1. COLLECTION OF DATA

In day to day life people collect some type of data like,

(i) Height of 20 students in your school.

(ii) Height of 15 plants in or around your school.

(iii) Income of parents of students of class IXth.

1.1 PRIMARY DATA

A data collected by an investigation or a group of investigation for a definite purpose is called primary data.

For e.g. Class teacher collects the data of heights of students of class IX.

1.2 SECONDARY DATA

When a data is collected from a source which already had the information stored, the data is called a secondary data.

For e.g. Data of date of birth collected by teacher from the class – register of class X of school.

1.3 RAW DATA/UNGROUPED DATA

The data obtained in the original form are called raw/ ungrouped data.

For e.g. (i) Grades of 20 students in an examination as
A,A, D,C, A,D, G,E,E,F, H,G,H,E, C,D,H, F,E,A.

(ii) Marks of 10 students in maths exam as
73, 36, 60, 25, 42, 75, 78, 62.

1.4. RANGE

The difference between the maximum value and the minimum value of the variable is known as range.

For e.g. Range of 5, 11, 17, 8, 4, 19 is (19 – 4) = 15.

1.5 FREQUENCY

The count of tally marks or the number of observations in a particular class is its frequency.

1.6 UNGROUPED FREQUENCY DISTRIBUTION TABLE

When number of observations in an experiment is large then to make the data more easily understandable, we write it in a table which is called ungrouped frequency distribution.

For e.g. Marks obtained by 20 students is 10, 20, 26, 92, 95, 40, 50, 56, 60, 70, 92, 88, 80, 70, 72, 70, 36, 40, 36, 40.

Table is

Marks	No. of Students
10	1
20	1
36	3
40	3
50	1
56	1
60	1
70	3
72	1

80	1
88	1
92	2
95	1
	20

- We can also write above example in discrete grouped frequency distribution as.

Class	Frequency	Tally marks
10 – 19	1	I
20 – 29	1	I
30 – 39	3	III
40 – 49	3	III
50 – 59	2	II
60 – 69	1	I
70 – 79	4	IIII
80 – 89	2	II
90 – 99	3	III
	20	

- We can also write above example as continuous grouped frequency distribution as.

Class	Frequency	Tally marks
10 – 20	1	I
20 – 30	1	I
30 – 40	3	III
40 – 50	3	III
50 – 60	2	II
60 – 70	1	I
70 – 80	4	IIII
80 – 90	2	II
90 – 100	3	III
	20	

- 10 – 19, 20 – 29, 30 – 39, ... 90 – 99 are called class-interval or classes.
- Size of class is called class-size or class width, which is 10.
- Lower number of class interval is called lower class limit while greater number is called upper class limit.
- Class mark $= \frac{\text{lower limit} + \text{upper limit}}{2}$ n

2. GRAPHICAL REPRESENTATION OF DATA

We have discussed the representation of data in tabular form. There is another representation known as graphical representation of data.

These representations become easier than tabular form. We have the following graphical representation :

(a) Bar graphs
(b) Histograms
(c) Frequency polygons.

2.1 BAR GRAPH

A bar graph is a pictorial representation of data in which usually bar of uniform width are drawn with equal spacing between them on one axis and values of variable are shown on other axis.

Each rectangle or bar represents only one value of the data. So, the number of rectangles will be exactly the same as the number of values in the numerical data. The height (or length in case the base is on a vertical line) of each bar is proportional to the numerical values of the data. The height/length of each bar represents the numerical values of the data on a scale, selected suitably.

2.2 HISTOGRAMS

This is a form of representation like the bar graph, but it is used for continuous class intervals.

For a continuous frequency distribution, a series of rectangles are constructed having their widths equal to the widths of the classes and heights (or lengths) are selected in such a way that, the areas of the rectangles are respectively proportional to the frequencies of the classes. In case, the widths of the classes are uniformly same, the heights of the rectangles are selected proportional to the corresponding frequencies of the classes. By selecting suitable scales on x-axis and y-axis, the rectangles are drawn leaving no gap in between consecutive rectangles. The figure drawn appears like a single solid figure and it is called a histogram.

2.3 FREQUENCY POLYGON

It is another representation in which we join upper mid point of all the rectangles. The polygon so formed is called Frequency polygon.

Frequency polygon of a given continuous frequency distribution can be drawn in two ways :

(i) With the help of the histogram of the given frequency distribution.
(ii) Without taking the help of the histogram.

3 MEASURES OF CENTRAL TENDENCY

The variable in frequency table can be either qualitative or quantitative. In the case of quantitative variables the information contained in the raw data, or in frequency table can be presented by means of few numerical values. Methods providing such values are called measures of location or measures of central tendency and the numerical value obtained is called an average or central value.

The commonly used measures of central tendency are (i) Arithmetic mean (ii) Geometric mean (iii) Harmonic mean (iv) Median (v) Mode.

We will study about arithmetic mean, median, mode etc.

3.1 MEAN

For ungrouped data :

If x_1, x_2, ….. x_n are n values of a variable x, then the arithmetic mean of these values is given by

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

$$= \frac{\sum_{i=1}^{n} x_i}{n}$$

For grouped data :

If a variate x has values x_1, x_2, ….. x_n with the corresponding frequencies f_1, f_2, ….. f_n respectively, then the arithmetic mean of these values is given by

$$X = \frac{x_1 f_1 + x_2 f_2 + ... + x_n f_n}{\sum f_i = N}$$

$$= \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i = N}$$

3.2 MEDIAN

If x_1, x_2, ….. x_n are n values of a variable arranged in descending or ascending order, then

$$\text{Median} = \text{value of } \left(\frac{n+1}{2}\right)\text{th observation if n is odd}$$

$$\text{Median} = \frac{\text{value of } \left(\frac{n}{2}\right)\text{th observation} + \text{value of } \left(\frac{n}{2}+1\right)\text{th observation if } n \text{ is even}}{2}$$

3.3 MODE

It is the value which occurs most frequently in a set of observation and around which the other items of the set cluster densely.