
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.

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.

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 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.

Merits:

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

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.

With the view of minimizing non-respondents and collecting correct information, the questionnaire should be carefully drafted. There is no hard and fast rule. But the following general principles may be helpful in framing the questionnaire. A covering letter and a self addressed and stamped envelope should accompany the questionnaire. The covering letter should politely point out the purpose of the survey and privilege of the respondent who is one among the few associated with the investigation. It should assure that the informations would be kept confidential and would never be misused. It may promise a copy of the findings or free gifts or concessions etc..

Characteristics of a good questionnaire:

1. Number of questions should be minimum.
2. Questions should be in logical orders, moving from easy to more difficult questions.

3. Questions should be short and simple. Technical terms and vague expressions capable of different interpretations should be avoided.
4. Questions fetching YES or NO answers are preferable. There may be some multiple choice questions requiring lengthy answers are to be avoided.
5. Personal questions and questions which require memory power and calculations should also be avoided.
6. Question should enable cross check. Deliberate or unconscious mistakes can be detected to an extent.
7. Questions should be carefully framed so as to cover the entire scope of the survey.
8. The wording of the questions should be proper without hurting the feelings or arousing resentment.
9. As far as possible confidential informations should not be sought.
10. Physical appearance should be attractive, sufficient space should be provided for answering each questions.

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.

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.

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.

Before the actual survey, a pilot survey is conducted. The questionnaire/Schedule is pre-tested in a pilot survey. A few among the people from whom actual information is needed are asked to reply. If they misunderstand a question or find it difficult to answer or do not like its wordings etc., it is to be altered. Further it is to be ensured that every questions fetches the desired answer.

Merits and Demerits of primary data:

1. The collection of data by the method of personal survey is possible only if the area covered by the investigator is small. Collection of data by sending the enumerator is bound to be expensive. Care should be taken twice that the enumerator record correct information provided by the informants.
2. Collection of primary data by framing a schedules or distributing and collecting questionnaires by post is less expensive and can be completed in shorter time.
3. Suppose the questions are embarrassing or of complicated nature or the questions probe into personnel affairs of individuals, then the schedules may not be filled with accurate and correct information and hence this method is unsuitable.
4. The information collected for primary data is mere reliable than those collected from the secondary data.

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:

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.

It should be noted that the publications mentioned above vary with regard to the periodically of publication. Some are published at regular intervals (yearly, monthly, weekly etc.,) whereas others are ad hoc publications, i.e., with no regularity about periodicity of publications.

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.

Merits and Demerits of Secondary Data:

1. Secondary data is cheap to obtain. Many government publications are relatively cheap and libraries stock quantities of secondary data produced by the government, by companies and other organisations.
2. Large quantities of secondary data can be got through internet.
3. Much of the secondary data available has been collected for many years and therefore it can be used to plot trends.
4. Secondary data is of value to:
 - The government – help in making decisions and planning future policy.
 - Business and industry – in areas such as marketing, and sales in order to appreciate the general economic and social conditions and to provide information on competitors.
 - Research organisations – by providing social, economical and industrial information.

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. 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

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 India, 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

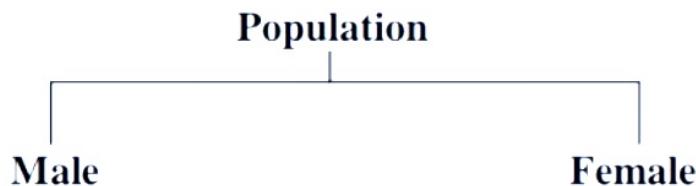
c) Qualitative classification:

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

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.

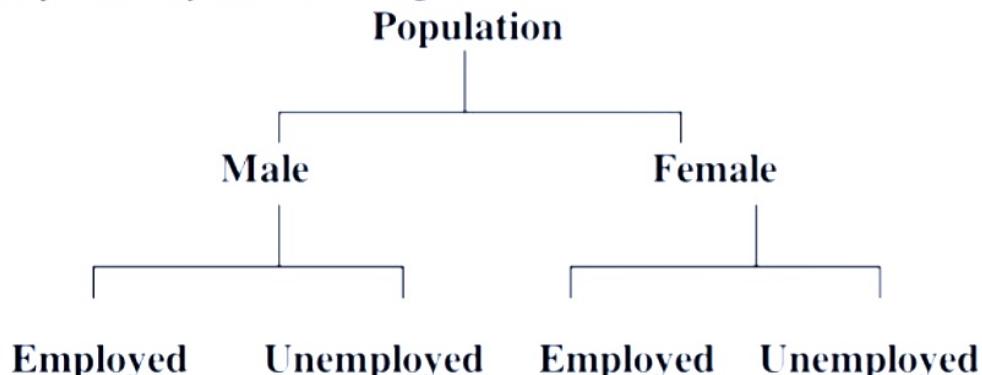
A simple classification may be shown as under



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 employment, then population are first classified with respect to 'sex' into 'males' and 'females'. Each of these classes may then be further classified into 'employment' and 'unemployment' on the basis of attribute 'employment' and as such Population are classified into four classes namely.

- (i) Male employed
- (ii) Male unemployed
- (iii) Female employed
- (iv) Female unemployed

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 students of a college may be classified according to weight as given below.