

3. COLLECTION OF DATA, CLASSIFICATION AND TABULATION

3.1 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. It is just because there is so much information available that people need to be able to absorb, select and reject it. In everyday life, in business and industry, certain statistical information is necessary and it is independent to know where to find it how to collect it. As consequences, everybody has to compare prices and quality before making any decision about what goods to buy. As employees of any firm, people want to compare their salaries and working conditions, promotion opportunities and so on. In time the firms on their part want to control costs and expand their profits.

One of the main functions of statistics is to provide information which will help on making decisions. Statistics provides the type of information by providing a description of the present, a profile of the past and an estimate of the future. The following are 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. The quality of data will greatly affect the conditions and hence at most importance must be given to this process and every possible precautions should be taken to ensure accuracy while collecting the data.

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. Spacio-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 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

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.

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 informations are collected are known as informants. The investigator personally meets them and asks questions to gather the necessary informations. It is the suitable method for intensive rather than extensive field surveys. It suits best for intensive study of the limited field.

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 preferred if the required information is on addiction or cause of fire or theft or murder etc., If a fire has broken out a certain place, the persons living in neighbourhood and witnesses are likely to give information on the cause of fire. In some cases, police interrogated third parties who are supposed to have knowledge of a theft or a murder and get some clues. Enquiry committees appointed by governments generally adopt this method and get people's views and all possible details of facts relating to the enquiry. This method is suitable whenever direct sources do not exist or cannot be relied upon or would be unwilling to part with the information.

The validity of the results depends upon a few factors, such as the nature of the person whose evidence is being recorded, the ability of the interviewer to draw out information from the third

parties by means of appropriate questions and cross examinations, and the number of persons interviewed. For the success of this method one person or one group alone should not be relied upon.

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. The advantage of this method is that it is cheap and appropriate for extensive investigations. But it may not ensure accurate results because the correspondents are likely to be negligent, prejudiced and biased. This method is adopted in those cases where informations are to be collected periodically from a wide area for a long time.

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. A covering letter accompanying the questionnaire explains the purpose of the investigation and the importance of correct informations and request the informants to fill in the blank spaces provided and to return the form within a specified time. This method is appropriate in those cases where the informants are literates and are spread over a wide area.

5. Schedules sent through Enumerators:

Under this method enumerators or interviewers take the schedules, meet the informants and filling their replies. Often distinction is made between the schedule and a questionnaire. A schedule is filled by the interviewers in a face-to-face situation with the informant. A questionnaire is filled by the informant which he receives and returns by post. It is suitable for extensive surveys.

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. According to W.A.Neiswanger, ‘ A primary source is a publication in which the data are published by the same authority which gathered and analysed them. A secondary source is a publication, reporting the data which have been gathered by other authorities and for which others are responsible’ .

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. There is a vast amount of published information from which statistical studies may be made and fresh statistics are constantly in a state of production. 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: Clinical and other personal records, death certificates, published mortality statistics, census publications, etc. Examples include:

1. Official publications of Central Statistical Authority
2. Publication of Ministry of Health and Other Ministries
3. News Papers and Journals.
4. International Publications like Publications by WHO, World Bank, UNICEF
5. Records of hospitals or any Health Institutions.

Note: A lot of secondary data is available in the internet. We can access it at any time for the further studies.

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. Such sources can also be used where necessary

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

Generally speaking, with secondary data, people have to compromise between what they want and what they are able to find.

3.4 Classification:

The collected data, also known as raw data or ungrouped data are always in an unorganised form and need to be organised and presented in meaningful and readily comprehensible form in order to facilitate further statistical analysis. It is, therefore, essential for an investigator to condense a mass of data into more and more comprehensible and assimilable form. 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.

For Example, letters in the post office are classified according to their destinations viz., Delhi, Madurai, Bangalore, Mumbai etc.,

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.

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.

Example 5:

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

Year	1970	1971	1972	1973	1974	1975	1976
Birth Rate	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, the production of paddy in different states in Iraq, production of wheat in different countries etc.,

Example 6:

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

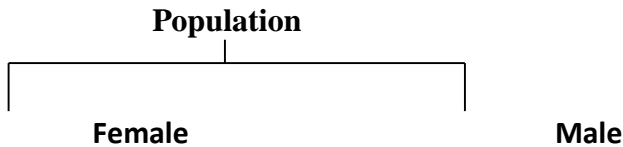
c) Qualitative classification:

In this type of classification data are classified on the basis of same attributes or quality like sex, literacy, religion, employment etc., Such attributes cannot be measured along with a scale.

For example, if the population to be classified in respect to one attribute, say sex, then we can classify them into two namely that of males and females. Similarly, they can also be classified into ‘married or ‘single’ on the basis of another attribute ‘marital status’.

Thus when the classification is done with respect to one attribute, which is dichotomous in nature, two classes are formed, one possessing the attribute and the other not possessing the attribute. This type of classification is called simple or dichotomous classification.

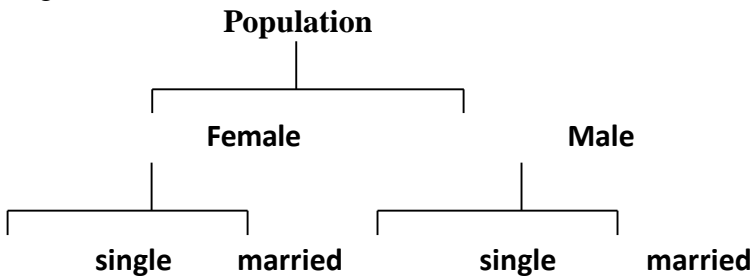
A simple classification may be shown as under



The classification, where two or more attributes are considered and several classes are formed, is called a manifold classification. For example, if we classify population simultaneously with respect to two attributes, e.g sex and marital status, then population are first classified with respect to ‘sex’ into ‘males’ and ‘females’. Each of these classes may then be further classified into ‘married’ and single on the basis of attribute ‘employment’ and as such Population are classified into four classes namely.

- (i) Male married
- (ii) Male single
- (iii) Female married
- (iv) Female single

Still the classification may be further extended by considering other attributes like marital status etc. This can be explained by the following chart



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 group of a children may be classified according to weight as given below.

Weight (in kg)	No of children
5-10	50
10-15	200
15-20	260
20-25	360
25-30	90
30-35	40
Total	1000

In this type of classification there are two elements, namely (i) the variable (i.e) the weight in the above example, and (ii) the frequency in the number of children. There are 50 children having weights ranging from 5 to 10 **kg**, 200 children having weight ranging between 10 to 15 **kg** and so on.

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. Thus, a statistical table makes it possible for the investigator to present a huge mass of data in a detailed and orderly form. It facilitates comparison and often reveals certain patterns in data which are otherwise not obvious. Classification and 'Tabulation', as a matter of fact, are not two distinct processes. Actually they go together. Before tabulation data are classified and then displayed under different columns and rows of table

Table 1: Overall immunization status of children in Adami Tullu

Woreda, Feb. 1995

Immunization status	Number	Percent
Not immunized	75	35.7
Partially immunized	57	27.1
Fully immunized	78	37.2
Total	210	100.0