

Sengamala Thayaar Educational Trust Women's College

(Affiliated to Bharathidasan University, Tiruchirapalli)

(Accredited with 'A' Grade {3.45/4.00} By NAAC)

(An ISO 9001: 2015 Certified Institution)

SUNDARAKKOTTAI, MANNARGUDI-614 016.

Thiruvarur (Dt.), Tamil Nadu, India.



BIOSTATISTICS

Dr. S.MANIKANDASELVI

ASSISTANT PROFESSOR

PG & RESEARCH DEPARTMENT OF BIOCHEMISTRY

I M.Sc., BIOCHEMISTRY

II SEMESTER

ELECTIVE COURSE I- BIostatISTICS-P16BCE1

CLASSIFICATION AND TABULATION OF DATA

- **Data** are characteristics or information, usually numerical, that are collected through observation.
- In a more technical sense, **data** is a set of values of qualitative or quantitative variables about one or more persons or objects, while a datum (singular of **data**) is a single value of a single variable.
- Data can be collected for different purposes.

The following are the three types of data:

1. Time series data
2. Spatial data
3. Spacio-temporal data

Time series data

It is a collection of a set of numerical values, collected over a period of time. The data might have been collected either at regular intervals of time or irregular intervals of time.

Spatial Data

If the data collected is connected with that of a place, then it is termed as spatial data.

Spacio Temporal Data

If the data collected is connected to the time as well as place then it is known as spacio temporal data.

Categories of data

Any statistical data can be classified under two categories depending upon the sources utilized. These categories are,

1. Primary data
2. Secondary data

Primary data

Primary data is the one, which is collected by the investigator himself for the purpose of a specific inquiry or study. Such data is original in character and is generated by survey conducted by individuals or research institution or any organisation.

The primary data can be collected by the following five methods.

1. Direct personal interviews.
2. Indirect Oral interviews.
3. Information from correspondents.
4. Mailed questionnaire method.
5. Schedules sent through enumerators.

Secondary source

A secondary source is a publication, reporting the data which have been gathered by other authorities and for which others are responsible

The sources of secondary data can broadly be classified as,

1. Published sources
2. Unpublished sources.

CLASSIFICATION OF DATA

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

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

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

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.

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

When data are classified on the basis of location or areas, it is called geographical classification.

Example: Classification of production of food grains in different states in India.

States	Production of food grains (in '000 tons)
Tamil Nadu	54500
Karnataka	44200
Andhra Pradesh	33600

b) Chronological Classification

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

Example: Sales reported by a departmental store

Month	Sales (Rs.) in lakhs
January	22
February	26
March	32
April	25

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.
- Example: 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.

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
Total Students	50

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

TABULATION

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

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

Classification of tables

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).

Parts of a Table,

(1) Table number: - each table should be numbered.

- There are different practices with regard to the place where this number is to be given.
- The number may be given either in the centre at the top above the title or inside of the title at the top or in the bottom of the table on the left hand side.
- However, if space permits the table number should be given in the centre as is shown in the specimen table given on page.
- When there so that easy reference to it is possible.

(2) Title of the table: - every table must be given suitable title.

- The title is a description of the contents of the table.
- A complete title has to answer the question what, where and when in that sequence. In other words:
 - (a) What precisely are the data in the table (i.e.) what categories of statistical data are shown?
 - (b) Where the data occurred i.e. the precise geographical, political or physical area covered?
 - (c) When the data occurred i.e. the specific time or period covered by the statistical materials in the table?

(3) Caption: - caption refers to the column headings.

- It explains what the column represents it may consist of one or more column headings.
- Under a column heading there may be sub-heads.
- The caption should be clearly defined and placed at the middle of the column, if the different columns are expressed in different units.
- The units should be mentioned with the captions.
- As compared with the main part of the table the caption should be shown in smaller letters.
- This helps in saving space.

(4) Stub: - as distinguished from caption, stubs are the designations of the rows or row heading.

- They are at the extreme left and perform the same function for the horizontal rows of numbers in the table as the column headings do for the vertical columns of numbers.
- The stubs are usually wide than column headings but should be kept as narrow as possible without sacrificing precision and clarity of statements.

(5) Body: - the body of the table contains the numerical information.

This is the most vital part of the table. Data presented in the body arranged according to description are classifications of the captions and stubs.

(6) Head note: - it is a brief explanatory statement applying to all or a major part of the material in the table, and is placed below the point centered and enclosed in brackets.

- It is used to explain certain points relating to the whole table that have not been included in the title nor in the captions or stubs.
- For example, the unit of measurement is frequently written as a head note, such as "in thousands" or "in million tonnes" or "in crores"

(7) Footnotes: - anything in a table which the reader may find difficult to understand from the title, captions and stubs should be explained in footnotes.

- If footnotes are needed they are placed directly below the body of the table.
- Footnotes are used for the following main purposes:
 - (a) To point out any exceptions as to the basis of arriving at the data
 - (b) Any special circumstances affecting the data, for example, strike, lock-out fire
 - (c) To clarify anything in the table
 - (d) To give the source in case of secondary data.

Format of a table

- Title
- Head Note
- Stub Heading
- Caption Heading-Column heading
- Stub Entries
- Body
- Footnotes
- Table Number

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.