UNIT 9 STATISTICS : AVERAGES, GRAPHIC REPRESENTATION AND CLASSIFICATION OF DATA

Structure

9.1 INTRODUCTION

The unit provides opportunities to involve pupils in enjoyable and challenging activities — collecting data, presenting them in the form of graphs and summarizing them in terms of average (mean), median, mode. Such data may be in respect of some activity in which pupils normally get interested. Statistics are a branch of mathematics which deal with collection, organization and interpretation of data. This unit also provides an opportunity to reinforce the study of related subject areas such as economics,. geography, science, health education etc. (from which data may be discussed) which are an important source of data and which rely heavily on data.

9.2 OBJECTIVES

At the end of this unit, you should be able to:

- describe briefly the meaning of statistics;
- illustrate the importance of statistics in every day life situations;
- organise projects of collecting, organising and interpreting data;
- decide which pictorial representation best suits a given set of data;
- represent given data pictorially;
- organise or arrange given data in suitable class intervals;
- calculate mean, median and mode of raw as well as grouped data;
- explain homogenous and heterogenous data from the statistical point of view;
- illustrate the ease in calculating mean by using an "assumed mean", and
- organise an exhibition in your school depicting pictorially varied characteristics of your school;

9.3 A BRIEF LOOK AT STATISTICS

9.3.1 What are Statistics?

Main Teaching Point : Statistics deal with the collection, presentation and interpretation of data.

Teaching-Learning Process : The unit should be introduced to students by raising questions which require the collection of data in a real life situation. The data collected would also need to be organised in some manner to be able to answer the questions raised. Here is an illustration

Ask : Suppose I need the following information about the students in your class:

a) the height of the tallest student

b) the height of the shortest student

c) the difference in their heights

d) the height which the maximum number of students have

e) the number of students who are taller than the height which the maximum number of students have

f) the number of students who are shorter than the height which the maximum number of students have.

The answer you may get is that they would make all the students stand in a line according to increasing or decreasing height and identify the tallest and shortest students. But then, can we answer questions (a) and (b) unless we measure their heights? Certainly not. To answer all the above questions, we need to measure the heights of all the students in the class.

You, as a teacher, should point out at this stage that the information we get by measuring their heights is called data in the language of statistics. Data are a number of facts. "Data" is the plural of the Latin word "datum" which means "fact". Measuring and noting down the heights of the students is called the process of **data collection.**

Suppose as a result of collecting the data, we get the heights of 45 students in a class as below:

Table 9.1

Heights (in cm) of a class of 45 students

140	142	143.5	140.5	150	149	148.5	148	141
148	148.5	152	153	150.5	147	146	142.5	148
152.5	154.5	155	153.5	152.5	149	147	146	156
152	148	143	149	144	150	152	153	154
156	153	154.5	154	147	148	146	153	154

Ask the students to answer the questions raised. It may still be difficult for them to answer these questions quickly. Lead the students to realise that they can answer the questions faster if the data are organised in some manner. Since the questions relate to a comparison of heights, the best way to organise the data is to write the heights in increasing or decreasing order. We then have the following table:

Table 9.2
Heights (in cm) of 45 students of a class
(arranged in ascending order)

140	140.5	141	142	142.5	143	143.5	144	146
146	146	147	147	147	148	148	148	148
148	148.5	148.5	149	149	149	150	150	150.5
151	152	152	152	152.5	152.5	153	153	153
153	153.5	154	154	154	154.5	154.5	155	156

Having organised the data as shown above, it is now easy to get the information sought Answering the questions and inferring other results from the organised data comes under the domain of interpretation of data.

Thus, **statistics are the branch of mathematics dealing with collecting, organising and interpreting data.** The teacher will do well to tell the students that interpreation of data involves much more than answering simple questions such as those raised above. It includes prediction, testing of assumptions made etc.

Methodology used : Mainly the lecture-cum-discussion method is used to discuss the nature of statistics.

Check Your Progress

Notes : a) Write your answers in the space given below.

b) Compare your answers with those given at the end of the unit.

1. The following data give the weights of 25 mangoes selected at random from a basket. Arrange the data in descending order and answer the questions given below:

Weight (in gm) :

45, 55, 50, 42, 32, 47, 51, 55, 65, 63, 55, 40, 35, 57, 55, 55, 38, 45, 48, 45, 45, 42, 45, 45, 46

a) What is the minimum weight of a mango?

....

b) What is the most common weight of a mango?

....

c) How many mangoes have the most common weight?

....

d) How many mangoes weigh 55 gm?

....

9.3.2 Why do We Use Statistics?

Main Teaching Point : Understanding the need to use statistics in various real life situations.

Teaching-Learning Process : Ask students to visualise the situations in which some data are needed to answer the questions posed. Let them start exploring their immediate environment : the school, the home, the community, the market place, busy thoroughfares, industry etc. Here are some situations:

1. Is the class doing better in mathematics than in English?
2. If the school is co-educational, are the girls doing better in mathematics than the boys in the class?
3. Is one section doing better than the other in some particular subject?
4. What is the average height of students in his/her class?
5. What is the average age of students in his/her class?
6. Has his/her school been improving its performance in the Board's examinations in the course of past ten years? (This will require a determination of indicators of performance and weights to be attached to these indicators.)
7. The expenditure in percent on different items in the family during the past one year (this will require a determination of broad categories under which expenditure should be recorded). If the expenditure is recorded monthly, does it show any trend?
8. The average number of children per family in the immediate neighbourhood (approximately 25 families may be considered).
9. Which size of shoes are sold the most (data may have to be collected for at least 10 to 15 days from a shop).

10. The number of vehicles per minute passing at a busy intersection of roads during different parts of the day (the survey may be extended to include different types of vehicles such as passenger cars, auto-rickshaws, tempos, two-wheelers, trucks, buses).
11. The average salary of workers in a factory.
12. The expenditure on advertisement and the corresponding sale over a period of time (this leads to the study of regression. Business houses use it to plan advertisement expenditure as a measure of sales promotion).

Listing various situations which require the use of statistics (collecting, presenting and interpreting data) could be a useful group activity where a group proposes a number of situations, discusses them and modifies them.

Methodology used : The discussion method is most suitable.

9.4 COLLECTING DATA

9.4.1 What are Data?

Main Teaching Point : Meaning of the word data.

Teaching-Learning Process : By now students would have realised that the situations listed above or those proposed by them require collection of some information to be able to answer the questions raised. Generally the information collected would be quantitative in nature but sometimes it might be qualitative also such as the degree of darkness of films put on glass panes, shades of colour of different objects from violet to red.

The information collected in each case constitutes the "data" for that particular study.

9.4.2 Sources of Data

Main Teaching Points : a) Collection of data.
b) Primary and secondary data.

Teaching-Learning Process : Explain to the students that **the set of objects or persons from whom data are collected forms the population for that study.** When data are collected directly from the elements of the population, it is called "primary data". However, many times the required data may already be available as part of data already collected for some bigger project or as a matter of routine exercise. The data obtained from a **secondary source**, not collected directly from the population, are called **secondary data.**

If we are assessing the relative performance of the students of a class in English and mathematics, then we may not go to the students to collect their marks because they would be available in the school records.

If we are collecting information about the sale of shoes of different sizes at a particular shoe store for a period of two weeks, we need not sit for two weeks at the store and record each sale. Instead, we can get the information from the stock register maintained at the store.

If we are making a study of the rural-urban population shift in the country over a period of fifty years, it would be impossible to collect the data from the primary source, but the information would be available in the records maintained with the Census Commissioner of India i.e., from a **secondary source.**

These examples illustrate the importance of secondary data. In fact, in modern days, every government lays emphasis on data collection. The population census helps the government to observe the rural-urban shift, the birth rate and the death rate, the growth rate of population, infant mortality, life expectancy etc. "specific death rates" help insurance agencies to determine the life insurance premium. A branch of mathematics titled "acturial science" is used to determine insurance premia.

Data on tax collection, exports and imports, production, GDP (Gross Domestic Product), etc., help the Government to plan strategies, the tax structure, incentives and taxation policies.

Divide the students in groups of four or five each. Let each group undertake a project on data collection. You should help each group refine its project and determine the line of action to be followed. A number of projects arise out of questions posed under sub-section 9.4.2. Why do we study statistics? Here are a few more possible projects.

1. The most common letter in five pages of your English textbook.
2. The most common length of words in say ten pages of a book.
3. Rolling a die 100 times and noting the number of times each score is obtained.

Methodology used : The lecture-cum-discussion method is used

9.5 ORGANISING DATA

9.5.1 Pictorial Presentation

Main Teaching Points : a) Bar chart
b) Pie chart
c) Pictogram

Teaching-Learning Process : Suppose a group of students undertake a project to find out how the students of their class travel to school. Their findings are :

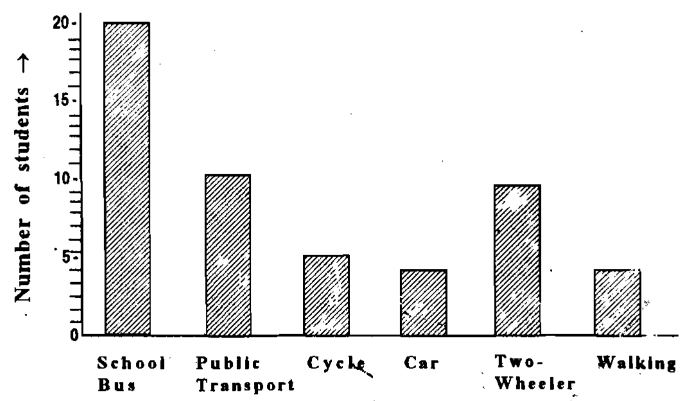
No. of students using the school bus	20
No. of students using public transport	10
No. of students cycling to their school	4
No. of students taken by their parents in a car	2
No. of students taken by their parents on a two-wheeler	7
No. of students walking to school	2
Total No. of Students	45

How best can these data be depicted pictorially?

Bar chart : One way is to use a bar chart. As the name suggests, the bar chart consists of bars of equal thickness (why?), with the length/height being proportional to the quantity the bars represent.

Here is the bar chart depicting the above information.

How students come to School



Note that :

i) The bar-chart has a heading.

ii) Sub-divisions are made on the vertical axis, using a suitable scale. (suitable means that the scale is so chosen that the space on the graph paper is optimally used by the given data).

iii) Bars are of equal thickness and do not touch each other (see next example).

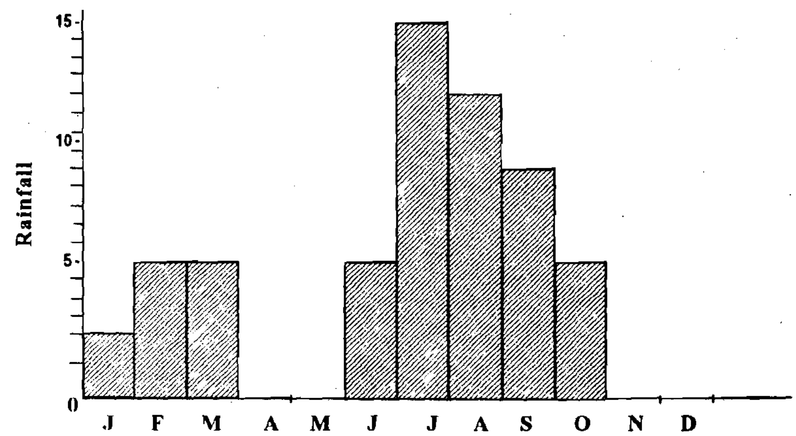
iv) Each bar is assigned a heading to denote what it represents.

v) The height of each bar represents the number it is supposed to represent.

Bars could be drawn horizontally also Then the numbering would be on the horizontal axis.

Here is another illustration.

Monthly rainfall in one year in a certain city



A city records rainfall (in cm) during 12 months of a certain year as follows:

January	2	July	15
February	5	August	12
March	5	September	8
April	-	October	5
May	-	November	-
June	5	December	-

Note that in this bar chart :

i) The bars are touching each other

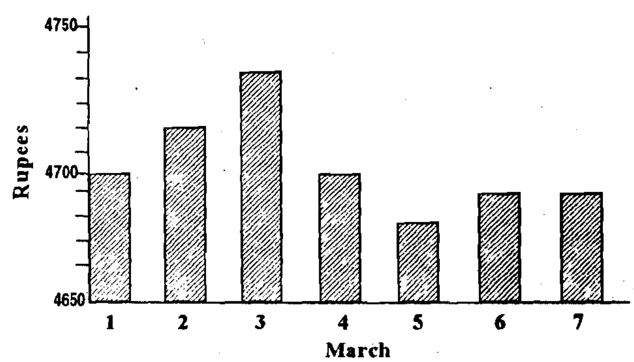
ii) The months showing no rainfall also find place in the bar chart.

In fact there is no hard and fast rule whether they should have intervening spaces. However, generally the bars are so drawn that they do not touch each other.

Sometimes the numerical scale may not begin from zero. This is particularly so if the change in the values of the variable is not large yet we want to depict the data vividly. Consider the day to day price variation for 10 gm of standard gold over a period of one week in a certain city.

Date	Price (per 10 gm)
1-3-95	Rs. 4,700
2-3-95	Rs. 4,710
3-3-95	Rs. 4,720
4-3-95	Rs. 4,700
5-3-95	Rs. 4,680
6-3-95	Rs. 4,690
7-3-95	Rs. 4,690

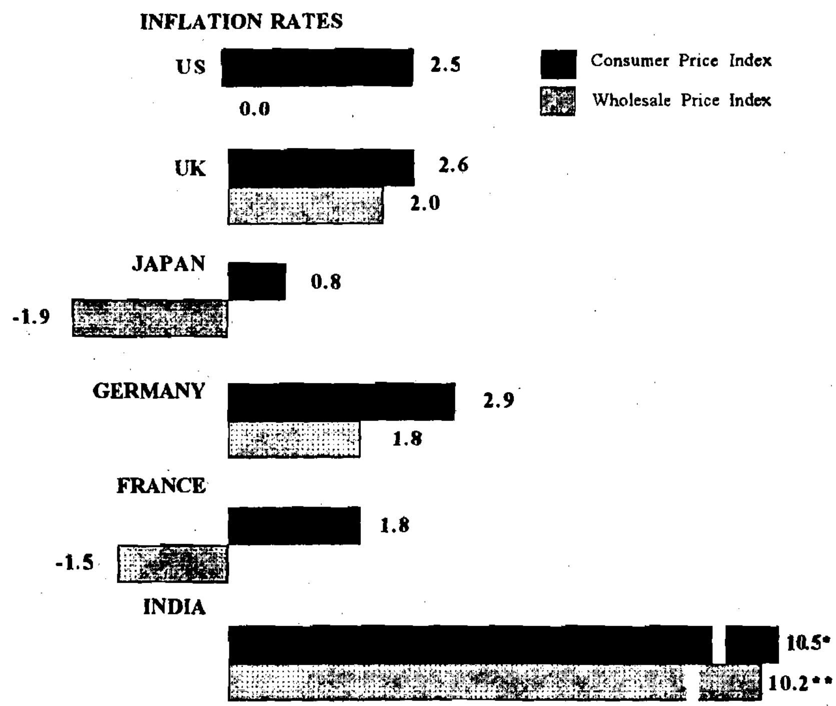
Price of 10 gm of standard gold



Ask the students to visualise what practical problem will arise if we were to start the scale from 0 instead of 4,750.

Sometimes, we need to make a comparative chart about two elements, such as exports/imports of a certain commodity over a period of time. Then two bars may be drawn side by side for each unit of time.

The accompanying diagram represents inflation rates in the consumer price index and the wholesale price index for a number of countries for June 1994 over June 1993.



Figures in percent for June 1994 over June 1993

***May 1994**

****Provisional**

Wholesale prices are stable in the US, and are falling in Japan and France. In India, however, they have risen by 10.2 percent when computed on an annual basis. Sure, the weak rupee is still protecting exports, but the exchange rate is also under strain at the moment.

● **Courtesy** : *Business Today*, Aug. 7-21, 1994.

An accompanying ability to draw bar charts is the ability to read and interpret a given bar chart. Present a few bar charts in the class and ask the students to describe what the bar charts show and what information can be derived from the same.

Pie Chart : Explain to the students that another mode of representing data is through the use of a pie chart (also called circle graph). The entire circle is taken to represent one whole and all the constituents of the entire data are represented propertionally in the pie chart.

Ask the students to suggest a method to determine the proportion.

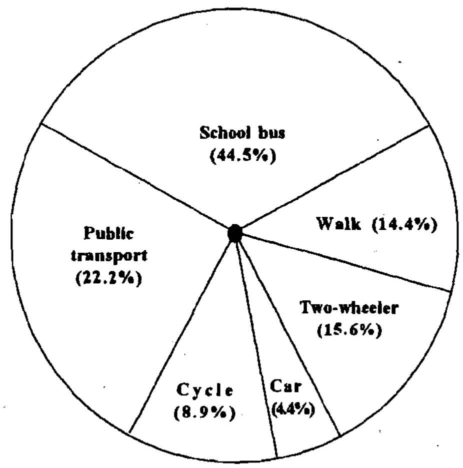
If an answer is not forthcoming straightaway, suggest that the angle around the centre of a circle is 360°. If we divide a circle into four equal parts by drawing two mutually perpendicular diameters, we get four quadrants. Each quadrant is 1/4 of the circle and the angle made by the two arms of a quadrant is also 1/4 of the entire angle, i.e., it is 90°. Similarly, a semi-circle divides the circle into two equal parts and the angle of 360° is also divided into two halves. Encourage the students to use their intuition to devise a method to make proportional parts of a circle.

The following table depicts to the "mode of transport to school" problem.

Mode of transport used	Number	Angle in the sector	Approximate percentage of the whole
Those using school bus	20	160°	44.5%
Those using public transport	10	80°	22.2%
Those using cycle transport	4	32°	8.9%
Those using parent's car	2	16°	4.4%
Those using parents two-wheeler	7	56°	15.6%
Those walking to the school	2	16°	4.4%
Total no. of students	45	360°	100%

A pie chart depicting the above information will look like the figure given below.

Students using different modes of transport



Note that while we construct a pie-chart using the idea of angles (since protractor helps us in constructing it), the magnitude of each item is described in terms of percent as shown.

Constructing a pie chart involves the following steps:

1. Work out the angle of the sector and approximate per cent of each item in the data.
2. Draw a circle and construct sectors using the sectoral angle.
3. Label each sector and write the per cent of the whole.

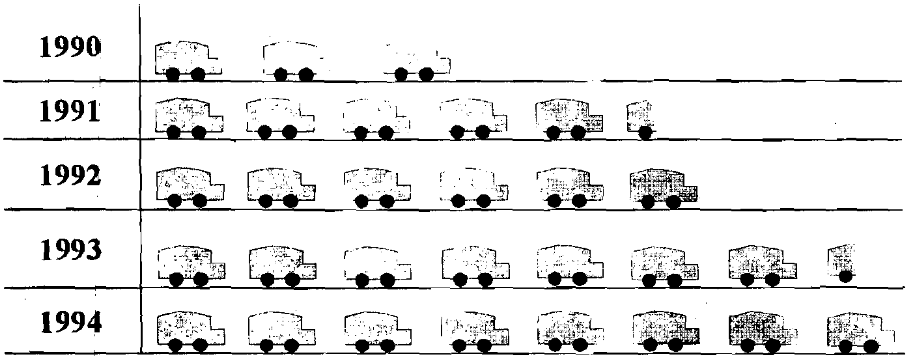
Ask the students to practice drawing pie charts for given data.

Pictogram : Sometimes the picture of an object is used to display information. For example, if we want to depict pictorially the production of cycles in the country over a number of years, we choose a suitable scale such as one cycle representing 1 lakh cycles and draw the required number of cycles. The last cycle representing some fraction of a lakh will be drawn only partly in the same proportion. Sometimes a dotted outline may be given to complete the picture. A pictorial representation of this type is called a pictogram.

Example : The production of motor cars in a certain country over a period of five years is shown by the following table:

Year	No. of Cars (in ten thousands)
1990	3
1991	5.5
1992	6
1993	7.5
1994	9

Taking one car to represent ten thousand, we can depict the above information in a pictogram as shown.



Production of Cars in the Country

( = ten thousand)

Methodology used : To draw figures is a skill to be developed by practice. A sufficient number of exercises should be provided to the students

Check Your Progress

Notes : a) Write your answers in the space given below.

b) Compare your answers with those given at the end of the unit.

2. The expenditure on health care by the Government of India during the first six five-year plans is shown below:

Plan	Expenditure (in crosses)
First	70
Second	140
Third	230
Fourth	340
Fifth	760
Sixth	1820

Construct a pie chart to depict the above information.

3. The number of schools of different types in Delhi in the year 1986 were as follows :

Primary	Middle	Secondary	Higher Secondary	Total
1838	366	259	663	3126

Illustrate by representing this information on a bar chart.

9.5.2 Handling Large Data

Main Teaching Points : a) Class intervals.

b) Constructing grouped frequency distribution and their graphical representation.

Teaching-Learning Process : So far we have restricted ourselves to representation of data which are not very large. You may now confront students with large data and ask them to represent them by a bar chart or in some other manner.

Ask : Suppose that a factory has 200 workers. You are provided with data about their monthly pay packet. You are now asked to represent it by a bar chart. How will you do it?

The students would obviously realise that it is futile to think of constructing 200 bars to represent pay packets of 200 persons. Therefore the next best option is to group together those persons whose salaries are close to some convenient figure, deviating from it by small amounts only. In other words, we choose a class of persons whose salaries lie in a small interval around that convenient figure. We make a number of such intervals. Now a host of questions arise. How do we go about constructing such class intervals? How many class intervals should we have? Once a person is included in a class interval, can we still get the information about his individual packet? The teacher will need to discuss these points one by one with the students.

Suppose the highest salary that a worker of the factory gets is Rs. 2,180 per month and the lowest salary a worker gets is Rs. 800. We need to divide this salary range into a suitable number of class intervals. If we round off the two salaries, we have the range from Rs. 800 to Rs 2200. The number of intervals should not be too small nor too large. Generally this number should lie between 5 and 15.

The intervals should be equal in size. From Rs. 800 to Rs 2200 is a range of Rs. 1400 We may make seven intervals of class size of Rs. 200 each. The number of workers who fall in one particular class interval is called the **frequency** of that class interval. To determine the frequency of each class interval we mark a tally for each person in that class, till all the data are exhausted. Explain to the students the mode of putting the tallies //// in groups of five tallies to facilitate counting.

Class mark, frequency, cumulative frequency

Suppose that the data about 200 workers of the factory is classified as below:

Class Interval	Tallies	Frequency	Cumulative Frequency
800-100	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 |||	33	33
1000-1200	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸	30	63
1200-1400	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 ||	32	95
1400-1600	𝍸 𝍸 𝍸 𝍸 𝍸 |||	28	123
1600-1800	𝍸 𝍸 𝍸 𝍸 𝍸	25	148
1800-2000	𝍸 𝍸 𝍸 𝍸 𝍸 ||	27	175
2000-2200	𝍸 𝍸 𝍸 𝍸 𝍸	25	200
	Total	200	

Note that :

i) the upper limit of each class interval appears again as the lower limit of the next interval. Whenever data are so classified, the overlapping limit is included only as the lower limit of the next interval so that there is no ambiguity.

ii) Cumulative frequency is the progressive total all the frequencies up to that class interval. Interpreted in physical terms, cumulative frequency such as 123 means that there are a total of 123 persons up to that class interval.

Ease of calculation vs loss of accuracy

Having classified all the data as above, an important question arises. From the data can we tell the salary of any particular worker of the factory?

The only thing we can say is that his salary lies in some particular income bracket. If we are required to calculate the entire monthly salary bill of the workers, we cannot do so unless we make an assumption. This important assumption is that all the persons in any particular class interval are supposed to be drawing a salary midway between the upper and lower limits of that class interval. This mid-value is called the "**class mark**" for that class.

Thus 33 persons are supposed to be drawing a salary of Rs. 800 + 1000/2 = Rs. 900 each; 30 persons are supposed to be drawing a salary of Rs. 1000 + 1200/2 = Rs. 1100 each, and so on.

Ask the students to guess how much inaccuracy this will introduce in the total salary. It will be difficult for them to answer it. The teacher should then point out that 33 persons in the first class interval will include persons some of whose salaries would be below the class mark and some whose salaries are above. The negative and positive deviations may balance each other to a good extent and only a small inaccuracy may enter in the total bill. This inaccuracy may be negligible compared to the total bill and the advantage of easy calculation may far outweigh the little error.

Histogram, frequency polygon, ogive

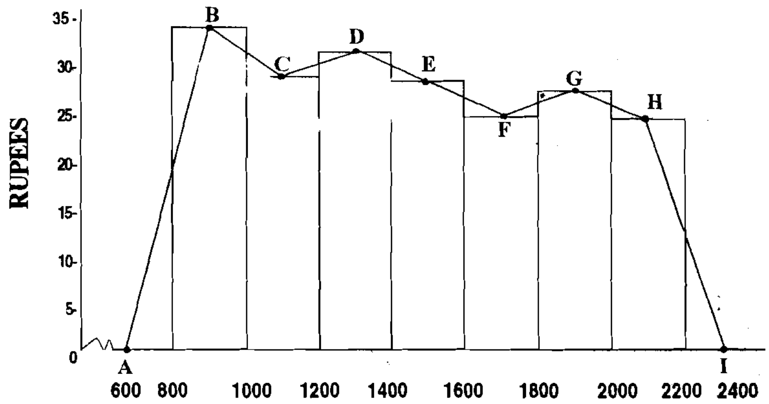
There are alternative modes of representation of data other than the bar chart, the pie chart and the pictogram. The histogram is closest to a bar graph. If the bars in a graph are drawn touching each other, then we get a histogram. Highlight the major differences between a bar chart and a histogram. They are:

i) The histogram is drawn when a frequency distribution is given.

ii) In a bar graph the lengths of the bars are proportional to the frequency, whereas in a histogram the area of the rectangle is proportional to the frequency. Clarify this using an example with unequal class intervals.

If the mid-points of the upper sides of the bars are joined by straight lines and the lines reach up to the base on either side, then we get a polygon. Each vertex of this polygon has coordinates equal to the mid-value of any interval and the corresponding frequency. In the adjoining diagram, B has coordinates (900, 33).

Histogram and Frequency Polygon



This represents the statement that the number of persons drawing a salary of Rs. 900 per month (as per our assumption) is 33. The polygon ABCDEFGHI is called "frequency polygon for this data (distribution)".

Note that :

i) B has been joined to A and H to I in the polygon. To locate A and I, we take the mid-points of the preceding and succeeding class intervals, i.e., of 600-800 and 2200-2400. Since the frequencies in these two intervals are zero, A and I lie on x-axis.

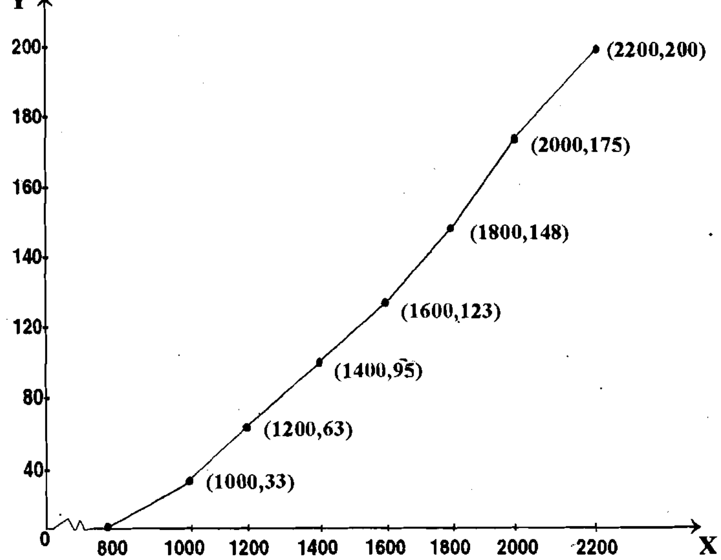
ii) A kink between O and M indicates that the OM length representing 0 to 600 has not been drawn to scale since that part is not required.

iii) If the points from A to I are joined by a free hand curve, the curve is called the frequency curve.

If both the histogram and the frequency polygon are to be drawn, then it is advisable first to draw the histogram and then join the mid-points of the tops of the rectangles of the histogram to get the frequency polygon as shown above. However if only the frequency polygon is to be drawn, then first represent the class marks along the x-axis and frequencies along the y-axis and then plot the corresponding points and finally join them.

Ogive : An ogive is a graph of cumulative frequency distribution. It is also called frequency distribution curve.

Cumulative frequency curve for wages of 200 factory workers.



To draw a cumulative frequency distribution curve, we plot the points with the upper limits of the classes on the x-axis and the corresponding cumulative frequencies on the y-axis.

Ask the students to explain why we take the upper limits and not the class-mark. Also ask them to explain why ogive is a rising curve.

The reason for taking the upper limit of each class is the **fact** that persons getting salaries within any class interval are actually spread over the whole class and we take all persons who are getting salaries upto the upper limit (cumulative frequency taken).

Ogive is a rising curve because the cumulative frequency goes on increasing with each class.

(**Note** : In the foregoing diagram, the ogive looks like a straight line since the frequencies in each class are very close to each other. But this may not always be the case).

Ask the students to practice drawing the ogive in all problems where they were asked to draw a histogram. They should practice both with or without the histogram.

Methodology used : The lecture-cum-discussion method is used with numerous illustrations.

Check Your Progress

Notes : a) Write your answers in the space given below.

b) Compare your answers with those given at the end of the unit.

4. Draw histogram and an ogive for the following data. Explain the steps involved in plotting histogram and ogive.

Marks	No. of Students
0-100	10
100-200	18
200-300	30
300-400	35
400-500	32
500-600	20
600-700	15
Total	160

9.6 INTERPRETING DATA

9.6.1 Measures of "Central Tendency"

Main Teaching Points : a) Mean
b) Mode
c) Median

Teaching-Learning Process : In the preceding sections 9.4 and 9.5, we discussed how we collect and display information. Let us now examine some of the ways of selecting a typical value to represent such information.

We go back to the first example in which we collected the heights of the 45 students of a class. Ask the students how they would choose one of the heights to represent all others.

There could be a number of suggetions such as:

i) Choose the height that most students have.

ii) Let all the students stand in increasing order and choose the middle-most student, i.e., the 23rd from either side. This suggestion is equivalent to arranging the numbers representing the heights in increasing order and then choosing the middle-most value.

iii) Add all the heights and divide the sum by the number of students, i.e., 45.

There might be a suggestion to choose the greatest height or the smallest height but hopefully it will not find favour with most students. If we look at the organised data, we notice that :

i) The height most students have is 148 cm;

ii) The middle-most height is 149 cm

iii) The sum of heights = 6709.5 and $\frac{6709.5}{45} = 149.1$

The teacher should explain that the value of the variable which occurs most number of times is called the "mode" of the data. The middle-most number after rearranging in increasing or decreasing order is called the "median" and the sum divided by the number of observations is called the "mean or arithmetic mean".

Pose the following problems:

a) If there were an even number of observations such as 40, how would you find the middle-most value?

b) If there were more than one number occurring an equal number of times, which one would you call mode?

c) If the data was grouped in class intervals, how would you calculate the mean of the data?

In case of even number of observations, we take the average of two middle-most values.

In case of most common values, we call it a bimodal distribution (bi means two).

In case of the grouped distribution, mean (which is denoted by $\bar{x}$) is given by $\bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i}$. Explain the meaning of $Ex_i f_i$.

Assign a few problems to the class to find mean, mode and median.

Explain the role of the "assumed mean" to facilitate calculation and explain $X = a + \frac{h \Sigma u_i f_i}{\Sigma f_i}$ where a = assumed mean, h is the class interval and $u_i = \frac{x_i - a}{h}$.

$x_i f_i$ as general notations for different values of x and f should be explained when they are first introduced.

Ask the students about how to choose the "assumed mean". Let them realise the importance of choosing a value somewhere mid-way in the grouped distribution data so that some u_i are negative and others are positive giving $\Sigma u_i f_i$ equal to a very small number.

Let the students discuss the following problems:

1. If each observation of some data is increased (or decreased) by 5, what happens to the mean? median?
2. If each observation of some data is multiplied (or divided) by 2, what happens to the mean?
3. If one of the observations of a data consisting of 10 values is wrongly copied as 65 in place of 25, will the true mean increase or decrease and by how much?
4. Show that :
   Mean, mode and median are called "measures of the central tendency" of a distribution

Methodology used : Mostly the lecture-cum-discussion method is used to transact the meaning and calculation of Central Tendencies. However, many examples should be given to illustrate the points.

Check Your Progress

Notes : a) Write your answers in the space given below.

b) Compare your answers with those given at the end of the unit.

How will you help the students in solving the following questions :

5. Find the mean of the first ten natural numbers.

....

....

....

6. Find the median of the following data:

a) 3, 10, 5, 9, 12, 15, 20, 13, 7, 1, 11

....

....

....

b) 2, 1, 3, 7, 9, 6, 5, 12

....

....

....

7. Find the mode of the following data:

3, 1, 2, 3, 5, 3, 1, 3, 6, 3, 9, 3, 3, 3, 7, 7, 3

....

....

....

8. Find the mean from the following data:

Marks	No. of Students
0-10	15
10-20	25
20-30	40
30-40	30
40-50	20
50-60	10

9.6.2 Measures of Dispersion : Spread, Deviation from Mean, Standard Deviation

Main Teaching Points : a) Spread
b) Deviation from mean
c) Standard deviation.

Teaching-Learning Process : Another important characteristic of any distribution is the spread or range of data.

Let the students consider the performance of two students A and B in their monthly tests in mathematics.

Students \ Test	I	II	III	IV	V	Total
A	15	24	40	9	32	120
B	28	30	24	18	20	120

The mean score of each of them is 24.

The highest score of A is 40 and the lowest 9.

The highest score of B is 30 and the lowest 18.

Thus the students are likely to observe that although A might be sharper, yet he is not consistent; B may not be as sharp as A but he is consistent.

Similarly, if two classes have the same mean score but the range (or spread) of the marks of one is large compared to that of the other, than the latter may be treated as a better class.

If the per capita income of two countries A and B is the same but there is a wide variation in the highest and lowest income group of country A than in country B, then A has more poor people and the disparities in income in A are higher than in B.

Range or spread is an important indicator of any distribution. It is called a "measure of dispersion" (i.e. how much the data is dispersed). We shall consider two more measures of dispersion.

Deviation from the mean : We have seen that the mean is an important representative of any given distribution. Some observations will be less than the mean and some will be more than it such that $\Sigma(x_i - \bar{x}) = 0$. To find out the average deviation from the mean we take the absolute deviation and find out its average. In case of ungrouped data

Mean deviation from the mean $= \frac{\Sigma (x_i - \bar{x})}{n}$

Standard deviation : Standard deviation is the most important and most frequently used measure of dispersion. It is denoted by σ (sigma) $\sigma = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}$

However the working formula is $\sigma = \sqrt{\frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2}$

It can be easily shown that the first formula simplifies into the second.

To find σ, we arrange the data in a tabular form as below .

Example : Find the standard deviation of the following distribution 6, 8, 9. 10, 12, 11. 7 8. 10, 9

X_i	X_i^2
6	36
8	64
9	81
10	100
12	144
11	121
7	49
8	64
10	100
9	81
90	840

$$\sigma = \sqrt{\frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2}$$

$$= \sqrt{\frac{840}{10} - \left(\frac{90}{10}\right)^2}$$

$$= \sqrt{84 - 9^2}$$

$$= \sqrt{84 - 81} = \sqrt{3}$$

$$= 1.732$$

The teacher should discuss with the students the method of finding standard deviation for grouped data and the short cut method to minimise lengthy calculations

Methodology used : Again we mostly depend on discussion with the students combined with the lecture method to teach the concept of dispersion and use the drill method to provide sufficient practice in the use of the formulae.

Check Your Progress

Notes : a) Write your answers in the space given below.

b) Compare your answers with those given at the end of the unit.

9. List the steps involved in finding the spread and standard deviation for the following data :

5, 10, 12, 7, 8, 6, 15, 17, 1, 19

....

....

....

....

....

....

....

....

....

....

....

10 Finding the standard deviation of the first ten natural numbers.

....

....

....

....

....

....

....

....

....

....

9.6.3 Inference and Prediction

The teacher should briefly introduce the students to the applications of statistics to set up relationships between two variables such as the relationship between advertisement expenditure made by a company and the sale proceeds of its products. Such a study comes under the title of correlation and regression. This brief mention should be treated only as a prelude to its study in post-secondary classes.

9.7 LET US SUM UP

This unit provides the teacher an opportunity to introduce the topic through everyday real life situations helping the students to develop a positive attitude towards the study of mathematics. The student sees the importance of collecting and organising data to be able to answer questions about the population from whom the data are collected. The students learn how to present data pictorially and to interpret such data appearing in books, newspapers, journals, etc.

The measures of the Central Tendency and of dispersion are introduced to the students as tools to make an objective assessment of the characteristics of any data. They learn how to determine these characteristics.

In the course of handling large data, the students learn that in order to make calculations less cumbersome, they have to make a sacrifice elsewhere (in accuracy). This is an attitude which often comes into play in real life situations also They are confronted with decision-making situations such as in deciding the number of class intervals or in making a choice of assumed mean.

Opportunities for logical thinking and deductive proof are provided in adequate measure. Transforming the formula for variation and standard deviation as obtained from its definition to a working formula provides a good challenge to the above average student. Inverse problems require that the students argue logically and deduce results.

9.8 UNIT-END ACTIVITIES

At the end of the unit let your students do the following :

1 Find out the age to the nearest month of each student in the class. Prepare a list in order of age and find the total

Next take your own age as a "reference point" or "zero". Write down the ages of everybody else as so many months older than or younger than yourself. Add these positive and negative values. Set up a formula to arrive at the previous total from this total.

2. Obtain the number of children in the families to which the students of your class belong. Make a bar chart with the number of children in the family along the x-axis and the number of families on the y-axis.

3. A family spends the following amounts monthly on the items listed below :

Food	Rs.	1,600/-
Clothing	Rs.	1,000/-
Housing	Rs.	2,400/-
Transport	Rs.	600/-
Education	Rs.	800/-
Misc. expenditure	Rs.	200/-
Savings	Rs.	600/-
Total	Rs.	7200/-

Represent the above information in a pie chart. Express the expenditure as per cent

4. The heights of 21 girls in centimeters were measured and the heights were :

157	160	163	157	162	162	164
165	155	147	162	162	164	156
158	156	161	164	159	162	159

Answer the following :

i) What is the greatest height?

ii) What is the least height?

iii) What is the median height?

iv) What is the mode?

v) What is the arithmetic mean?

5. Twenty-five students of a class were found to be absent for the following number of days during a month.

0, 0, 0, 0, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4, 5, 5, 6, 19, 20, 21, 21

i) What is the mode of this distribution?

ii) What is the median?

iii) Why is the mean not a good representative value for the distribution? Give reasons for your answer.

6. Fill in the blanks using mean, mode, median (there may be more than one correct statement).

i) .... is one of the actual readings.

ii) .... can be the greatest or least reading.

iii) When the readings are arranged in order of size, ... has as many numbers before it as after it.

iv) One or two extreme values may unduly bias .. . .

v) .... can have more than the value.

7. The following table represents marks obtained by 12 students in mathematics and physics respectively

Serial No.	1	2	3	4	5	6	7	8	9	10	11	12
Marks (Maths)	53	54	32	30	60	45	28	25	48	72	33	65
Marks (Physics)	55	41	48	49	27	24	23	20	28	60	43	67

In which subject is the level of achievement higher?

8. The following table gives the distribution of total household expenditure (in rupees) of factory workers in a city:

Expenditure	Frequency
600-700	24
700-800	40
800-900	33
900-1000	28
1000-1100	30
1100-1200	22
1200-1300	16
1300-1400	7
Total	200

i) Find the average expenditure per household using $\bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i}$

ii) If any assumed mean is used, what value should be taken as assumed mean?

Calculate using $\bar{x} = a + \frac{h \Sigma U_i f_i}{\Sigma f_i}$

where a = assumed mean

h = class size and $U_1 = \frac{X_1 - a}{h}$

9. Calculate deviation from the mean for the data given in Q 7. Calculate the standard deviation also for this data.

10. A school has 4 sections in class IX having 40, 35, 45 and 42 students. The mean marks obtained in the chemistry test by three of the sections are 50, 60, and 55 respectively. If the overall average of marks per student for all the section is 52.3, calculate the mean marks obtained by the fourth section.

11. Select any three concepts from the following:

a) Mean b) Mode c) Median d) Standard deviation

How will you teach these concepts to your students? Illustrate the methodology to explain the context.

9.9 ANSWERS TO CHECK YOUR PROGRESS

1. Weights (in gm) in decending order are (in gm) :

65, 63, 57, 55, 55, 55, 55, 55, 51, 50, 48, 47, 46, 45, 45, 45, 45, 45, 45, 42, 42, 40, 38, 35, 32.

a) Minimum weight = 32 gm

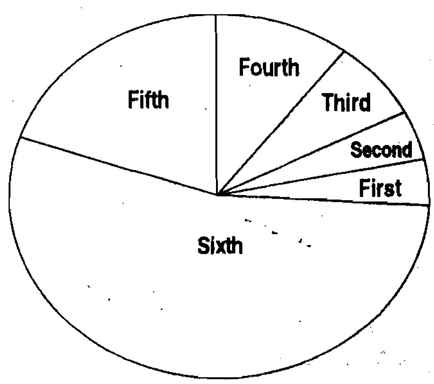
b) 45gm

c) 6

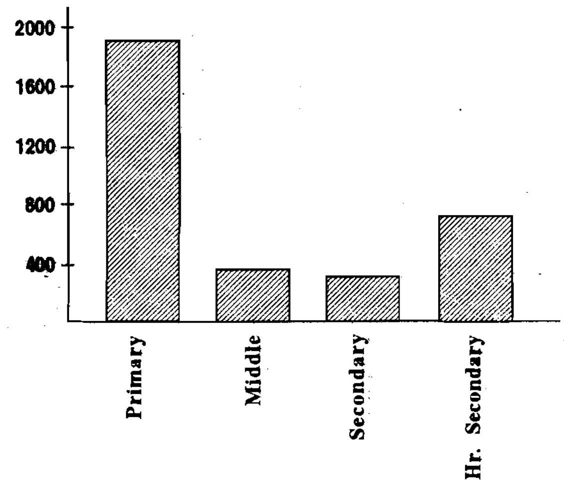
d) 5

2.

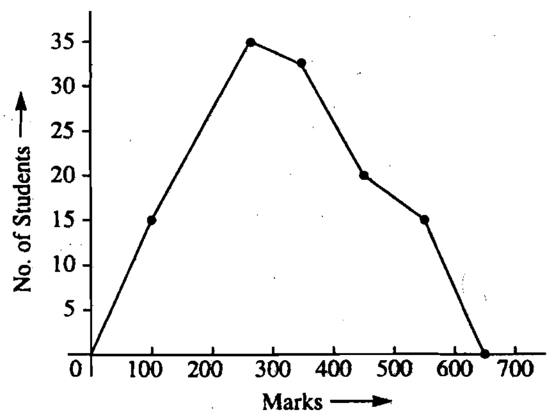
Plan	Expenditure (in crores)	Angle of Sector	%
First	10	7.5	2.08
Second	140	15.0	4.17
Third	230	24.5	6.85
Fourth	340	36.5	10.12
Fifth	760	81.5	22.62
Sixth	1820	195.0	54.16
Total	3360	360.0	100.00

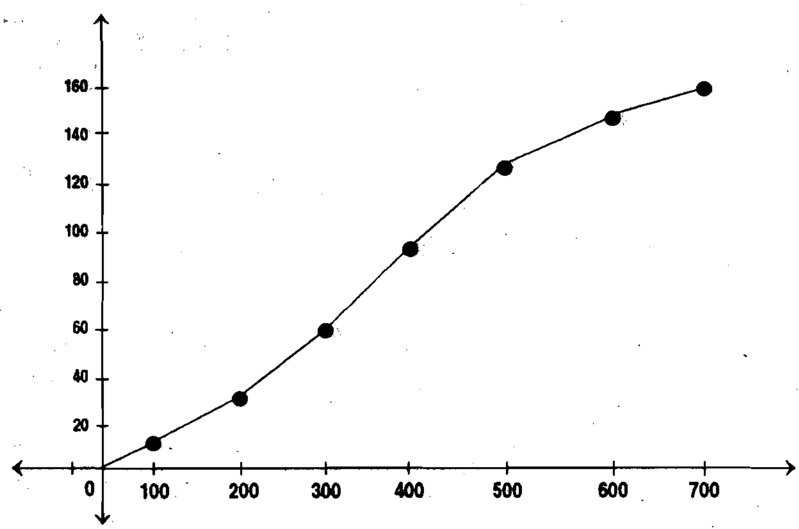


3.



4. **Histogram**





Ogive

5. Mean = 5.5

6. a) Median = 10 b) Median = 5.5

7. Mode = 3

8.

Marks	x	f	a = 35 x − a	h = 10 u = x − a/h	u.f
0-10	5	15	−30	−3	−45
10-20	15	25	−20	−2	−50
20-30	25	40	−10	−1	−40
30-40	35	30	0	0	0
40-50	45	20	10	1	20
50-60	55	10	20	2	20
Total	-	140			−95

$$\text{Mean} = a + \frac{h\Sigma uf}{\Sigma f}$$

$$= 35 + \frac{10 \times (-95)}{140}$$

$$= 35 - 6.79$$

$$= 28.21$$

9. Spread = 19 − 1 = 18

S.D.

X = 10

X : 5, 10, 12, 7, 8, 6, 15, 17, 1, 19

$d = X - \overline{X}$: −5, 0, 2, −3, −2, −4, 5, 7, −9, 9

d^2 : 25, 0, 4, 9, 4, 16, 25, 49, 81, 81

$$\text{S.D.} = \sqrt{\frac{\Sigma d^2}{N}} = \sqrt{\frac{294}{10}} = 5.42$$

10. $$\text{S.D.} = \sqrt{\frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2}$$

$$= \sqrt{385/10 - (55/10)^2}$$

$$= \sqrt{38.5 - 30.25}$$

$$= \sqrt{8.25}$$

$$= 2.87$$

9.10 SUGGESTED READINGS

M.J. Moroney (1961) : *Facts from Figures*, Penguin Books, Baltimore, USA.

William L. Hays, Holt, Rinehart and Winston (1965) : *Statistics for Psychologists*, New York.

Richard Goodman (1967) : *Teach Yourself Statistics*, The English Language Book Society, London.

J.K. Backhouse (1967) : *Statistics*, Longman, London.

The Teaching of Secondary School Mathematics (1970) : XXXIII Yearbook of NCTM Washington.

W. Servais and T. Varga (1971) : *Teaching School Mathematics* : *A UNESCO Source Book*, Penguin Books, UNESCO.