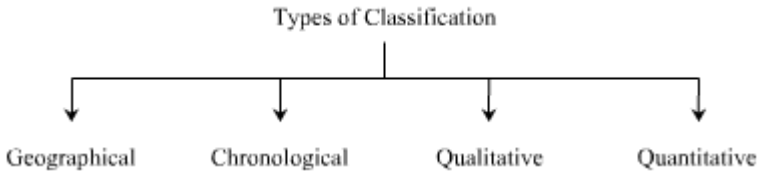
CHAPTER 3

ORGANISATION OF DATA

- **Organisation of data** refers to the act of arranging figures in such a way that comparison and further analysis can be made possible.

- **Advantages of Classification**
  - Saves our time and energy
  - Easy to fetch information
  - Easy to draw conclusions and inferences

- **Types of Classification**



  - **Geographical Classification**-The classification of data on the basis of geographical locations such as, states, countries, cities, districts, etc. is called **geographical classification**.
  - **Chronological Classification-** The classification of data on the basis of their occurrence at different time period is called **chronological classification**.
  - **Qualitative Classification-** The classification of data on the basis of the qualities or the attributes of data is called **qualitative classification**.
  - **Quantitative Classification-** The classification of data on the basis of the numerical values into different classes or groups is known as **quantitative classification**.

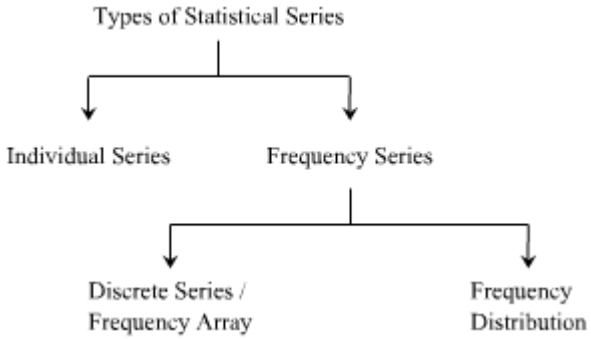
- **Variable**- A measurable characteristic whose value changes overtime is called a **variable**.

- **Types of Variable**
  A variable can either be **discrete** or **continuous.**
  - A variable that takes only whole number as its value is called discrete variable. For example, number of people in a family, number of students in a class, etc.
  - A variable that can take any value within a reasonable limit is called a continuous variable. For example, age, height, weight, etc.

- **Raw Data-** A mass data in its crude form is called **raw data**. It is classified in the form of series.

- **Series-** It refers to those data that are presented in some specific order and sequence.

Types of Statistical Series



- **Individual Series** are the series that consists of single set of values.
  For example:

Series showing marks obtained by 10 students in a class										
S. No.	1	2	3	4	5	6	7	8	9	10
X	30	25	25	15	17	19	34	46	50	46

- **Frequency Array or Discrete Series** are those series in which data is presented in the form of exact measurement.
  For example:

Series showing marks obtained by 20 students in a class out of 30	
Marks (*X*)	**Frequency (*F*)**
25	2
26	3
27	5
28	3
29	2
30	5
Total	**20**

- **Continuous series** are those series in which items assume a range of values and are placed within those ranges (or class intervals).
  For example:

Series showing marks obtained by 20	
Marks (*C.I*)	**Frequency (*F*)**
20 – 25	2
25 – 30	3
30 – 35	5
35 – 40	3
40 – 45	2
45 – 50	5
Total	**20**

- **Frequency** is the number of times an item occurs in the series.

- **Class frequency** is the number of times an item repeats itself corresponding to a range of value.

- **Size of class** refers to the width of the class.

$$\text{Size of Class} = \frac{\text{Range}}{\text{Number of Classes}}$$

- **Class** is the range of values that incorporates a set of items.

  For example, 5 – 10, 10 – 15, 15 – 20.

- The extreme values (the lowest value and the highest value) of a class are called **limits** of that particular class. For example, in class 5 – 10, 5 is the lower limit and 10 is the upper limit.

- The difference between the upper limit and lower limit is called **magnitude of that class interval.**

  Algebraically,

  $$i = l_2 - l_1$$

  where,
  i = magnitude of class interval
  l_2 = upper limit of class interval
  l_1 = lower limit of class interval

- **Mid-value** is the average value of upper and lower limit.

$$\text{Mid-value} = \frac{\text{Upper Limit} + \text{Lower Limit}}{2}$$

For example, mid-value of class interval $10 - 20$ is

$$\text{Mid-value} = \frac{20 + 10}{2} = 15$$

- **Univariate Distribution Series** are defined as the series of statistical data of only one variable.

- **Bivariate Distribution Series** are defined as the series of statistical data of two variables.

- **Multivariate Distribution Series** are defined as the series of statistical data of many variables simultaneously.

- **Exclusive series** are those series in which the upper limit of one class becomes the lower limit of the following (next) class.

For example:

Marks (*C.I*)	**Frequency (*F*)**
20 – 25	2
25 – 30	3
30 – 35	5
35 – 40	3
40 – 45	2
45 –50	5
Total	**20**

- **Inclusive series** are those series in which both the upper limit and the lower limit of a class interval are included in that particular class interval.
  For example:

Marks (*C.I*)	**Frequency (*F*)**
1 – 5	2
6 – 10	3
11 – 15	5
16 – 20	3
21 – 25	2
26 – 30	5
Total	**20**

- **Open-ended series** are those series in which the lower limit of the first class interval and the upper limit of the last class interval are missing.
  For example:

Marks (*C.I*)	**Frequency (*F*)**
Below 20	2
25 – 30	3
30 – 35	5
35 – 40	3
40 – 45	2
45 and above	5
Total	**20**