

9.1 INTRODUCTION

One day Somu visited his mathematics teacher at his home. At that time his teacher was busy in compiling the information which he had collected from his ward for the population census of India.

Somu : Good evening sir, it seems you are very busy. Can I help you in your work, Sir?

Teacher : Somu, I have collected the household information for census i.e. number of family members, their age group, family income, type of house they live and other data.

Somu : Sir, what is the use of this information?

Teacher : This information is useful for government in planning the developmental programs and allocation of resources.

Somu : How does government use this large information?

Teacher : The Census Department compiles this massive data and by using required data handling tools analyse the data and interprets the results in the form of information. Somu, you must have learnt basic statistic (data handling) in your earlier classes, didn't you?



Like Somu we too come across a lot of situations where we see information in the form of facts, numerical figures, tables, graphs etc. These may relate to price of vegetables, city temperature, cricket scores, polling result and so on. These facts or figures which are numerical or otherwise collected with a definite purpose are called '**data**'. Extraction of meaning from the data is studied in a branch of mathematics called **statistics**.

Lets us first revise what we have studied in statistic (data handling) in our previous classes.

9.2 Collection of Data

The primary activity in statistics is to collect the data with some purpose. To understand this let us begin with an exercise of collecting data by performing the following activity.

ACTIVITY

Divide the students of your class into four groups. Allot each group the work of collecting one of the following kinds of data:

- i. Weights of all the students in your class.
- ii. Number of siblings that each student have.
- iii. Day wise number of absentees in your class during last month.
- iv. The distance between the school and home of every student of your class.

Let us discuss how these students have collected the required information?

1. Have they collected the information by enquiring each student directly or by visiting every house personally by the students?
2. Have they got the information from source like data available in school records?

In first case when the information was collected by the investigator (student) with a definite objective, the data obtained is called **primary data** (as in (i), (ii), (iv)).

In the above task (iii) number of absentees in the last month could only be known by school attendance register. So here we are using data which is already collected by class teachers. This is called secondary data. The information collected from a source, which had already been recorded, say from registers, is called **secondary data**.

Do This

Which of the following are primary and secondary data?

- i. Collection of the data about enrollment of students in your school for a period from 2001 to 2010.
- ii. Height of students in your class recorded by physical education teacher.

9.3 Presentation of Data

Once the data is collected, the investigator has to find out ways to present it in the form which is meaningful, easy to understand and shows its main features at a glance. Let us take different situation where we need to present the data.

Consider the marks obtained by 15 students in a mathematics test out of 50 marks:

25, 34, 42, 20, 39, 50, 28, 30, 50, 11, 20, 42, 45, 40, 7.

The data in this form is called raw data.

From the given data you can easily identify the minimum and maximum marks. You also remember that the difference between the minimum and maximum marks is called the **range** of given data.

Here minimum and maximum marks are 7 and 50 respectively.

$$\text{So the range} = 50 - 7 = 43,$$

From the above we can also say that our data lies between 7 to 50.

Now let us answer the following questions from the above data.

- Find the middle value of the given data.
- Find how many children got 60% or more marks in the mathematics test.

Discussion

(i) Ikram said that the middle value of the data is 25 because the exam was conducted for 50 marks. What do you think?

Mary said that it is not the middle value of the data. In this case we have marks of 15 students as raw data, so after arranging the data in ascending order,

7, 11, 20, 20, 25, 28, 30, 34, 39, 40, 42, 42, 45, 50, 50

we can say that the 8th term is the middle term and it is 34.

(ii) You already know how to find 60% of 50 marks (i.e. $\frac{60}{100} \times 50 = 30$).

You find that there are 9 students who got 60% or more marks (i.e. 30 marks or more).

When the number of observations in a data are too many, presentation of the data in ascending or descending order can be quite time consuming. So we have to think of an alternative method.

See the given example.

Example-1. Consider the marks obtained by 50 students in a mathematics test for a total marks of 10.

5, 8, 6, 4, 2, 5, 4, 9, 10, 2, 1, 1, 3, 4, 5,
8, 6, 7, 10, 2, 1, 1, 3, 4, 4, 5, 8, 6, 7, 10,
2, 8, 6, 4, 2, 5, 4, 9, 10, 2, 1, 1, 3, 4, 5,
8, 6, 4, 5, 8

| Marks | Tally Marks | No of students |
|-------|-------------|----------------|
| 1 | | 6 |
| 2 | | 6 |
| 3 | | 3 |
| 4 | | 9 |
| 5 | | 7 |
| 6 | | 5 |
| 7 | | 2 |
| 8 | | 6 |
| 9 | | 2 |
| 10 | | 4 |
| | Total | 50 |

The data is tabulated by using the tally marks, as shown in table.

Recall that the number of students who have obtained a certain number of marks is called the frequency of those marks. For example, 9 students got 4 marks each. So the frequency of 4 marks is 9.

Here in the table, tally marks are useful in tabulating the raw data.

Sum of all frequencies in the table gives the total number of observations of the data.

As the actual observations of the data are shown in the table with their frequencies, this table is called ‘**Ungrouped Frequency Distribution Table**’ or ‘**Table of Weighted Observations**’.

ACTIVITY



Make frequency distribution table of the initial letters of that denotes surnames of your classmates and answer the following questions.

- (i) Which initial letter occurred mostly among your classmates?
- (ii) How many students names start with the alphabat ‘I’?
- (iii) Which letter occurred least as an initial among your classmates?

Suppose for specific reason, we want to represent the data in three categories (i) how many students need extra classes, (ii) how many have an average performance and (iii) how many did well in the test. Then we can make groups as per the requirement and grouped frequency table as shown below.

| Class interval (marks) | Category | Tally marks | No. of students |
|------------------------|--------------------|------------------|-----------------|
| 1 - 3 | (Need extra class) | III III III | 15 |
| 4 - 5 | (Average) | III III III I | 16 |
| 6 - 10 | (Well) | III III III IIII | 19 |

To classify the data according to the requirement or if there are large number of observations. We make groups to condense it. Lets take one more example in which group and frequency make us easy to understand the data.

Example-2. The weight (in grams) of 50 oranges, picked at random from a basket of oranges, are given below:

35, 45, 55, 50, 30, 110, 95, 40, 70, 100, 60, 80, 85, 60, 52, 95, 98, 35, 47, 45, 105, 90, 30, 50, 75, 95, 85, 80, 35, 45, 40, 50, 60, 65, 55, 45, 30, 90, 115, 65, 60, 40, 100, 55, 75, 110, 85, 95, 55, 50

To present such a large amount of data and to make sense of it, we make groups like 30-39, 40-49, 50-59, 100-109, 110-119. (since our data is from 30 to 115). These groups are called 'classes' or class-intervals, and their size is called length of the class or class width, which is 10 in this case. In each of these classes the least number is called the lower limit and the greatest number is called the upper limit, e.g. in 30-39, 30 is the 'lower limit' and 39 is the 'upper limit'.

| (Oranges weight) Class interval | Tally marks | (Number of oranges) Frequency |
|------------------------------------|-------------|----------------------------------|
| 30 - 39 | | 6 |
| 40 - 49 | | 8 |
| 50 - 59 | | 9 |
| 60 - 69 | | 6 |
| 70 - 79 | | 3 |
| 80 - 89 | | 5 |
| 90 - 99 | | 7 |
| 100-109 | | 3 |
| 110 - 119 | | 3 |
| | Total | 50 |

Presenting data in this form simplifies and condenses data and enables us to observe certain important features at a glance. This is called a **grouped frequency distribution table**.

We observe that the classes in the table above are non-overlapping i.e. 30-39, 40-49 ... no number is repeating in two class intervals. Such classes are called inclusive classes. Note that we could have made more classes of shorter size, or lesser classes of larger size also. Usually if the raw data is given the range is found (Range = Maximum value – Minimum value). Based on the value of ranges with convenient, class interval length, number of classes are formed. For instance, the intervals could have been 30-35, 36-40 and so on.

Now think if weight of an orange is 39.5 gm. then in which interval will we include it? We cannot include 39.5 either in 30-39 or in 40-49.

In such cases we construct real limits (or boundaries) for every class.

Average of upper limit of a class interval and lower limit of the next class interval becomes the upper boundary of the class. The same becomes the lower boundary of the next class interval.

| Class limit | Class boundaries |
|-------------|------------------|
| 20 - 29 | |
| 30 - 39 | 29.5 - 39.5 |
| 40 - 49 | 39.5 - 49.5 |
| 50 - 59 | 49.5 - 59.5 |
| 60 - 69 | 59.5 - 69.5 |
| 70 - 79 | 69.5 - 79.5 |
| 80 - 89 | 79.5 - 89.5 |
| 90 - 99 | 89.5 - 99.5 |
| 100 - 109 | 99.5-109.5 |
| 110 - 119 | 109.5-119.5 |
| 120 - 129 | |

Similarly boundaries of all class intervals are calculated. By assuming a class interval before the first class and next class interval after the last class, we calculate the lower boundary any of the first and upper boundary any of the last class intervals.

Again a problem arises that whether 39.5 has to be included in the class interval 29.5-39.5 or 39.5 - 49.5? Here by convention, if any observation is found to be equivalent to upper boundary of a particular class, then that particular observation is considered under next class, but not that of the particular class.

So 39.5 belongs to 39.5 - 49.5 class interval.

The classes which are in the form of 30-40, 40-50, 50-60, are called over lapping classes and called as exclusive classes.

If we observe the boundaries of inclusive classes, they are in the form of exclusive classes. The difference between upper boundary and lower boundary of particular class given the length of class interval. Length class interval of 90, 99 is (i.e. $99.5 - 89.5 = 10$) 10.

Example-3. The relative humidity (in %) of a certain city for a September month of 30 days was as follows:

98.1 98.6 99.2 90.3 86.5 95.3 92.9 96.3 94.2 95.1
 89.2 92.3 97.1 93.5 92.7 95.1 97.2 93.3 95.2 97.3
 96.0 92.1 84.9 90.0 95.7 98.3 97.3 96.1 92.1 89

- (i) Construct a grouped frequency distribution table with classes 84-86, 86,-88 etc.
- (ii) What is the range of the data?

Solution : (i) The grouped frequency distribution table is as follows-

| Relative humidity | Tally marks | Number of days |
|-------------------|-------------|----------------|
| 84-86 | | 1 |
| 86-88 | | 1 |
| 88-90 | | 2 |
| 90-92 | | 2 |
| 92-94 | | 7 |
| 94-96 | | 6 |
| 96-98 | | 7 |
| 98-100 | | 4 |

[Note:- 90 comes in interval 90-92 likewise 96 comes in 96-98 class interval]



- (ii) Range $99.2 - 84.9 = 14.3$ (vary for different places).

EXERCISE - 9.1



1. Write the mark wise frequencies in the following frequency distribution table.

| Marks | Up to 5 | Up to 6 | Up to 7 | Up to 8 | Up to 9 | Up to 10 |
|----------------|---------|---------|---------|---------|---------|----------|
| No of students | 5 | 11 | 19 | 31 | 40 | 45 |

2. The blood groups of 36 students of IX class are recorded as follows.

A O A O A B O A B A B O
 B O B O O A B O B AB O A
 O O O A AB O A B O A O B

Represent the data in the form of a frequency distribution table. Which is the most common and which is the rarest blood group among these students?

3. Three coins were tossed 30 times simultaneously. Each time the number of heads occurring was noted down as follows;

1 2 3 2 3 1 1 1 0 3 2 1
 2 2 1 1 2 3 2 0 3 0 1 2
 3 2 2 3 1 1

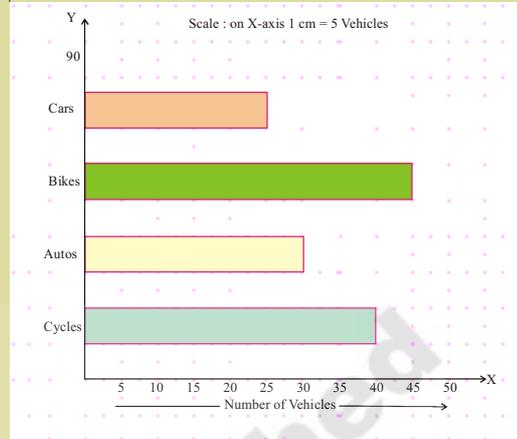
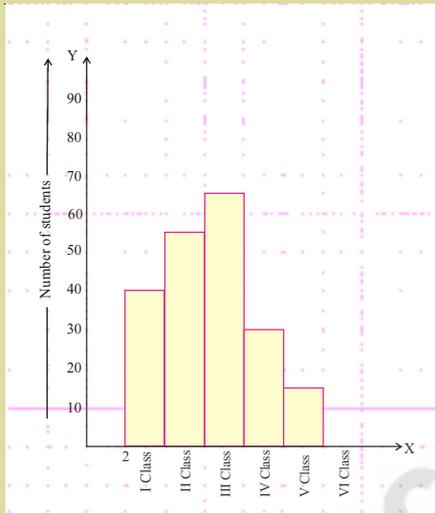
Prepare a frequency distribution table for the data given above.

4. A TV channel organized a SMS (Short Message Service) poll on prohibition on smoking, giving options like A – complete prohibition, B – prohibition in public places only, C – not necessary. SMS results in one hour were A B A B C B

A B B A C C B B A B
 B A B C B A B C B A
 B B A B B C B A B A
 B C B B A B C B B A
 B B A B B A B C B A
 B B A B C A B B A

Represent the above data as grouped frequency distribution table. How many appropriate answers were received? What was the majority of peoples' opinion?

5. Represent the data in the adjacent bar graph as frequency distribution table.



6. Identify the scale used on the axes of the adjacent graph. Write the frequency distribution from it.

7. The marks of 30 students of a class, obtained in a test (out of 75), are given below:

42, 21, 50, 37, 42, 37, 38, 42, 49, 52, 38, 53, 57, 47, 29
59, 61, 33, 17, 17, 39, 44, 42, 39, 14, 7, 27, 19, 54, 51.

Form a frequency table with equal class intervals. (**Hint** : one of them being 0-10)

8. The electricity bills (in rupees) of 25 houses in a locality are given below. Construct a grouped frequency distribution table with a class size of 75.

170, 212, 252, 225, 310, 712, 412, 425, 322, 325, 192, 198, 230, 320, 412,
530, 602, 724, 370, 402, 317, 403, 405, 372, 413

9. A company manufactures car batteries of a particular type. The life (in years) of 40 batteries were recorded as follows:

| | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|
| 2.6 | 3.0 | 3.7 | 3.2 | 2.2 | 4.1 | 3.5 | 4.5 |
| 3.5 | 2.3 | 3.2 | 3.4 | 3.8 | 3.2 | 4.6 | 3.7 |
| 2.5 | 4.4 | 3.4 | 3.3 | 2.9 | 3.0 | 4.3 | 2.8 |
| 3.5 | 3.2 | 3.9 | 3.2 | 3.2 | 3.1 | 3.7 | 3.4 |
| 4.6 | 3.8 | 3.2 | 2.6 | 3.5 | 4.2 | 2.9 | 3.6 |

Construct a grouped frequency distribution table with exclusive classes for this data, using class intervals of size 0.5 starting from the interval 2 - 2.5.

9.4 MEASURES OF CENTRAL TENDENCY

Consider the following situations:

Case-1 : In a hostel 50 students usually eat 200 idlies in their breakfast. How many more idlies does the mess incharge make if 20 more students joined the hostel.

Case-2 : Consider the wages of staff at a factory as given in the table. Which salary figure represents the whole staff:

| | | | | | | | | | | |
|-------------------------------|----|----|----|----|----|----|----|----|----|----|
| Staff | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Salary in ₹ (in thousands) | 12 | 14 | 15 | 15 | 15 | 16 | 17 | 18 | 90 | 95 |

Case-3 : The different forms of transport in a city are given below. Which is the popular means of transport?

1. Car 15%
2. Train 12%
3. Bus 60%
4. Two wheeler 13%



In the first case, we will usually take an average (mean), and use it to resolve the problem. But if we take average salary in the second case then it would be 30.7 thousands. However, verifying the raw data suggests that this mean value may not be the best way to accurately reflect the typical salary of a worker, as most workers have their salaries between 12 to 18 thousands. So, median (middle value) would be a better measure in this situation. In the third case mode (most frequent) is considered to be a most appropriate option. The nature of the data and its purpose will be the criteria to go for average or median or mode among the measures of central tendency.

THINK, DISCUSS AND WRITE



1. Give 3 situations, where mean, median and mode are separately appropriate and counted.

Consider a situation where fans of two cricketers Raghu and Gautam claim that their star score better than other. They made comparison on the basis of last 5 matches.

| Matches | | 1 st | 2 nd | 3 rd | 4 th | 5 th |
|-----------|--------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Runs made | Raghu | 50 | 50 | 76 | 31 | 100 |
| | Gautam | 65 | 23 | 100 | 100 | 10 |

Fans of both the players added the runs and calculated the averages as follows.

$$\text{Raghu's average score} = \frac{307}{5} = 61.4$$

$$\text{Gautam average score} = \frac{298}{5} = 59.6$$

Since Raghu's average score was more than Gautam's, Raghu fan's claimed that Raghu performed better than Gautam, but Gautam fans did not agree. Gautam fan's arranged both their scores in descending order and found the middle score as given below:

| | | | | | |
|--------|-----|-----|----|----|----|
| Raghu | 100 | 76 | 50 | 50 | 31 |
| Gautam | 100 | 100 | 65 | 23 | 10 |

Then Gautam fan's said that since his middle-most score is 65, which is higher than Raghu middle-most score, i.e. 50 so his performance should be rated better.'

But we may say that Gautam made two centuries in 5 matches and so he may be better.

Now, to settle the dispute between Raghu's and Gautam's fans, let us see the three measures adopted here to make their point.

The average score they used first is the mean. The 'middle' score they used in the argument is the Median. Mode is also a measure to compare the performance by considering the scores repeated many times. Mode score of Raghu is 50. Mode score of Gautam is 100. Of all these three measures which one is appropriate in this context?

Now let us first understand mean in details.

9.4.1 Mean

Mean is the sum observation of a data divided by the number of observations. We have already discussed about computing arithmetic mean for a raw data.

$$\text{Mean } \bar{x} = \frac{\text{Sum of observations}}{\text{Number of observations}} \text{ or } \bar{x} = \frac{\sum x_i}{n}$$

9.4.1.1 Mean of Raw Data

Example-4. Rain fall of a place in a week is 4cm, 5cm, 12cm, 3cm, 6cm, 8cm, 0.5cm. Find the average rainfall per day.

Solution : The average rainfall per day is the arithmetic mean of the above observations.

Given rainfall through a week are 4cm, 5cm, 12cm, 3cm, 6cm, 8cm, 0.5cm.

Number of observations (n) = 7

$$\text{Mean } \bar{x} = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \quad \text{Where } x_1, x_2, \dots, x_n \text{ are } n \text{ observations}$$

$$\text{and } \bar{x} \text{ is their mean} = \frac{4 + 5 + 12 + 3 + 6 + 8 + 0.5}{7} = \frac{38.5}{7} = 5.5 \text{ cm.}$$

Example-5. If the mean of 10, 12, 18, 13, P and 17 is 15, find the value of P.

Solution : We know that Mean $\bar{x} = \frac{\sum x_i}{n}$

$$15 = \frac{10 + 12 + 18 + 13 + P + 17}{6}$$

$$90 = 70 + P$$

$$P = 20.$$



9.4.1.2 Mean of Ungrouped frequency distribution

Consider this example; Weights of 40 students in a class are given in the following frequency distribution table.

| | | | | | | |
|--------------------|----|----|----|----|----|----|
| Weights in kg (x) | 30 | 32 | 33 | 35 | 37 | 41 |
| No of students (f) | 5 | 9 | 15 | 6 | 3 | 2 |

Find the average (mean) weight of 40 students.

From the table we can see that 5 students weigh 30 kg., each. So sum of their weight is $5 \times 30 = 150$ kg. Similarly we can find out the sum of weights with each frequency and then their total. Sum of the frequencies gives the number of observations in the data.

$$\text{Mean } (\bar{x}) = \frac{\text{Sum of all the observations}}{\text{Total number of observations}}$$

$$\text{So Mean} = \frac{5 \times 30 + 9 \times 32 + 15 \times 33 + 6 \times 35 + 3 \times 37 + 2 \times 41}{5 + 9 + 15 + 6 + 3 + 2} = \frac{1336}{40} = 33.40 \text{ kg.}$$

If observations are $x_1, x_2, x_3, x_4, x_5, x_6$ and corresponding frequencies are $f_1, f_2, f_3, f_4, f_5, f_6$ then we may write the above expression as

$$\text{Mean } \bar{x} = \frac{f_1x_1 + f_2x_2 + f_3x_3 + f_4x_4 + f_5x_5 + f_6x_6}{f_1 + f_2 + f_3 + f_4 + f_5 + f_6} = \frac{\sum f_i x_i}{\sum f_i}$$

$$\therefore \bar{x} = \frac{\sum f_i x_i}{\sum f_i}$$

Example-6. Find the mean of the following data.

| | | | | | |
|-----|---|----|----|----|----|
| x | 5 | 10 | 15 | 20 | 25 |
| f | 3 | 10 | 25 | 7 | 5 |

Solution :

Step-1 : Calculate $f_i \times x_i$ of each row

Step-2 : Find the sum of frequencies ($\sum f_i$)

and sum of the $f_i \times x_i$ ($\sum f_i x_i$)

Step-3 : Calculate $\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{755}{50} = 15.1$

| x_i | f_i | $f_i x_i$ |
|-------|-----------------|----------------------|
| 5 | 3 | 15 |
| 10 | 10 | 100 |
| 15 | 25 | 375 |
| 20 | 7 | 140 |
| 25 | 5 | 125 |
| | $\sum f_i = 50$ | $\sum f_i x_i = 755$ |

Example-7. If the mean of the following data is 7.5 , then find the value of ‘A’.

| | | | | | | |
|-----------------|---|----|----|---|---|----|
| Marks | 5 | 6 | 7 | 8 | 9 | 10 |
| No. of Students | 3 | 10 | 17 | A | 8 | 4 |

Solution :

Sum of frequencies ($\sum f_i$) = 42 + A

Sum of the $f_i \times x_i$ ($\sum f_i x_i$) = 306 + 8A

$$\text{Mean } \bar{x} = \frac{\sum f_i x_i}{\sum f_i}$$

Given Arithmetic Mean = 7.5

$$\text{So, } 7.5 = \frac{306 + 8A}{42 + A}$$

$$306 + 8A = 315 + 7.5A$$

| Marks (x_i) | No. of Students (f_i) | $f_i x_i$ |
|-----------------|---------------------------|-----------|
| 5 | 3 | 15 |
| 6 | 10 | 60 |
| 7 | 17 | 119 |
| 8 | A | 8A |
| 9 | 8 | 72 |
| 10 | 4 | 40 |
| | 42+A | 306+8A |

$$\begin{aligned}8A - 7.5A &= 315 - 306 \\0.5A &= 9 \\A &= 18\end{aligned}$$

9.4.1.3 Mean of ungrouped frequency Distribution by Deviation method

Example-8. Find the arithmetic mean of the following data:

| | | | | | | | |
|-----|----|----|----|----|----|----|----|
| x | 10 | 12 | 14 | 16 | 18 | 20 | 22 |
| f | 4 | 5 | 8 | 10 | 7 | 4 | 2 |

Solution :

(i) Simple Method

Thus in the case of ungrouped frequency distribution, you can use the formula,

$$\bar{x} = \frac{\sum_{i=1}^7 f_i x_i}{\sum_{i=1}^7 f_i} = \frac{622}{40} = 15.55$$

(ii) Deviation Method

In this method we assume one of the observations which is convenient as assumed mean. Suppose we assume '16' as a mean, be $A = 16$ the deviation of other observations from the assumed mean are given in table.

Sum of frequencies = 40

Sum of the $f_i \times d_i$ products = $-60 + 42$

$$\sum f_i d_i = -18$$

$$\begin{aligned}\text{Mean } \bar{x} &= A + \frac{\sum f_i d_i}{\sum f_i} = 16 + \frac{-18}{40} \\&= 16 - 0.45 \\&= 15.55\end{aligned}$$

| x_i | f_i | $f_i x_i$ |
|-------|-------------------------|------------------------------|
| 10 | 4 | 40 |
| 12 | 5 | 60 |
| 14 | 8 | 112 |
| 16 | 10 | 160 |
| 18 | 7 | 126 |
| 20 | 4 | 80 |
| 22 | 2 | 44 |
| | $\sum_{i=1}^7 f_i = 40$ | $\sum_{i=1}^7 f_i x_i = 622$ |

| x_i | f_i | $d_i = x_i - A$ | $f_i d_i$ |
|-------|-------|-----------------|------------------|
| 10 | 4 | -6 | -24 |
| 12 | 5 | -4 | -20 |
| 14 | 8 | -2 | -16 |
| 16 A | 10 | 0 | 0 |
| 18 | 7 | +2 | +14 |
| 20 | 4 | +4 | +16 |
| 22 | 2 | +6 | +12 |
| | 40 | | $-60 + 42 = -18$ |

9.4.2 Median

Median is the middle observation of a given raw data, when it is arranged in an order (ascending / descending). It divides the data into two groups of equal number, one part comprising all values greater than median and the other part comprising values less than median.

We have already discussed in the earlier classes that median of a raw data with observations, arranged in order is calculated as follows.

When the data as ‘n’ number of observations and if ‘n’ is odd, median is $\left(\frac{n+1}{2}\right)^{\text{th}}$ observation.

When n is even, median is the average of $\left(\frac{n}{2}\right)^{\text{th}}$ and $\left(\frac{n}{2}+1\right)^{\text{th}}$ observations

TRY THESE



1. Find the median of the scores 75, 21, 56, 36, 81, 05, 42
2. Median of a data, arranged in ascending order 7, 10, 15, x, y, 27, 30 is 17 and when one more observation 50 is added to the data, the median has become 18 Find x and y.

9.4.2.1 Median of a frequency distribution

Let us now discuss the method of finding the median for a data of weighted observations consider the monthly wages of 100 employees of a company.

| | | | | | | | |
|------------------|------|------|------|------|------|-------|-------|
| Wages (in ₹) | 7500 | 8000 | 8500 | 9000 | 9500 | 10000 | 11000 |
| No. of employees | 4 | 18 | 30 | 20 | 15 | 8 | 5 |

How to find the median of the given data? First arrange the observations given either in ascending or descending order. Then write the corresponding frequencies in the table and calculate less than cumulative frequencies. The cumulative frequency upto a particular observation is the progressive sum of frequencies upto that particular observation.

| Wages (x) | No. of employees (f) | Cumulative frequency (cf) |
|-----------|----------------------|---------------------------|
| 7500 | 4 | 4 |
| 8000 | 18 | 22 |
| 8500 | 30 | 52 |
| 9000 | 20 | 72 |
| 9500 | 15 | 87 |
| 10000 | 8 | 95 |
| 11000 | 5 | 100 |
| | 100 | |

Find $\frac{N}{2}$ and identify the median class, whose cumulative frequencies just exceeds $\frac{N}{2}$, where N is sum of the frequencies.

Here $N = 100$ even so find $\left(\frac{N}{2}\right)^{th}$ and $\left(\frac{N}{2} + 1\right)^{th}$ observations which are 50 and 51 respectively.

From the table corresponding values of 50th and 51st observations is the same, falls in the wages of 8500. So the median class of this distribution is 8500.

TRY THESE



1. Find the median marks in the data.

| | | | | | |
|----------------|----|----|----|----|---|
| Marks | 15 | 20 | 10 | 25 | 5 |
| No of students | 10 | 8 | 6 | 4 | 1 |

2. In finding the median, which column in the frequency distribution must be in order? Why?

9.4.3 Mode

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

Example-9. The following numbers are the sizes of shoes sold by a shop in a particular day. Find the mode.

6, 7, 8, 9, 10, 6, 7, 10, 7, 6, 7, 9, 7, 6.

Solution : First we have to arrange the observations in order 6, 6, 6, 6, 7, 7, 7, 7, 7, 8, 9, 9, 10, 10 to make frequency distribution table

| | | | | | |
|-------------------|---|---|---|---|----|
| Size | 6 | 7 | 8 | 9 | 10 |
| No. of shoes sold | 4 | 5 | 1 | 2 | 1 |

Here No. 7 occurred most frequently. i.e, 5 times.

\therefore Mode (size of the shoes) of the given data is 7. This indicates the shoes of size No. '7' is a fast selling items.

THINK AND DISCUSS



- Classify your class mates according to their heights and find the mode of it.
- If shopkeeper has to place a order for shoes, which number shoes should he order more?

Example-10. Test scores out of 100 for a class of 20 students are as follows:

93, 84, 97, 98, 100, 78, 86, 100, 85, 92, 55, 91, 90, 75, 94, 83, 60, 81, 95

- (a) Make a frequency table taking class interval as 91-100, 81-90,
- (b) Find the modal class. (The “Modal class” is the class containing the greatest frequency).
- (c) find the interval that contains the median.

Solution :

(a)

| Test Scores | Frequency | Greater than Cumulative frequency |
|--------------|-----------|-----------------------------------|
| 91-100 | 9 | 20 |
| 81-90 | 6 | 11 |
| 71-80 | 3 | 5 |
| 61-70 | 0 | 2 |
| 51-60 | 2 | 2 |
| Total | 20 | |

- (b) 91-100 is the modal class. This class has the maximum frequency.
- (c) The middle of 20 is 10. If I count from the top, 10 will fall in the class interval 81-90. If I count from the bottom and go up, 10 will fall in the class interval 81-90. The class interval that contains the median is 81-90.

9.5 DEVIATION IN VALUES OF CENTRAL TENDENCY

What will happen to the measures of central tendency if we add the same amount to all data values, or multiply each data value by the same amount.

Let us observe the following table

| Particular | Data | Mean | Mode | Median |
|----------------------------------|--|------|------|--------|
| Original Data Set | 6, 7, 8, 10, 12, 14, 14, 15, 16, 20 | 12.2 | 14 | 13 |
| Add 3 to each data value | 9, 10, 11, 13, 15, 17, 17, 18, 19, 23 | 15.2 | 17 | 16 |
| Multiply 2 times each data value | 12, 14, 16, 20, 24, 28, 28, 30, 32, 40 | 24.4 | 28 | 26 |

After observing the table, we see that

When added : Since all values are shifted by the same amount, the measures of central tendency are all shifted by the same amount. If 3 is added to each data value, the mean, mode and median will also increase by 3.

When multiplied : Since all values are affected by the same multiplicative values, the measures of central tendency will also be affected similarly. If each observation is multiplied by 2, the mean, mode and median will also be multiplied by 2.

EXERCISE - 9.2



1. Weights of parcels in a transport office are given below.

| | | | | | | |
|---------------|----|----|----|----|-----|-----|
| Weight (kg) | 50 | 65 | 75 | 90 | 110 | 120 |
| No of parcels | 25 | 34 | 38 | 40 | 47 | 16 |

Find the mean weight of the parcels.

2. Number of families in a village in correspondence with the number of children are given below:

| | | | | | | |
|----------------|----|----|----|----|---|---|
| No of children | 0 | 1 | 2 | 3 | 4 | 5 |
| No of families | 11 | 25 | 32 | 10 | 5 | 1 |

Find the mean number of children per family.

3. If the mean of the following frequency distribution is 7.2 find value of 'K'.

| | | | | | | |
|-----|---|---|----|----|----|----|
| x | 2 | 4 | 6 | 8 | 10 | 12 |
| f | 4 | 7 | 10 | 16 | K | 3 |

4. Number of villages with respect to their population as per India census 2011 are given below.

| | | | | | | |
|---------------------------|----|----|----|----|----|---|
| Population (in thousands) | 12 | 5 | 30 | 20 | 15 | 8 |
| Villages | 20 | 15 | 32 | 35 | 36 | 7 |

Find the average population in each village.

5. AFLATOUN social and financial educational program initiated savings program among the high school children in Hyderabad district. Mandal wise savings in a month are given in the following table.

| Mandal | No. of schools | Total amount saved (in rupees) |
|--------------|----------------|--------------------------------|
| Amberpet | 6 | 2154 |
| Thirumalgiri | 6 | 2478 |
| Saidabad | 5 | 975 |
| Khairathabad | 4 | 912 |
| Secundrabad | 3 | 600 |
| Bahadurpura | 9 | 7533 |

Find arithmetic mean of school wise savings in each mandal. Also find the arithmetic mean of saving of all schools.

6. The heights of boys and girls of IX class of a school are given below.

| | | | | | | |
|-------------|-----|-----|-----|-----|-----|-----|
| Height (cm) | 135 | 140 | 147 | 152 | 155 | 160 |
| Boys | 2 | 5 | 12 | 10 | 7 | 1 |
| Girls | 1 | 2 | 10 | 5 | 6 | 5 |

Compare the heights of the boys and girls

[Hint : Find median heights of boys and girls]

7. Centuries scored and number of cricketers in the world are given below.

| | | | | | |
|-------------------|----|----|----|----|----|
| No. of centuries | 5 | 10 | 15 | 20 | 25 |
| No. of cricketers | 56 | 23 | 39 | 13 | 8 |

Find the mean, median and mode of the given data.

8. On the occasion of New year's day a sweet stall prepared sweet packets. Number of sweet packets and cost of each packet are given as follows.

| | | | | | | |
|-----------------------|-----|-----|-----|------|------|------|
| Cost of packet (in ₹) | ₹25 | ₹50 | ₹75 | ₹100 | ₹125 | ₹150 |
| No of packets | 20 | 36 | 32 | 29 | 22 | 11 |

Find the mean, median and mode of the data.

9. The mean (average) weight of three students is 40 kg. One of the students Ranga weighs 46 kg. The other two students, Rahim and Reshma have the same weight. Find Rahims weight.

10. The donations given to an orphanage home by the students of different classes of a secondary school are given below.

| Class | Donation by each student (in ₹) | No. of students donated |
|-------|---------------------------------|-------------------------|
| VI | 5 | 15 |
| VII | 7 | 15 |
| VIII | 10 | 20 |
| IX | 15 | 16 |
| X | 20 | 14 |

Find the mean, median and mode of the data.

11. There are four unknown numbers. The mean of the first two numbers is 4 and the mean of the first three is 9. The mean of all four number is 15, if one of the four number is 2 find the other numbers.

WHAT WE HAVE DISCUSSED



- Representation of the data with actual observations with frequencies in a table is called ‘**Ungrouped Frequency Distribution Table**’ or ‘**Table of Weighted Observations**’
- Representation of a large data in the form of a frequency distribution table enables us to view the data at a glance, to find the range easily, to find which observation is repeating for how many times, to analyse and to interpret the data easily.
- A measure of central tendency is a typical value of the data around which other observations congregate.
- Types of measure of central tendency : Mean, Mode, Median.
- Mean is the sum observation of a data divided by the number of observations.

$$\text{Mean} = \frac{\text{Sum of observations}}{\text{Number of observations}} \quad \text{or} \quad \bar{x} = \frac{\sum x_i}{n}$$

- For an ungrouped frequency distribution arithmetic mean $\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$.

- By deviation method, arithmetic mean $= A + \frac{\sum fd}{\sum n}$ where A is assumed mean and where $\sum f$ is the sum of frequencies and $\sum f d$ is the sum of deviations.
- Median is the middle observation of a data, when arranged in order (ascending / descending).
- When number of observations 'n' is odd, median is $\left(\frac{n+1}{2}\right)^{\text{th}}$ observation.
- When number of observations 'n' is even, median is the average of $\left(\frac{n}{2}\right)^{\text{th}}$ and $\left(\frac{n}{2} + 1\right)^{\text{th}}$ observations
- Median divides the data into two groups of equal number, one part comprising all values greater and the other comprising all values less than median.
- Mode is the value of the observation which occurs most frequently, i.e., an observation with the maximum frequency is called mode.



Brain teaser

In a row of students, Gopi is the 7th boy from left and Shankar is the 5th boy from the right. If they exchange their seats, Gopi is the 8th boy from the right. How many students are there in the row?

A boy chaitanya carved his name on the bark of a tree at a height of 1.5 m tall. The tree attains a height of a 6.75 m from the ground at what height from the ground will chaitanya's name can be located now?

Give reason to your answer.