

# Chapter 3

## 3. COLLECTION OF DATA, CLASSIFICATION AND TABULATION

# Introduction:

Everybody collects, interprets and uses information, much of it in a numerical or statistical forms in day-to-day life. It is a common practice that people receive large quantities of information everyday through conversations, televisions, computers, the radios, newspapers, posters, notices and instructions

# some of the objectives of collecting statistical information.

1. To describe the methods of collecting primary statistical information.
2. To consider the status involved in carrying out a survey.
3. To analyse the process involved in observation and interpreting.
4. To define and describe sampling.
5. To analyse the basis of sampling.
6. To describe a variety of sampling methods.

# Statistical investigation

is a comprehensive and requires systematic collection of data about some group of people or objects, describing and organizing the data, analyzing the data with the help of different statistical method, summarizing the analysis and using these results for making judgements, decisions and predictions.,

The validity and accuracy of final judgement is most crucial and depends heavily on **how well the data was collected in the first place.**

## 3.2 Nature of data:

It may be noted that different types of data can be collected for different purposes. The data can be collected in connection with time or geographical location or in connection with time and location. The following are the three types of data:

- 1. Time series data.**
- 2. Spatial data**
- 3. Specie-temporal data.**

## 3.2.1 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.

### Example 1:

The following is the data for the three types of expenditures in RS for a family for the four years 2001,2002,2003,2004.

Year	Food	Education	Others	Total
2001	3000	2000	3000	8000
2002	3500	3000	4000	10500
2003	4000	3500	5000	12500
2004	5000	5000	6000	16000

## 3.2.2 Spatial Data:

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

- 1- District wise rainfall in Tamilnadu .
- 2- Prices of silver in four metropolitan cities

### Example 2:

The population of the southern states of India in 1991.

State	Population
Tamilnadu	5,56,38,318
Andhra Pradesh	6,63,04,854
Karnataka	4,48,17,398
Kerala	2,90,11,237
Pondicherry	7,89,416

## 3.2.3 Spacio Temporal Data:

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

### Example 3:

State	Population	
	1981	1991
Tamil Nadu	4,82,97,456	5,56,38,318
Andhra Pradesh	5,34,03,619	6,63,04,854
Karnataka	3,70,43,451	4,48,17,398
Kerala	2,54,03,217	2,90,11,237
Pondicherry	6,04,136	7,89,416



## 3.3 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

## 3.3.1 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.

### Example 4:

If a researcher is interested to know the impact of noon meal scheme for the school children, he has to undertake a survey and collect data on the opinion of parents and children by asking relevant questions. Such a data collected for the purpose is called primary data.

## **3.3.1 Primary data:**

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.

# 1. Direct personal interviews:

The persons from whom information's are collected are known as informants. The investigator personally meets them and asks questions to gather the necessary information's. It is the suitable method for intensive rather than extensive field surveys.

# 1. Direct personal interviews:

## Merits:

1. People willingly supply informations because they are approached personally. Hence, more response noticed in this method than in any other method.
2. The collected informations are likely to be uniform and accurate. The investigator is there to clear the doubts of the informants.
3. Supplementary informations on informant's personal aspects can be noted. Informations on character and environment may help later to interpret some of the results.
4. Answers for questions about which the informant is likely to be sensitive can be gathered by this method.
5. The wordings in one or more questions can be altered to suit any informant. Explanations may be given in other languages also. Inconvenience and misinterpretations are thereby avoided.

# 1. Direct personal interviews:

## Limitations:

1. It is very costly and time consuming.
2. It is very difficult, when the number of persons to be interviewed is large and the persons are spread over a wide area.
3. Personal prejudice and bias are greater under this method

## 2. Indirect Oral Interviews:

Under this method the investigator contacts witnesses or neighbours or friends or some other third parties who are capable of supplying the necessary information.

This method is suitable whenever direct sources do not exist or cannot be relied upon or would be unwilling to part with the information.

### **3. Information from correspondents:**

The investigator appoints local agents or correspondents in different places and compiles the information sent by them.

Informations to Newspapers and some departments of Government come by this method.



## 4. Mailed questionnaire method:

Under this method a list of questions is prepared and is sent to all the informants by post.

- The list of questions is technically called **questionnaire**.

### **Merits:**

1. It is relatively cheap.
2. It is preferable when the informants are spread over the wide area.

# 4. Mailed questionnaire method:

## Limitations:

1. The greatest limitation is that the informants should be literates who are able to understand and reply the questions.
2. It is possible that some of the persons who receive the questionnaires do not return them.
3. It is difficult to verify the correctness of the informations furnished by the respondents.

[Characteristics Good Questioner ... click](#)

# 5. Schedules sent through Enumerators:

Under this method enumerators or interviewers take the schedules, meet the informants and filling their replies.

- **Merits:**

1. It can be adopted even if the informants are illiterates.
2. Answers for questions of personal and pecuniary nature can be collected.
3. Non-response is minimum as enumerators go personally and contact the informants.
4. The informations collected are reliable. The enumerators can be properly trained for the same.
5. It is most popular methods.

# 5. Schedules sent through Enumerators:

- **Limitations:**

1. It is the costliest method.
2. Extensive training is to be given to the enumerators for collecting correct and uniform informations.
3. Interviewing requires experience. Unskilled investigators are likely to fail in their work

## 3.3.2 Secondary Data:

- Secondary data are those data which have been already collected and analysed by some earlier agency for its own use; and later the same data are used by a different agency
- **Sources of Secondary data:**

In most of the studies the investigator finds it impracticable to collect first-hand information on all related issues and as such he makes use of the data collected by others..

- **The sources of secondary data can broadly be classified under two heads:**
  1. Published sources, and
  2. Unpublished sources.

# 1. Published sources

The various sources of published data are:

1. Reports and official publications of
  - (i) International bodies such as the International Monetary Fund, International Finance Corporation and United Nations Organisation.
  - (ii) Central and State Governments such as the Report of the Tandon Committee and Pay Commission.
2. Semi-official publication of various local bodies such as Municipal Corporations and District Boards.
3. Private publications-such as the publications of –
  - (i) Trade and professional bodies such as the Federation of Indian Chambers of Commerce and Institute of Chartered Accountants.
  - (ii) Financial and economic journals such as ‘ Commerce’ , ‘ Capital’ and ‘ Indian Finance’ .
  - (iii) Annual reports of joint stock companies.
  - (iv) Publications brought out by research agencies, research scholars, etc.

# Assignment 3-1

Visit :

[http://www.secondarydata.com/secdata/index.  
asp](http://www.secondarydata.com/secdata/index.asp)

- collect some published secondary data

## 2. Unpublished Sources

All statistical material is not always published.

There are various sources of unpublished data such as records maintained by various Government and private offices, studies made by research institutions, scholars, etc.



# Precautions in the use of Secondary data

- The following are some of the points that are to be considered in the use of secondary data
  1. How the data has been collected and processed
  2. The accuracy of the data
  3. How far the data has been summarized
  4. How comparable the data is with other tabulations
  5. How to interpret the data, especially when figures collected for one purpose is used for another

## 3.4 Classification:

The process of grouping into different classes or sub classes according to some characteristics is known as classification, tabulation is concerned with the systematic arrangement and presentation of classified data.

Thus classification is the first step in tabulation.

# 3.4 Classification:

## Objects of Classification:

The following are main objectives of classifying the data:

1. It condenses the mass of data in an easily assimilable form.
2. It eliminates unnecessary details.
3. It facilitates comparison and highlights the significant aspect of data.
4. It enables one to get a mental picture of the information and helps in drawing inferences.
5. It helps in the statistical treatment of the information collected.

## 3.4 Classification:

- **Types of classification:**

Statistical data are classified in respect of their characteristics. Broadly there are four basic types of classification namely

- a) Chronological classification
- b) Geographical classification
- c) Qualitative classification
- d) Quantitative classification

# a) Chronological classification:

In chronological classification the collected data are arranged according to the order of time expressed in years, months, weeks, etc., The data is generally classified in ascending order of time. For example, the data related with population, sales of a firm, imports and exports of a country are always subjected to chronological classification.

## Example 5:

The estimates of birth rates in India during 1970 – 76 are

<b>Year</b>	1970	1971	1972	1973	1974	1975	1976
<b>Birth Rate</b>	36.8	36.9	36.6	34.6	34.5	35.2	34.2

## b) Geographical classification:

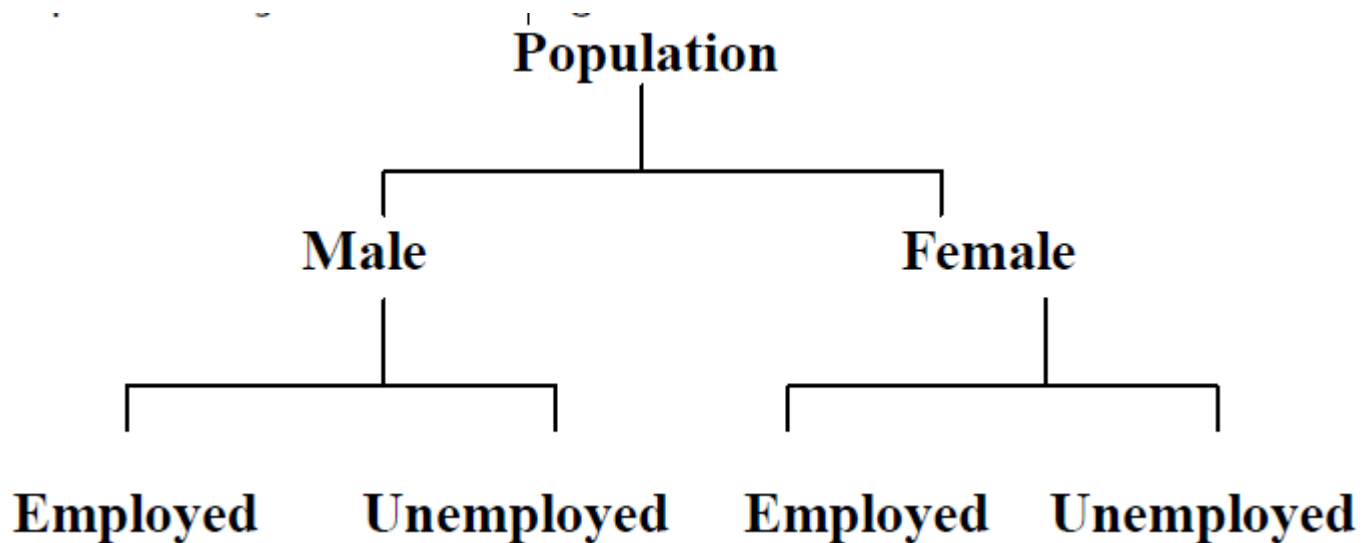
In this type of classification the data are classified according to geographical region or place. For instance, production of wheat in different countries etc.,

### Example 6:

Country	America	China	Denmark	France	India
Yield of wheat in (kg/acre)	1925	893	225	439	862

## c) Qualitative classification:

In this type of classification data are classified on the basis of same attributes or quality like sex, literacy, religion, employment etc.,



## d) Quantitative classification:

Quantitative classification refers to the classification of data according to some characteristics that can be measured such as height, weight, etc.,

For example the students of a college may be classified according to weight as given below

<b>Weight (in lbs)</b>	<b>No of Students</b>
90-100	50
100-110	200
110-120	260
120-130	360
130-140	90
140-150	40
Total	1000



## 3.5 Tabulation:

Tabulation is the process of summarizing classified or grouped data in the form of a table so that it is easily understood and an investigator is quickly able to locate the desired information.

A table is a systematic arrangement of classified data in columns and rows.

# 3.5 Tabulation:

## Advantages of Tabulation:

Statistical data arranged in a tabular form serve following objectives:

1. It simplifies complex data and the data presented are easily understood.
2. It facilitates comparison of related facts.
3. It facilitates computation of various statistical measures like averages, dispersion, correlation etc.
4. It presents facts in minimum possible space and unnecessary repetitions and explanations are avoided.

Moreover, the needed information can be easily located.

5. Tabulated data are good for references and they make it easier to present the information in the form of graphs and diagrams

# 3.5 Tabulation:

- **Type of Tables:**
- Tables can be classified according to their purpose, stage of enquiry, nature of data or number of characteristics used. On the basis of the number of characteristics, tables may be classified as follows:
  1. Simple or one-way table
  2. Two way table
  3. Manifold table

# 3.5 Tabulation:

## Type of Tables:

- **Simple or one-way Table:**

A simple or one-way table is the simplest table which contains data of one characteristic only.

**The number of adults in different occupations in a locality**

<b>Occupations</b>	<b>No. Of Adults</b>
Total	

# 3.5 Tabulation:

## Type of Tables:

- **Two-way Table:**

A table, which contains data on two characteristics, is called a two way table. In such case, therefore, either stub or caption is divided into two co-ordinate parts..

The umber of adults in a locality in respect of occupation and sex

Occupation	No. of Adults		Total
	Male	Female	
Total			

# 3.5 Tabulation:

## Type of Tables:

- **Manifold Table:**

Thus, more and more complex tables can be formed by including other characteristics.

Occupation	No. of Adults						Total
	Male			Female			
	<b>M</b>	<b>U</b>	<b>Total</b>	<b>M</b>	<b>U</b>	<b>Total</b>	
<b>Total</b>							

# Assignment

- Collect a primary data about the mode of transport of your school students. Classify the data and tabulate it.

*Exercise*

[Click here](#)