

UNIT 5:

Statistics

What is Statistics?

Statistics is simply defined as the study and manipulation of data. As we have already discussed in the introduction that statistics deals with the analysis and computation of numerical data. Let us see more definitions of statistics given by different authors here.

Statistics is the study of the collection, analysis, interpretation, presentation, and organization of data. In other words, it is a mathematical discipline to collect, summarize data. Also, we can say that statistics is a branch of applied mathematics. However, there are two important and basic ideas involved in statistics; they are uncertainty and variation. The uncertainty and variation in different fields can be determined only through statistical analysis. These uncertainties are basically determined by the [probability](#) that plays an important role in statistics.

Statistics Examples

Some of the real-life examples of statistics are:

- To find the mean of the marks obtained by each student in the class whose strength is 50. The average value here is the statistics of the marks obtained.
- Suppose you need to find how many members are employed in a city. Since the city is populated with 15 lakh people, hence we will take a survey here for 1000 people (sample). Based on that, we will create the data, which is the statistic.

Basics of Statistics

The basics of statistics include the measure of central tendency and the measure of dispersion. The central tendencies are [mean, median and mode](#) and dispersions comprise variance and standard deviation.

Mean is the average of the observations. Median is the central value when observations are arranged in order. The mode determines the most frequent observations in a data set.

Variation is the measure of spread out of the collection of data. Standard deviation is the measure of the dispersion of data from the mean. The square of standard deviation is equal to the variance.

Methods in Statistics

The methods involve collecting, summarizing, analyzing, and interpreting variable numerical data. Here some of the methods are provided below.

- Data collection
- Data summarization
- Statistical analysis

What is Data in Statistics?

Data is a collection of facts, such as numbers, words, measurements, observations etc.

Types of Data

1. **Qualitative data**- it is descriptive data.
 - Example- She can run fast, He is thin.
2. **Quantitative data**- it is numerical information.
 - Example- An Octopus is an Eight legged creature.

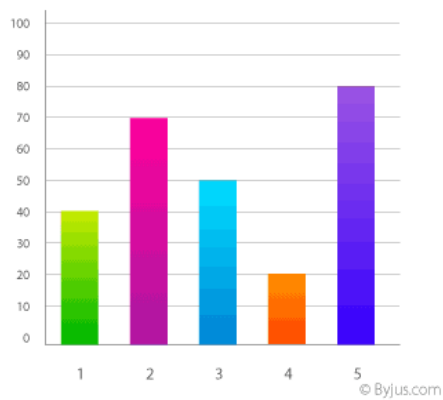
Types of quantitative data

1. **Discrete data**- has a particular fixed value. It can be counted
2. **Continuous data**- is not fixed but has a range of data. It can be measured.

Representation of Data

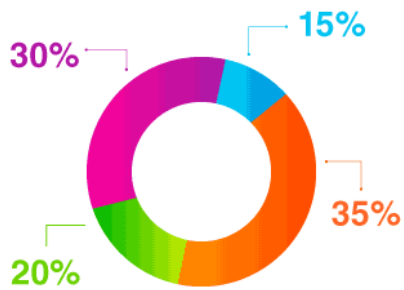
There are different ways to represent data such as through graphs, charts or tables. The general representation of statistical data are:

- Bar Graph
- Pie Chart
- Line Graph
- Pictograph
- Histogram
- Frequency Distribution



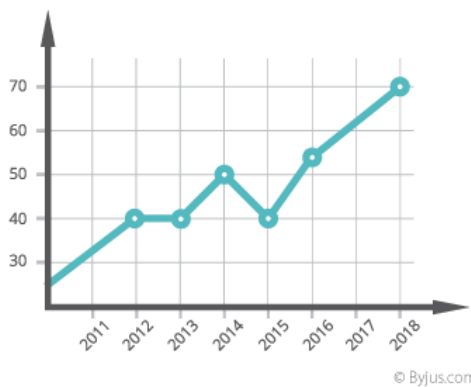
Bar Graph

A Bar Graph represents grouped data with rectangular bars with lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally.



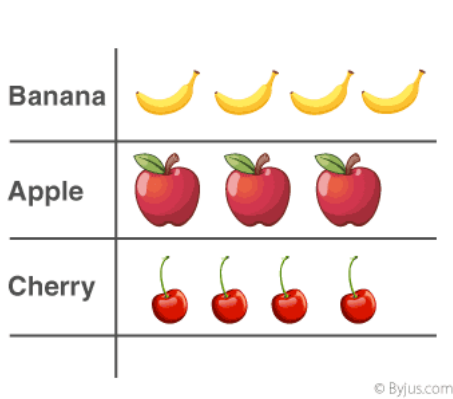
Pie Chart

A type of graph in which a circle is divided into **Sectors**. Each of these sectors represents a proportion of the whole.



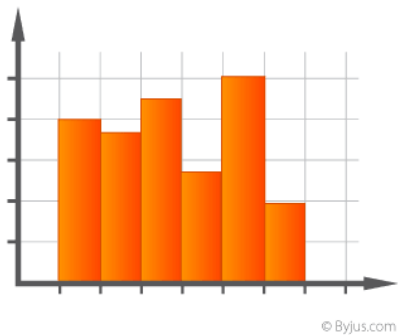
Line graph

The line chart is represented by a series of data points connected with a straight line. The series of data points are called 'markers.'



Pictograph

A pictorial symbol for a word or phrase, i.e. showing data with the help of pictures. Such as Apple, Banana & Cherry can have different numbers, and it is just a representation of data.



Histogram

A diagram is consisting of rectangles. Whose area is proportional to the frequency of a variable and whose width is equal to the class interval.

Marks obtained	Frequency
5	4
6	3
7	6
8	5
9	3
10	1

Frequency Distribution

The frequency of a data value is often represented by "f." A frequency table is constructed by arranging collected data values in ascending order of magnitude with their corresponding frequencies.

Measures of Central Tendency

In Mathematics, statistics are used to describe the central tendencies of the grouped and ungrouped data. The three measures of **central tendency** are:

- Mean
- Median
- Mode

All three measures of central tendency are used to find the central value of the set of data.

Measures of Dispersion

In statistics, the dispersion measures help interpret data variability, i.e. to understand how homogenous or heterogeneous the data is. In simple words, it indicates how squeezed or scattered the variable is. However, there are two types of dispersion measures, absolute and relative. They are tabulated as below:

Absolute measures of dispersion	Relative measures of dispersion
1. Range	1. Co-efficient of Range
2. Variance	2. Co-efficient of Variation
3. Standard deviation	3. Co-efficient of Standard Deviation
4. Quartiles and Quartile deviation	4. Co-efficient of Quartile Deviation
5. Mean and Mean deviation	5. Co-efficient of Mean Deviation

Skewness in Statistics

Skewness, in statistics, is a measure of the asymmetry in a probability distribution. It measures the deviation of the curve of the normal distribution for a given set of data.

The value of skewed distribution could be positive or negative or zero. Usually, the bell curve of normal distribution has zero skewness.

ANOVA Statistics

ANOVA Stands for Analysis of Variance. It is a collection of statistical models, used to measure the mean difference for the given set of data.

Degrees of freedom

In statistical analysis, the degree of freedom is used for the values that are free to change. The independent data or information that can be moved while estimating a parameter is the degree

of freedom of information.

Applications of Statistics

Statistics have huge applications across various fields in Mathematics as well as in real life. Some of the applications of statistics are given below:

- Applied statistics, theoretical statistics and mathematical statistics
- Machine learning and data mining
- Statistics in society
- Statistical computing
- Statistics applied to the mathematics of the arts