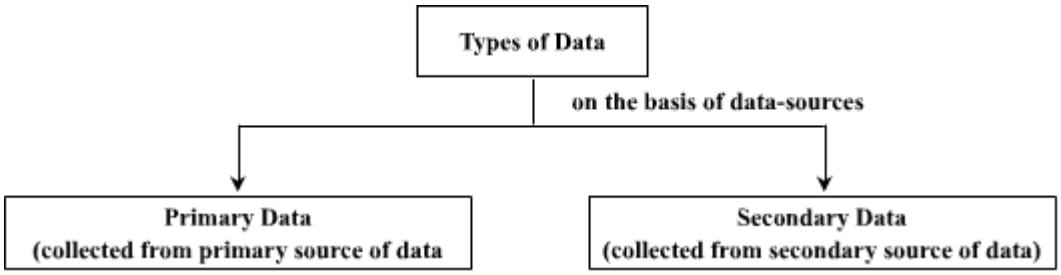
CHAPTER 2

COLLECTION OF DATA

- **Types of Data-** ***On the basis of Data-Sources***

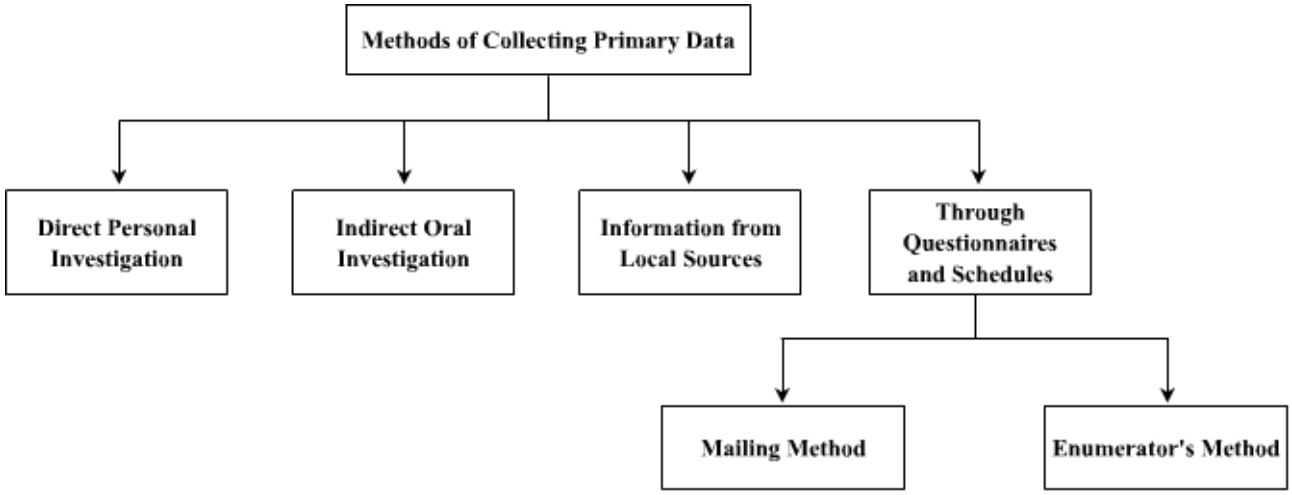


- **Primary Data-** The data that is collected by the investigator for the first time is known as **primary data**. It is collected from the original source (or the primary source) of information.
- **Secondary Data**: The collection of data that has already been collected by somebody in the past is known as secondary data. These data are collected and modified (as per the need of the present research) from the published or unpublished sources. For example, Handbook of Indian Statistics at RBI's website, etc.

- **Difference between Primary and Secondary data**

Basis	Primary Data	Secondary Data
Originality	Primary data are original in nature as the investigator himself collects data from the origin.	Secondary data are not original as the data has already been collected by someone in the past.
Cost	Primary data are comparatively costlier in terms of money and efforts than secondary data.	Secondary data are collected from the published reports or articles, so these are comparatively less expensive.
Objectives	Primary data are in accordance with the objective of the researcher.	Secondary data being collected by some other people has to be adjusted to suit the objective of the researcher.

Methods of Collecting Primary Data

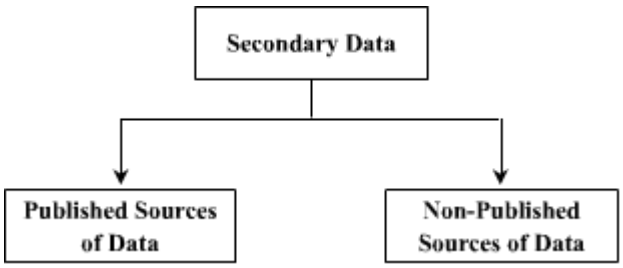


- **Direct Personal Investigation** is the method under which the investigator directly collects data from the informants.
- **Indirect Oral Investigation** is the method by which information is obtained orally from those persons who are expected to possess the necessary information.
- **Information from Local Sources** is the method in which the investigator appoints local correspondents at different places to collect information.
- **Information through Questionnaires and Schedules-** There are two methods of collecting information on the basis of questionnaires and schedules
    - ***Mailing Method***- In this method, the questionnaires are mailed to the informants, and informants fill up the questions secretly.
    - ***Enumerator's Method***- In this method, the enumerator himself approaches the informants with the questionnaire, and based on their response, fill up the questionnaire.

Characteristics of Good Questionnaire

- It should have limited number of questions.
- It should be asked in simple and easy (preferably local) language.
- The questions should have proper order.
- It should be free from undesirable questions.
- It should be pre-tested.

Collection of Secondary Data



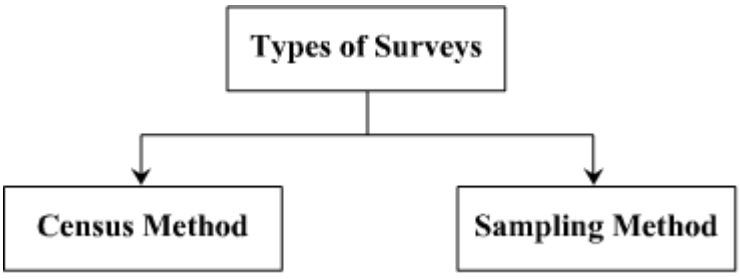
- ***Published Sources of Data***
  - ♦ Government publication
  - ♦ Semi-government publications
  - ♦ Report of committees and commissions
  - ♦ Publications of trade associations
  - ♦ Publication of research Institutions
- ***Non-Published Sources of Data***
  Data collected by government for their office use, which is not meant to be published.

- **Two Important Sources of Secondary Data**
  - ***Census of India-*** It is published every 10 years by the Registers General and Census Commission, India. It covers all the aspects of population parameters such as demographic distribution, birth rate, death rate, etc. in India.
  - ***Reports of National Sample Survey Organisation (NSSO)*** - NSSO is a government organisation operating under the supervision of Ministry of Statistics and Programme Implementation. The organisation conducts regular sample survey concentrating on workforce, employment, etc.

- **Universe or Population-** It refers to the aggregate of all items to be studied for an investigation. For example, if IQ level is to be checked for any 10 students of Class XI (consisting of 60 students), then the total number of students in the class is referred to as the population of the study.

- **Sample-** It is a subset of the population or universe. A small population of items selected from the universe for statistical study is referred to as a sample. In the above example, the selected 10 students are the sample that is taken out of the universe of 60 students in the class.

- **Types of Surveys**



- **Census Method-** It is defined as that method in which the data is collected from each and every unit of the universe related to the investigation.

- **Sampling Method-** It refers to the process of selection of a subset of items from a population. Only a part of universe is studied and conclusions are drawn on the basis of that study.

- **Techniques/Methods of Sampling**

- **Random Sampling-** A method of sampling in which each and every item of the universe has equal chance of being selected in the sample is called as random sampling.

- **Non-random Sampling**- A method of sampling in which each and every item of the universe does not have equal chance of being selected in the sample is called as non-random sampling.

- **Purposive Sampling-** A method of sampling in which the investigator himself chooses the sample items according to his opinion in such a way that the selected sample items represent the whole universe is called as purposive sampling.

- **Stratified or Mixed Sampling-** A method of sampling that involves selection of items from each stratum of different characteristics in such a way that the entire population is represented by the selected items of different stratum is called stratified or mixed sampling.

- **Quota Sampling-** A method of selecting sample items from the population that is divided into different groups according to different characteristics of the population is called quota sampling.

- **Convenience Sampling**- A method of sampling in which the sample is selected as per the convenience of the investigator is called convenience sampling. For example, to collect data on income earning capacity of the poor people, the investigator may conduct his research at a place that is near to his residence. Such type of sampling is called convenience sampling.

- **Sampling Errors-** These are referred to those errors that arise due to the size or nature of the sample selected. Such errors emerge due to the divergence between the estimated and the actual value of a parameter of a small sized sample population.
  For example, errors due to personal bias, misinterpretation of results, miscalculations, etc. lead to sampling errors.

- **Non-sampling Errors-** These are referred to those errors that arise due to the faulty means of collection of data**.** It includes:
  - errors of measurement
  - errors of non response
  - errors of misinterpretation
  - errors of calculations
  - errors of sampling bias

- **Non-sampling Errors** are more serious than the sampling errors. The sampling errors can be minimised by increasing the size of a small sample, so that the difference between the actual and the estimated value is reduced. But the non-sampling errors are difficult to rectify as it would require selection of a new sample and conducting a fresh survey. Thus, non-sampling errors are more serious than the sampling errors.

- **Advantages of Sampling Method over Census Method**
  - Provides reliable and accurate results
  - Consumes less time and energy
  - Cost efficient
  - Probability of non-sampling errors is low as it involves smaller sample size.
  - More efficient as it involves small teams of enumerators that can be easily trained and well supervised.