute measure of dispersion and is expressed in the unit of measurement of values of a distribution. Hence, it cannot be used to compare two distributions expressed in different units.

To overcome this difficulty, we need a relative measure which is independent of the units of measurement. This relative measure, called the coefficient of range, is defined as follows:

$$\text{Coefficient of Range} = \frac{\text{Range}}{\text{Sum of the largest and the smallest values}}$$

i.e., $\text{Coefficient of range} = \frac{L-S}{L+S} \times 100$

where, L is the largest value and S is the smallest value

Example 5. Following are the wages of 10 workers expressed in INR. 45, 72, 78, 90, 65, 20, 90, 65, 50, 70. Find the coefficient of range.

(a) 60 (b) 59 (c) 63.63 (d) None of these

Sol. (c) Given: The wages of 10 workers in Rs. are 45, 72, 78, 90, 65, 20, 90, 65, 50, 70

Maximum wage (L) = 90

Minimum wage (S) = 20

So, the range = L – S = 90 – 20 = 70

$$\text{Coefficient of range} = \frac{90-20}{90+20} \times 100$$

$$= \frac{70}{110} \times 100 = \frac{7}{11} \times 100 = 63.63$$

Hence, the correct answer is option (c).

Example 6. The following data represents the weights (in kilograms) of a group of individuals in a gym:

Weight (kg)	Frequency
41 – 50	6
51 – 60	12
61 – 70	15
71 – 80	8
81 – 90	5

Find the coefficient of range for the given grouped frequency data

(a) 1.13% (b) 3.82% (c) 4.29% (d) None of these

Sol. (b) According to the data given,

$S = 41 - 0.5 = 40.5$

$L = 90 + 0.5 = 90.5$

Thus, Coefficient of range $= \frac{L - S}{L + S} \times 100$

$$= \frac{90.5 - 40.5}{90.5 + 40.5} \times 100 = \frac{50}{131} \times 100 = 3.82\%$$

Hence, the correct option is (b) i.e., 3.82%.

Example 7. What is the coefficient of range for the following distribution?

Class Interval	10-19	20-29	30-39	40-49	50-59
Frequency	11	25	16	7	3

(a) 22 (b) 50 (c) 72.46 (d) 75.82

Sol. (c) The data is exclusive data,

First we find the boundary point for the first class & last class

The boundary points for the first class will be 9.5 – 19.5

So, the minimum value (S) = 9.5

The boundary points for last class will be 49.5 – 59.5

So, the maximum value (L) = 59.5

Thus, Coefficient of range $= \frac{L - S}{L + S} \times 100 = \frac{59.5 - 9.5}{59.5 + 9.5} \times 100 = \frac{50}{69} \times 100 = 72.46$

Hence, the correct answer is option (c).

PROPERTIES OF RANGE

Range remains unaffected due to a change of origin but affected in the same ratio due to a change in scale i.e., for any two constants a and b, two variables x and y are related by $y = a + bx$

Then, the range of y is given by $R_y = |b| \times R_x$

Example 8. If the relationship between x and y is given by 2x + 5y = 10 and the range of x is 5, what would be the range of y?

(a) 1 (b) 2 (c) 3 (d) 4

Sol. (b) As we know,

Range is given by the formula $R_y = |b| \times R_x$

As, 2x + 5y

$\Rightarrow 5y = 10 - 2x$

$\Rightarrow y = 2 - \frac{2x}{5}$

Now, $R_y = \frac{2}{5} \times 5$

$R_y = 2$

Hence, the correct option is (b).

Example 9. If R_x and R_y denote the range of x and y respectively where x and y are related by 3x + 2y + 10 = 0, What would be the relation between x and y?

(a) $R_x = R_y$ (b) $2R_x = 3R_y$

(c) $3R_x = 2R_y$ (d) $R_x = 2R_y$

Sol. (c) The relation of x and y is given by 3x + 2y + 10 = 0

$2y = -3x - 10$

$y = \frac{-3x-10}{2}$

So, the relation between their ranges will be

$R_y = \left|\frac{-3}{2}\right| R_x$

$3R_x = 2R_y$

Hence, the correct answer is option (c).

PRACTICE QUESTIONS (PART A)

1. The range of 15, 12, 10, 9, 17, 20 is

(a) 5 (b) 12 (c) 13 (d) 11

2. What is the coefficient of range for the wages of 8 workers?

₹80, ₹65, ₹90, ₹75, ₹70, ₹72, ₹85

(a) ₹ 30 (b) ₹ 20 (c) 30 (d) 20

3. Find the range of the daily wages of 10 persons given below:

₹240, ₹180, ₹250, ₹160, ₹200, ₹280, ₹220, ₹170, ₹210 and ₹270

(a) ₹130 (b) ₹140 (c) ₹120 (d) ₹150

4. If the range of x is 2, what would be the range of −3x + 50?

(a) 2 (b) 6 (c) −6 (d) 44

5. Find the range for the following frequency distribution:

x	3–5	6–8	9–11	12–14
f	3	2	2	3

(a) 12 (b) 12.5 (c) 11 (d) None of these

6. The following are the prices of shares of a company from Monday to Saturday:

Days	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Price (in ₹)	55	54	52	53	56	58

Calculate the range and the coefficient of range.

(a) 5 and 0.064 (b) 6 and 0.054 (c) 8 and 0.034 (d) None of these

7. Find the coefficient of range for the following frequency distribution:

x	3–5	6–8	9–11	12–14
f	3	2	2	3

(a) 60.5% (b) 65.80% (c) 70.59% (d) None of these

Answer Key

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

MERITS

The range possesses the following merits

1. It is simple to understand and easy to calculate.
2. It requires minimum time to calculate the value of range.
3. It is useful in studying fluctuations in the share prices.

DEMERITS

The range has the following drawbacks:

1. It is not based on all the observations.
2. Range is a poor measure of variation. It considers only the extreme values and tells us nothing about the distribution of numbers in between.
3. It is very much affected by fluctuations of sampling. Its value varies widely from sample to sample
4. It cannot be calculated for grouped frequency distribution with open-end classes.
5. It is not suitable for further mathematical treatment.

CHARACTERISTICS FOR AN IDEAL MEASURE OF DISPERSION

An ideal measure of dispersion should be

- Properly defined

- Easy to comprehend
- Simple to compute
- Based on all the observations
- Unaffected by sampling fluctuations and amenable to some desirable mathematical treatment.

MEAN DEVIATION

As we saw, range is not based on all the observations. Moreover, it does not show any scatterness around an average. If we wish to measure variation in the sense of showing the scatter around an average, we must include the deviations of each and every item from an average.

- Mean deviation or the Average deviation helps us in achieving this goal. As the name suggests, this measure of dispersion is obtained by taking the average (arithmetic mean) of the deviations of the given values from a measure of central tendency.
- Mean Deviation (about an average A) = $\frac{1}{n}\Sigma\left|x_i - A\right|$ where, $\left|x_i - A\right|$ is the modulus value or the absolute value of the deviation from A, ignoring ± signs.

Usually, we obtain mean deviation about any one of the three averages Mean (M), Median (M_d) or Mode (M_O). As we know mode is generally ill-defined, in practice, mean deviation is computed about mean or median and if we calculate mean deviation about median, it will be much beneficial because the sum of the deviations of items from median is least when signs are ignored. But mostly, the mean is more frequently used in computing the average deviation and this is the reason why it is more commonly referred to as mean deviation.

PROCEDURE FOR COMPUTING THE MEAN DEVIATION

We now outline the procedure for computing the mean deviation:

Step 1. Calculate the average A about which mean deviation is to be computed, by the methods discussed earlier.

Step 2. Find the deviation of each observation X from A and denote it by D. That is, find D = X – A.

Step 3. Find the absolute value of the deviation of each observation from A ignoring the signs and denote it by $|D|$

Step 4. Find the sum of all absolute deviations obtained in Step 3 to get $\Sigma|D|$

Step 5. Divide the sum obtained in Step 4 by the number of observations to get the required mean deviation about the average A.

Example 10. What is the mean deviation for the following numbers?

15, 18, 20, 9, 12, 16

(a) 2 (b) 3 (c) 7 (d) 8

Sol. (b) Given numbers: 15, 18, 20, 9, 12, 16

Mean $(\bar{x}) = \frac{15+18+20+9+12+16}{6} = 15$

Now, mean deviation about mean

$$= \frac{|15-15|+|18-15|+|20-18|+|9-18|+|12-15|+|16-15|}{6}$$

$$= \frac{0+3+5+6+3+1}{6} = \frac{18}{6} = 3$$

Hence, the correct option is (b).

Example 11. What is the value of mean deviation about mean for the following numbers? 5, 8, 6, 3, 4 **(ICAI)**

(a) 5.20 (b) 7.20 (c) 1.44 (d) 2.23

Sol. (c) Given data: 5, 8, 6, 3, 4

Here, n = 5

Mean of the observations $(\bar{x}) = \frac{(5+8+6+3+4)}{5} = \frac{26}{5} = 5.2$

$|\text{Deviation from mean}| = |x_i - \bar{x}|$

$= |5-5.2|, |8-5.2|, |6-5.2|, |3-5.2|, |4-5.2|$

$= 0.2, 2.8, 0.8, 2.2, 1.2$

Thus, sum of deviations:

$\sum|x_i - \bar{x}| = 0.2 + 2.8 + 0.8 + 2.2 + 1.2 = 7.2$

Mean deviation about mean $= \frac{\sum|x_i - \bar{x}|}{n} = \frac{7.2}{5} = 1.44$

Hence, the correct answer is option (c).

Example 12. Given the observations: 4, 9, 11, 14, 37. The mean deviation about the median is

(a) 11 (b) 8.5 (c) 7.6 (d) 7.45

Sol. (c) Given observations, X_i : 4, 9, 11, 14, 37

Here, n = 5

Since, n is odd

Thus, median of observation $= \frac{n+1}{2} = \frac{5+1}{2} = \frac{6}{2} = 3^{rd}$ Observation = 11

$\therefore |d| = |X_i - \text{Median}| = |X_i - 11|$

Xi	$|d| = |X_i - 11|$
4	7
9	2
11	0

14	3
13	26
	$\sum\lvert d\rvert = 38$

Mean deviation about the median = $\frac{\sum(xi - \text{median})}{n} = \frac{\sum\lvert d\rvert}{n} = \frac{38}{5} = 7.6$

Hence, the correct option is (c) i.e.7.6.

Example 13. What is the value of mean deviation about mean for the following observation?
50, 60, 50, 50, 60, 60, 60, 50, 50, 50, 60, 60, 50 **(ICAI)**

(a) 5 (b) 7 (c) 35 (d) 10

Sol. (a) Given data,

50, 60, 50, 50, 60, 60, 60, 50, 50, 50, 60, 60, 60, 50

The mean of the data

$$= \frac{50+60+50+50+60+60+60+50+50+50+60+60+60+50}{14} = \frac{770}{14} = 55$$

Now, the sum of deviation from mean

= 5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 + 5 +5 + 5 +5 = 70

$\therefore$ Mean deviation about mean = $\frac{\text{Sum of deviation from mean}}{\text{Mean}} = \frac{70}{14} = 5$

Therefore, mean deviation about mean is 5.

Hence, the correct answer is option (a).

Example 14. The mean deviation about mode for the numbers $\frac{4}{11}, \frac{6}{11}, \frac{8}{11}, \frac{9}{11}, \frac{12}{11}, \frac{8}{11}$ is

(a) $\frac{1}{6}$ (b) $\frac{1}{11}$ (c) $\frac{6}{11}$ (d) $\frac{5}{11}$ **(ICAI)**

Sol. (a) Given data (X_i): $\frac{4}{11}, \frac{6}{11}, \frac{8}{11}, \frac{9}{11}, \frac{12}{11}, \frac{8}{11}$

Since, $\frac{}{11}$ is occurring most frequent (2 times), thus

Mode of the data = $\frac{8}{11}$

Now, the deviation from mode $\lvert X_i - \text{Mode}\rvert = \left\lvert X_i - \frac{8}{11}\right\rvert = \frac{4}{11}, \frac{2}{11}, 0, \frac{1}{11}, \frac{4}{11}, 0$

$\therefore$ Mean deviation about mode

$$= \frac{\sum(X_i - \text{Mode})}{n} = \frac{\frac{4}{11} + \frac{2}{11} + 0 + \frac{1}{11} + \frac{4}{11} + 0}{6} = \frac{\frac{11}{11}}{6} = \frac{1}{6}$$

Hence, the correct answer is option (a) i.e., $\frac{1}{6}$.

COMPUTATION OF MEAN DEVIATION – DISCRETE SERIES

- In case of discrete series where the variable X takes the values $X_1, X_2, \ldots, X_n$ with respective frequencies $f_1, f_2, \ldots\ldots, f_n$, the mean deviation about an average A is given by
- Mean Deviation (about an average A) = $\dfrac{\Sigma f_i \left|X_i - A\right|}{N}$

PROCEDURE FOR COMPUTING THE MEAN DEVIATION

Step 1. Calculate the average A about which mean deviation is to be computed.

Step 2. Take the deviation of each observation from A and denote it by D. That is, find D = X − A.

Step 3. Find the absolute value of the deviation of each observation from A ignoring ± signs and denote it by $|D|$.

Step 4. Multiply each absolute deviation $|D|$ by the corresponding frequency f_i to get $\Sigma|D|$.

Step 5. Add all the products obtained in Step 4 to get $\Sigma f_i |D|$.

Step 6. Divide the sum obtained in Step 5 by N, the total frequency, to get the required mean deviation.

Example 15. Calculate the mean deviation about the mean for the following data:

X	10	11	12	13	14	Total
f	3	12	18	12	3	48

(a) 12 (b) 0.75 (c) 15.5 (d) None of these

Sol. (b) According to the given data, we have

X	f	fX	$D = X - \bar{X}$	$\|D\|$	$f\|D\|$
10	3	30	−2	2	6
11	12	132	−1	1	12
12	18	216	0	0	0
13	12	156	1	1	12
14	3	42	2	2	6
	$N = \Sigma f = 48$	$\Sigma fX = 576$		$\Sigma f\|D\| = 36$	

Mean $\bar{X} = \dfrac{\Sigma fX}{\Sigma f} = \dfrac{576}{48} = 12$

Thus, Mean Deviation about mean $= \dfrac{\Sigma f|D|}{N} = \dfrac{36}{48} = 0.75$

Hence, the correct option is (b).

COMPUTATION OF MEAN DEVIATION: CONTINUOUS SERIES

- The computation of the mean deviation in the case of continuous series is exactly the same as discussed above for discrete series.

- The only difference is that here we have to obtain the class marks (or mid-values) of the various classes and take absolute deviations of these values from the average A.
- Thus, if $x_1, x_2, ..., x_n$ are the class marks (or mid-values) of a set of grouped data with corresponding class frequencies $f_1, f_2, ..., f_n$, then the mean deviation about an average A is given by:

  Mean Deviation (about an average A) = $\frac{\Sigma\ f_i|x_i - A|}{N}$

Example 16. Calculate mean deviation from the median for the following data:

Marks	10–20	20–30	30–40	40–50	50–60	60–70	70–80	80–90
No. of students	2	6	12	18	25	20	10	7

(a) 15.57 (b) 12.95 (c) 16.25 (d) None of these

Sol. (b) According to the given data, we have

Marks	Mid-Value X	No. of students (f)	Less than cf	D	|D|	f|D|
10-20	15	2	2	–39.8	39.8	79.6
20-30	25	6	8	–29.8	29.8	178.8
30-40	35	12	20	–19.8	19.8	237.6
40-50	45	18	38	–9.8	9.8	176.4
50-60	55	25	63	0.2	0.2	5.0
60-70	65	20	83	10.2	10.2	204.0
70-80	75	10	93	20.2	20.2	202.0
80-90	85	7	100	30.2	30.2	211.4
		N = Σ f = 100				Σ f |D| = 1294.8

We have, $\frac{N}{2} = 50$

Thus, the c.f just greater than or equal to 50 is 63 which corresponds to class interval 50-60.

Therefore, Median = $l + \frac{\frac{N}{2} - c}{f} \times h = 50 + \frac{50 - 38}{25} \times 10 = 54.8$

Mean deviation = $\frac{\Sigma f|D|}{N} = \frac{1294.8}{100}$ = 12.9948 = 12.95 (approx).

Hence, the correct option is (b).

Example 17. What is the coefficient of mean deviation for the following distribution of heights? Take deviation from AM. **(ICAI)**

Height in inches	60-62	63-65	66-68	69-71	72-74
No. of students	5	22	28	17	3

(a) 2.30 inches (b) 3.45 inches (c) 3.82 inches (d) 2.48 inches

Sol. (b) According to the given data,

C.I.	Class boundary	Frequency (f_i)	x_i	$u_i = \frac{x_i - A}{h}$ A = 67, h = 3	$f_i u_i$
60-62	59.5 - 62.5	5	61	-2	-10
63-65	62.5 - 65.5	22	64	-1	-22
66-68	65.5 - 68.5	28	67 = A	0	0
69-71	68.5 - 71.5	17	70	1	17
72-74	71.5 - 74.5	3	73	2	6
		75			

Thus, Mean $(\overline{x}) = A + \frac{\sum f_i u_i}{n} \times h \Rightarrow \overline{x} = 67 + \frac{-9}{75} \times 3$

$\Rightarrow \overline{x} = 67 - 0.36 \Rightarrow \overline{x} = 66.64$

Now,

d = $\lvert x_i - \text{mean} \rvert$	$f_i .d$
5.64	28.20
2.64	58.08
0.36	10.08
3.36	57.12
6.36	19.08
	172.56

$$\text{Mean Deviation about mean} = \frac{\sum_{i=1}^{n} f_i \left|(x_i - \overline{x})\right|}{n} = \frac{172.56}{75} = 2.3008$$

$$\text{Coefficient of Mean Deviation} = \frac{\text{Mean Deviation}}{\overline{x}} \times 100$$

$$= \frac{2.3008}{66.64} \times 100 \approx 3.45 \text{ inches}$$

Hence, the correct option is (b).

COEFFICIENT OF MEAN DEVIATION

The relative measure corresponding to the mean deviation, called the coefficient of mean deviation, is given by

- Coefficient of M.D. = $\frac{1}{n}\Sigma|x_i - A|$
- Coefficient of M.D. about mean $= \frac{1}{n}\sum|x_i - \bar{x}|$ where, $\bar{x}$ is mean
- Coefficient of M.D. about median = $\frac{1}{n}\Sigma|x_i - \text{Median}|$

  Coefficient of mean deviation is a pure number independent of the units of measurement and can be used to compare two distributions expressed in different units.

PROPERTIES

- Mean deviation takes its minimum value when the deviations are taken from the median.
- Also mean deviation remains unchanged due to a change of origin but changes in the same ratio due to a change in scale i.e., if y = a + bx, a and b being constants, then MD of y = $|b|$ × MD of x

Example 18. If the relation between x and y in 5y – 3x = 10 and the mean deviations about mean for x is 12, then the mean deviation of y about mean is

(a) 7.20 (b) 6.80 (c) 20 (d) 18.80

Sol. (a) Given, the relation is 5y – 3x = 10

$$\Rightarrow y = \frac{(10 + 3x)}{5}$$

$$\Rightarrow y = \frac{3}{5}x + 2$$

Mean deviation about mean of y = $\left|\frac{3}{5}\right|$ × mean deviation about mean of x

$$= \left|\frac{3}{5}\right| \times 12 = \frac{3}{5} \times 12 = 7.2$$

Hence, the correct option is (a) i.e.,7.2.

MERITS AND DEMERITS OF MEAN DEVIATION

Merits

1. It is easy to understand and simple to calculate.
2. It is based on each and every item of the data.
3. It is rigidly defined.
4. As compared to standard deviation, it is less affected by extreme observations.
5. Since deviations are taken from a central value, comparison about formation of different distributions can easily be made.

Demerits

1. The major drawback of mean deviation is that algebraic signs are ignored while taking the deviations of the items.
2. It is not suitable for further mathematical treatment.
3. it cannot be computed for distribution with open-end classes.
4. It is rarely used in sociological studies.

Example 19. If two variables x and y are related by the equation $2x - 3y + 4 = 0$ and the mean and mean deviation about mean of x are 4 and 0.6 respectively, then the coefficient of mean deviation of y about its mean is

(a) 5 (b) 8 (c) 10 (d) None of these

Sol. (c) Given,

The relation is $2x - 3y + 4 = 0$

$\Rightarrow \ 3y = 2x + 4$

$\Rightarrow \ y = \frac{2}{3}x + \frac{4}{3}$

Thus, the mean of $y = \frac{2}{3}(x) + \frac{4}{3}$

$\Rightarrow \ y = \frac{2}{3}(4) + \frac{4}{3}$

$\Rightarrow \ y = 4$

Mean deviation of $y = \left|\frac{2}{3}\right| \times$ mean deviation of x

$\Rightarrow \ y = \frac{2}{3}(0.6)$

$\Rightarrow \ y = 0.4$

Thus, Coefficient of mean deviation $= \frac{\text{mean deviation}}{\text{mean}} \times 100$

$= \frac{0.4}{4} \times 100 = 10$

Hence, the correct option is (c) i.e., 10.

Example 20. The coefficient of mean deviation about mean for the first 9 natural number is

(a) $\frac{200}{9}$ (b) 80 (c) $\frac{400}{9}$ (d) 50 **(ICAI)**

Sol. (c) We know that,

First 9 natural numbers are 1, 2, 3, 4, 5, 6, 7, 8 and 9.

Sum of first n natural numbers $= \frac{n(n+1)}{2}$

Then, sum of first 9 natural numbers $= \frac{9(9+1)}{2} = 45$

Mean of these numbers $= \frac{45}{9} = 5$

$\therefore \Sigma |\text{deviation from mean}| = 4 + 3 + 2 + 1 + 0 + 1 + 2 + 3 + 4 = 20$

The mean deviation about mean $= \frac{20}{9}$

$\therefore$ Coefficient of mean deviation $= \frac{\text{Mean deviation}}{\text{Mean}} \times 100$

$$= \frac{20}{9 \times 5} \times 100$$

$$= \frac{400}{9}$$

Hence, the correct answer is option (c).

Example 21. If two variables x and y are related by $2x + 3y - 7 = 0$ and the mean and mean deviation about mean of x are 1 and 0.3 respectively, then the coefficient of mean deviation of y about its mean is **(ICAI)**

(a) -5 (b) 12 (c) 50 (d) 4.

Sol. (b) Given,

The relation is $2x + 3y - 7 = 0$

$\Rightarrow 3y = 7 - 2x$

$\Rightarrow y = \frac{7}{3} - \frac{2}{3}x$

Thus, the mean of $y = \frac{7}{3} - \frac{2}{3}$ (mean of x)

$\Rightarrow y = \frac{7}{3} - \frac{2}{3}$ (1)

$\Rightarrow y = \frac{5}{3}$

Mean deviation of $y = \left|\frac{-2}{3}\right|$ mean deviation of x

$\Rightarrow y = \frac{2}{3}(0.3)$

$\Rightarrow y = 0.2$

Coefficient of mean deviation $= \frac{\text{mean deviation}}{\text{mean}} \times 100$

$= \frac{0.2}{\frac{5}{3}} \times 100 = 12$

Hence, the correct answer is option (b).

PRACTICE QUESTIONS (PART B)

1. If x and y are related as $3x + 4y + 7 = 0$ and mean deviation of x is 6.40, what is the mean deviation of y?
   (a) 2.52 (b) 4.18 (c) 6.40 (d) None of these
2. Find the mean deviation about the median for the following data:
   82, 56, 75, 70, 52, 80, 68
   (a) 21.5 (b) 12.45 (c) 17.68 (d) None of these
3. Calculate mean deviation about the mean for the following data:

X	10	11	12	13	14	Total
Y	3	12	18	12	3	48

   (a) 0.25 (b) 0.3 (c) 0.75 (d) 8.0
4. Compute the mean deviation about the arithmetic mean for the following data:

X	1	3	5	7	9
Y	5	8	9	2	1

   (a) 1.72 (b) 1.38 (c) 1.55 (d) 1.80
5. The mean deviation of weights about median for the following data: **(ICAI)**

Weight (lb)	131-140	141-150	151-160	161-170	171-180	181-190
No. of persons	3	8	13	15	6	5

   (a) 10.97 (b) 8.23 (c) 9.63 (d) 11.45
6. Calculate the mean deviation from the mean for the following data:

Marks:	0-10	10-20	20-30	30-40	40-50	50-60	60-7
No. of Students	6	5	8	15	7	6	3

   (a) 12.52 (b) 13.18 (c) 7.55 (d) 8.50
7. Calculate the coefficient of mean deviation for the following data:

Marks	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
No. of students	2	6	12	18	25	20	10	7

   (a) 0.52 (b) 0.18 (c) 0.24 (d) 0.50
8. What is the mean deviation about median for the following data? **(ICAI)**

X	3	5	7	9	11	13	15
F	2	8	9	16	14	7	4

   (a) 2.50 (b) 2.46 (c) 2.43 (d) 2.37

9. Calculate coefficient of mean deviation from the median for the following data:

Marks	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
No. of students	2	6	12	18	25	20	10	7

(a) 0.24 (b) 0.95 (c) 1.50 (d) None of these

Answer Key

1. (b) 2. (b) 3. (c) 4. (a) 5. (a) 6. (b) 7. (c) 8. (d) 9. (a)

QUARTILE DEVIATION OR SEMI INTER- QUARTILE RANGE

Inter-quartile Range is an absolute measure of dispersion defined by the formula:

Inter-quartile range = $Q_3 - Q_1$, where Q_1 and Q_3 are the first (lower) and the third (or upper) quartiles respectively.

Quartile deviation, also called semi inter-quartile range, is an absolute measure of dispersion defined by the formula:

$$\text{Quartile Deviation } (Q.D.) = \frac{Q_3 - Q_1}{2}$$

- In a symmetrical distribution, the two quartiles Q_1 and Q_3 are equidistant from the median, meaning their difference is equal.
- The quartile deviation provides a measure of the spread or dispersion of the data around the median.
- The quartile deviation can be used to estimate the spread of data and compare the variability between different datasets.
- The range of values from (median – Q.D.) to (median + Q.D.) covers exactly 50% of the data, making it useful for analysing the central half of the distribution.

COEFFICIENT OF QUARTILE DEVIATION

The coefficient of quartile deviation is a relative measure of dispersion that provides a standardized measure of the spread of data. It is defined by the formula:

$$\text{Coefficient of Quartile Deviation} = \frac{\text{Quartile Deviation}}{\text{Median}} \times 100 = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$$

The coefficient of quartile deviation is a pure number that is independent of the units of measurement used for the data. This makes it useful for comparing the variability of different distributions, even if they are expressed in different units.

Therefore, Quartile deviation can be computed from the distribution having open-end classes. It is affected considerably by the sampling fluctuations. Quartile deviation remains affected due to change of origin but is affected in the same ratio due to change in scale.

Example 22. When 1^{st} quartile = 20, 3^{rd} quartile = 30, the value of quartile deviation is

(a) 7 (b) 4 (c) –5 (d) 5 **(ICAI)**

Sol. (d) Given: $Q_1 = 20$ and $Q_3 = 30$

We know that,

$$\text{Quartile deviation} = \frac{Q_3 - Q_1}{2} = \frac{30 - 20}{2} = \frac{10}{2} = 5$$

Therefore, the value of quartile deviation is 5.

Hence, the correct answer is option (d) i.e., 5.

Example 23. The first quartile of a dataset is 25 and the third quartile is 75. What is the interquartile range of the dataset?

(a) 25 (b) 50 (c) 75 (d) 100

Sol. (b) To find the interquartile range, we need to calculate the difference between the third quartile and the first quartile.

Given:

$Q_1 = 25$

$Q_3 = 75$

Therefore, Interquartile range = 75 – 25 = 50

Therefore, the interquartile range of the dataset is 50.

Hence, the correct option is (b).

Example 24. Quartile Deviation for the data 1, 3, 4, 5, 6, 6, 10 is **(ICAI)**

(a) 3 (b) 1 (c) 6 (d) 1.5

Sol. (d) As we know that,

Quartile Deviation is given by $\frac{Q_3 - Q_1}{2}$

where, Q_3 = third quartile and Q_1 = first quartile

We have, 1, 3, 4, 5, 6, 6, 10

Given series of numbers is already arranged in ascending order and the number of observations is odd i.e. n = 7.

$$\text{First Quartile } Q_1 = \left(\frac{n+1}{4}\right)^{th} \text{ observations}$$

$$= \left(\frac{7+1}{4}\right)^{th} \text{ observations} = 2^{nd} \text{ observation} = 3$$

$$\text{Third Quartile } Q_3 = \frac{3(n+1)^{th}}{4} \text{ observations}$$

$$= \left(\frac{24}{4}\right)^{th} \text{ observations} = 6^{th} \text{ observation} = 6$$

$$\text{Therefore, Quartile Deviation} = \frac{6 - 3}{2} = 1.5$$

Hence, the correct option is (d) i.e., 1.5.

Example 25. If median = 5, Quartile deviation = 1.5, then the coefficient of quartile deviation is

(a) 33 (b) 35 (c) 30 (d) 20 **(ICAI)**

Sol. (c) Given, Median = 5, Quartile deviation = 1.5

We know that,

$$\text{Coefficient of Quartile Deviation} = \frac{\text{Quartile deviation}}{\text{median}} \times 100$$

$$= \frac{1.5 \times 100}{5} = 30$$

Therefore, the coefficient of quartile deviation is 30.

Hence, the correct option is (c) i.e., 30.

Example 26. The quartile deviation for the data is:

x	2	3	4	5	6
y	3	4	8	4	1

(a) $\frac{1}{4}$ (b) $\frac{1}{2}$ (c) 1 (d) 0

Sol. (c) We know that,

Quartile deviation is given by Q. D. = $\frac{1}{2}(Q_3 - Q_1)$

where $Q_1 = \left(\frac{N+1}{4}\right)^{th}$ observation and $Q_3 = \frac{3(N+1)^{th}}{4}$ observation

Calculating cumulative frequency, we get

x	2	3	4	5	6
f	3	4	8	4	1
cf	3	7	15	19	20

Here, $N = \Sigma x_i = 20$

So, $Q_1 = \frac{(20+1)^{th}}{4} = 5.25^{th}$observation = 3

Similarly, $Q_3 = \frac{3(20+1)^{th}}{4} = 15.75^{th}$ observation = 5

Therefore, Q.D. = $\frac{1}{2}(5 - 3) = 1$

Hence, the correct option (c) is i.e. 1.

Example 27. Following are the marks of the 10 students: 56, 48, 65, 35, 42, 75, 82, 60, 55, 50. Find the coefficient of quartile deviation.

(a) 16.50 (b) 18 (c) 18.42 (d) None of these

Sol. (c) On arranging the marks in ascending order, we get

35, 42, 48, 50, 55, 56, 60, 65, 75, 82

Now,

First Quartile $Q_1 = \left(\frac{10+1}{4}\right)^{th}$ observations

$= \left(\frac{11}{4}\right)^{th}$ observations = 2.75^{nd} observation

= 2^{nd} observation + 0.75 × difference between 3^{rd} and 2^{nd} observation

= 42 + 0.75(48 − 42) = 46.50

Third Quartile $Q_3 = \frac{3(n+1)^{th}}{4}$ observations

$= \left(\frac{33}{4}\right)^{th}$ observations = 8.25^{th} observation

= 8^{th} observation + 025 × difference between 9^{th} and 8^{th} observation

= 65 + 0.25(75 − 65) = 67.50

Thus, the coefficient of quartile deviation is given by $\frac{(Q_3 - Q_1)}{(Q_3 + Q_1)} \times 100$

$= \frac{67.50 - 46.50}{67.50 + 46.50} \times 100 = \frac{21}{114} \times 100 = 18.42$

Hence, the correct option is (c) i.e. 18.42.

Example 28. If x and y are related as 2x + 5y = 30 and the quartile deviation of x is 10, then the quartile deviation of y is:

(a) 2 (b) 4 (c) 5 (d) 6

Sol. (b) Given relation: 2x + 5y = 30

Also, quartile deviation of x = 10

Thus, 5y = 30 − 2x

$y = 6 - \frac{2}{5}x$

Therefore, Quartile deviation of y = $\left|-\frac{2}{5}\right|$ × Quartile deviation of x

Put the value of quartile deviation of x to find quartile deviation of y i.e.,

$y = \frac{2}{5}(10)$

y = 4

Therefore, the quartile deviation of y is 4.

Hence, the correct option is (b).

PRACTICE QUESTIONS (PART C)

1. Following are the marks of the 10 students: 34, 28, 45, 26, 24, 47, 38, 36, 49, 50. Find the quartile deviation.

(a) 20 (b) 30
(c) 10 (d) None of these

2. The quartiles of a variable are 45, 52 and 65 respectively. Its quartile deviation is

(a) 10 (b) 20 (c) 25 (d) 8.30 **(ICAI)**

3. Find the interquartile range for the following dataset representing the scores of 10 students in a mathematics test: 35, 42, 48, 55, 60, 63, 68, 70, 72, 78

(a) 12 (b) 24
(c) 46.5 (d) None of these

4. If the first quartile is 56 and the third quartile is 77, then the coefficient of quartile deviation is

(a) 18.09 (b) 15.79 (c) 63.80 (d) 56.71

5. Find the value of the third quartile if the values of first quartile and quartile deviation are 90 and 20 respectively.

(a) 100 (b) 130 (c) 110 (d) 220

6. If the first quartile is 48 and quartile deviation is 6, find the median. (assuming the distribution to be symmetrical).

(a) 54 (b) 48 (c) 42 (d) 6

7. Find the interquartile range and the coefficient of quartile deviation from the following data:

Marks less than	10	20	30	40	50	60	70	80
No. of Students	4	16	40	76	96	112	120	125

(a) 50 and 25 (b) 20 and 25 (c) 50 and 30 (d) None

8. If the quartile deviation of x is 5 and 2x + 5y = 10, what is the quartile deviation of y?

(a) 2 (b) 4 (c) 1 (d) cannot be determined

9. The quartiles of a variable are 45, 52 and 65 respectively. Its quartile deviation is

(a) 10 (b) 20 (c) 25 (d) 8.30.

10. If x and y are related as 3x + 4y = 20 and the quartile deviation of x is 12, then the quartile deviation of y is

(a) 16 (b) 14 (c) 10 (d) 9

11. Calculate quartile deviation from the following distribution:

X	5-7	8-10	11-13	14-16	17-19
Frequency	14	24	38	20	4

(a) 1.55 (b) 2.27 (c) 3.05 (d) None of these

12. Find the value of the third quartile if the values of first quartile and quartile deviation are 104 and 18 respectively.

(a) 100 (b) 110 (c) 120 (d) 140

Answer Key									
1. (c)	**2.** (a)	**3.** (b)	**4.** (b)	**5.** (b)	**6.** (a)	**7.** (b)	**8.** (a)	**9.** (a)	**10.** (d)
11. (b)	**12.** (d)								

MERITS

- Quartile deviation provides the best measure of dispersion for open-end classification. In fact, it is the only measure of dispersion which can be obtained while dealing with a distribution having open-end classes.
- It is also less affected due to sampling fluctuations. It is not affected at all by extreme observations as it ignores 25% of the data from the beginning of the distribution and another 25% of the data from the top end.
- Quartile deviation is useful especially when it is desired to study variability in the central half part of the data.
- Like other measures of dispersion, quartile deviation remains unaffected due to a change of origin but is affected in the same ratio due to change in scale.

DEMERITS

1. Quartile deviation is not based on all the observations. In fact, it ignores 25% of the data at the lower end and 25% of the data at the upper end. Hence it cannot be considered as a good measure of dispersion.
2. Quartile deviation is not suitable for further mathematical treatment.
3. It is affected considerably by sampling fluctuations.

STANDARD DEVIATION

The standard deviation, abbreviated as S.D. of a given set of observations is defined as the positive square root of the arithmetic mean of the squares of deviations of the observations from their arithmetic mean. It is denoted by the Greek letter (read as sigma).

Thus, standard deviation of a set of n observations $X_1, X_2, \ldots, X_n$ is given by

$$\sigma = \sqrt{\frac{\sum (X - \bar{X})^2}{N}} \text{ where } \bar{X} = \frac{\sum X}{n}$$

If $X_1, X_2, \ldots\ldots, X_n$ are the class marks of a set of grouped data with class frequencies $f_1, f_2, \ldots f_n$, then the standard deviation is given by

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

Example 29. The heights (in centimeters) of a group of students are as follows: 150, 160, 165, 155, 170, 155, 160, 155. What is the standard deviation of the heights?

(a) 4.16 (b) 5.20 (c) 6.62 (d) 7.07

Sol. (d) Given observations: 150, 160, 165, 155, 170, 155, 170, 155

∴ Sum of the observations =150 + 160 + 165 + 155 + 170 + 155 + 170 + 155 = 1280

Now, mean of the data $= \frac{1280}{8} = 160$

Thus, the sum of squares of the deviation from mean

$$\sum_{i=0}^{N}\left(x_i - \bar{x}\right)^2 = 10^2 + 0^2 + 5^2 + 5^2 + 10^2 + 5^2 + 10^2 + 5^2$$

= 100 + 0 + 25 + 25 + 100 + 25 + 100 + 25 = 400

$$\text{Therefore, S.D} = \sqrt{\frac{\sum_{i=0}^{N}\left(x_i - \bar{x}\right)^2}{n}}$$

$$\text{S.D} = \sqrt{\frac{400}{8}}$$

$$\text{S.D} = \sqrt{50} = 7.07$$

Hence, the correct option is (d).

Example 30. What is the standard deviation of 5, 5, 9, 9, 9, 10, 5, 10, 10? **(ICAI)**

(a) $\sqrt{14}$ (b) $\frac{\sqrt{42}}{3}$ (c) 4.50 (d) 8

Sol. (b) Given observations: 5, 5, 9, 9, 9, 10, 5, 10, 10

Sum of the observations = 5 + 5 + 5 + 10 + 10 + 10 + 9 + 9 = 72

Now, mean of the data

Thus, the sum of squares of the deviation from mean n

$$\sum_{i=0}^{N}(x_i - \bar{x})^2 = 3^2 + 3^2 + 3^2 + 2^2 + 2^2 + 2^2 + 1^2 + 1^2 + 1^2$$

= 9 + 9 + 9 + 1 + 1 + 1 + 4 + 4 + 4 = 42

$$\text{Therefore, S.D} = \sqrt{\frac{\sum_{i=0}^{N}(x_i - \bar{x})^2}{n}}$$

$$\text{S.D} = \sqrt{\frac{42}{9}}$$

$$\text{S.D} = \frac{\sqrt{42}}{3}$$

Hence, the correct answer is option (b) i.e., $\frac{\sqrt{42}}{3}$.

Example 31. Calculate the standard deviation for the following data:

X	20	30	40	50	60	70
Frequency	8	12	20	10	6	4

(a) 13.75 (b) 14.50 (c) 10.15 (d) None of these

Sol. (a) According to the given data,

X	f	fX	$x - \bar{x}$	$(x - \bar{x})^2$	$f(x - \bar{x})^2$
20	8	160	–21	441	3528
30	12	360	–11	121	1452
40	20	800	–1	1	20
50	10	500	9	81	810
60	6	360	19	361	2166
70	4	280	29	841	3364
	N = 60	$\sum fx = 2460$			$\sum f(x - \bar{x})^2 = 11340$

Therefore, $\bar{X} = \frac{\sum fx}{N} = \frac{2460}{60} = 41$

Now, Standard deviation $(\sigma) = \sqrt{\frac{\sum f(X - \bar{X})^2}{N}} = \sqrt{\frac{11340}{60}} = \sqrt{189} = 13.75$

Hence, the correct option is (a).

Example 32. Calculate the standard deviation from the following data:

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of students	10	15	25	25	10	10	5

(a) 11.65 (b) 13.39 (c) 14.40 (d) 15.94

Sol. (d) According to the given data,

Marks	Mid-value (X)	f	fX	$x - \bar{x}$	$(x - \bar{x})^2$	$f(x - \bar{x})^2$
0-10	5	10	50	–26	676	6760
10-20	15	15	225	–16	256	3840
20-30	25	25	625	–6	36	900
30-40	35	25	875	4	16	400
40-50	45	10	450	14	196	1960
50-60	55	10	550	24	576	5760
60-70	65	5	325	34	1156	5780

		N = 100	$\sum fx = 3100$			$\sum f(x - \bar{x})^2 = 25400$

Therefore, $\bar{X} = \frac{\sum fx}{N} = \frac{3100}{100} = 31$

Now, Standard deviation = $(\sigma) = \sqrt{\frac{\sum f(X - \bar{X})^2}{N}} = \sqrt{\frac{25400}{100}} = \sqrt{254} = 15.94$

Hence, the correct option is (d).

DIFFERENT METHODS OF CALCULATING STANDARD DEVIATION: UNGROUPED DATA

The standard deviation of an ungrouped data consisting of N observations $X_1, X_2, ..., X_n$ is given by

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

The computation of standard deviation is very effective if $\bar{X}$ is an integer. However, if X comes out to be in fraction, its computation becomes very cumbersome and time-consuming. In that case we apply the following short-cut method which is very effective and reduces the numerical calculations to a great extent.

DIFFERENT METHODS OF CALCULATING STANDARD DEVIATION – GROUPED DATA

All the methods discussed earlier for calculating standard deviation in the case of ungrouped data can also be used in the case of grouped data. However, in practice it is the step deviation method that is mostly used.

Standard deviation is given by: $\sigma = \sqrt{\frac{\sum fx^2}{N} - \left(\frac{\sum fx}{N}\right)^2}$

Example 33. From the following information, find standard deviations of X and Y variables:

$\sum X = 235, \sum X = 250, \sum X^2 = 6750, \sum Y^2 = 6840, N = 10$

(a) 11.08 and 7.68
(b) 12.55 and 8.06
(c) 29.50 and 16.76
(d) None of these

Sol. (a) Given: $\sum X = 235, \sum X = 250, \sum X^2 = 6750, \sum Y^2 = 6840, N = 10$

We know that, $\sigma_x = \sqrt{\frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2} = \sqrt{\frac{6750}{10} - \left(\frac{235}{10}\right)^2} = \sqrt{675 - (23.5)^2}$

$= \sqrt{675 - 552.25} = \sqrt{122.75} = 11.08$

$$\sigma_Y = \sqrt{\frac{\sum Y^2}{N} - \left(\frac{\sum Y}{N}\right)^2} = \sqrt{\frac{6840}{10} - \left(\frac{250}{10}\right)^2} = \sqrt{684-(25)^2} = \sqrt{684-625}$$

$$= \sqrt{59} = 7.68$$

Hence, the correct option is (a).

Example 34. From the following data, calculate the standard deviation:

X	10	11	12	13	14	15	16	17	18
Frequency	2	7	10	12	15	11	10	6	3

(a) 1.50 (b) 1.986 (c) 2.576 (d) None of these

Sol. (b) According to the given data, we have

X	f	d = X – 14	fd	fd²
10	2	–4	–8	32
11	7	–3	–21	63
12	10	–2	–20	40
13	12	–1	–12	12
14	15	0	0	0
15	11	1	11	11
16	10	2	20	40
17	6	3	18	54
18	3	4	12	48
	N = 76		$\sum fd = 0$	$\sum fd^2 = 300$

Therefore, Standard deviation $= \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} = \sqrt{\frac{300}{76}} = 1.986$

Hence, the correct option is (b).

VARIANCE

The variance of a given set of observations is defined as the square of its standard deviation and is denoted by σ^2.

Thus,

For individual observations, $\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n}$ or $\sigma^2 = \frac{\sum x^2}{n} - (\bar{x})^2$

For Ungrouped/Grouped Observations, $\sigma^2 = \frac{\sum f_i (x_i - \bar{x})^2}{n}$

Example 35. $\sum x^2 = 3390$, $n = 30$, $\sigma = 7$, then $\bar{x}$ = ________

(a) 113 (b) 210 (c) 8 (d) None

Sol. (c) We know that, $\sigma^2 = \frac{\sum x^2}{n} - (\bar{x})^2$

Given that, $\sum x^2 = 3390$ $n = 30$, $\sigma = 7$

Putting the values, we get

$$\sigma^2 = \frac{3390}{30} - (\bar{x})^2$$

$$\Rightarrow 49 = 113 - (\bar{x})^2$$

$$\Rightarrow (\bar{x})^2 = 64 \Rightarrow \bar{x} = 8$$

Hence, the correct option is (c) i.e. 8.

Example 36. Find variance if $\Sigma D^2 = 150$ and $N = 6$ Deviations are taken from actual mean.

(a) 5 (b) 25 (c) 36 (d) None of these

Sol. (b) Given: $\Sigma D^2 = 150$ and $N = 6$

Thus, Variance $= \frac{\sum (x - \bar{x})^2}{N} = \frac{\sum D^2}{N} = \frac{150}{6} = 25$

Therefore, the variance is 25.

Hence, the correct option is (b).

Example 37. If the standard deviation for the marks obtained by a student in monthly test is 36, then the variance is

(a) 36 (b) 6 (c) 1296 (d) None of these

Sol. (c) We know that,

Variance is the square of standard deviation.

Since, SD (σ) = 36,

thus Variance = $\sigma^2 = (36)^2 = 1296$

Hence, option (c) is correct i.e. 1296.

CORRECTING INCORRECT VALUES OF MEAN AND STANDARD DEVIATION

Example 38. The mean and variance of 100 items were worked out as 40 and 25 respectively by a student. By mistake an item 50 was wrongly taken as 5 in calculating the above. You are required to find the correct mean and correct standard deviation.

(a) 2.54 (b) 3.05 (c) 3.68 (d) None of these

Sol. (c) Given,

Number of observations, $n = 100$,

Wrong mean, $\bar{x}_w = 40$

Variance, $\sigma^2 = 25$

Now, $\bar{x} = \frac{\sum x}{n}$

$\Rightarrow \bar{x}_w = n\bar{x} = 100(40) = 4000$

Now, $\sum x^2 = n(\sigma^2 + \bar{x}^2)$ $\left[\because \sigma^2 = \frac{\sum x^2}{n} - (\bar{x})^2\right]$

= 100(25 + 1600)= 16250

Since, wrong observation 5 is to be replaced by correct observation 50, then

Thus: Corrected $\sum x = 4000 - 5 + 50 = 4045$

Corrected $\sum x^2 = 162500 - 5^2 + (50)^2 = 164975$

$$\text{Corrected } \bar{x} = \frac{\text{corrected} \sum x}{n} = \frac{4045}{100} = 40.45$$

$$\text{Corrected } \sigma = \sqrt{\frac{\text{corrected} \sum x^2}{n} - (\text{corrected } \bar{x})^2} = \sqrt{\frac{164975}{100} - (40.45)^2}$$

$= \sqrt{13.55} = 3.68$

Hence, the correct option is (c).

COMBINED STANDARD DEVIATION

If two sets of data contain n_1 and n_2 observations having means $\bar{x}_1$ and $\bar{x}_2$, and standard deviations σ_1 and σ_2 respectively, then the standard deviation, σ, of the combined data with $n_1 + n_2$ observations is given by:

- $$\sigma = \sqrt{\frac{n_1\left(\sigma_1^2 + d_1^2\right) + n_2\left(\sigma_2^2 + d_2^2\right)}{n_1 + n_2}}$$

- $$\bar{X} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2}$$

The result can be generalized to more than two sets of data.

E.g.: if $\bar{X}_1, \bar{X}_2, \ldots \bar{X}_k$ be the means, σ_1, σ_2, ... σ_k be the standard deviations and n_1, n_2, ..., n_k be the number of observations in each set, then the standard deviation of the combined data with $n_1 + n_2 + \ldots\ldots + n_k$ observations is given by

- $$\sigma = \sqrt{\frac{n_1\left(\sigma_1^2 + d_1^2\right) + n_2\left(\sigma_2^2 + d_2^2\right) + \ldots\ldots\ldots + n_k\left(\sigma_k^2 + \sigma_k^2\right)}{n_1 + n_2 + \ldots\ldots + n_k}}$$

where $d_1 = \bar{X}_1 - \bar{X}, d_2 = \bar{X}_2 - \bar{X}, \ldots, d_k = \bar{X}_k - \bar{X}$

- $\bar{X} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \ldots + n_k\bar{x}_k}{n_1 + n_2 + \ldots + n_k}$ is the combined mean.

Example 39. If two samples of sizes 30 and 20 have means as 55 and 60, and variances as 16 and 25 respectively, then what would be the SD of the combined sample of size 50?

(a) 5.00 (b) 5.06 (c) 5.23 (d) 5.35

Sol. (b) According to the question, Variances, $\sigma_1^2 = 16$ and $\sigma_2^2 = 25$

$n_1 = 30$ and $n_2 = 20$

$\bar{x}_1 = 55$ and $\bar{x}_2 = 60$

Thus, combined mean $\bar{x} = \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} = \dfrac{32(55) + 20(60)}{30 + 20} = \dfrac{1650 + 1200}{50}$

$= \dfrac{2850}{50} = 57$

Thus, $d_1 = \bar{x}_1 - \bar{x} = 55 - 57 = -2$

$d_2 = \bar{x}_2 - \bar{x} = 60 - 57 = 3$

We know that,

Combined SD, $\sigma = \sqrt{\dfrac{n_1\sigma_1^2 + n_2\sigma_2^2 + n_1 d_1^2 + n_2 d_2^2}{n_1 + n_2}}$

$= \sqrt{\dfrac{30(16) + 20(25) + 30(4) + 20(9)}{30 + 20}}$

$= \sqrt{\dfrac{480 + 500 + 120 + 180}{50}} = \sqrt{\dfrac{1280}{50}} = 5.06$

Hence, the correct option is (b).

COEFFICIENT OF VARIATION

- The standard deviation is an absolute measure of dispersion, depending upon the units of measurement. It does not tell us much about the variability of a single set of data. The coefficient of standard deviation, based on standard deviation, is a relative measure of dispersion.
- This is a pure number independent of the units of measurement and hence can be used to compare the variability of two distributions expressed in different units.
- Perhaps a more appropriate measure is the coefficient of variation (C.V.), defined by

  Coefficient of Variation $= \dfrac{\text{S.D.}}{\bar{x}} \times 100$

  The above expression expresses the standard deviation as a percentage of the mean.
- Since C.V. is a measure of relative variation expressed as a percent, the coefficient of variation can be used to compare the variability of two or more sets of data even when observations are expressed in different units of measurement.
- A distribution for which the coefficient of variation is smaller is said to be less variable or more consistent, more uniform, more stable or more homogeneous.
- On the other hand, the distribution for which the coefficient of variation is greater is said to be more variable or less consistent, less uniform, less stable or less homogeneous.

Example 40. If Mean = 5, Standard deviation = 2.6, then the coefficient of variation is

(a) 49 (b) 51 (c) 50 (d) 52 **(ICAI)**

Sol. (d) We have, Mean = 5, Standard deviation = 2.6

We know that,

$$\text{Coefficient of variation} = \frac{\text{Standard deviation} \times 100}{\text{mean}}$$

$$= \frac{2.6 \times 100}{5} = 52$$

Hence, the correct option is (d) i.e.,52 .

Example 41. If the coefficient of variation and standard deviation are 30 and 12 respectively, then the arithmetic mean of the distribution is **(Dec 2022)**

(a) 40 (b) 36 (c) 25 (d) 19

Sol. (a) Given: Coefficient of variation = 30, Standard deviation (σ) = 12

We know that,

$$\text{Coefficient of variation} = \frac{\sigma}{\bar{X}} \times 100$$

$$30 = \frac{12}{X} \times 100$$

$$\Rightarrow \bar{X} = \frac{12}{30} \times 100 \Rightarrow \bar{X} = 40$$

Therefore, the arithmetic mean of the distribution is 40.

Example 42. What is the coefficient of variation of the following numbers?

(a) 8.09 (b) 18.08 (c) 20.23 (d) 20.45

Sol. (a) Given observations:

$\therefore$ Mean of the data

The sum of the squares of deviation from mean

The sum of the squares of deviation from mean $\sum_{i=0}^{N}(x_i - \bar{x})^2$

$= 5^2 + 6^2 + 3^2 + 2^2 + 6^2 = 25 + 36 + 9 + 4 + 36 = 110$

We know that.

$$\text{Standard deviation,S.D.} = \sqrt{\frac{\sum_{i=0}^{N}(x_i - \bar{x})^2}{n}}$$

$$\Rightarrow \text{S.D} = \sqrt{\frac{110}{5}} \Rightarrow \text{S.D} = \sqrt{22}$$

$$\text{Coefficient of variation} = \frac{\text{S.D}}{\bar{x}} \times 100 = \frac{\sqrt{22}}{58} \times 100 = 8.09$$

Hence, the correct answer is option (a) i.e., 8.09.

Example 43. The sum of squares of deviation from mean of 10 observations is 250. Mean of the datais 10. Find the co-efficient of variation.

(a) 10% (b) 25% (c) 50% (d) 0%

Sol. (c) Given, N ;N = 10; $\sum(x-\bar{x})^2 = 250$ and Mean = 10

We know that,

Standard Deviation, $\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{N}} = \sqrt{\frac{250}{10}} = 5$

Thus,

Coefficient of variance $= \frac{\sigma}{N} \times 100 = \frac{5}{10} \times 100 = 50\%$

Hence, the correct option is (c) i.e. 50%.

Example 44. Mean of a series is equal to 100, coefficient of variation is 45% then the S.D. is

(a) 45 (b) 0.45 (c) 4.5 (d) 40.5

Sol. (a) We know that, $(C.V.) = \frac{\sigma}{\bar{x}} \times 100$

$\bar{X} = 100, C.V = 45$

Given that $45 = \frac{\sigma}{100} \times 100$

Now, $\sigma = 45$

Hence, the correct option is (a) i.e., 45.

PROPERTIES FOR STANDARD DEVIATION

Standard deviation is independent of change of origin but not of scale, i.e., if there are two variables x and y related as y = a + bx for any two constants a and b, then SD of y is given by:

$$s_y = |b| \times s_x$$

Example 45. If x and y are related by 2x + 3y + 4 = 0 and SD of x is 6, then SD of y is

(a) 22 (b) 4 (c) 5 (d) 9

Sol. (b) Given relation. 2x + 3y + 4 = 0 and S.D. of x = 6

$\Rightarrow 3y = -2x - 4$

$\Rightarrow y = \frac{-2x-4}{3}$

$\Rightarrow y = \frac{-2x}{3} - \frac{4}{3}$

Thus, SD of $y = \left|\frac{-2}{3}\right| \times SD = \frac{2}{3} \times 6 = 4$

Hence, option (b) is the correct answer.

Example 46. If AM and C.V of a random variable x are 10 & 40 respectively, then the variance of $\left(-15+\frac{3x}{2}\right)$

(a) 64 (b) 81 (c) 49 (d) 36

Sol. (d) We know that,

Coefficient of variance $(C.V)=\frac{\sigma}{\bar{x}}\times 100$

Given, $C.V=40, \bar{X}=10$

then, $40=\frac{\sigma}{10}\times 100$

Thus, $\Rightarrow \sigma=\frac{400}{100}=4$

Thus, $S.D\left(-15+\frac{3x}{2}\right)=\frac{3}{2}\times SD(X)$

$\Rightarrow S.D\left(-15+\frac{3X}{2}\right)=\frac{3}{2}\times 4=6$

Therefore, variance of $\left(-15+\frac{3x}{2}\right)=6^2=36$

Hence, option (d) is correct i.e. 36 .

Note:

- Standard deviation is suitable for further mathematical treatment. For instance, if we know the sizes, means and standard deviations of two or more sets of data, then we can obtain the standard deviation of the combined data.
- The standard deviation of first natural numbers is $\sqrt{\frac{n^2-1}{12}}$.
- If all the observations assumed by a variable are constant i.e., equal, then the SD is zero. This means that if all the values taken by a variable x is k, say, then s = 0. This result applies to range as well as mean deviation.

Example 47. If the S.D. of the 1st n natural numbers is $\sqrt{30}$, then the value of n is

(a) 19 (b) 20 (c) 21 (d) None of these

Sol. (a) We know that, SD for first n natural numbers is given by

$$SD=\sqrt{\frac{n^2-1}{12}}=\sqrt{30}\Rightarrow \frac{n^2-1}{12}=30$$

$\Rightarrow n^2=360+1$

$\Rightarrow n^2=361$

$\Rightarrow n = 19$

Therefore, the value of n is 19.

Hence, option (a) is correct i.e.19 .

Example 48. The mean and SD for a, b and 2 are 3 and $\frac{2}{\sqrt{3}}$ respectively, The value of ab would be **(ICAI)**

(a) 5 (b) 6 (c) 11 (d) 3

Sol. (c) Given: Mean of a, b and 2 is 3.

Thus, $\frac{a+b+2}{3} = 3$

$\Rightarrow a + b + 2 = 9$

$\Rightarrow a + b = 7$

Also, S.D $= \frac{2}{\sqrt{3}}$

$\Rightarrow$ Variance $= \frac{4}{3}$

We know that, Variance $(V(X)) = \frac{\sum_{i=1}^{n} X_i^2}{n} - \left(\frac{\sum_{i=1}^{n} X_i}{n}\right)^2$

$\Rightarrow \frac{4}{3} = \frac{\sum_{i=1}^{n} X_i^2}{n} - \left(\frac{\sum_{i=1}^{n} X_i}{n}\right)^2 \Rightarrow \frac{4}{3} = \frac{a^2 + b^2 + 4}{3} - 9$

$\Rightarrow \frac{4}{3} = \frac{a^2 + b^2 + 4 - 27}{3}$

$\Rightarrow 4 = a^2 + b^2 - 23$

$\Rightarrow a^2 + b^2 = 27$

$\Rightarrow (a + b)^2 - 2ab = 27$ [$\because a^2 + b^2 + 2ab = (a + b)^2$]

$\Rightarrow (7)^2 - 2ab = 27$ [$\because$ From (i)]

$\Rightarrow 49 - 27 = 2ab$

$\Rightarrow 2ab = 22$

$\Rightarrow ab = 11$

Hence, the correct answer is option (c) i.e., 11.

Example 49. Which of the following companies A or B is more consistent so far as the payment of dividend is concerned? **(ICAI)**

Dividend paid byA	5	9	6	12	15	10	8	10
Dividend paid by B	4	8	7	15	18	9	6	6

(a) A (b) B (c) Both A & B (d) Neither A nor B

Sol. (a) We are given that

Dividend paid by A	5	9	6	12	15	10	8	10
Dividend paid by B	4	8	7	15	18	9	6	6

To check the consistency for the payment of dividend by A and B; we will find the coefficient of variation of A & B.

Let dividend paid by A be x, dividend paid by B be y

Then, $\sum X = 5 + 9 + 6 + 12 + 15 + 10 + 8 + 10 = 75$

$\Rightarrow \overline{X} = \frac{75}{8} = 9.375$

Also, $\sum X^2 = 5^2 + 9^2 + 6^2 + 12^2 + 15^2 + 10^2 + 8^2 + 10^2 = 775$

Now, SD is given by (σ_A)

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

$$\sqrt{\frac{775}{8} - \left(\frac{75}{8}\right)^2} = \sqrt{\frac{775}{8} - \frac{5625}{64}} = 2.99 \sim 3$$

Thus, $C.V_A = \frac{\sigma_A}{\overline{X}} \times 100 = \frac{3}{9.375} \times 100 = 32$

Similarly, $\sum Y = 73; \overline{Y} = \frac{73}{8} = 9.125, \sum Y^2 = 831$

$$\Rightarrow \sigma_B^2 = \frac{831}{8} - \left(\frac{73}{8}\right)^2 = \frac{831}{8} - \frac{5329}{64} = 20.61$$

Now, $C.V_B = \frac{4.54}{9.125} \times 100 = 49.75$

Since, $C.V_A < C.V_B$

Therefore, company A is more consistent than company B.

Hence, option (a) is correct.

PRACTICE QUESTIONS (PART D)

1. The variance of given data is 12 and their mean value is 40, what is the coefficient of variation (CV)?

(a) 5.66% (b) 6.66% (c) 7.50% (d) 8.65%

2. For any two numbers, SD is always **(ICAI)**

(a) Twice the range
(b) Half of the range
(c) Square of the range
(d) None of these

3. If all the observations are increased by 10, then **(ICAI)**

(a) SD would be increased by 10
(b) Mean deviation would be increased by 10

(c) Quartile deviation would increase by 10
(d) All these three remain unchanged.

4. If all the observations are multiplied by 2, then **(ICAI)**
(a) New SD would be also multiplied by 2
(b) New SD would be half of the previous SD
(c) New SD would be increased by 2
(d) New SD would be increased by 2

5. If the SD of x is 3, what is the variance of (5 – 2x)?
(a) 36 (b) 6 (c) 1 (d) 9

6. If x and y are related by y = 2x + 5 and the SD and AM of x are known to be 5 and 10 respectively, then the coefficient of variation is
(a) 25 (b) 30 (c) 40 (d) 20

7. Calculate the arithmetic mean and standard deviation from the following series:

Class Interval	5-15	15-25	25-35	35-45	45-50
Dividend paid by B	8	12	15	9	9

(a) 12.33 (b) 14.53 (c) 18.75 (d) None of these

8. The monthly test marks obtained by a student in Business Mathematics (out of 10) are 5, 9, 6, 12, 15, 10, 8, 10. The standard deviation for the student's marks is
(a) 5.05 (b) 1.80 (c) 2.99 (d) 3.51

9. The mean and standard deviation of 100 items is found to be 40 and 10. If at the time of calculations, two items are wrongly taken as 30 and 70 instead of 3 and 27, find the corrected mean and corrected standard deviation?
(a) 40 and 10 (b) 10.80 and 8.65
(c) 39.30 and 10.24 (d) 30.50 and 9.50

10. What is the standard deviation from the following data relating to the age distribution of 200 persons? **(ICAI)**

Age (year):	20	30	40	50	60	70	80
No. of people:	13	28	31	46	39	23	20

(a) 15.29 (b) 16.87 (c) 18.00 (d) 17.52

11. What is the coefficient of variation for the following distribution of wages? **(ICAI)**

Daily Wages (₹)	30-40	40-50	50-60	60-70	70-80	80-90
No of workers	17	28	21	15	13	6

(a) ₹14.73 (b) 14.73 (c) 26.53 (d) 20.82

12 The mean and SD of a sample of 100 observations were calculated as 40 and 5.1 respectively by a CA student who took one of the observations as 50 instead of 40 by mistake. The current value of SD would be **(ICAI)**
(a) 4.90 (b) 5.00 (c) 4.88 (d) 4.85.

Answer Key

1. (d) 2. (b) 3. (d) 4. (a) 5. (a) 6. (c) 7. (a) 8. (c) 9. (c) 10. (b) 11. (c) 12. (b)

PRACTICE QUESTIONS (PART E)

1. The sum of mean and SD of series is a + b, if we add 2 to each observation of the series then the sum of mean and SD is
   (a) a + b + 2 (b) 6 − a + b (c) 4 + a − b (d) a + b +4
2. If the coefficient of quartile deviation is $\frac{1}{4}$, then $\frac{Q_3}{Q_1}$ is
   (a) 5/3 (b) 4/3 (c) 3/4 (d) 3/5
3. For the distribution:

X	1	2	3	4	5	6
F	6	9	10	14	12	8

   The value of median is
   (a) 3.5 (b) 3 (c) 4 (d) 5
4. For a symmetric distribution
   (a) Mean = Median = Mode (b) Mode = 3 median − 2 mean
   (c) Mode $= \frac{1}{3}$ median $= \frac{1}{2}$ Mean (d) None of these
5. If the profits of a company remain some for the last ten months, then the S.D of profit of the company would be:
   (a) Positive (b) Negative (c) Zero (d) (a) or (b)
6. S.D of first five consecutive natural numbers is
   (a) $\sqrt{10}$ (b) $\sqrt{8}$ (c) $\sqrt{3}$ (d) $\sqrt{2}$
7. Which of the following is positional average?
   (a) Median (b) GM (c) HM (d) AM
8. The Q.D of 6 numbers 15, 8, 36, 40, 38, 41 is equal to
   (a) 12.4 (b) 25 (c) 13.5 (d) 37
9. Standard deviation is __________ times of $\sqrt{MD \times QD}$.
   (a) $2\sqrt{3}$ (b) 4/5 (c) $\sqrt{\frac{15}{8}}$ (d) $\sqrt{\frac{8}{15}}$
10. If the points of inflexion of a normal curve are 40 and 60 respectively, then its mean deviation is
   (a) 8 (b) 45 (c) 50 (d) 60

11. In a moderately skewed distribution the values of mean & median are 12 & 8 respectively. The value of mode is

(a) 0 (b) 12 (c) 15 (d) 30

12. The AM of 15 observations is 9 and the AM of first 9 observations is 11 and then AM of remaining observation is

(a) 11 (b) 6 (c) 5 (d) 9

13. If $\sigma^2 = 100$ and coefficient of variation = 20% then $\bar{X} =$

(a) 60 (b) 70 (c) 80 (d) 50

14. If the mean of the following distribution is 6 then the value of P is

X	2	4	6	10	P + 5
F	3	2	3	1	2

(a) 7 (b) 5 (c) 8 (d) 11

15. Find the interquartile range for the following dataset representing the scores of 10 students in a mathematics test:

35, 42, 48, 55, 60, 63, 68, 70, 72, 78

(a) 12 (b) 24 (c) 46.5 (d) None of these

16. If in a moderately skewed distribution the values of mode and mean are 32.1 and 35.4 respectively, then value of the median is

(a) 34.3 (b) 33.3 (c) 34 (d) 33

17. Standard deviation for the marks obtained by a student in monthly test in mathematics (out of 50) as 30, 35, 25, 20, and 15 is

(a) 25 (b) $\sqrt{50}$ (c) $\sqrt{30}$ (d) 50

18. If the variation of 5.7 and 11 is 4, then the coefficient of variation is :

(a) 15 (b) 25 (c) 17 (d) 19

19. If the range of a set of values is 65 and maximum values in the set is 83 , then the minimum values in the set is

(a) 74 (b) 9 (c) 18 (d) None of these

20. Which one of the following is not a central tendency?

(a) Mean Deviation (b) Arithmetic mean

(c) Median (d) Mode

21. The Algebraic sum of the deviation of a set of values from their arithmetic mean is

(a) greater than 0 (b) equal to 0 (c) less than 0 (d) None of these

22. The Geometric mean of 3, 6, 24, and 48 is

(a) 8 (b) 12 (c) 24 (d) 20

23. The mean of 20 items of a data is 5 and If each item is multiplied by 3, then the new mean will be

(a) 5 (b) 10 (c) 15 (d) 20

24. The median of the data 5, 6, 7, 7, 8, 9, 10, 11, 11, 12, 15, 18, 18, and 19 is

(a) 10.5 (b) 10 (c) 11 (d) 11.5

25. The average of a series of overlapping averages, each of which is based on a certain number of items within a series is known as

(a) Moving average (b) Weighted average
(c) Simple average (d) None of these

26. For 899, 999, 391, 384, 390, 480, 485, 760, 111, 240 then rank of median is

(a) 2.7 (b) 5.5 (c) 8.25 (d) None

27. If each item is reduced by 15, then A.M is

(a) Reduced by 15 (b) Increased by 15 (c) Reduced by 10 (d) None

28. $\frac{(Q_3 - Q_1)}{(Q_3 + Q_1)}$ is known as

(a) Coefficient of Range. (b) Coefficient of Q.D.
(c) Coefficient of S.D (d) Coefficient of M. D.

29. Mean deviation is the least when deviation are taken from

(a) Mean (b) Median (c) Mode (d) Harmonic mean

30. If the mean value of seven numbers 7, 9, 12, x, 4, 11, and 5 is 9, then the missing number x will be:

(a) 13 (b) 14 (c) 15 (c) 8

Answer Key

1. (a)	2. (a)	3. (c)	4. (a)	5. (c)	6. (d)	7. (a)	8. (c)	9. (c)	10. (a)
11. (a)	12. (b)	13. (d)	14. (a)	15. (b)	16. (a)	17. (b)	18. (b)	19. (c)	20. (a)
21. (b)	22. (b)	23. (c)	24. (a)	25. (a)	26. (a)	27. (a)	28. (b)	29. (b)	30. (c)

SUMMARY

MEASURES OF CENTRAL TENDENCY

- **Arithmetic mean:** Sum of the observations divided by the number of observations

$$\text{Arithmetic Mean (Mean)} = \frac{\text{Sum of observations}}{\text{Number of observations}}$$

Methods to find mean:

(i) **Direct method:** $\text{Mean} = \frac{f_1x_1 + f_2x_2 + \ldots + f_nx_n}{f_1 + f_2 + \ldots + f_n} = \frac{\sum_{i=1}^{n} f_ix_i}{\sum_{i=1}^{n} f_i}$

(ii) **Step-deviation method:** $\text{Mean} = a + \frac{\sum_{i=1}^{n} f_iu_i}{\sum_{i=1}^{n} f_i} \times h$

- **Geometric mean:**
  (i) For individual series: $GM = (x_1 x_2 x_3 \ldots\ldots\ldots\ldots x_n)^{1/n}$

  (ii) For discrete series: $G.M. = AL\left[\frac{1}{N}\sum f \log X\right]$

  (iii) For continuous series: $G.M. = AL\left[\frac{1}{N}\sum f \log X\right]$

- **Harmonic mean:**

  (i) For individual series: $H.M. = \dfrac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\ldots+\frac{1}{x_n}}$

  (ii) For discrete series: $HM = \dfrac{n}{\sum\left[\frac{f}{X}\right]}$

  (iii) For continuous series: $HM = \dfrac{n}{\sum\left[\frac{f}{X}\right]}$

- **Mode:** The value that occurs the maximum number of times.

  For continuous frequency distribution: $Mode = l + \dfrac{f_0 - f_{-1}}{2f_0 - f_{-1} - f_1} \times c$

- **Median:**
  (i) **For individual series:** Arrange the dataset in ascending order or descending order, then

  (a) If n is odd, the value is given by: $\left(\frac{n+1}{2}\right)$th observation

  (b) If n is even, the value is given by: $\dfrac{\frac{n}{2}\text{th observation} + \left(\frac{n}{2}+1\right)\text{th observation}}{2}$

  (ii) **For discrete series:**

  Median is the size of $\left(\frac{N+1}{2}\right)$th observation

  (iii) **For continuous series:**

  $$Median = l + \frac{\frac{N}{2} - c.f}{f} \times h$$

- **Quartiles** divide the entire dataset into four equal parts. So: there are three quartiles; first second and third represented by Q_i, Q_2 and Q_3 respectively. Q_l is the lower quartile and median of the lower half of the data set. Q_2 is the median.
- **Deciles** divide the entire dataset into ten equal parts. There are 9 deciles
- **Percentiles** divide the entire dataset into hundred equal parts. There are 99 percentiles.

TABLE FOR PARTITION VALUE

Calculation for	Quartiles	Deciles	Percentiles
Individual Observations	Q_k = size of $\frac{k(n+1)}{4}$th observation	$D_k = \frac{k(n+1)}{10}$th observation	$P_k = \frac{k(n+1)}{100}$th observation
Discrete Series	The value of X corresponding to the c.f. just greater than or equal to $\frac{k(n+1)}{4}$	The value of X corresponding to the c.f. just greater than or equal to $\frac{k(n+1)}{10}$	The value of X corresponding to the c.f. just greater than or equal to $\frac{k(n+1)}{100}$
Continuous Series	$Q_k = l + \frac{\frac{KN}{4} - C}{f} \times h$ = lower limit of Qk class C = c.f. of the class preceding the Qk class f = frequency of the Q_k class, h = size or width of Q_k class.	$D_k = l + \frac{\frac{KN}{10} - C}{f} \times h$	$P_k = l + \frac{\frac{KN}{100} - C}{f} \times h$

- Algebraic sum of deviations of a set of observations from their AM is zero i.e. $\Sigma(x_i - \overline{x}) = 0$
- AM is affected due to a change of origin and/or scale which implies that if the original variable is changed to another variable by effecting a change of origin,
- For given two positive numbers, A.M. × H.M.= $(G.M)^2$
- AM ≥ GM ≥ HM The equality sign occurs, as we have already seen, when all the observations are equal.
- Mode = 3 median – 2 mean

MEASURES OF DISPERSION

Range, interquartile range, and standard deviation are the three commonly used measures of dispersion.

- **Range** is the difference between the largest and the smallest observation in the data.
  - (i) For individual series: Range = Largest value – Smallest value
  - (ii) For grouped frequency distribution: Range = Upper class boundary (U.C.B) – Lower class boundary (L.C.B)
- **Coefficient of Range**:

$$\text{Coefficient of Range} = \frac{\text{Range}}{\text{Sum of the l arg est and the lowest values}}$$

$$\text{i.e., Coefficient of range} = \frac{L-S}{L+S} \times \ddot{u}$$

- **Quartile Deviation or** Semi Inter-Quartile Range:
  - **(i) Interquartile range** is defined as the difference between the 25th and 75th percentile (also called the first and third quartile). Hence the interquartile range describes the middle 50% of observations. If the interquartile range is large it means that the middle 50% of observations are spaced wide apart.

    Inter-quartile range = $Q_3 - Q_1$
  - (ii) **Quartile deviation**, also called semi inter-quartile range, is an absolute measure of dispersion defined by the formula:

$$\text{Quartile Deviation (Q.D.)} = \frac{Q_3 - Q_1}{2}$$

  - **(iii) Coefficient of Quartile Deviation**

$$\text{Coefficient of Quartile Deviation} = \frac{\text{Quartile Deviation}}{\text{Median}} \times 100 = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100$$

- **Mean Deviation:**
  - (i) Discrete Series: Mean Deviation (about an average A) $= \frac{\sum f_i \, |x_i - A|}{N}$
  - (ii) Continuous Series: Mean Deviation (about an average A) $= \frac{\sum f_i \, |x_i - A|}{N}$ where x_1, x_2, ……, x_n are the class marks (or mid-values) of a set of grouped data

Coefficient of Mean Deviation

The relative measure corresponding to the mean deviation, called the coefficient of mean deviation, is given by

1. Coefficient of M.D. = $\frac{1}{n}\sum |x_i - A|$

2. Coefficient of M.D. about mean $\frac{1}{n}\sum |x_i - \bar{x}|$ where, $\bar{x}$ is mean

3. Coefficient of M.D. about median = $\frac{1}{n}\sum |x_i - \text{Median}|$

- **Standard Deviation** (SD) is the most commonly used measure of dispersion. It is a measure of spread of data about the mean. SD is the square root of the sum of squared deviation from the mean divided by the number of observations.

  (i) For individual series: $\sigma = \sqrt{\frac{\sum (X - \bar{X})^2}{N}}$ where $\bar{X} = \frac{\sum X}{n}$

  (ii) For grouped data: $\sigma = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{n}}$

  (iii) For ungrouped data: $\aleph \sqrt{\frac{\sum x^2}{N} \quad \left(\frac{\sum x}{N}\right)^2}$

  Standard deviation of first natural numbers $= \sqrt{\frac{n^2 - 1}{12}}$

- **Variance:** It is the square of the standard deviation.

  (i) For individual observations, $\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n}$ or $\sigma^2 = \frac{\sum x^2}{n} - (\bar{x})^2$

  (ii) For Ungrouped/Grouped Observations, $\sigma = \frac{\sum f_i (x_i - \bar{x})}{n}$

- ***Coefficient of Variation:***

  Coefficient of Variation $= \frac{S.D.}{\bar{x}} \times 100$

❑❑❑