

## Lecture 2

### CLASSIFICATION AND TABULATION

*Nariman Yahya Othman*

#### **Classification and Tabulation**

The data collected for the purpose of a statistical inquiry some times consists of a few fairly simple figures, which can be easily understood without any special treatment. But more often there is an overwhelming mass of raw data without any structure. Thus, unwieldy, unorganised and shapeless mass of collected is not capable of being rapidly or easily associated or interpreted. Unorganised data are not fit for further analysis and interpretation. In order to make the data simple and easily understandable the first task is not condense and simplify them in such a way that irrelevant data are removed and their significant features are stand out prominently. The procedure adopted for this purpose is known as method of classification and tabulation. Classification helps proper tabulation.

“Classified and arranged facts speak themselves; unarranged, unorganised they are dead as mutton”.

- Prof. J.R. Hicks

#### ◆ **Meaning of Classification**

Classification is a process of arranging things or data in groups or classes according to their resemblances and affinities and gives expressions to the unity of attributes that may subsist among a diversity of individuals.

#### ◆ **Definition of Classification**

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

- Secrist

The process of grouping large number of individual facts and observations on the basis of similarity among the items is called classification.

- Stockton & Clark

#### **Characteristics of classification**

- a) Classification performs homogeneous grouping of data
- b) It brings out points of similarity and dissimilarities.
- c) The classification may be either real or imaginary
- d) Classification is flexible to accommodate adjustments

## **Objectives / purposes of classifications**

- i) To simplify and condense the large data
- ii) To present the facts to easily in understandable form
- iii) To allow comparisons
- iv) To help to draw valid inferences
- v) To relate the variables among the data
- vi) To help further analysis
- vii) To eliminate unwanted data
- viii) To prepare tabulation

## **Guiding principles (rules) of classifications**

Following are the general guiding principles for good classifications

- a) *Exhaustive*: Classification should be exhaustive. Each and every item in data must belong to one of class. Introduction of residual class (i.e. either, miscellaneous etc.) should be avoided.
- b) *Mutually exclusive*: Each item should be placed at only one class
- c) *Suitability*: The classification should confirm to object of inquiry.
- d) *Stability*: Only one principle must be maintained throughout the classification and analysis.
- e) *Homogeneity*: The items included in each class must be homogeneous.
- f) *Flexibility*: A good classification should be flexible enough to accommodate new situation or changed situations.

## **Modes / Types of Classification**

Modes / Types of classification refers to the class categories into which the data could be sorted out and tabulated. These categories depend on the nature of data and purpose for which data is being sought.

## **Important types of classification**

- a) Geographical (i.e. on the basis of area or region wise)
- b) Chronological (On the basis of Temporal / Historical, i.e. with respect to time)
- c) Qualitative (on the basis of character / attributes)
- d) Numerical, quantitative (on the basis of magnitude)

**a) Geographical Classification**

In geographical classification, the classification is based on the geographical regions.

**Ex:** Sales of the company (In Million Rupees) (region – wise)

<b>Region</b>	<b>Sales</b>
North	285
South	300
East	185
West	235

**b) Chronological Classification**

If the statistical data are classified according to the time of its occurrence, the type of classification is called chronological classification.

Sales reported by a departmental store

<b>Month</b>	<b>Sales (Rs.) in lakhs</b>
January	22
February	26
March	32
April	25
May	27
June	30

**c) Qualitative Classification**

In qualitative classifications, the data are classified according to the presence or absence of attributes in given units. Thus, the classification is based on some quality characteristics / attributes.

**Ex:** Sex, Literacy, Education, Class grade etc.

Further, it may be classified as

- a) Simple classification
- b) Manifold classification

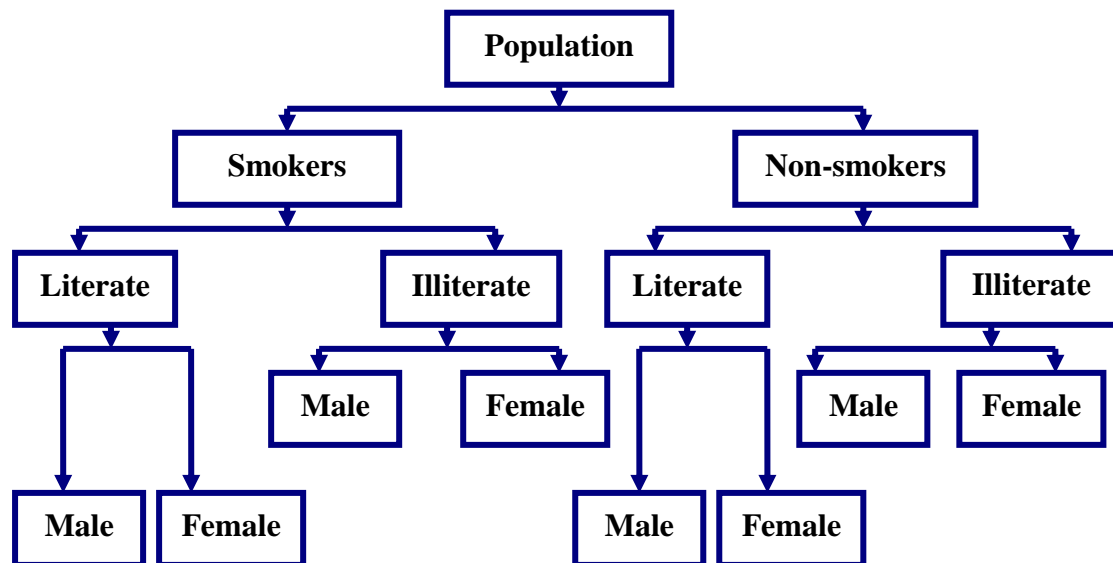
i) *Simple classification:* If the classification is done into only two classes then classification is known as simple classification.

**Ex:** a) Population in to Male / Female

b) Population into Educated / Uneducated

ii) *Manifold classification:* In this classification, the classification is based on more than one attribute at a time.

Ex:



d) **Quantitative Classification:** In Quantitative classification, the classification is based on quantitative measurements of some characteristics, such as age, marks, income, production, sales etc. The quantitative phenomenon under study is known as variable and hence this classification is also called as classification by variable.

Ex:

For a 50 marks test, Marks obtained by students as classified as follows

Marks	No. of students
0 – 10	5
10 – 20	7
20 – 30	10
30 – 40	25
40 – 50	3
<b>Total Students = 50</b>	

In this classification marks obtained by students is variable and number of students in each class represents the frequency.

## Tabulation

### Meaning and Definition of Tabulation

Tabulation may be defined, as systematic arrangement of data in column and rows. It is designed to simplify presentation of data for the purpose of analysis and statistical inferences.

## Major Objectives of Tabulation

1. To simplify the complex data
2. To facilitate comparison
3. To economise the space
4. To draw valid inference / conclusions
5. To help for further analysis

## Differences between Classification and Tabulation

1. First data are classified and presented in tables; classification is the basis for tabulation.
2. Tabulation is a mechanical function of classification because in tabulation classified data are placed in row and columns.
3. Classification is a process of statistical analysis while tabulation is a process of presenting data in suitable structure.

## Classification of tables

Classification is done based on

1. Coverage (Simple and complex table)
2. Objective / purpose (General purpose / Reference table / Special table or summary table)
3. Nature of inquiry (primary and derived table).

**Ex:**

- a) *Simple table*: Data are classified based on only one characteristic

Distribution of marks

<b>Class Marks</b>	<b>No. of students</b>
30 – 40	20
40 – 50	20
50 – 60	10
<b>Total</b>	<b>50</b>

b) *Two-way table*: Classification is based on two characteristics

Class Marks	No. of students		
	Boys	Girls	Total
30 – 40	10	10	20
40 – 50	15	5	20
50 – 60	3	7	10
<b>Total</b>	<b>28</b>	<b>22</b>	<b>50</b>

### Frequency Distribution

Frequency distribution is a table used to organize the data. The left column (called classes or groups) includes numerical intervals on a variable under study. The right column contains the list of frequencies, or number of occurrences of each class/group. Intervals are normally of equal size covering the sample observations range.

It is simply a table in which the gathered data are grouped into classes and the number of occurrences, which fall in each class, is recorded.

#### ◆ Definition

A frequency distribution is a statistical table which shows the set of all distinct values of the variable arranged in order of magnitude, either individually or in groups with their corresponding frequencies.

- Croxton and Cowden

A frequency distribution can be classified as

- a) Series of individual observation
- b) Discrete frequency distribution
- c) Continuous frequency distribution

#### a) *Series of individual observation*

Series of individual observation is a series where the items are listed one after the each observation. For statistical calculations, these observation could be arranged is either ascending or descending order. This is called as array.

*Ex:*

<b>Roll No.</b>	<b>Marks obtained in statistics paper</b>
1	83
2	80
3	75
4	92
5	65

The above data list is a raw data. The presentation of data in above form doesn't reveal any information. If the data is arranged in ascending / descending in the order of their magnitude, which gives better presentation then, it is called arraying of data.

### **Discrete (ungrouped) Frequency Distribution**

If the data series are presented in such away that indicating its exact measurement of units, then it is called as discrete frequency distribution. Discrete variable is one where the variants differ from each other by definite amounts.

*Ex:*

Assume that a survey has been made to know number of post-graduates in 10 families at random; the resulted raw data could be as follows.

0, 1, 3, 1, 0, 2, 2, 2, 2, 4

This data can be classified into an ungrouped frequency distribution. The number of post-graduates becomes variable (x) for which we can list the frequency of occurrence (f) in a tabular form as follows;

<b>Number of post graduates (x)</b>	<b>Frequency (f)</b>
0	2
1	2
2	4
3	1
4	1

The above example shows a discrete frequency distribution, where the variable has discrete numerical values.

## Continuous frequency distribution (grouped frequency distribution)

Continuous data series is one where the measurements are only approximations and are expressed in class intervals within certain limits. In continuous frequency distribution the class interval theoretically continuous from the starting of the frequency distribution till the end without break. According to Boddington 'the variable which can take very intermediate value between the smallest and largest value in the distribution is a continuous frequency distribution.

*Ex:*

Marks obtained by 20 students in students' exam for 50 marks are as given below – convert the data into continuous frequency distribution form.

18	23	28	29	44	28	48	33	32	43
24	29	32	39	49	42	27	33	28	29

By grouping the marks into class interval of 10 following frequency distribution tables can be formed.

Marks	No. of students
0 - 5	0
5 - 10	0
10 - 15	0
15 - 20	1
20 - 25	2
25 - 30	7
30 - 35	4
35 - 40	1
40 - 45	3
45 - 50	2