

# Statistics & Data Representation

## Statistics

Statistics is the process of drawing data. It includes collecting, classifying, summarising, analysing and interpreting numerical facts.

## Collection of Data

Collection of data is the first step in statistics towards achieving the goal on conclusion. There are different sources of collection of data.

*The two main types of data on the basis of collection are*

- (i) **Primary Data** The data collected actually in the process of investigation by the investigator is known as primary data.
- (ii) **Secondary Data** Data which is already collected by other persons is called as secondary data.  
e.g., As investigator collects data related to industries through the government publications.

## Important Definitions

### Frequency

The number of times a particular observation occurs is called frequency.

## Class Frequency

The frequency of a class in a continuous frequency distribution is called as class frequency.

As the frequency of class 50-60 is 6 and of 60-70 is 7.

## Class Marks

It is the mid value of the class interval.

Class mark

$$= \frac{\text{Lower limit of class} + \text{Upper limit of class}}{2}$$

## Range

It is the difference between the highest and the lowest values of the observation.

## Types of Frequency Distribution

Frequency distribution are of two types

### (i) Discrete Frequency Distribution

A frequency distribution is called a discrete frequency distribution if data are presented in such a way that exact measurements of the units are clearly shown.

Marks	Number of Students (Frequency)
40	7
60	3
80	3
100	2
<b>Total</b>	<b>15</b>

## (ii) Continuous Frequency Distribution

A frequency distribution in which data are arranged in classes or groups which are not exactly measurable.

Marks	Number of Students (Frequency)
10-20	1
20-30	2
30-40	3
40-50	4
50-60	6
60-70	7
<b>Total</b>	<b>23</b>

## Arithmetic Mean

The sum of all the observations is divided by the number of observations is called arithmetic mean.

### Mean of the Individual Data

If  $x_1, x_2, x_3, \dots, x_n$  are the  $n$  values of variable  $x$ , then the arithmetic mean (AM) is given by  $\bar{x}$ .

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

## Mean ( $\bar{x}$ ) for a Grouped Data

If each observation  $x_1, x_2, \dots, x_n$  have corresponding frequencies  $f_1, f_2, \dots, f_n$ ,

$$\text{then } \bar{x} = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i}$$

### Example 1 Find the mean of

$x, x+2, x+4, x+6$  and  $x+8$ .

- (a)  $x+3$  (b)  $x+4$   
(c)  $x+5$  (d) None of these

**Sol. (b)** Given,  $n = 5$

$$\begin{aligned} \therefore \bar{x} &= \frac{x + (x+2) + (x+4) + (x+6) + (x+8)}{5} \\ &= \frac{5x + 20}{5} = \frac{5(x+4)}{5} = x+4 \end{aligned}$$

**Example 2** If the arithmetic mean of 6, 8, 5, 7,  $p$  and 4 is 7. Find the value of  $p$ .

- (a) 12 (b) 11 (c) 10 (d) 9

**Sol. (a)** Given,  $\bar{x} = 7$  and  $n = 6$

$$\begin{aligned} \therefore \bar{x} &= \frac{6 + 8 + 5 + 7 + p + 4}{6} \\ \Rightarrow 7 &= \frac{30 + p}{6} \Rightarrow 42 = 30 + p \\ \Rightarrow p &= 42 - 30 \\ \therefore p &= 12 \end{aligned}$$

## Combined Mean

If  $\bar{x}_1$  and  $\bar{x}_2$  are the means of two groups computed from  $n_1$  and  $n_2$  values, then the mean ( $\bar{x}$ ) of the values of the two groups respectively taken together is given by

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

If there are  $k$  groups with means  $\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_k$  computed from  $n_1, n_2, n_3, \dots, n_k$  values respectively, then

$$\bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + n_3 \bar{x}_3 + \dots + n_k \bar{x}_k}{n_1 + n_2 + n_3 + \dots + n_k}$$

**Example 3** The mean monthly salary paid to 75 employees in a company is ₹ 1420. The mean salary of 25 of them is ₹ 1350 and that of 30 other is ₹ 1425. What is the mean salary of the remaining?

- (a) 1500 (b) 1400  
(c) 1450 (d) None of these

**Sol. (a)**  $n = n_1 + n_2 + n_3$

$$\Rightarrow 75 = 25 + 30 + n_3$$

$$\Rightarrow n_3 = 75 - (25 + 30) = 20$$

Now,  $\bar{x}_1 = 1350, \bar{x}_2 = 1425, \bar{x}_3 = ?$

$$\therefore \bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2 + n_3 \bar{x}_3}{n_1 + n_2 + n_3}$$

$$\Rightarrow 1420 = \frac{25 \times 1350 + 30 \times 1425 + 20 \times \bar{x}_3}{75}$$

$$\begin{aligned}\Rightarrow 1420 \times 75 &= 33750 + 42750 + 20\bar{x}_3 \\ \Rightarrow 20\bar{x}_3 &= 106500 - 76500 = 30000 \\ \therefore \bar{x}_3 &= \frac{30000}{20} = 1500\end{aligned}$$

## Median

The median is that value of the variable which divides the distribution into two equal parts such that the number of observations above is equal to the number of observations below it. Thus, after arranging the given data in an ascending or a descending order of magnitude, the value of the middle most observation, is called the median of the data.

### Median of Individual Data

Suppose, there are  $n$  observations, first arrange the given observations in ascending or descending order.

1. If  $n$  is odd, then Median = Value of  $\left(\frac{n+1}{2}\right)^{\text{th}}$  observation
2. If  $n$  is even, then

### Median

$$= \frac{1}{2} \left[ \text{Value of } \left(\frac{n}{2}\right)^{\text{th}} \text{ and } \left(\frac{n}{2} + 1\right)^{\text{th}} \text{ observations} \right]$$

**Example 4** The percentage of marks obtained by students of a class in Mathematics are 65, 35, 37, 23, 0, 18, 82, 92, 72, 32, 10, 5. Find the median.

- (a) 34.5                      (b) 33.5  
(c) 33                        (d) None of these

**Sol.** (b) First, we write the given data in ascending order.

0, 5, 10, 18, 23, 32, 35, 37, 65, 72, 82, 92

Here, the number of observations is even, i.e.  $n = 12$

$\therefore$  There are two middle terms  $\frac{12}{2}$ th and  $\left(\frac{12}{2} + 1\right)$ th,

i.e. 6th and 7th term.

So, the median is the mean of the values of the 6th and 7th terms.

$$\begin{aligned}\therefore \text{Median} &= \frac{32 + 35}{2} \\ & \quad [\because \text{6th term} = 32, \text{7th term} = 35] \\ &= \frac{67}{2} = 33.5\end{aligned}$$

## Mode

Mode is the value of the variable which occurs most frequently, i.e. an observation with the maximum frequency, is called the mode of the data.

**Example 5** Find the mode of the following marks (out of 10) obtained by 20 students.

4, 6, 5, 9, 3, 2, 7, 7, 6, 5, 4, 9, 10, 10, 3, 4, 7, 6, 9, 9

- (a) 9                      (b) 10                      (c) 7                      (d) 2

**Sol.** (a) On arranging the given data in ascending order, we get

2, 3, 3, 4, 4, 4, 4, 5, 5, 6, 6, 6, 6, 7, 7, 7, 9, 9, 9, 9, 10, 10

Here, 9 occurs most frequently, i.e. four times.

So, the required mode is 9.

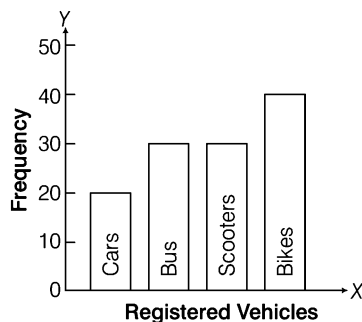
## Graphical Representation of Data

Data Interpretation as the same suggests tests your skills to understand data presented in different forms like bar graphs, tables, charts, line graphs etc. The questions are given with a set of data and candidate is required to deduce the required results from the set of data.

### A Bar Graph

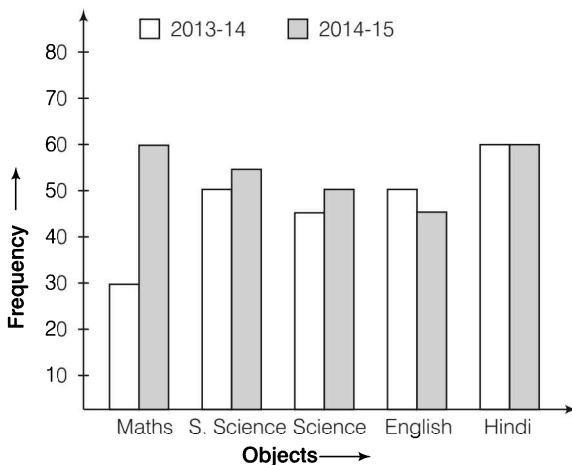
A pictorial representation of numerical data in the form of rectangles (or bars) of uniform width and various heights is called a bar graph, where the class intervals (width) are shown on the horizontal axis and the heights of the bars (rectangles) show the frequency of the class interval.

e.g. The registration of vehicles verses number of vehicles shown through the bar graph.



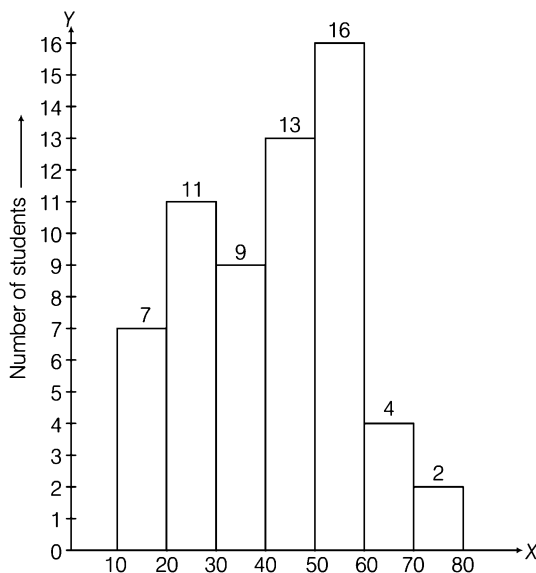
## Double Bar Graph

Through double bar graph (as shown above), we can easily find the information about the improved performance of the student in a particular subject and also can find the subject in which the performance of the student is deteriorated.



## Histogram

Histogram is the graphical representation of a grouped frequency distribution in exclusive form with continuous classes in the form of rectangles with class intervals as bases and the corresponding frequencies as heights. There is no gap between any two consecutive rectangles. The shape of the graph is shown below.



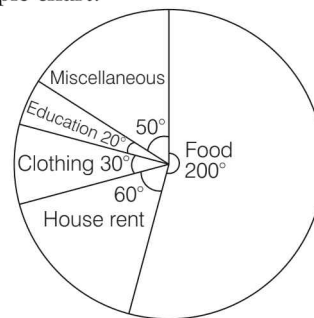
## Circle Graph (Pie chart)

A pie diagram or a pie chart is a pictorial representation of the numerical data by non-intersecting adjacent sectors of the circle such that area of each sector is proportional to the magnitude of the data represented by the sector.

Central angle of a component

$$= \left( \frac{\text{Value of the component}}{\text{Sum of the all component values}} \times 360 \right)^\circ$$

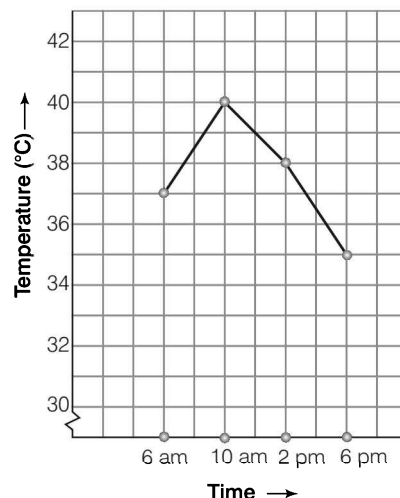
e.g. The different items such as clothing, food, house rent, education, miscellaneous as shown through pie chart.



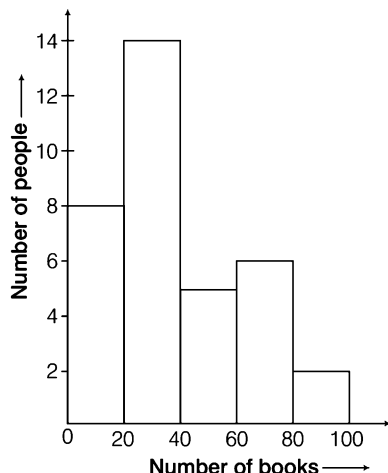
## Line Graph

A line graph is a graph, which is used to displays data that changes continuously over periods of time.

e.g. A line graph given below shows temperature of a day every four hours.



**Example 6** Histogram given below shows the number of people owning the different number of books. Answer the questions based on it.



- (i) The total number of people surveyed is  
 (a) 33 (b) 34  
 (c) 35 (d) 36

**Sol.** (c) Total number of people surveyed  
 $= 8 + 14 + 5 + 6 + 2 = 35$

- (ii) The number of people owning books more than 60 is  
 (a) 8 (b) 7  
 (c) 6 (d) 5

**Sol.** (a) The number of people owning books more than 60  $= 6 + 2 = 8$

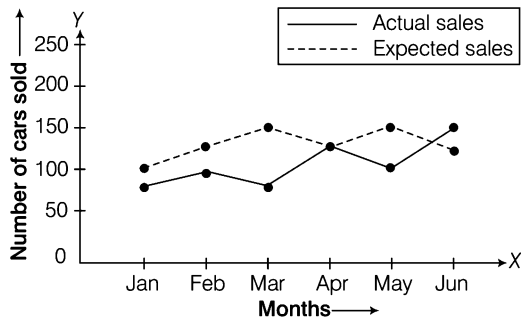
- (iii) The number of people owning books less than 40 is  
 (a) 23 (b) 22  
 (c) 21 (d) 20

**Sol.** (b) The number of people owning books less than 40  $= 8 + 14 = 22$

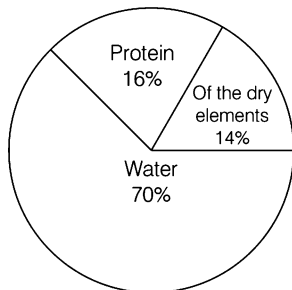
## Practice Exercise

- A geometric representation showing the relationship between a whole and its parts, is a  
 (a) pie chart (b) histogram  
 (c) bar graph (d) pictograph
- In a frequency distribution with class 0-10, 10-20 etc., the size of the class intervals is 10. The lower limit of fourth class is  
 (a) 40 (b) 50 (c) 20 (d) 30
- A graph showing two sets of data simultaneously is known as  
 (a) pictograph (b) histogram  
 (c) pie chart (d) double bar graph
- Data represented using circle is known as  
 (a) bar graph (b) histogram  
 (c) pictograph (d) pie chart
- Which of the following is not a random experiment?  
 (a) Tossing a coin  
 (b) Rolling a die  
 (c) Choosing a card from a deck of 52 cards  
 (d) Throwing a stone from the roof of a building
- Tally marks are used to find  
 (a) class intervals (b) range  
 (c) frequency (d) upper limit
- The mid value of a class intervals is 42. If the class size is 10, then the upper and lower limits of the class are  
 (a) 37.5 and 47.5 (b) 47 and 37  
 (c) 37 and 47 (d) 47.5 and 37.5
- The range of the data 30, 61, 55, 56, 60, 20, 26, 46, 28, 56 is  
 (a) 26 (b) 30  
 (c) 41 (d) 61
- In a grouped frequency distribution, the class intervals are 1-10, 11-20, 21-30, ... . Find the class width.  
 (a) 9 (b) 10  
 (c) 12 (d) 10.5

**Directions (Q. Nos. 10-12)** The graph given below gives the actual and expected sales of cars of a company for 6 months. Study the graph and answer the questions that follow.

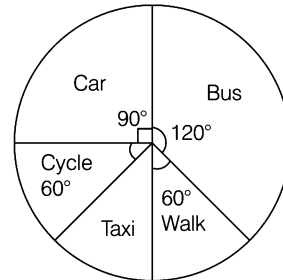


10. What was the total sales of cars in the months–January, February and March?  
 (a) 240 (b) 250  
 (c) 255 (d) 260
11. What is the average sales of cars in the last three months?  
 (a) 124 (b) 125  
 (c) 126 (d) 127
12. Find the ratio of sales in the first three months to the last three months.  
 (a) 2 : 3  
 (b) 2 : 5  
 (c) 3 : 2  
 (d) None of these
13. The following pie chart gives the distribution of constituents in the human body. The central angle of the sector showing the distribution of protein and other constituents is



- (a)  $108^\circ$  (b)  $54^\circ$   
 (c)  $30^\circ$  (d)  $216^\circ$

**Directions (Q. Nos. 14-16)** The pie chart given below shows the result of survey carried out to find the modes of travel used by children to go to school. Study the pie chart and answer the questions that follow.



14. What is the most common mode of transport?  
 (a) Bus (b) Car (c) Taxi (d) Cycle
15. What fraction of children travel by car?  
 (a)  $\frac{1}{4}$  (b)  $\frac{1}{3}$   
 (c)  $\frac{1}{2}$  (d) None of these
16. If 18 children travel by car, how many children took part in the survey?  
 (a) 70 (b) 72  
 (c) 73 (d) 74
17. Find the arithmetic mean for the scores 8, 6, 10, 12, 1, 3, 4, 4  
 Also, find the range of the data  
 (a) 12 (b) 11  
 (c) 9 (d) 8
18. Find median and mode of following data.  
 6, 9, 12, 15, 9, 3, 6, 9, 12, 6, 10, 3, 6, 15, 6  
 (a) 9, 6 (b) 9, 9  
 (c) 6, 10 (d) None of these
19. If each observation of the data is decreased by 5, then what is the effect on the mean?  
 (a) Increase by 5  
 (b) Decrease by 5  
 (c) Do not change  
 (d) None of these
20. If the mean of 6, 8, 9,  $x$  and 13 is 10, then find the value of  $x$ .  
 (a) 14 (b) 13  
 (c) 12 (d) 15

21. The mean score of 12 students, in an examination was 52.5. Later on scrutiny, the scores of two students were 800 of which 350 were female. The total number of male workers would be  
 (a) 52.6 (b) 52.8  
 (c) 53 (d) 53.5
22. The mean of the first 8 observations is 18 and last 8 observations is 20. If the mean of all 15 observations is 19, then find the 8th observation.  
 (a) 19 (b) 18 (c) 17 (d) 16
23. Find the median of following data.  
 17, 23, 57, 46, 33, 29, 28, 30, 34  
 If observation 23 is removed from data, then find the new median.  
 (a) 30, 31.5  
 (b) 32, 31.7  
 (c) 31, 31.5  
 (d) None of these
24. 10 is the mean of set of 7 observations and 5 is the mean of a set of 3 observations. The mean of the combined set is given by  
 (a) 15 (b) 10 (c) 8.5 (d) 7.5
25. In a class of 50 students, 10 have failed and their average marks are 28. The total marks obtained by the entire class are 2800. The average marks of those who have passed are  
 (a) 43 (b) 53 (c) 63 (d) 70
26. A, B, C are three sets of values of x  
 A : 2, 3, 7, 1, 3, 2, 3  
 B : 7, 5, 9, 12, 5, 3, 8  
 C : 4, 4, 11, 7, 2, 3, 4  
 Select the correct statement from the following  
 (a) Mean of A and B are same  
 (b) Mean of A > Mean of C  
 (c) Mean of B > Mean of A  
 (d) None of the above

## Answers

1	(a)	2	(d)	3	(d)	4	(d)	5	(d)	6	(c)	7	(c)	8	(c)	9	(b)	10	(b)
11	(b)	12	(a)	13	(a)	14	(a)	15	(a)	16	(b)	17	(b)	18	(a)	19	(b)	20	(a)
21	(c)	22	(a)	23	(a)	24	(c)	25	(c)	26	(c)								

## Hints & Solutions

- (a) Data can also be represented by using a pie chart (circle graph). It shows the relationship between a whole and its parts.
- (d) Given classes are 0-10 and 10-20.  
 As, the class of given classes is 10, so the next classes will be 20-30 and 30-40.  
 As, the fourth class is 30-40  
 Hence, the lower limit of fourth class is 30.
- (d) A graph showing two sets of data simultaneously is known as double bar graph.
- (d) Data represented using circles is known as pie chart.  
**Note Bar graph** Data using bars of different height in a graphical display is known as bar graph (bar chart).
- Histogram** Grouped data can be represented a histogram.
- Pictograph** Data using pictures and symbols to represent the statistical information is known as pictograph.
- (d) Tossing a coin, rolling a die and choosing a card from a deck of 52 cards are the random experiments, as we don't have an idea about the output of these experiments. But if we throw a stone from the roof of a building, we know the output, it will fall on the ground.
- (c) Tally marks are used to find the frequency of the observations.
- (c) Let the lower limit be x. Then, the upper limit of Class interval =  $x + 10$   

$$\therefore \frac{x + (x + 10)}{2} = 42$$

$$\Rightarrow 2x + 10 = 84 \Rightarrow x = 37$$

$$\therefore \text{Lower limit} = 37$$

$$\text{and upper limit} = 37 + 10 = 47$$

8. (c) Range of data = Maximum value – Minimum value =  $61 - 20 = 41$
9. (b) This is a discontinuous class interval, so both 1 and 10 are included in the interval.  
Hence, the class width is 10.
10. (b) The total sales of cars in the months January, February and March was  $(75 + 100 + 75)$  i.e. 250.
11. (b) The average sales of cars in the last three months is 125 i.e.  $\frac{125 + 100 + 150}{3} = 125$ .
12. (a) The number of sales of car in the first three months = 250 and the number of sales of car in the last three months = 375  
 $\therefore$  The required ratio is  $250 : 375$  i.e.  $2 : 3$
13. (a) Distribution of protein and other constituents in human body =  $16 + 14 = 30\%$   
Central angle of the sector showing the distribution of protein and other constituents  
 $= \frac{30}{100} \times 360^\circ = 108^\circ$
14. (a) The central angle is maximum for bus, hence bus is the most common mode of transport.
15. (a) Fraction of children travelled by car  
 $= \frac{\text{Central angle}}{360^\circ} = \frac{90^\circ}{360^\circ} = \frac{1}{4}$
16. (b) We know that, fraction of children travel by car  
 $= \frac{1}{4}$   
Hence, total number of children travelled by car  
 $\text{car} = \frac{1}{4} \times \text{Total number of children}$   
 $\Rightarrow 18 = \frac{1}{4} \times \text{Total number of children}$   
 $\therefore \text{Total number of children} = 18 \times 4 = 72$
17. (b) Arithmetic mean  
 $= \frac{8 + 6 + 10 + 12 + 1 + 3 + 4 + 4}{8} = \frac{48}{8} = 6$   
Range = Value of highest observation – Value of lowest observation =  $12 - 1 = 11$

18. (a) Firstly, arrange the given data in ascending order

i.e. 3, 3, 6, 6, 6, 6, 6, 9, 9, 9, 10, 12, 12, 15, 15

Here,  $n = 15$  [odd]

$$\therefore \text{Median} = \left( \frac{n+1}{2} \right) \text{th term} = \frac{15+1}{2} = 8 \text{th term} = 9$$

From the data we see that 6 occurs maximum five times.

$$\therefore \text{Mode} = 6$$

19. (b) Mean is also decreased by 5.

20. (a) Mean of the given numbers

$$= \frac{6 + 8 + 9 + x + 13}{5} = \frac{36 + x}{5}$$

$$\therefore \text{Mean} = 10 \text{ [given]}$$

$$\therefore \frac{36 + x}{5} = 10 \Rightarrow 36 + x = 50$$

$$\Rightarrow x = 50 - 36 = 14$$

21. (c) Sum of scores of 12 students

$$= 12 \times 52.5 = 630$$

$$\text{New sum of score} = 630 - 45 - 51 + 40 + 62 = 636$$

$$\therefore \text{Correct mean} = \frac{636}{12} = 53$$

22. (a) Given, mean of the first 8 observations = 18

$$\Rightarrow \frac{\text{Sum of first 8 observations}}{8} = 18$$

$$\Rightarrow \text{Sum of first 8 observations} = 18 \times 8 = 144$$

$$\text{Also given, the mean of last 8 observations} = 20$$

$$\Rightarrow \text{Sum of last 8 observations} = 20 \times 8 = 160$$

$$\text{Again, mean of all 15 observations} = 19 \text{ [given]}$$

$$\Rightarrow \text{Sum of all 15 observations} = 19 \times 15 = 285$$

$$\therefore 8 \text{th observation} = (144 + 160) - 285 = 304 - 285 = 19$$

23. (a) After writing the given data in ascending order, we get

17, 23, 28, 29, 30, 33, 34, 46, 57

Here  $n = 9$ , [odd]

$$\therefore \text{Median} = \left( \frac{n+1}{2} \right) \text{th term} = \left( \frac{9+1}{2} \right) \text{th term}$$

$$= \frac{10}{2} \text{th term} = 5 \text{th term} = 30$$



When 23 is removed from given data, then given data becomes 17, 28, 29, 30, 33, 34, 46, 57.

Here,  $n = 8$

$\therefore$  Median = Average of 4th and 5th terms

$$= \frac{30 + 33}{2} = \frac{63}{2} = 31.5$$

$$\left[ \because \text{if } n \text{ is even, then } \frac{1}{2} \left\{ \left( \frac{n}{2} \right) \text{th term} + \left( \frac{n+1}{2} \right) \text{th term} \right\} \right]$$

Hence, the new median is 31.5.

**24.** (c) Given,  $n_1 = 7, \bar{x}_1 = 10, n_2 = 3, \bar{x}_2 = 5$

$$\begin{aligned} \therefore \text{Combined mean} &= \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2} \\ &= \frac{7 \times 10 + 3 \times 5}{7 + 3} = \frac{85}{10} = 8.5 \end{aligned}$$

**25.** (c) Total numbers of students passed

$$= 50 - 10 = 40$$

Total marks obtained by failed students

$$= 28 \times 10 = 280$$

$\therefore$  Total marks obtained by passed i.e. 40 students

$$= 2800 - 280 = 2520$$

$\therefore$  Required average of passed students

$$= \frac{2520}{40} = 63$$

**26.** (c) Mean of A =  $\frac{2 + 3 + 7 + 1 + 3 + 2 + 3}{7}$

$$= \frac{21}{7} = 3$$

$$\text{Mean of B} = \frac{7 + 5 + 9 + 12 + 5 + 3 + 8}{7} = \frac{49}{7} = 7$$

$$\text{Mean of C} = \frac{4 + 4 + 11 + 7 + 2 + 3 + 4}{7}$$

$$= \frac{35}{7} = 5$$

$\therefore$  Mean of B > Mean of A