

Economics
Class - XI
Organisation of Data - Notes

Organisation of data refers to the systematic arrangement of collected figures (raw data), so that the data becomes easy to understand and convenient for further statistical treatment.

Classification is the process of arranging data into sequences and groups according to their common characteristics or separating them into different but related parts.

Objectives of classification

- To simplify and condense the mass of data:** In classification, the aim is to eliminate unnecessary details and make the huge mass of complex data in simple, condensed, logical and comprehensible form. It helps in highlighting the significant features of the data.
- To explain similarity and dissimilarity of data:** Classification facilitates the grouping of data according to certain similarities (affinities) and dissimilarities (diversities).
- To facilitate comparisons:** Classification enables us to make meaningful comparisons, draw inferences and locate facts.
- To study the relationships:** Classification helps in finding out cause and effect relationship based on some criteria between the items of data.
- To prepare the data for tabulation:** Only classified data can be presented in tabular form. Classification thus provides a basis for tabulation and further statistical processing.

Methods of classification:

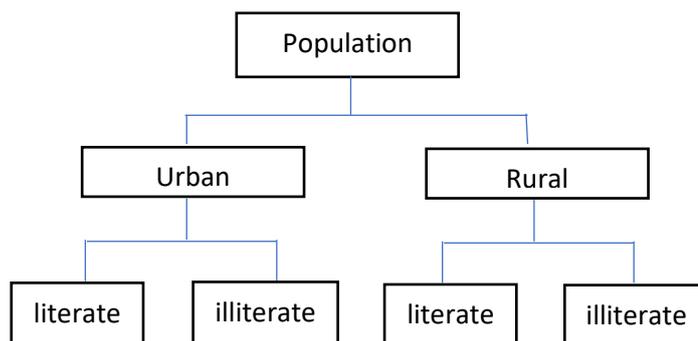
- 1. Chronological or Temporal classification:** In such a classification, the collected data is classified with respect to different periods of time (such as years, quarters, months, weeks etc). Data can be ordered either in ascending or in descending order of time. For example:

Year	Production of rice (in tonnes)
2015	345
2016	500
2017	457

- 2. Geographical or Spatial classification:** In this type of classification, the collected data is classified according to geographical location or region (such as countries, states, cities etc). For example:

State	Production of wheat (kg per hectare)
Punjab	2345
Haryana	3775
Uttar Pradesh	2500

- 3. Qualitative classification:** In this type of classification, data is classified on the basis of descriptive characteristics or on the basis of attributes like gender, literacy, employment status etc. which cannot be quantified. *An attribute refers to qualitative characteristics of an object, person or place which cannot be expressed in numerical terms.* For example:



In the above given example, the first stage of classification is based on the presence of an attribute *region* i.e urban or not urban (rural), the second stage of classification is based on the presence of a second attribute *literacy* i.e. literate or absence of it (illiterate).

4. **Quantitative classification:** In this classification, data is classified on the basis of some characteristics which can be measured and expressed numerically such as height, weight, income, expenditure, production, or sales. For example, if we classify the ages of students of a school, we can express the ages numerically, i.e. in terms of numbers. For example:

Age (in years)	Number of students
10-12	150
12-14	130
14-16	100
16-18	120
Total Students	500

Important terms

- **Raw Data:** A mass of data in its original form is called **raw data**.
- **Variable:** A variable refers to a quantity or characteristic whose value varies from one investigation to another. Variables are of two types.
 - a) **Discrete Variable:** Discrete variables are those variables which are capable of taking only exact or finite values and not any fractional value. It increases in jumps or in complete numbers. For example: Number of students in a class could be 15, 20, 25, but cannot be 15.5 or 20.2.
 - b) **Continuous Variable:** Continuous variables are those variables which can take all the possible values (integral and fractional) in a given specified range. For example: Wages of workers in a factory, height or weight of individuals etc.

Statistical series

The arrangement of classified data in some logical order, like according to the size, according to the time of occurrence or according to some other measurable or non-measurable characteristics, is known as Statistical Series.

Kinds of statistical series

A. On the Basis of Characteristics

- **Time Series:** If the different values that a variable has taken in a period of time are arranged in a chronological order, the series so obtained is called a time series. For example: Population of India (1971-2011).
- **Spatial Series:** In this series, data is arranged according to location or geographical considerations. For example: Population of BRICS countries in 2019.
- **Condition Series:** In this series, data is classified according to the changes occurring under certain conditions. For example: Age distribution of 100 students of a school.

B. On the Basis of Construction

- **Individual Series :** It refers to those series in which items are listed singly i.e. each item is given a separate value of measurement. This series consists of ungrouped data. For example :

S.No.	1	2	3	4	5	6	7	8	9	10
Marks	6	8	10	10	11	11	12	13	17	20

It can either be **organised (ordered)** in ascending or descending order or may be **unorganised (unordered)**.

• **Discrete Series (Frequency array):**

A discrete series is a series where individual values of the variable differ from each other by a definite or integral value. It is a series which represents a discrete variable which does not take intermediate fractional values. In this series, various values of the variable are shown along with their corresponding frequencies. For example :

Marks	6	8	10	11	12
No. of Students (f)	1	1	2	2	1

• **Continuous Series (Grouped frequency distribution):**

A continuous series is that series which represents continuous variables, showing range of values of different items of the series. Data is divided into different classes and expressed in **class-intervals**. In a continuous series, the class intervals are shown along with their corresponding frequencies. For example :

Marks	0-10	10-20	20-30	30-40	40-50
No. of Students (f)	5	6	8	7	4

Note: Data is grouped in both discrete and continuous series with the use of frequencies.

Important terms associated with continuous series

1. **Class:** Class means a group of numbers in which items are placed such as 0-10, 10-20, 20-30, etc. Classes should be mutually exclusive, so that any value of the variable corresponds to one and only one of the classes.
2. **Class mid-point or class mark:** It is the middle value of a class. It lies halfway between the lower limit and the upper limit of a class and can be ascertained in the following manner.

$$\text{Class mid-point} = \frac{\text{Upper limit} + \text{Lower limit}}{2}$$

3. **Class frequency:** It means the number of observations corresponding to a particular class.
4. **Class limits:** These are the lowest and highest values of a variable within a class. The lowest value is called **lower limit** and highest value is called **upper limit**.
5. **Class size / interval / width:** It is the difference between the upper limit and lower limit of a class.

$$\text{Class size / interval / width} = \text{Upper limit} - \text{Lower limit}$$

Different Forms of Continuous Series

(a) **Inclusive and Exclusive Series:**

Inclusive series		Exclusive series	
In such a series both upper limit and lower limit is included in the class.		In such a series only one of limits, usually the lower limit, is included in the class while the upper limit is excluded.	
The upper limit of one class is not equal to the lower limit of the next class.		The upper limit of one class is equal to the lower limit of the next class.	
For example:	Marks 10-19 20-29 30-39	For example:	Marks 10-20 20-30 30-40

Steps to convert Inclusive Series into an Exclusive Series

For example: In the above given inclusive series:

- i. Find the difference between the upper limit of first class and lower limit of the next class.
(20 - 19 = 1)
- ii. Add half of this difference (0.5) to the upper limit of each class-interval and subtract remaining half from the lower limit of each class-interval.
- iii. Thus, the exclusive series would be **9.5-19.5, 19.5-29.5, 29.5-39.5** etc.

(b) **Open-Ended Distribution:** (when the lower limit of the first class or the upper limit of the last class or both are not defined or given). For example:

Marks	No. of Students
Below 50	20
50 -100	10
100 -150	18
Above 150	13

(c) **Cumulative Frequency Distribution:** (It is obtained by adding up or cumulating frequencies). There are two types of cumulative frequency distributions:

- a. **Less than cumulative frequency distribution:** In this distribution, the cumulative frequencies of each class interval is obtained by adding up or cumulating the frequencies successively from top to bottom. For example:

Marks (Classes)	Frequency	Marks (Less than upper limit)	Cumulative frequency
10-20	7	Less than 20	7
20-30	9	Less than 30	7+9= 16
30-40	5	Less than 40	7+9+5= 21
40-50	6	Less than 50	7+9+5+6= 27
50-60	3	Less than 60	7+9+5+6+3= 30
	Σ f = 30		

- b. **More than cumulative frequency distribution:** In this distribution, the cumulative frequencies of each class interval is obtained by adding up or cumulating the frequencies starting from highest value of the variable to the lowest value. For example:

Marks (Classes)	Frequency	Marks (More than lower limit)	Cumulative frequency
10-20	7	More than 10	3+6+5+9+7= 30
20-30	9	More than 20	3+6+5+9= 23
30-40	5	More than 30	3+6+5= 14
40-50	6	More than 40	3+6= 9
50-60	3	More than 50	3
	Σ f = 30		

(d) Unequal Class Interval Series: (when class size or width varies in a distribution)

For example:

Marks	Frequency	Class Size
10-20	7	10
20-40	9	20
40-70	5	30

(e) Mid-value Series: (when class marks or midpoints are given along with frequencies in a series) For example:

Mid-value/ Class mark	Frequency	Classes
15	7	10-20
25	9	20-30
35	5	30-40
45	6	40-50
55	3	50-60

Steps to convert mid-value or midpoint series into a simple frequency distribution

For example: In the above given mid-value series:

- 1) Calculate the size of the class by finding the **difference between two successive midpoints** ($25-15=10=A$)
- 2) Add half of this class size ($A/2=5$) to the midpoint ($15+5=20$) to get the **upper limit** and subtract from the midpoint ($15-5=10$) to get the **lower limit** of the class.
- 3) Thus, the first class for the above given series will be : **10-20**. Similarly, find the other classes.

Loss of information

The classification of data as a frequency distribution has an inherent shortcoming. While it summarizes the raw data making it concise and comprehensible, it does not show the details that are found in raw data. Once the data is grouped into classes, an individual observation has no significance in further statistical calculations. Hence, this leads to a loss of information while classifying raw data.

Univariate and Bivariate Distributions

Univariate frequency distribution	Bivariate frequency distribution
When data is classified on the basis of single variable, the distribution is known as univariate frequency distribution.	When data is classified on the basis of two variables, the distribution is known as bivariate frequency distribution.
It is also known as one-way frequency distribution.	It is also known as two-way frequency distribution.
Example: Distribution showing height of students in a class.	Example: Distribution showing height and weight of students in a class.

Recap

- Classification brings order to raw data.
- A Frequency Distribution shows how the different values of a variable are distributed in different classes along with their corresponding class frequencies.
- Both the upper and the lower-class limits are included in the Inclusive Method.
- When data is classified on the basis of single variable, the distribution is known as univariate frequency distribution.
- When data is classified on the basis of two variables, the distribution is known as bivariate frequency distribution.

Click on the following link for further explanation of the topics discussed above:

<https://www.youtube.com/watch?v=Y6CpMQN2UoM> (Types of frequency distribution)

SUMMARY

