

# STATISTICAL DESCRIPTION OF DATA

## INTRODUCTION OF STATISTICS

Latin	▪ Status
Italian	▪ Statista
German	▪ Statistik (statistic)
French	▪ Statistique

### Definition of Statistics

statistics either in a singular sense or in a plural sense  
Statistics, when used as a plural noun, may be defined as data qualitative as well as quantitative, that are collected, usually with a view of having statistical analysis.

### Limitations of Statistics

- ❖ Statistics deals with the aggregates.
- ❖ Statistics is concerned with quantitative data.
- ❖ Future projections of sales, production, price and quantity etc. are possible under a specific set of conditions.
- ❖ The theory of statistical inferences is built upon random sampling.

# STATISTICAL DESCRIPTION OF DATA

## COLLECTION OF DATA

We can broadly classify data as

- (a) Primary;
- (b) Secondary.

Collection of data plays the very important role for any statistical analysis. The data which are collected for the first time by an investigator or agency are known as primary data whereas the data are known to be secondary if the data, as being already collected, are used by a different person or agency.

Collection of Primary Data The following methods are employed for the collection of primary data:

- i. Interview method;
- ii. Mailed questionnaire method;
- iii. Observation method;
- iv. Questionnaires filled and sent by enumerators.

# STATISTICAL DESCRIPTION OF DATA

## Interview method again could be divided into

- A. Personal Interview method,
- B. Indirect Interview method and
- C. Telephone Interview method.

## Sources of Secondary Data

There are many sources of getting secondary data. Some important sources are listed below:

- ❖ International sources like WHO, ILO, IMF, World Bank etc.
- ❖ Government sources like Statistical Abstract by CSO, Indian Agricultural Statistics by the Ministry of Food and Agriculture and so on.
- ❖ Private and quasi-government sources like ISI, ICAR, NCERT etc.
- ❖ Unpublished sources of various research institutes, researchers etc.

# STATISTICAL DESCRIPTION OF DATA

## Difference between Primary data and Secondary data

<u>Primary data</u>	<u>Secondary data</u>
<ul style="list-style-type: none"><li>• Primary data is the collection of original data For the first time.</li></ul>	<ul style="list-style-type: none"><li>• Secondary data are basically compensation of Existing data</li></ul>
<ul style="list-style-type: none"><li>• It is collected by the investigator himself</li></ul>	<ul style="list-style-type: none"><li>• It is collected by the person other than who requires it.</li></ul>
<ul style="list-style-type: none"><li>• It is usually directly suitable to the purpose Of enquiry.</li></ul>	<ul style="list-style-type: none"><li>• Its collections are relatively less costly.</li></ul>
<ul style="list-style-type: none"><li>• There is possibility of personal bias in its Collection.</li></ul>	<ul style="list-style-type: none"><li>• It may or may not be directly suitable to the Purpose of enquiry .</li></ul>

## Scrutiny of Data

the statistical analyses are made only on the basis of data, it is necessary to check whether the data under consideration are accurate as well as consistence.

$$\text{Density} = \frac{\text{Area}}{\text{Population}} \text{ holds.}$$

# STATISTICAL DESCRIPTION OF DATA

## PRESENTATION OF DATA

data are collected and verified for their homogeneity and consistency, we need to present them in a neat and condensed form highlighting the essential features of the data.

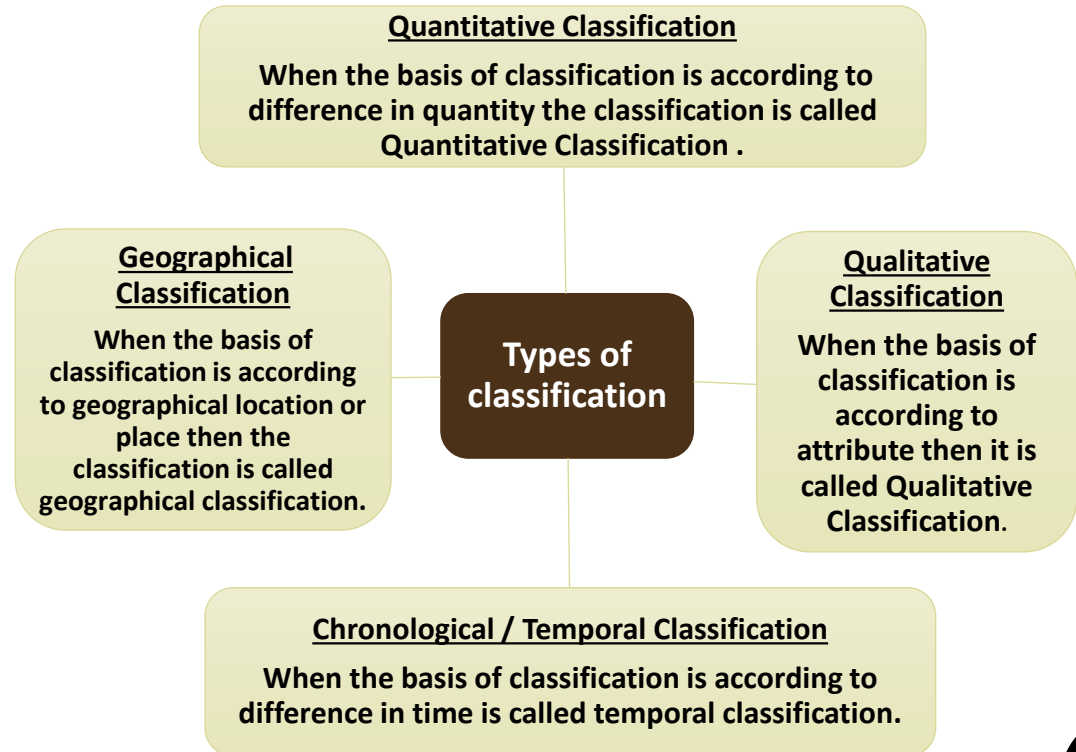
### Classification or Organization of Data

- It puts the data in a neat, precise and condensed form so that it is easily understood and interpreted.
- It makes comparison possible between various characteristics, if necessary, and thereby finding the association or the lack of it between them.
- Statistical analysis is possible only for the classified data.
- It eliminates unnecessary details and makes data more readily understandable.

### Data may be classified as -

- Chronological or Temporal or Time Series Data;
- Geographical or Spatial Series Data;
- Qualitative or Ordinal Data;
- Quantitative or Cardinal Data.

# STATISTICAL DESCRIPTION OF DATA



### Mode of Presentation of Data

- Textual presentation
- Tabular presentation or Tabulation
- Diagrammatic representation of data
- Graphical Representation of a Frequency distribution

# STATISTICAL DESCRIPTION OF DATA

## □ Textual presentation

This method comprises presenting data with the help of a paragraph or a number of paragraphs. The official report of an enquiry commission is usually made by textual presentation.

## □ Tabular presentation or Tabulation

Tabulation may be defined as systematic presentation of data with the help of a statistical table having a number of rows and columns and complete with reference number, title, description of rows as well as columns and foot notes, if any.

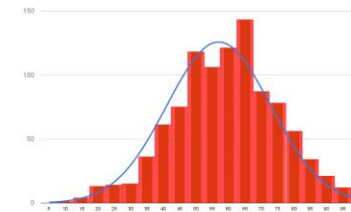
## □ Diagrammatic representation of data

Another alternative and attractive representation of statistical data is provided by charts, diagrams and pictures. Unlike the first two methods of representation of data, diagrammatic representation can be used for both the educated section and uneducated section of the society.

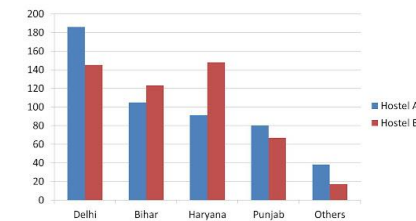
# STATISTICAL DESCRIPTION OF DATA

We are going to consider the following types of diagrams:

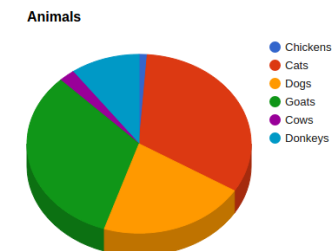
## △ Line diagram or Histogram;



## △ Bar diagram;



## △ Pie chart.

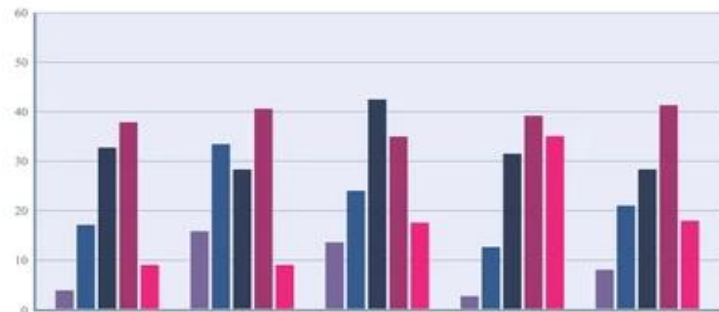


# STATISTICAL DESCRIPTION OF DATA

## □ GRAPHICAL REPRESENTATION OF A FREQUENCY DISTRIBUTION

We consider the following types of graphical representation of frequency distribution:

### ⊡ Histogram or Area diagram



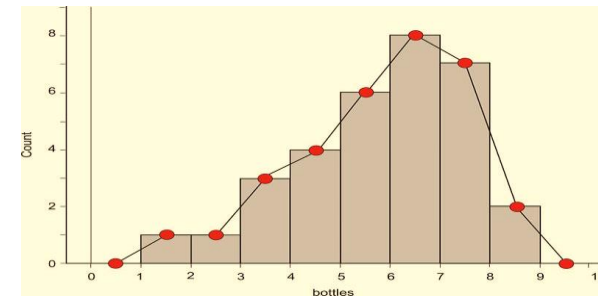
This is a very convenient way to represent a frequency distribution. Histogram helps us to get an idea of the frequency curve of the variable under study. Some statistical measure can be obtained using a histogram.

### ⊡ Frequency Polygon

Usually, frequency polygon is meant for single frequency distribution. However, we also apply it for

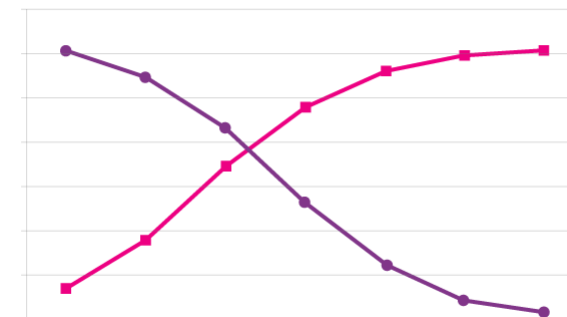
# STATISTICAL DESCRIPTION OF DATA

grouped frequency distribution provided the width of the class intervals remains the same. A frequency curve can be regarded as a limiting form of frequency polygon.



### ⊡ Ogives or Cumulative Frequency Graph

By plotting cumulative frequency against the respective class boundary, we get ogives.



© Byjus.com



## STATISTICAL DESCRIPTION OF DATA

As such there are two ogives - less than type ogives, obtained by taking less than cumulative frequency on the vertical axis and more than type ogives by plotting more than type cumulative frequency on the vertical axis and thereafter joining the plotted points successively by line segments.

### FREQUENCY DISTRIBUTION

frequency data occur when we classify statistical data in respect of either a variable or an attribute. A frequency distribution may be defined as a tabular representation of statistical data, usually in an ascending order, relating to a measurable characteristic according to individual value or a group of values of the characteristic under study.

Some important terms associated with a frequency distribution :

#### Class Limit (CL)

the class limits may be defined as the minimum value and the maximum value the class interval may contain. The minimum value is known as the lower-class limit (LCL)

## STATISTICAL DESCRIPTION OF DATA

and the maximum value is known as the upper class limit (UCL).

#### Class Boundary (CB)

Class boundaries may be defined as the actual class limit of a class interval.

$$LCB = LCL - \frac{D}{2}$$

$$\text{and } UCB = UCL + \frac{D}{2}$$

#### Mid-point or Mid-value or class mark

Corresponding to a class interval, this may be defined as the total of the two class limits or class boundaries to be divided by 2.

$$\begin{aligned} \text{mid-point} &= \frac{LCL+UCL}{2} \\ &= \frac{LCB+UCB}{2} \end{aligned}$$

# STATISTICAL DESCRIPTION OF DATA

## Cumulative Frequency

The cumulative frequency corresponding to a value for a discrete variable and corresponding to a class boundary for a continuous variable may be defined as the number of observations less than the value or less than or equal to the class boundary.

## Frequency Curve

A frequency curve is a smooth curve for which the total area is taken to be unity. It is a limiting form of a histogram or frequency polygon. The frequency curve for a distribution can be obtained by drawing a smooth and free hand curve through the mid-points of the upper sides of the rectangles forming the histogram.

There exist four types of frequency curves namely

- (a) Bell-shaped curve;
- (b) U-shaped curve;
- (c) J-shaped curve;
- (d) Mixed curve