Economics

(Statistics)

Chapter 2: Collection of Data

Collection of Data

Definition -

Data collection is defined as the procedure of collecting, measuring and analyzing accurate insights for research using standard validated techniques. A researcher can evaluate their hypothesis on the basis of collected data. In most cases, data collection is the primary and most important step for research, irrespective of the field of research. The approach of data collection is different for different fields of study, depending on the required information.

Data is a tool which helps in reaching a sound conclusion by providing information therefore. For statistical investigation, collection of data is the first and foremost.

- **Sources of Data:**
  - Primary Source
  - Secondary Sources
  - Published sources
  - Un-published sources
- **Primary Data:** Data originally collected in the process of investigation are known as primary data. This is original form of data which are collected for the first time.It is collected directly from its source of origin.

Methods of collecting primary data –

There are three basic ways of collecting data:

- Personal interview OR Direct Personal Investigation.
- Mailing (questionnaire surveys).
- Telephone interviews.
- Indirect verbal investigation.
- Information from local sources.
- Enumerator method.

Secondary data It refers to collection of data by some agency, which already collected the data and processed. The data thus collected is called secondary data.

Point of difference between Primary and Secondary data -

- Accuracy,
- Originality,
- Cost,
- Need of modification

BASIS FOR COMPARISON	PRIMARY DATA	SECONDARY DATA
Meaning	Primary data refer to the first hand data gathered by the researcher himself.	Secondary data means data collected by someone else earlier.
Data	Real time data	Past data
Process	Very involved	Quick and easy
Source	Surveys, observations, experiments, questionnaire, personal interview, etc.	Government publications, websites, books, journal articles, internal records etc.
Cost effectiveness	Expensive	Economical
Collection time	Long	Short
Specific	Always specific to the researcher's needs.	May or may not be specific to the researcher's need.
Accuracy and Reliability	More	Relatively less

Sources of secondary data –

Secondary Data: Using existing data generated by large government Institutions, healthcare facilities etc. as part of organizational record keeping. The data is then extracted from more varied datafiles.

Published sources:

- Govt. publication.

- semi-Govt. Publication.
- Reports of committees & commissions.
- Private publications e.g., Journals and News papers research institute, publication of trade association.
- International publications.

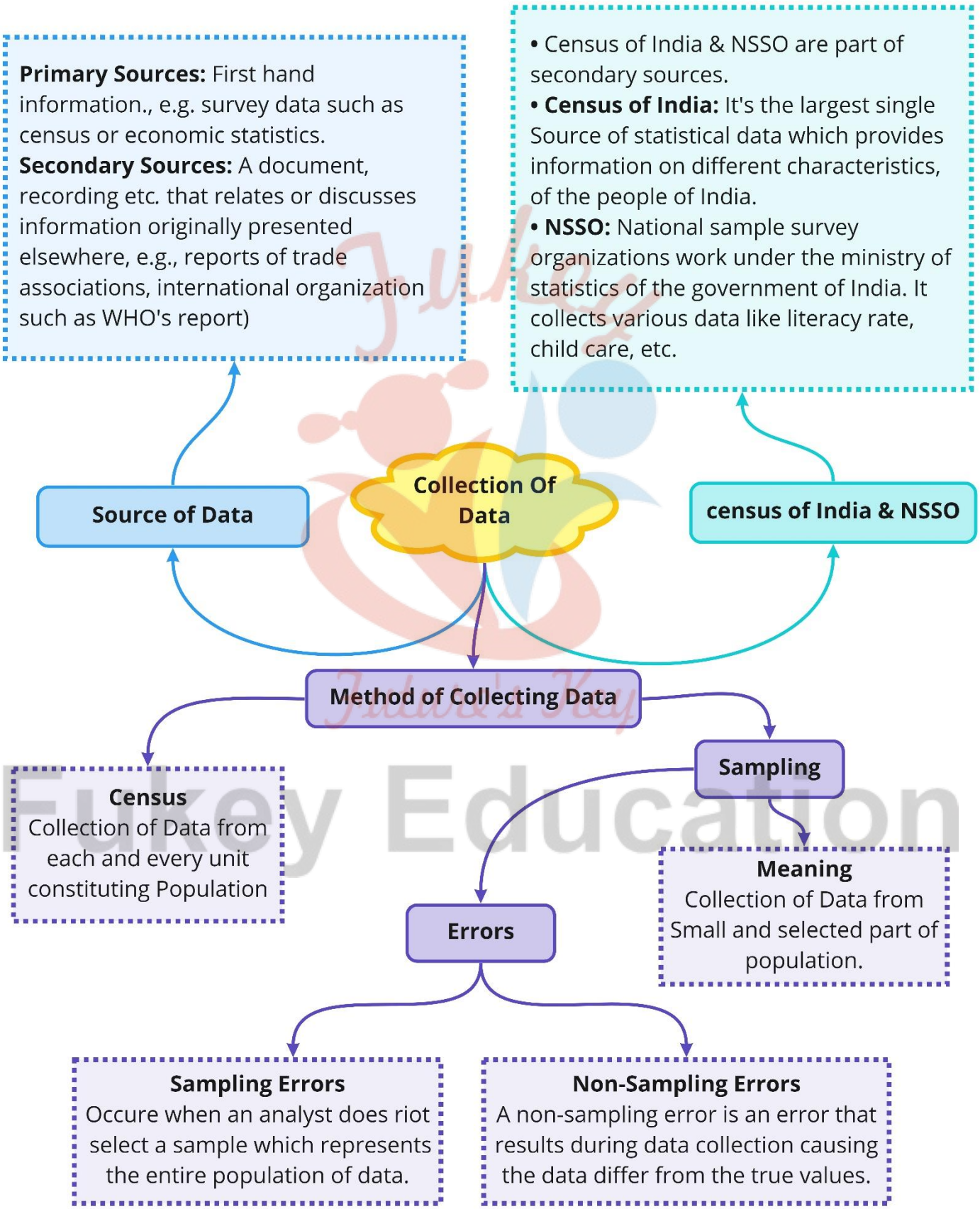
Unpublished Sources -

The statistical data needn't always be published. There are various sources of unpublished statistical material such as the records maintained by private firms, business enterprises, scholars, research workers, etc. They may not like to release their data to any outside agency.

Other source : web-site

Pilot Survey: Before sending the questionnaire to the information. It should be pretested. As a result of its short comings if any, can be removed. Such pretesting named as pilot survey

Class : 11th Economics (Statistics)
Chapter-2 : Collection Of Data



Important Questions

Multiple Choice Questions-

1. Stratified sample is preferred where:
   (a) Population is perfectly homogeneous
   (b) Population is non-homogeneous
   (c) Random sampling is not possible
   (d) Small samples are required
2. Data collected for the first time from the source of origin is called:
   (a) primary data
   (b) secondary data
   (c) internal data
   (d) none of these
3. Census method is suitable for that investigation in which:
   (a) the size of population is large
   (b) high degree of accuracy is not required
   (c) there are widely diverse items
   (d) intensive examination of diverse items is not required
4. Which of the following factor(s) are considered when comparison between sampling and census method is made?
   (a) Area of survey
   (b) Accuracy of data
   (c) Cost of collection
   (d) All of these
5. Reliability of sampling data depends on:
   (a) size of sample
   (b) method of sampling
   (c) training of enumerators
   (d) all of these

6. Under random sampling, each item of the universe has __ chance of being selected.

(a) equal

(b) unequal

(c) zero

(d) none of these

7. Which of the following methods is used for the estimation of population in a country?

(a) Census method

(b) Sampling method

(c) Both fa) and (b)

(d) None of these

8. Personal bias is possible under:

(a) random sampling

(b) purposive sampling

(c) stratified sampling

(d) quota sampling

9. If the investigator wants to select a sample on the basis of diverse characteristics of the population, which method should he use?

(a) Convenience sampling method

(b) Quota sampling method

(c) Stratified sampling method

(d) Both (b) and (c)

10. For drawing lottery _____ sampling is used.

(a) random

(b) purposive

(c) stratified

(d) quota

11. Which of the following methods is used for the estimation of population in country`

(a) Sampling Method

(b) Census Method

(c) Both (a) and (b)

(d) Neither (a) nor (b)

12. What kind of data are contained in the census of population and national income estimates, for the government?

(a) Primary data

(b) Secondary data

(c) Internal data

(d) None of these

13. The data collected on the height of a group of students after recording their heights with a measuring tape are:

(a) Primary data

(b) Continuous data

(c) Discrete data

(d) Secondary data

14. Which of the following is a method of secondary data collection?

(a) Direct personal investigation

(b) Direct oral investigation

(c) Collection of information through questionnaire

(d) None of these

15. In random sampling:

(a) Each element has equal chance of being selected

(b) Sample is always full of bias

(c) Cost involved is very less

(d) Cost involved is high

Very Short Questions:

1. Define primary data.
2. Define secondary data.
3. What are the two sources of data?
4. Mention two sources of secondary data.

5. In what parameters is the statistical information published in the census of India?
6. Mention two demerits of indirect oral investigation.
7. The progress report of a railway published by the railway department is what kind of data?
8. When is a direct personal investigation suitable for primary data collection?
9. When are the qualities of a good Questionnaire?
10. Why is a pilot survey important?
11. What is the universe in statistics?
12. Define sample.
13. Define the census method.
14. Explain the sample method.
15. What do you mean by random sampling?
16. What is purposive or deliberate sampling?
17. Define stratified and mixed sampling?
18. Explain systematic sampling.
19. What is quota sampling?
20. What is convenience sampling?

Short & Long Questions:

1. Primary Source?
2. Secondary Source?
3. Principal Differences between Primary and Secondary Data?
4. Direct Personal Investigation?
5. Indirect Oral Investigation
6. Difference between Direct Personal Investigation and Indirect Oral Investigation
7. Information from Local Sources or Correspondents
8. Collection of secondary data?
9. Statistical Errors: Sampling and Non-Sampling Errors

ANSWER KEY

Multiple Choice Answers-

1. B
2. A
3. C
4. D
5. D
6. A
7. A
8. B
9. D
10. A
11. B
12. B
13. A
14. D
15. A

Very Short Answers:

1. Primary data is the collection of data collected by the investigator for his own purpose for the first time. These are collected from the source of origin.
2. According to Wessel, “Data collected by another person is known as secondary data”. It is known as secondary data as it has already been collected by somebody else. These data are accessible in the form of a published and unpublished report.
3. The two sources of data are:
   - Primary source
   - Secondary source
4. The two sources of secondary data are:

- Government publication
- Semi-government publication

5. The statistical information is published in the following parameters in the census of India
   - Population projection
   - Sex composition of a population
   - Density of population
   - Size, growth rate, and distribution of people in India
6. The two demerits of indirect oral investigation are:
   - Less accurate
   - Biased
   - Doubtful conclusion
7. The progress report of a railway published by the railway department is secondary data.
8. The direct personal investigation method is suitable for collecting primary data only on the following situations:
   - When the investigation is confined and less
   - When an authentic and accurate information is required
   - When the data is to be kept secret
   - When the direct contact with information is needed
9. A good Questionnaire should have the following qualities:
   - Less number of Questions
   - Should be clear
   - Proper order of Question
   - Non-controversial
   - Questions related to the topic
   - Request for return
10. A pilot survey is essential because of the following:
    - It helps in assessing the quality and suitability of Questions.

- It evaluates the performance of enumerators.
- It helps in designing a set of rules for the investigator.
- It estimates the time and cost involved in the final survey.

11. In statistics, the term universe or population indicates an aggregate of items studied for investigators.

12. Sample is a collection of an item from the population that represents the characteristics of the population.

13. It is a method of collecting data where each item related to the problem of the investigation is collected.

14. It is a process of collecting data in which the sample of a group of items are examined, and conclusions are drawn on their basis.

15. In this method, every item of the universe has an equal chance of being selected in the sample.

16. It is a sampling method where the investor chooses the sampling items according to his opinion, and it is the best for the population.

17. In this method, the universe is divided into two groups having different characteristics, and the items are selected for each group, hence the entire group is represented.

18. In systematic sampling, population units are arranged according to the alphabets, numbers, and geography. Here, every nth numerical item is selected as a sample.

19. Here, the universe is divided into two sections or groups in terms of their characteristics.

20. In this method, sampling is done according to the investigator's convenience.

Short & Long Answers:

1. You want to know about the quality of life of the people in your town. You may like to ascertain the quality of life in terms of per capita expenditure of different households in your town. You decide to collect the basic data yourself through statistical survey(s), of course with the help of investigators or field workers. While doing this exercise you are relying on primary source of the data. Thus, primary source of data implies collection of data from its source of origin. It offers you firsthand quantitative information relating to your statistical study. You or your team of investigators are contacting the respondents (people offering basic information) and obtaining the desired quantitative information on per capita expenditure of different households in your town.

Primary source of data implies collection of data from its source of origin. It offers you first-hand quantitative information relating to your statistical study.

2. Secondary Source of collection of data implies obtaining the relevant statistical information from an agency, or an institution which is already in possession of that information. To continue with the previous example, data relating to the quality of life of the people of your town (or the data on per capita expenditure) may have already been collected by the State Government. You can simply approach the concerned Government department and request for the desired information. This will be a Secondary Source of data for you. Thus, secondary source implies that the desired statistical information already exists and you are simply to collect it from the concerned agency or the department. You are not to conduct statistical survey(s) yourself and you are not to contact the respondents (people offering basic information). OT course, you are not getting first hand information relating to your statistical study. You are simply relying on the information which is already existing.

Secondary source of data implies collection of data from some agency or institution which already happens to have collected the data through statistical survey(s). It does not offer you first-hand information relating to your statistical study. You are to rely on the information which is already existing.

3. The following are some principal differences between primary and secondary data:

1) **Difference in Originality:** Primary data are original because these are collected by the investigator from the source of their origin. Against this, secondary data are already in existence and therefore, are not original.

2) **Difference in Objective:** Primary data are always related to a specific objective of the investigator. These data, therefore, do not need any adjustment for the concerned study. On the other hand, secondary data have already been collected for some other purpose. Therefore, these data need to be adjusted to suit the objective of study in hand.

3) **Difference in Cost of Collection:** Primary data are costlier in terms of time, money and efforts involved than the secondary data. This is because primary data are collected for the first time from their source of origin. Secondary data are simply collected from the published or unpublished reports. Accordingly, these are much less expensive.

Of course, it may be noted that, there are no fundamental differences between primary data and secondary data. Data are data, whether primary or secondary. These are classified as primary or secondary just on the basis of their collection: first-hand or second-hand. Thus, a particular set of data when collected by the investigator for a specific purpose from the source of origin, would be primary data. And the same set of data, when used by some other investigator for his own purpose, would be known as secondary data. Thus, Secrist has rightly pointed out, "The distinction between primary and secondary data is one of the degree. Data

which are primary in the hands of one party may be secondary in the hands of other.''

Primary and Secondary Data—The Basic Difference

- If we are collecting data from its source of origin, for the first time, it is primary data.
- If we are using data which have already been collected by somebody else, it is secondary data.

Note: If you are getting data from somebody else who collected it from its source of origin but did not use it for his own study, it will be deemed as primary data.

4. The direct personal investigation is the method by which data are personally collected by the investigator from the informants. In other words, the investigator establishes direct relation with the persons from whom the information is to be obtained. The success of this method, however, requires that the investigator should be very diligent, efficient, impartial and tolerant.

Direct contact with the workers of an industry to obtain information about their economic conditions is an example of this method.

Suitability

This method of collecting primary data is suitable particularly when:

(i) the field of investigation is limited or not very large.

(ii) a greater degree of originality of the data is required.

(iii) information is to be kept secret.

(iv) accuracy of data is of great significance, and

(v) when direct contact with the informants is required.

Merits

Data, thus, collected have the following merits:

(i) Originality: Data have a high degree of originality.

(ii) Accuracy: Data are fairly accurate when personally collected.

(iii) Reliability: Because the information is collected by the investigator himself, reliability of the data is not doubted.

(iv) Related Information: When in direct contact with the informants, the investigator may obtain other related information as well.

(v) Uniformity: There is a fair degree of uniformity in the data collected by the investigator himself from the informants. It facilitates comparison.

(vi) Elastic: This method is fairly elastic because the investigator can always make necessary adjustments in his set of questions.

Demerits

However, the method of direct personal investigation suffers from certain demerits, as **under:**

(i) Difficult to Cover Wide Areas: Direct personal investigation becomes very difficult when the area of the study is very wide.

(ii) Personal Bias: This method is highly prone to personal bias of the investigator. As a result, the data may lose their credibility.

(iii) Costly: This method is very expensive in terms of the time, money and efforts involved.

(iv) Limited Coverage: In this method, area of investigation is generally small. The results are, therefore, less representative. This may lead to wrong conclusions.

5. Indirect oral investigation is the method by which information is obtained not from the persons regarding whom the information is needed. It is collected orally from other persons who are expected to possess the necessary information, these other persons are known as witnesses. For example, by this method, the data on the economic conditions of the workers may be collected from their employers rather than the workers themselves.

Suitability

This method is suitable particularly when:

(i) the field of investigation is relatively large.

(ii) it is not possible to have direct contact with the concerned informants.

(iii) the concerned informants are not capable of giving information because of their ignorance or illiteracy.

(iv) investigation is so complex in nature that only experts can give information.

This method is mosdy used by government or non-government committees or commissions.

Merits

Some of the notable merits of this method are as under:

(i) Wide Coverage: This method can be applied even when the field of investigation is very wide.

(ii) Less Expensive: This is relatively a less expensive method as compared to Direct Personal Investigation.

(iii) Expert Opinion: Using this method an investigator can seek opinion of the experts and thereby can make his information more reliable.

(iv) Free from Bias: This method is relatively free from the personal bias of the investigator.

(v) Simple: This is relatively a simple approach of data collection.

Demerits:

However, there are some demerits, as under:

(i) Less Accurate: The data collected by this method are relatively less accurate. This is because the information is obtained from persons other than the concerned informants.

(ii) Biased: There is possibility of personal bias of the witnesses giving information.

(iii) Doubtful Conclusions: This method may lead to doubtful conclusions due to carelessness of the witnesses.

6. The difference between direct personal investigation and indirect oral investigation is as under:

i. In the case of direct personal investigation, the investigator establishes direct contact with the informants. On the other hand, in the case of indirect oral investigation, information is obtained by contacting other than those about whom information is sought.

ii. Direct Personal Investigation is generally possible when the field of investigation is small. On the other hand, indirect oral investigation is generally preferred when the field of investigation is relatively large.

iii. In the Direct Personal Investigation, the investigator must be well versed in the language and cultural habits of the informants. There is no such requirement in the case of Indirect Oral Investigation.

iv. Direct investigation is relatively costlier than the indirect investigation.

7. Under this method, the investigator appoints local persons or correspondents at different places. They collect information in their own way and furnish the same to the investigator.

Suitability

This method is suitable particularly when:

(i) regular and continuous information is needed.

(ii) the area of investigation is large.

(iii) the information is to be used by journals, magazines, radio, TV, etc. and

(iv) a very high degree of accuracy of information is not required.

Merits

Principal merits of this method are as under:

(i) Economical: This method is quite economical in terms of time, money or efforts involved.

(ii) Wide Coverage: This method allows a fairly wide coverage of investigation.

(iii) Continuity: The correspondents keep on supplying almost regular information.

(iv) Suitable for Special Purpose: This method is particularly suitable for some specialpurpose investigations, e.g., price quotations from the different grain markets for the construction of Index Number of agricultural prices.

Demerits

Following are some notable demerits of this method:

(i) Loss of Originality: Originality of data is sacrificed owing to the lack of personal contact with the respondents.

(ii) Lack of Uniformity: There is lack of uniformity of data. This is because data is collected by a number of correspondents.

(iii) Personal Bias: This method suffers from the personal bias of the correspondents.

(iv) Less Accurate: The data collected by this method are not very accurate.

(v) Delay in Collection: Generally, there is a delay in the collection of information through this method.

8. There are two main sources of secondary data:

(1) Published Sources

Some of the published sources of secondary data are:

(i) Government Publications: Ministries of the Central and State Governments in India publish a variety of Statistics as their routine activity. As these are published by the Government, data are fairly reliable. Some of the notable Government publications on Statistics are: Statistical Abstract of India, Annual Survey of Industries, Agricultural Statistics of India, Report on Currency and Banking, Labour Gazette, Reserve Bank of India Bulletin, etc.

(ii) Semi-Government Publications: Semi-Government bodies (such as Municipalities and Metropolitan Councils) publish data relating to education, health, births and deaths. These data are also fairly reliable and useful.

(iii) Reports of Committees and Commissions: Committees and Commissions appointed by the Government also furnish a lot of statistical information in their reports. Finance Commission, Monopolies Commission, Planning Commission are some of the notable commissions in India which supply detailed statistical information in their reports.

(iv) Publications of Trade Associations: Some of the big trade associations, through their statistical and research divisions, collect and publish data on various aspects of trading activity. For example, Sugar Mills Association publishes information regarding sugar mills in India.

(v) Publications of Research Institutions: Various universities and research institutions publish information as findings of their research activities. In India, for example, Indian Statistical Institute, National Council of Applied Economic Research publish a variety of statistical data as a regular feature.

(vi) Journals and Papers: Many newspapers such as 'The Economic Times' as well as magazines such as Commerce, Facts for You also supply a large variety of statistical information.

(vii) Publications of Research Scholars: Individual research scholars also sometimes publish their research work containing some useful statistical information.

(viii) International Publications: International organisations such as UNO, IMF, World Bank, ILO, and foreign governments etc., also publish a lot of statistical information. These are used as secondary data.

(2) Unpublished Sources

There are some unpublished secondary data as well. These data are collected by the government organisations and others, generally for their self use or office record. These data are not published. This unpublished numerical information may, however, be used as secondary data.

A Note of Caution for the Users of Secondary Data Users of secondary data must check:

(i) reliability of data,

(ii) suitability of data, and

(iii) adequacy of data.

9. Statistical errors are broadly classified as (i) sampling errors, and (ii) non-sampling errors. Following are the details:

(i) Sampling Errors: These are related to the size or nature of the sample selected for the study. Due to a very small size of the sample selected for study or due to nonrepresentative nature of the sample, the estimated value may differ from the actual value of a parameter. The error thus emerging, is called sampling error. For example, if the estimated value of a parameter is found to be 10 while the actual/true value is 20 then, the sampling error = estimated value of the parameter – true value of the parameter = 10-20 = -10.

(ii) Non-sampling Errors: These are errors related to the collection of data. These are of the following types:

Error of Measurement: Error of measurement may occur due to.- (a) difference in the scale of measurement, and (b) difference in the rounding off procedure adopted by different investigators.

Error of Non-response: This arises when the respondents do not offer the required information. Error of Misinterpretation: This arises when the respondent fails to interpret the questions in the questionnaire.

Error of Calculation or Arithmetical Error: It occurs in the course of addition, subtraction or multiplication of data.

Error of Sampling Bias: It occurs when, for some reason or the other, a part of target population, cannot be included in the choice of a sample.

Larger the field of investigation or larger the population size, greater is the possibility of errors related to the collection of data, or data acquisition. It must be noted here that a non-sampling error is more serious than a sampling error. Because a sampling error can be minimised by opting for a larger sample size. No such possibility exists in case of nonsampling errors.