Statistics

Data

A systematic record of facts or different values of a quantity is called **data**.

Data is of two types - Primary data and Secondary data.

Primary Data: The data collected by a researcher with a specific purpose in mind is called primary data.

Secondary Data: The data gathered from a source where it already exists is called secondary data

Features of the data

- Statistics deals with collection, presentation, analysis and interpretation of numerical data.
- Arranging data in a order to study their salient features is called presentation of data.
- Data arranged in ascending or descending order is called arrayed data or an array
- **Range** of the data is the difference between the maximum and the minimum values of the observations
- Table that shows the frequency of different values in the given data is called a **frequency distribution table**
- A frequency distribution table that shows the frequency of each individual value in the given data is called an ungrouped frequency distribution table.
- A table that shows the frequency of groups of values in the given data is called a grouped frequency distribution table
- The groupings used to group the values in given data are called classes or class-intervals. The number of values that each class contains is called the class size or class width. The lower value in a class is called the lower class limit. The higher value in a class is called the upper class limit.
- **Class mark** of a class is the mid value of the two limits of that class.
- A frequency distribution in which the upper limit of one class differs from the lower limit of the succeeding class is called an **Inclusive or discontinuous Frequency Distribution**.
- A frequency distribution in which the upper limit of one class coincides from the lower limit of the succeeding class is called an **exclusive or continuous Frequency Distribution**
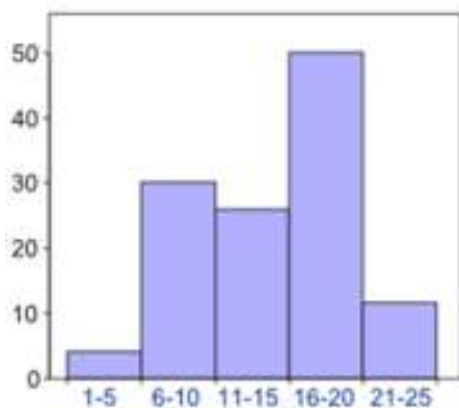
Graph's

Bar Graph

A bar graph is a pictorial representation of data in which rectangular bars of uniform width are drawn with equal spacing between them on one axis, usually the x axis. The value of the variable is shown on the other axis that is the y axis.

Histogram

A histogram is a set of adjacent rectangles whose areas are proportional to the frequencies of a given continuous frequency distribution

Mean

The mean value of a variable is defined as the sum of all the values of the variable divided by the number of values.

$$a_m = \frac{a_1 + a_2 + a_3 + a_4}{4} = \frac{\sum_0^n a}{n}$$

Median

The **median** of a set of data values is the middle value of the data set when it has been arranged in ascending order. That is, from the smallest value to the highest value

Median is calculated as

$$\frac{1}{2}(n+1)$$

Where n is the number of values in the data

If the number of values in the data set is even, then the **median** is the average of the two middle values.

Mode
Mode of a statistical data is the value of that variable which has the maximum frequency.