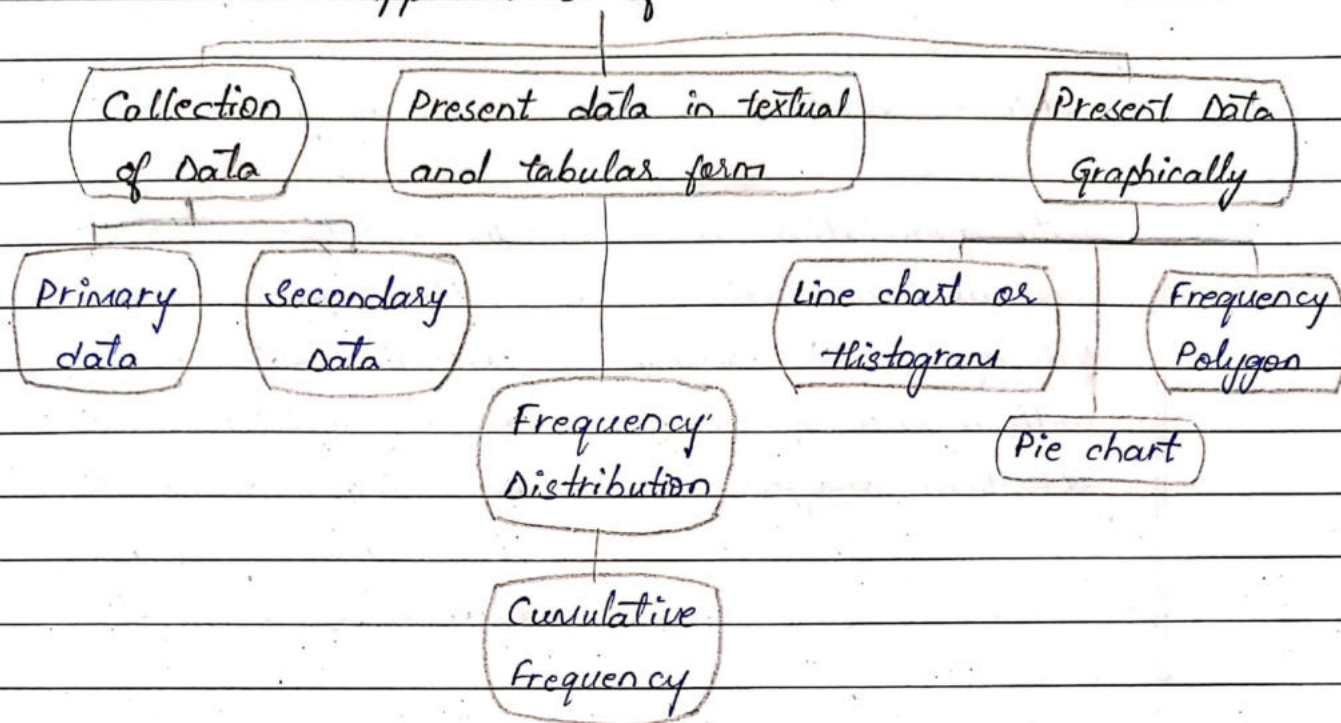


13. unit-1: Statistical Description of Data

Applications of Statistics



Introduction of statistics:

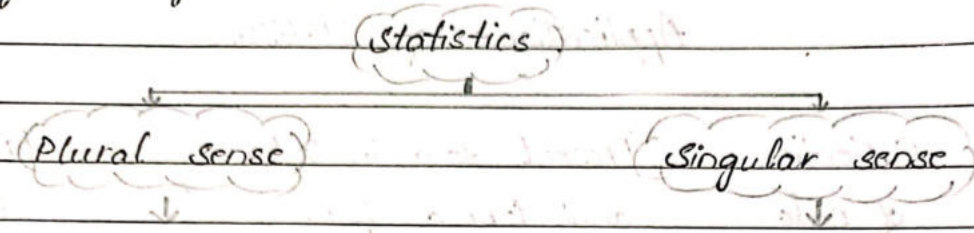
- The modern development in the field of not only management, Commerce, Economics, social sciences, Mathematics, and so on but also in our life like public services, defence, banking, insurance sector, tourism and hospitality, police and military etc. are dependent on a particular subject known as statistics.
- Statistics does play a vital role in enriching a specific domain.
- Statistics has almost a universal application.

History of statistics:

→ The origin of the word "statistics"

- Latin word - status
- Italian word - statista
- German word - statistik believed by scholars
- French word - statistique

Definition of statistics:



Data qualitative as well as quantitative, that are collected usually with a view of having statistical analysis.

The scientific method that is employed for collecting, analysing & presenting data, leading finally to drawing statistical inferences about some imp characteristics it means it is "science of counting" or "science of averages".

Application of Statistics:

⇒ Economics: Modern development in Economics have roots in statistics. In fact, Economics and statistics are closely associated. Time series, Analysis, Index Numbers, Demand Analysis etc are some overlapping areas of Economics & statistics.

- Econometrics - a branch of Economics that interact with statistics in a very positive way.
- Application of statistics in economics is for making future projection of demand of goods, sales, prices, quantities etc.

⇒ Business Management:

- statistical decision theory is another component of statistics that focuses on the analysis of complicated business strategies with a list of alternatives - their merits as well as demerits.

⇒ Statistics in Commerce and Industry

- The industrialists and the businessmen are expanding their horizon of industries and businesses with the help of statistical procedures.

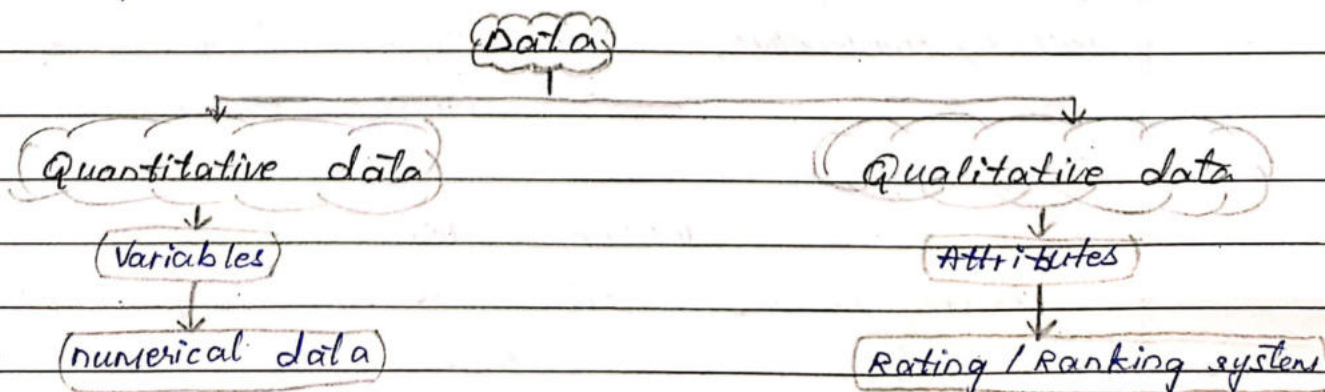
- Data on previous sales, raw materials, wages and salaries, production of identical nature of other factories etc are collected, analysed and experts are consulted in order to maximise profits.
- Measures of central tendency and dispersion, correlation and regression analysis, time series analysis, index numbers, sampling, statistical quality control are some of the statistical methods employed in commerce and industry.

Limitations of Statistics.

- Statistics deals with the ^{average} aggregates.
- It is concerned with quantitative data.
- Future projection of sales, production, price & quantity etc. are possible under a specific set of conditions. If any of these condⁿ is violated, projections are likely to be inaccurate.
- The theory of statistical inferences is built upon random sampling. If the rules for random sampling is not strictly adhered to, the conclusion drawn on the basis of these unrepresentative sample would be ^{wrong} erroneous.
- The experts should be consulted before deciding the sampling scheme.

Collection of Data

Data: It may be defined as quantitative information about some particular characteristics under consideration.



Types of variables

Discrete variable

Continuous variable

a finite or a countably infinite no. of isolated values
Ex: no. of petals in flower, the no. of misprints of a book...

when variable can assume any value from a given interval
Ex: height, weight, sale, profit and so on

Sources of Data

Primary sources

Secondary sources

The data collected for the first time by investigator or agency is known as primary data

The data as being already collected, and used by a different person or agency is known as secondary data

Collection of Primary Data

Sources of secondary data

- Interview method
- Mailed questionnaire method
- Observation method
- Questionnaires filled and sent by enumerators

- International sources like WHO, ILO...
- Govt. sources like Statistical Abstract by CSO, etc.
- Private & quasi-govt. sources like ISI.
- Unpublished sources of various research institutes, researchers etc.

Interview Method

Personal interview method

Indirect interview method

Telephone interview method

Scrutiny of Data: Checking the quality of data. Since 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 consistent.

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

PRESENTATION OF DATA:

Once ^{the} data are collected and verified for their ^{same type} homogeneity and consistency, we need to present them. Any statistical analysis is dependent on a proper presentation of the data.

→ Objectives of classification of data

- simplification and brevity
- Comparability
- Statistical analysis
- makes data more understandable.

→ Data may be classified as

- chronological or Temporal or Time series data
- Geographical or Spatial series data
- Qualitative or ordinal data (ranks)
- Quantitative or cardinal data (numbers)

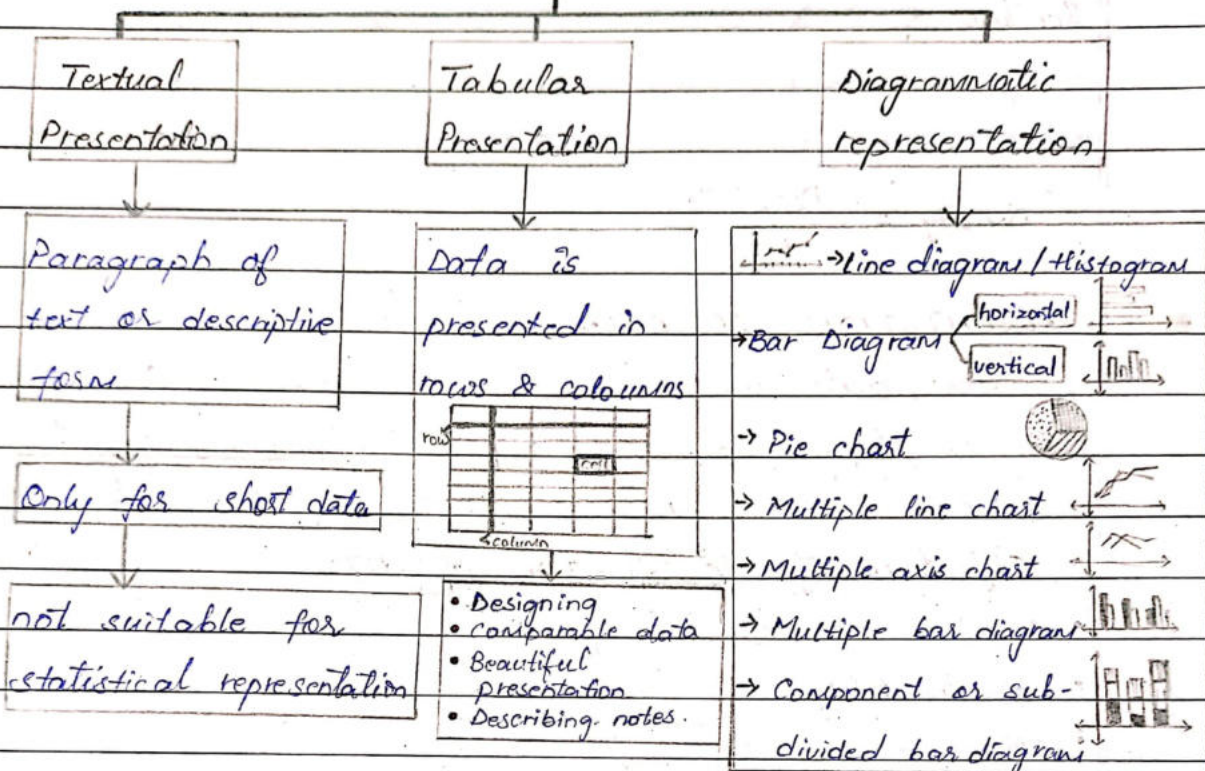
Further data may be classified as frequency data and non-frequency data.

Frequency data: The frequency of a particular value is the number of times that value occurs in a dataset.

Non-frequency data: It refers to data where values or observations are not repeated frequently. Time series data and geographical data belong to the non-frequency group.

Note: The qualitative as well as quantitative data belong to the frequency group.

Modes of Presentation of Data



Pie chart:

Angle of circle = $360^\circ = 100\%$

$1\% = 3.6^\circ$, $10\% = 36^\circ$

Calculation: $10 \times \frac{360}{100} = 36$

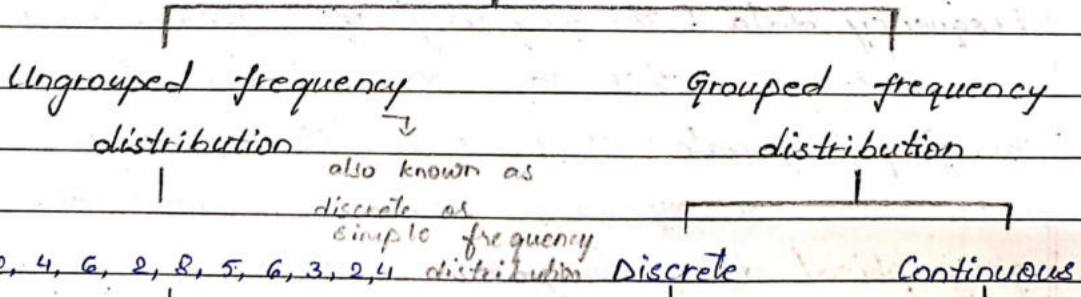
To convert in percentage

$$\frac{\text{Favourable value}}{\text{Total value}} \times 100$$

To convert percentage in degree

(Given %) $\times 3.6^\circ$ or $\frac{360}{100}$

Frequency Distribution



	Tally mark	frequency
0		1
1		1
2		2
3		3
4		4
5		4
6		4
7		4
8		4
9		4
10		4

1-10
11-20
21-30

0-10
10-20
20-30

Important terminologies:

- **Class limit (CL):** Corresponding to a class interval the class limit may be defined as the min value and the max value the class interval may contain. The min value is known as the lower class limit (LCL) and the max value is known as the upper class limit (UCL).

- **Class boundary:** class boundary may be defined as the actual class limit of a class interval.

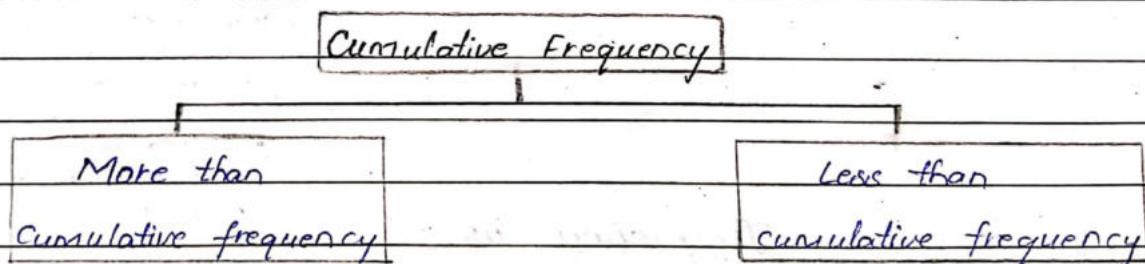
$$LCB = LCL - \frac{D}{2}, \quad UCB = UCL + \frac{D}{2} \quad D = UCL - LCL$$

- **Mid-point or Mid-value or class mark:** This may be defined as the total of the two class limits or class boundaries to be divided by 2.

$$\text{mid-point} = \frac{UCB + LCB}{2} \quad \text{or} \quad \frac{LCL + UCL}{2}$$

- **Width or size of a class interval:** The width of a class interval may be defined as the difference between the UCB and the LCB of that class interval. $UCB - LCB$

- **Cumulative Frequency:** Cumulative frequency refers to ^{the} running total of frequencies, it's the sum of all frequencies up to a certain point or class interval in a data set.



- **Frequency density of a class interval:** It may be defined as the ratio of the frequency of that class interval to the corresponding class length.

$$\text{Frequency Density} = \frac{\text{Class Frequency}}{\text{Class length}}$$

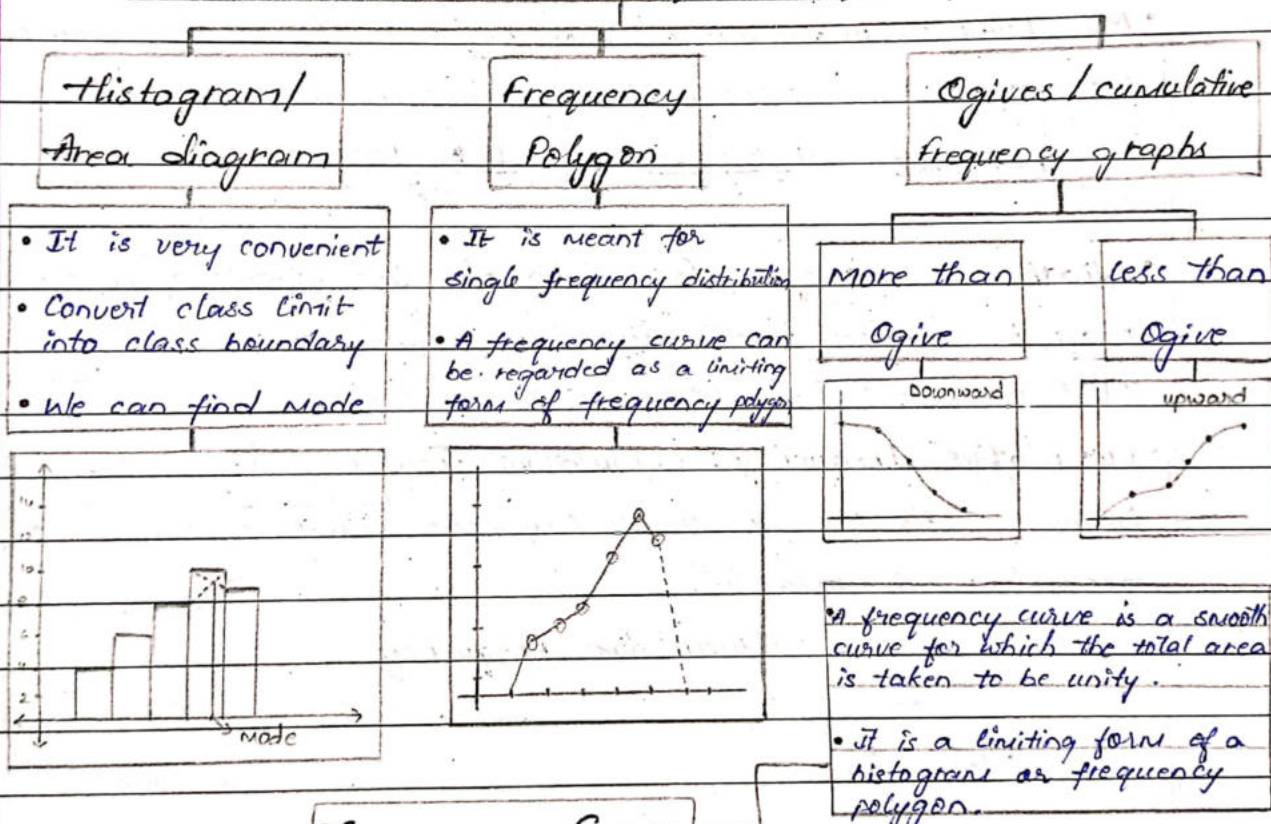
• Relative frequency of a class interval: It may be defined as the ratio of the class frequency to the total frequency.

$$\text{Relative frequency} = \frac{\text{Frequency of class interval}}{\text{Total frequency}}$$

• Percentage frequency of a class interval: It may be defined as the ratio of class frequency to the total frequency, expressed as a percentage.

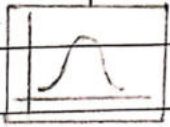
$$\text{Percentage frequency} = \text{Relative frequency} \times 100$$

Graphical Representation of a Frequency Distribution

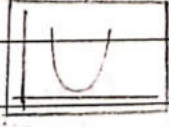


Frequency Curve

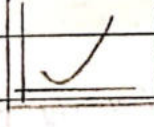
Bell shaped curve



U-Shaped Curve



J-Shaped Curve



Mixed Curve

