

# Collection of Data

---

## Meaning and Sources of Data

### Objective

After going through this lesson, you shall be able to understand the following concepts.

- Meaning of Data and
- Sources of Collecting Data

### Introduction

For conducting any research or study what is the first thing that is required? It is a set of facts or numerical figures. Any study cannot be initiated without the availability of a suitable data. For example, a study on the performance of the students in a school requires data related to the marks of the students. Similarly, a study on unemployment in a country requires numerical facts relating to unemployment.

Thus, it can be said that a suitable and appropriate data forms the base of any study and research. The required data must be collected from one source or the other. Thus, we can say that collection of data is the first step to commence any research or study.

### Meaning of Data

Data refers to the numerical facts and figures that provide information relating to a particular problem or subject matter.

In order to make an analysis of any theoretical information, we first need to convert that information into some quantitative numbers. Such quantitative information if presented in an organised way is known as 'Data'. Data basically refers to the abstract form of the detailed information presented in a summarised way.

Data can be categorised into the following two broad categories.

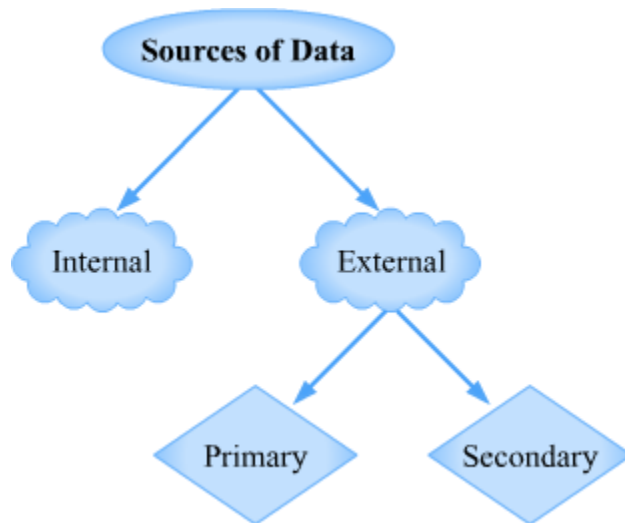
- Qualitative Data:** It comprises of those variables that cannot be measured in numerical units. For example, intelligence, appearance, etc.
- Quantitative Data:** It refers to those variables that can be expressed in numerical terms. For example, income, marks, etc.

The following points highlight the difference between qualitative and quantitative data.

| Basis of Difference | Quantitative Data  | Qualitative Data  |
|---------------------|--|---|
| Definition          | The data that are expressed in numerical terms are called quantitative data. | The qualitative attributes that cannot be measured or expressed in numerical terms are called qualitative data. |
| Deals with          | These variables deal with numbers.   | These variables deal with quality and characteristics.  |
| Example             | Income, expenditure, investment, etc.  | Intelligence, beauty, appearance, etc.  |

## Sources of Data

For collection of data, the source from which it is to be collected must be appropriately decided. The following diagram describes the various sources of data.



The sources of data can broadly divided into the following two categories.

- Internal Sources of Data
- External Sources Data

**Internal Source of Data:** Internal sources of data refers to the data that is generated within the organisation. Most of the organisation as a regular practice gather data for the purpose of their internal use. For example, data relating to sales, revenue, profit, etc. are collected by every organisation on a monthly or annual basis.

Such data are used by the organisation for the purpose of management and future planning. Similarly, the government generates internal data in its various department that is used in the process of planning.

**External Source of Data:** As against internal sources, external sources of data lie outside the domain of a particular organisation. Collection of information from the external sources requires investigations and surveys. In other words, the data is collected from certain outside agencies. External sources of data can be further classified into the following two categories.

- Primary Source of Data
- Secondary Source of Data

**Primary Source of Data:** The data that are collected by an investigator himself to carry out a specific research are known as primary data. Such data are collected by the investigator for the first time. They are collected from the original source of information. It gives the investigator a firsthand information relating to his study.

For example, data regarding the fees charged by various private schools can be obtained by conducting a survey among the students for the same. Primary sources of data forms a reliable and accurate source of data since the data is collected by the investigator in accordance with the needs and requirements of the study.

**Secondary Source of Data:** When the information is collected from the sources that already possess the required information, the source is known as secondary source. Data collected from such sources are not original as they have not been collected directly by the investigator rather they had been collected by someone else in the past. The secondary sources of data can be in published or unpublished form.

#### *Published Secondary Data*

Published data can be of the following form.

*i. Government publications:* The Central Statistical Organisation, the Central and the State Government Ministries along with some of the apex institutions such as RBI, etc. compile and publish data in their various publications. These data are published periodically, i.e., annually, half-yearly, quarterly and monthly.

These include Annual Survey of Industries, Statistical Abstract of India, Report on Currency and Banking, etc. Some of the important publications are **Reserve Bank of India Bulletin, Indian Trade Journals, Reports on Currency and Finance, etc.**

*ii. Semi-government publications:* The Municipalities and Metropolitan Councils publish various data related to education, health, births, deaths and other important fields.

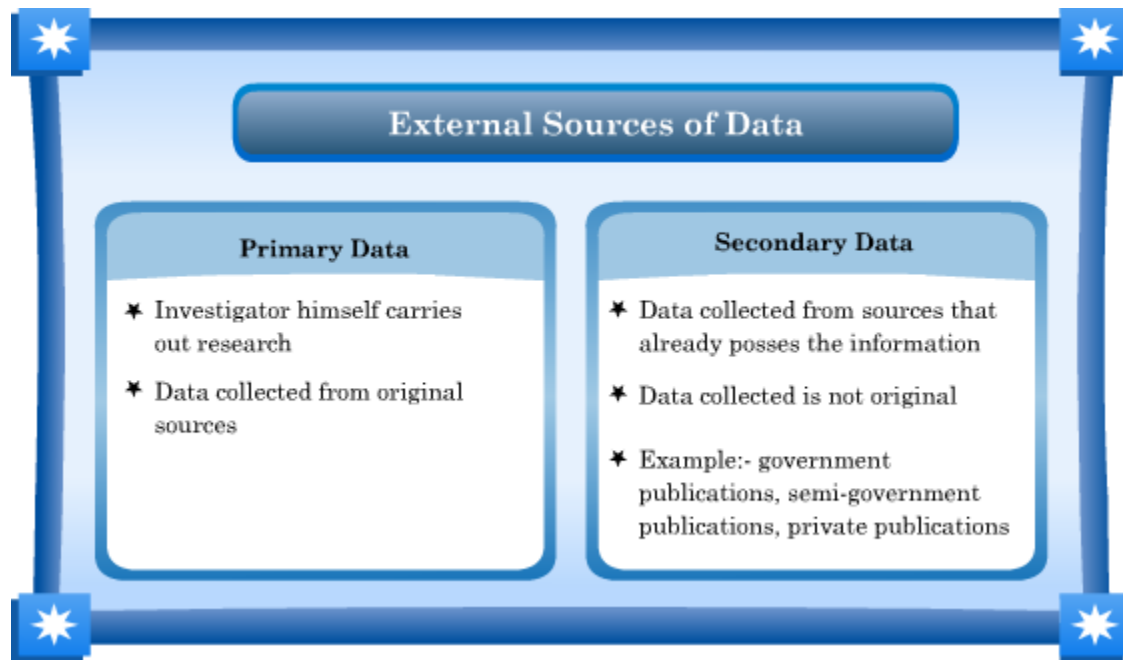
*iii. Publications by international bodies:* Some of the prime international organisations such as WHO, UNO and governments of all countries publish periodic data that are of international importance. These reports contain data related to different countries

and their various economic variables such as price, interest rate, poverty, population, literacy rate, etc.

*iv. Private publications:* Private publications include data published by the newspapers and journals, research institutes, reports of companies, articles, etc.

### *Unpublished Secondary Data*

Sometimes the data collected by the government or other organisations remain unpublished. However, such data can be used as secondary source.



### **Types of Data**

Based on the source from which the data is collected, data can be classified into the following two categories.

- Primary Data
- Secondary Data

**(i) Primary data:** Primary data refers to the data that is collected from primary sources. That is such data are collected by the investigator for the first time. The investigator collects such data in accordance with the objectives of the research. Since they are collected by the researcher they are more reliable. However, the collection of such data involves huge cost in terms of money, time and effort of the researcher.

(ii) **Secondary data:** Secondary data refers to the data that are collected from secondary sources. Such data have already been collected by somebody in the past and are collected and modified by the researcher as per the objectives of the study. As against the primary data, such data are comparatively less expensive to collect.

### ***Difference between Primary and Secondary Data***

The following are the basic points of difference between primary data and secondary data.

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

## **Method of Data Collection**

### **Objectives**

In this lesson, we will study the following methods of collecting primary data. We will study the following methods.

- Direct Personal Investigation
- Indirect Oral Investigation
- Telephonic Interview
- Information from Local Sources
- Information Through Questionnaire and Schedules

### **Introduction**

In the previous lesson, we studied about the two types of data- primary data and secondary data. In this lesson, we explore the various methods of collecting primary data.

### Methods of Collecting Primary Data

Some common methods of collecting primary data are:

- i. Direct Personal Investigation
- ii. Indirect Oral Investigation
- iii. Telephonic Interview
- iv. Information from Local Sources
- v. Information Through Questionnaire and Schedules



Let us understand each one of them in detail.

**I. Direct Personal Investigation:** Under this method the researcher directly collects information from the respondents. The researcher asks questions related to the study and the answers of the respondents are recorded. For example, if data is to be collected regarding the working conditions in a factory, the researcher can go to the factory and interview the workers. However, this method requires the researcher to be unbiased.

This method is suitable in those cases where the area of research or investigation is small and a high degree of originality and accuracy is required in the research.

The merits and demerits of using direct personal investigation are:

### **Merits**

- i. The data collected in this manner is **original** and commands high degree of **accuracy**.
- ii. The researcher can mould the interview according to the respondents. For example, language barriers, difficulty in understanding questions, etc. can be overcome by the researcher. Thus, this method is **flexible**.
- iii. As the information is collected directly by the researcher, it is highly **reliable**.
- iv. **Personal reactions** of the respondents to various questions can be easily observed under this method.
- v. **Uniformity** of the data can be maintained by the researcher.
- vi. Besides getting information on the question directly asked by the researcher, this method facilitates the collection of **related information** which might prove helpful in the research process.
- vii. This method is most **appropriate under limited field of study**.

### **Demerits**

- i. There exist chances of **personal bias** on part of the researcher.
- ii. This method involves huge **cost in terms of money, time and effort** of the researcher.
- iii. It proves **unsuitable when the field of investigation is large**.
- iv. This method **requires the researcher to be trained** so as to obtain appropriate results.

**II. Indirect Oral Investigation:** Under this method, the information is obtained orally from those persons who are expected to possess the necessary information. In other words, the information is not directly collected from the concerned person. It is rather collected from other people who are likely to have the required information. This method proves helpful in the cases where the person to be interviewed, is not in a position to give the interview or is reluctant to give the information.

For example, in cases of child marriage, drug addiction, etc. the person involved may not divulge the necessary details. Studies of such cases require the collection of data from other people who are related to the concerned person.

The merits and demerits of using indirect oral investigation are:

### **Merits**

- i. This method is ***suitable where the field of investigation is large.***
- ii. ***Different aspects of a situation can be analysed*** by interviewing different individuals who are related to the concerned person.
- iii. This is a relatively ***time and money saving*** method.
- iv. This method ***does not involve personal bias*** of the researcher.
- v. As this method involves collection of information from more than one related sources, ***complex information can also be collected.***

### **Demerits**

- i. Since the data is not collected directly from the concerned person, there ***exists a chance of the information being wrong or biased.***
- ii. This method of collecting data can lead to ***inappropriate and wrong conclusions.***

**III. Telephonic Investigation:** Under this method, the investigator interviews the respondents over telephone. This method is one of the simplest methods of data collection.

This method involves the following merits and demerits.

### **Merits**

- i. It is a ***cheap and convenient*** method of collecting the required information.
- ii. This method ***proves useful when the data is to be collected in a short period of time.***
- iii. This method is ***suitable in the cases where the respondent is hesitant in answering the questions in direct investigation.***



### *Demerits*

- i. ***The reactions of the respondents cannot be observed*** by the investigator.
- ii. This method ***can be used only where the respondents possess a telephone***.

**IV. Information from Local Sources:** In this method, the investigator appoints local correspondents at different places who collect the required information for the investigator. This method of investigation is used when the area of investigation is large. For example, a news channel has their correspondents at different places to get the information on various happenings.

This method involves the following merits and demerits

### ***Merits***

- i. This method has a ***large coverage*** as the correspondents are spread at different places.
- ii. This method is suitable in the cases where ***continuous flow of information*** is required.

### ***Demerits***

- i. ***Originality of the data is lost*** as the information is collected by the correspondents.
- ii. As the information is collected by more than one correspondent, the information ***lacks uniformity***.
- iii. As the information is first collected by the correspondent and then given to the investigator, the information may get ***delayed***.
- iv. The data may ***involve the bias of the correspondents***.

**V. Information Through Questionnaire and Schedules:** In this method, the investigator prepares a questionnaire or schedule that comprises of a list of relevant questions. The respondents answers are then recorded in the schedules. The the two methods of collecting information on the basis of questionnaires and schedules are:

- Mailing Method
- Enumerators Method

**(i) Mailing method:** In this method, the questionnaires are mailed to the informants, who then answer the questions secretly. After the respondents complete the questionnaire, they mail it back to the investigator. The investigator often instructs the respondents to return the questionnaire within a specified period of time.

The merits and demerits of using this method are:

### **Merits**

- i. This method is quite easy to use and **does not involve much money and effort**.
- ii. This method proves to be **helpful in the cases where the field of study is large** or where the information is required to be collected from different places.
- iii. As the respondents themselves fill up the questionnaire, the information provided is **reliable**.
- iv. It encourages the **respondent to freely express their opinion**.
- v. This method is **free from personal bias of the investigator**.

### **Demerits**

- i. This method requires the respondents to have a basic level of literacy and education. Thus, this method **cannot be used where the target population is illiterate**.
- ii. Sometimes it can happen that the respondents are **not able to understand** certain questions. In such cases, they often leave the questionnaire **incomplete**.
- iii. There can be **delay on part of the respondents in returning** the schedule to the investigator.
- iv. The schedule or the **questionnaire may even get lost** in the process.
- v. This method lacks **flexibility** in the sense that the questions cannot be reframed or moulded as per the respondents.

**(ii) Enumerators method:** In this method, the enumerator himself approaches the informants with the questionnaire, and based on their response, fills up the questionnaire. This method is able to overcome the demerits of mailing method of questionnaire and schedule. However, some merits and demerits of using this method are:

### **Merits**

- i. As the enumerator himself fills up the questionnaire or schedule, it ***can be used even in those cases where the target population is not literate.***
- ii. The ***complex and difficult questions can also be handled*** easily by the enumerators.
- iii. It leaves ***little scope for the questionnaire being incomplete.***
- iv. As against the mailing method, there are ***no chances of delay or the questionnaire being lost.***

### **Demerits**

- i. It ***involves huge time, effort and money.***
- ii. The success of this method ***depends on the availability of trained and competent enumerators.***
- iii. It ***may involve the bias of enumerators.***

### **Characteristics of a Good Questionnaire**

For effective collection of data, the questionnaire should be designed in an appropriate manner. Some essential characteristics of a good questionnaire are:

- i. The questionnaire should have ***limited number of questions.***
- ii. The questions should be ***simple and easy to understand.***
- iii. The questions should be ***arranged in a proper logical order.***
- iv. ***Undesirable questions*** such as those involving personal queries ***must be avoided.***
- v. The questionnaire ***should be tested*** by the enumerator before sending it to the respondents.
- vi. As far as possible, ***questions involving some mathematical or algebraic calculations should be avoided.***
- vii. ***Proper instructions must be provided*** for filling the questionnaire.
- viii. The questionnaire ***should also state the objective and aim of the survey.***

ix. The questions should be kept as short as possible. Besides, it is suitable to provide close ended questions, which require the respondents to provide **a yes/no for an answer**.

x. **Multiple choice questions should be preferred** over questions involving descriptive answers.

xi. **Doubtful or ambiguous** questions should be avoided.

xii. As far as possible the questions should not start with a negative word. For example, questions such as "don't you think....?" should be avoided.



## Survey: Census and Sample

### Objective

In this lesson, we will study and compare the two methods of survey, i.e. census method and sampling method.

### Introduction

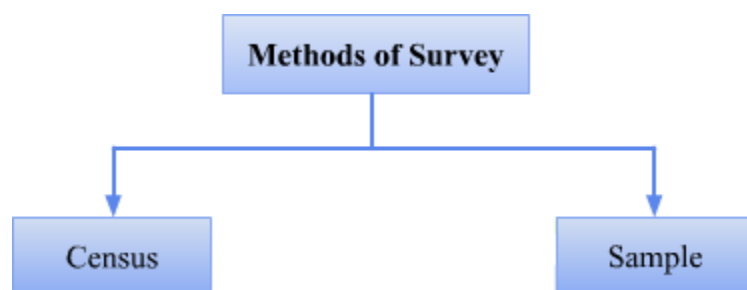
We know that every research or study requires a data. In the previous lessons, we have studied about the two types of data (primary data and secondary data) and how they can be collected.

An important question that arises is whether the data, for a research, should be collected from the whole of target population or only from a section or part of the population. For example, suppose a survey is conducted to judge the performance of the students in a school.

Now, the information related to this can be collected in two ways. One way is to collect the marks of the students of the entire school. The other way is choose some students and consider their marks to represent that of the entire school.

In other words, there are two methods of collecting data for the purpose of investigation- one by considering the whole of population and the other by considering a sample (part) of population. The former is called the census method and the latter is called the sample method of survey.

The following diagram depicts the two methods of survey.



### **Census Method**

Under the census method of survey, each and every item of the population is taken into account. That is, the information is collected from every single element of the population. Census method is also known as 'the method of complete enumeration'.

This method is most suitable in the cases where the target population size is limited. However, in some cases where exhaustive and complete information is required, this method is used even if the population size is large. The population census conducted by the Government of India is the most common example of the census method.

The following are some of the merits and demerits of census method.

### ***Merits***

- i. As census method takes into account each and every element of the investigation, it provides a ***detailed analysis*** of the respective fields of investigation.

- ii. This method ***does not involve any discretion*** of the investigator.
- iii. Data collected from this method possesses ***high degree of accuracy and reliability***.
- iv. This method is most ***suitable in the cases where diverse characteristics of each element of the population are required to be studied***.
- v. The data collected from this method is ***complete and exhaustive*** in all respects.

### ***Demerits***

- i. This method can ***prove to be cumbersome in the cases where the field of study is very large***.
- ii. Census method ***involves huge cost, time and effort***.
- iii. The success of this method ***requires the availability of a large number of trained and competent enumerators***.
- iv. Since a large volume of data is involved in this method, it ***increases the possibility of various types of errors***- calculation errors, careless handling of the schedule and questionnaire, errors in recording the responses, etc.

### ***Sample Method***

Sometimes, taking the entire population into account is not required and a section or sample of population can be taken to represent the entire population. This method, where only a part of the entire target population is studied and conclusions are drawn on the basis of that study, is known as the sample method.

This method is widely used in our daily lives. For example, while preparing rice, a sample of grains is taken and tested to check whether the rice is cooked. This method proves suitable in the cases where the field of investigation is large and the elements of the population are fairly homogenous.

### ***Features of a Good Sample***

The conclusion drawn from a sample can be generalised for the entire population, only if the chosen sample possesses certain essential characteristics. The following are some of the essential characteristics of a good sample.

- i. A sample ***must be based on the Principle of Statistical Regularity***. This principle states that the items in the sample should be randomly chosen. In other

words, each element in the population should have an equal chance of being selected in the sample.

ii. It **must be based on the Principle of Inertia of Large Numbers**. This principle states that the sample size should be fairly large. This is because a large sample is less prone to any changes.

iii. The **elements in the sample must be independent** of each other. That is, the selection of one item should not influence the selection or rejection of other items in the sample.

iv. The sample should be **representative of the entire population**.

v. The sample should be of **appropriate size**. It should be in proportion to the size of the population.

The following are some of the merits and demerits involved in using the sampling method of survey.

### **Merits**

i. Since only a section of the population is studied, the sampling method **economises on time, effort and money**.

ii. If the survey is carried out by trained and efficient enumerators, this method can also **provide reliable results**.

iii. It is **comparatively simple and easy to carry out** than the census method.

iv. **Errors arising due to large and voluminous data are avoided** in this method.

v. This method is most **suitable in cases that involve large population**.

### **Demerits**

i. This method **can provide misleading results if the sample size is not appropriately chosen**.

ii. In cases where the population is not homogenous, a sample **may not be a true representative** of the population.

iii. The results obtained on the basis of a sample **lack accuracy**.

iv. Sampling method is **not suitable in cases where the information regarding the whole population is required**.

v. The execution of survey using this method ***requires trained experts.***

### **Census Method versus Sampling Method**

Both census method and sampling method have their respective merits and demerits. Their suitability differs from one situation to another. Let us closely look at the points of difference between the two methods.

| <b>Basis of Difference</b> | <b>Census Method</b>  | <b>Sample Method</b>                                     |
|----------------------------|---|--|
| Field of study             | The whole population is the field of study.                   | Only a part of population is the field of study.         |
| Time and Money             | It is costly in terms of time as well as money.               | It is both time and money saving.                        |
| Suitability                | It is suitable when the field of study is small.              | It is suitable when the field of study is large.         |
| Investigators              | This method involves a larger number of investigators         | This method requires comparatively lesser investigators. |
| Accuracy                   | High degree of accuracy is maintained.                        | Comparatively lower degree of accuracy is maintained.    |
| Nature of field of study   | This method can also be used when the items are heterogeneous | It is not suitable when the items are heterogeneous      |

### **Sampling: Concept and Types**

#### **Objective**

In this lesson, we will study the following method of sampling.

- Random Sampling
  - Unrestricted Sampling- Lottery Method and Table of Random Numbers
  - Restricted Sampling- Stratified Sampling, Systematic Sampling and Cluster Sampling
- Non-Random Sampling
  - Judgement Sampling
  - Quota Sampling
  - Convenience Sampling

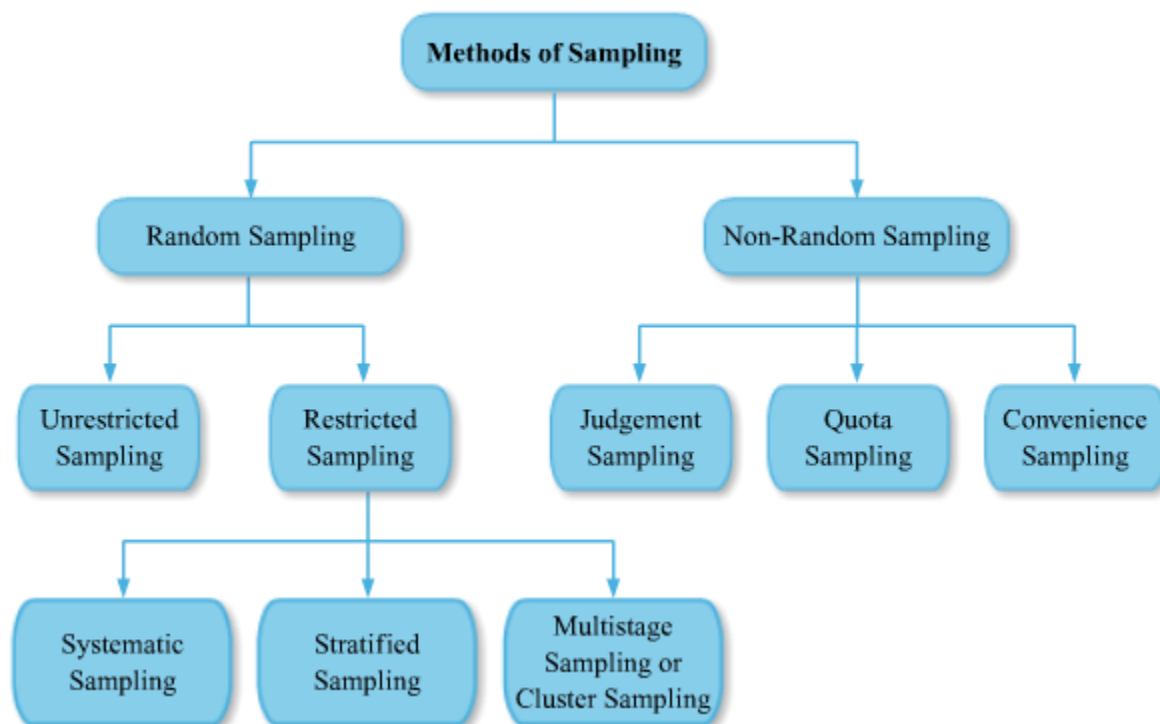


## Introduction

In the previous lesson, we studied about the two methods of survey i.e. sampling method and the census method. There are various methods of following a sampling method. In this lesson, we will explore some of the important methods of sampling.

## Methods of Sampling

The following diagram depicts the various methods of sampling.



The methods of sampling can be divided into the following two broad categories.

I. Random sampling

II. Non-Random sampling

Let us understand the two methods in detail.

**I. Random sampling:** In random sampling method, the items in the sample are chosen at random. In other words, each and every item of the universe has an equal chance of being selected in the sample. The items that would be chosen are not pre-planned or predictable. Thus, the selection of items in the sample is purely based on chance or on the concept of probability. Since, the items are chosen randomly, this method avoids personal bias. Random sampling can be further classified into the following two categories.

## ***Types of Random Sampling***

- i. Unrestricted Random Sampling
- ii. Restricted Random Sampling

**(i) *Unrestricted random sampling:*** Unrestricted sampling also known as simple random sampling is purely based on the idea of probability. Under this method, the chance of selection of each item has equal probability of being chosen. There is no discretion on the part of the investigator in selecting the items for the sample.

Unrestricted or random sampling can be done in the following two ways.

- a. Lottery Method
- b. Table of Random Numbers

a. *Lottery Method:* Under this method, each item in the population is given a unique number. Now, slips of paper are made for the number of each item. The slips are then mixed in a bowl or a container and one by one a slip is drawn. These items which bear the number of the drawn slips, are selected. However, caution must be taken that the slips of paper are identical in all respects (size, colour, etc.).

b. *Table of Random Numbers:* When the size of the population is too large, then the lottery method becomes every complex and difficult. In such cases, table of random numbers are used. There are three common random number tables.

These are (a) Tippet's table of random numbers (b) Fisher and Yate's number and (c) Kendall and Babington Smith numbers. The Tippet's random number tables is the most popular of the three tables. It consists of 10,400 four digit numbers.

For using the random number tables, various items of the population are first arranged in ascending order. Then, a page of the table is consulted and the required sample is taken by selecting successive numbers in the table either horizontally or vertically. Some of the numbers in the table are represented below.

### **Tippett's Numbers**

|      |      |      |      |      |
|------|------|------|------|------|
| 4357 | 2648 | 7453 | 2983 | 1786 |
| 6129 | 1166 | 1502 | 3675 | 1349 |
| 4227 | 0233 | 6487 | 4153 | 1248 |
| 1395 | 1598 | 3698 | 2496 | 1129 |

Some of the merits and demerits of random sampling are highlighted in the following points.

### **Merits**

- i. This method of sampling is **economical** and saves money, time and effort.
- ii. This is a **scientific method** for selecting a sample and is **free from any personal bias**.
- iii. Sample selected from this method is a better **representative of the population**.
- iv. It is **based on the theory of probability**.

### **Demerits**

- i. This method is **not suitable when a small sample size is to be selected**. The small sample selected by this method is not a representative of the population.
- ii. Use of this method **requires a list of the entire population**, which at times is difficult to obtain.
- iii. This method is **not suitable when the items in the population are heterogeneous**.
- iv. Numbering of each item and preparation of the slips can prove to be **cumbersome**.
- v. This method **cannot be used when some of the items are necessarily required to be included in the sample**.

**(ii) Restricted random sampling:** In restricted random sampling, although the selection of sample is random, it is restricted by following a particular procedure of selection. Restricted random sampling can be of the following three types.

- a. Stratified Sampling
- b. Systematic Sampling
- c. Cluster Sampling

a. **Stratified sampling:** In stratified sampling, the entire population is divided into small groups (*strata*) of homogenous items. A sample is then drawn at random from each strata. Different samples drawn from different strata are then grouped together to get the required sample for the population.

This method proves most useful in the cases where the population is heterogeneous. In such cases, different items of the population are grouped according to their

characteristics. For example, suppose a survey is to be conducted to know the pocket money of the children in India.

In this case, the strata can be formed based on the region, namely children of north India, children of west India, children of south India and children of east India.

Stratified sampling is also known as *mixed sampling*. This is because it involves the techniques of both restricted sampling and unrestricted sampling. Division of the population in a strata is restricted while the selection of sample from each strata is purely unrestricted.

The following are some of the merits and demerits of stratified random sampling.

#### *Merits*

- i. This method is *comparatively easy to implement* as the population is divided into different groups.
- ii. As the sample is selected from different groups having different characteristics, the sample is a *better representative* of the population.
- iii. This method is most *suitable in cases where the population is heterogeneous*.

#### *Demerits*

- i. The *division of population in strata may involve bias* of the investigator.
- ii. The division of strata requires trained personnel.
- iii. This method cannot be used in cases where the population size is small because further division of a small population becomes difficult.

b. *Systematic sampling*: In systematic sampling, the whole population is arranged systematically in an order (numerically, alphabetically, etc.). Then, an item is randomly chosen. Subsequently, every  $n^{\text{th}}$  item in the population is selected. For example, suppose from a population comprising of 5,000 items a sample of 500 is to be selected.

Here, the sampling interval would be  $10 (= \frac{5000}{500})$ . Now suppose the first item chosen is 7. After this, every  $10^{\text{th}}$  item would be selected in the sample, i.e.  $17^{\text{th}}$ ,  $27^{\text{th}}$ ,  $37^{\text{th}}$  and so on.

Systematic sampling has the following merits and demerits.

#### *Merits*

- i. This method is the *simplest* to implement.
- ii. It is *economical* and saves time and money.

#### *Demerits*

- i. The *selection of items in the sample depends on the first item chosen*.
- ii. This method *requires proper arrangement of all the items* of the population.

c. **Cluster sampling:** When the population is very large as well as spread, then the method of cluster sampling is used. In this method, the population is divided into groups and then, further into sub-groups. Selection of sampling is done in stages. For example, if a country wide survey is to be conducted, then the country can be divided into zones. These zones can be divided into states, then in cities and then further in villages. So, first a random selection of the zones is done. Then, from the selected zones, a sample of states is chosen and so on.

The following are some of the merits and demerits of cluster sampling.

#### *Merits*

- i. This method of sampling is *convenient in case of very large population* where even the stratified sampling proves lengthy.
- ii. This method is a more *organised* method of sampling.

#### *Demerits*

- i. This method is quite *complicated* as it involves sampling at a number of stages.
- ii. Its accuracy cannot be tested due to large population size.

**II. Non-random sampling:** As against random sampling, under non-random sampling each and every item of the universe does not have equal chance of being selected in the sample. This method of sampling is not based on the concept of probability. Non-random sampling can be further classified into the following three categories.

- i. Judgement Sampling/Purposive Sampling
- ii. Quota Sampling
- iii. Convenience Sampling

**(i) Judgement Sampling/Purposive Sampling:** In this method, the investigator himself chooses the sample items according to his own judgement or opinion in such a way that the selected sample items represent the whole universe. In other words, this method involves complete bias of the investigator.

Using this method, those items are taken in the sample that are necessarily required so that the sample truly represents the population. This method proves to be suitable in cases where a pilot survey is to be carried out. Judgement sampling involves the following merits and demerits.

### **Merits**

- i. This method involves **flexibility**. In other words, those items that hold special importance can be chosen by the investigator. However, in case of pure random sampling such items may or may not get selected.
- ii. This technique of sample selection is **simple**.

### **Demerits**

- i. It involves complete **bias** of the investigator.
- ii. As the selection of the items is at the discretion of the investigator, so the results are **not reliable**.

**(ii) Quota Sampling:** Under this method, the population is divided into different groups according to various characteristics of the population. The required sample is then selected from different groups. Each investigator works according to the quotas allotted to them.

For example suppose an investigator is asked to interview 100 children. He/She is further instructed that out of these 25% in class 6, 25% in class 8, 25% should be in class 10 and 25% in class 12. The selection of the required sample for the interview within the quota is done at the discretion of the investigator.

The following points highlight some of the merits and demerits of quota sampling.

### **Merits**

- i. This method **involves less money and time**.
- ii. This method is useful in cases of public opinion surveys.

### **Demerits**

- i. This method ***requires the investigators to be trained*** properly.

**(iii) Convenience Sampling:** As the name suggests, this method of sampling completely relies on the convenience of the investigator. In this method, the sample is selected as per the convenience of the investigator. 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. The following points highlight some of the merits and demerits of this method of sampling.

### ***Merits***

- i. This method is the ***simplest and easiest*** to use.
- ii. Sample can be obtained very ***quickly***.

### ***Demerits***

- i. It involves complete ***bias*** of the investigator.
- ii. The sample as chosen by the investigator ***may not be a true representative*** of the population.

## **Sampling and Non-sampling Errors**

### **Objective**

In this lesson, we will study about the following two types of statistical errors.

- Sampling Errors
- Non-Sampling Errors

### **Introduction**

While conducting a survey or a research either by census method or by sample method, there are chances of occurrence of some kind of errors. Besides the usual errors similar to that in mathematics, such as errors in calculation or wrong application of a formula, wrong interpretation, etc. errors of other kinds can also occur in statistics.

These errors lead to a difference between the true value/characteristic of a population and its estimated or approximated value. In this lesson, we will study the two kinds of statistical errors, namely, sampling errors and non-sampling errors.

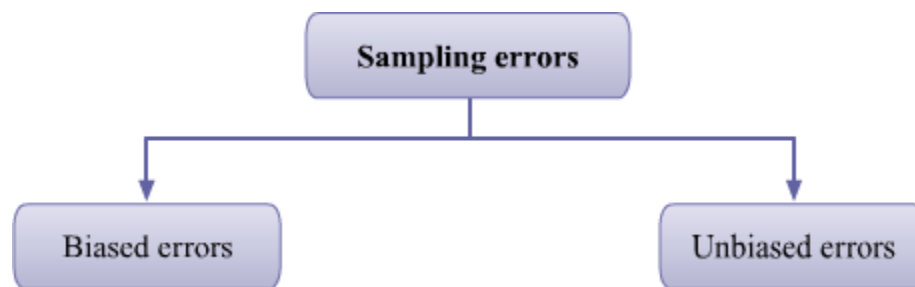
### **Sampling Errors**

For the purpose of conducting a research or a survey, often a sample is drawn. This sample is taken to be a representative of the population. Based on the sample, inferences and conclusions are drawn for the entire population. However, what would happen if the sample selected is not actually a true representative of the population, or if the sample selected is very small in proportion to the size of the population. In such cases, an error is said to occur.

There arises a divergence between the estimated value of a parameter/characteristic (such as average) and its actual value. Such errors that arise due to the size or the nature of the sample selected are called sampling errors. Sampling errors can be classified into the following two categories.

i. Biased Errors

ii. Unbiased Errors



(i) Biased sampling errors: When errors occur on account of the bias or the discretion of the investigator, they are known as biased sampling errors. Such errors occur when non-random sampling methods are used. For instance, an investigator may select a sample which is not a true representative of the population. Thus, based on such a sample, the inferences for the entire population can be misled.

(ii) Unbiased errors: When errors occur due to pure chance factor rather than the bias of the investigator, they are known as unbiased errors. For instance, some significant item or information required in the sample to make it a true representative of the population, might happen to be left out. In such cases, the conclusion drawn would not be reliable. Now, the question that arises is that how can the sampling errors be reduced. It is found that there exists a negative relationship between the sample size and the probability of occurrence of sampling errors. Thus, sampling errors can be reduced by increasing the sample size.

### **Non-Sampling Errors**

These are referred to those errors that arise due to the adoption of faulty means of collection of data. Such errors can arise in both the census as well as the sample methods of survey. Some of the common non-sampling errors are:



(i) **Errors in response:** Sometimes it happens that the respondent does not provide correct information to the interviewer or may not respond at all to a particular question.

(ii) **Errors in arithmetical calculation:** Errors might occur in the mathematical calculation procedures such as addition, multiplication, errors in rounding off the digits, etc.

(iii) **Errors in interpretation:** Sometimes the investigator may misinterpret the data.

Unlike sampling errors, non-sampling errors cannot be eliminated even by increasing the sample size. In other words, non-sampling errors are more serious than the sampling errors. Rectification of non-sampling errors requires selection of a new sample and conducting a fresh survey.

## **Census of India and NSSO**

### **Objective**

In this lesson, you will study about two government bodies engaged in conducting surveys and collecting data in India. These two are:

- Census of India
- National Sample Survey Organisation (NSSO)

### **Introduction**

In the previous lesson, we studied about the census and the sampling methods of survey. In India the process of collection and tabulation of statistical data is done by various national and state level bodies such as Census of India, National Sample Survey Organisation (NSSO), Labour Bureau, Registrar of India, etc. In this lesson, we will study about two of these bodies the Census of India and the National Sample Survey Organisation (NSSO).

### **Census of India**

The Census of India is responsible for the collection of data related to different demographic characteristics. Data is collected for characteristics such as birth rate, death rate, size of the population and composition of population, fertility rate, mortality rate, etc. The information on all population related aspects are collected by the investigators appointed by the government.

They collect this information from each and every household of the country. In India, this exercise is carried out every 10 years. The first census was conducted in the year 1872 and the next in 1881. Since then, India has been conducting census every 10

years. The census conducted by India is the largest such exercise in the world. India completed its 15<sup>th</sup> National Census in the year 2011.

The data collected by the Census of India acts as a mirror for assessing the impact of various programmes and development policies initiated by the government. It also helps in identifying the improvement areas and making the future policy.

### **National Sample Survey Organisation (NSSO)**

In the year 1950, the National Sample Survey (NSS) was established under the Ministry of Finance. It was established on the proposal of Prof. P.C. Mahalanobis. The objective of this organisation is to collect data on various socio-economic matters such as unemployment, literacy, utilisation of educational services provided by the government, etc. Such data is used to support the process of economic planning and policy making.

In 1970, all aspects of NSS were brought under NSSO (National Sample Survey Organisation). NSSO functions under the direction of a Governing Council comprising of 18 experts from various government and non-government departments.

The NSSO has four broad divisions, namely- Survey Design and Research Division (SDRD), Field Operation Division (FDO), Data Processing Division (DPD) and Coordination Public Division (CPD). The data collected by the NSSO is released through its reports and journals known as '*Sarvekshana*'.

### ***Role of NSSO***

The following are some of the important function areas of NSSO.

- i. Collection of data related to organised industrial sector in India.
- ii. Survey on demographic, health and family planning.
- iii. Conducting surveys on employment, unemployment, rural labor and consumer expenditure
- iv. Collecting data related to land holdings and livestock enterprises.
- v. Survey on assets, debt and investment status.
- vi. Collection of the retail prices in rural and urban areas for the construction of Consumer Price Index.