INTRODUCTION

By the term investigation (or enquiry) we mean the search for information or knowledge. Statistical investigation, thus, implies search for knowledge with the help of statistical devices like collection, classification, analysis and interpretation, etc. According to Griffin, “Statistical enquiries have always required considerable skill on the part of the statistician, rooted in a broad knowledge of the subject matter area and combined with considerable ingenuity in overcoming practical difficulties.” So to apply statistical methods to any problem it is necessary to collect the numerical facts since statistical analysis is not possible without them.

For example, if an investigation is made into accounts of a college hostel, then the investigation will mainly cover:

1. Income from residents as seat rent, meal charge, any grant from college.
2. Expenditure as hostel rent, overhead charges.
3. Expenditure on dry rations, broad meals, including special meals.
4. Expenditure for annual functions.

ORGANISATION OF STATISTICAL INVESTIGATION

Statistical investigation is a long and comprehensive process. It extends over various stages from initial planning to the final preparation of the report. The various stages are:

1. Planning of statistical investigation.
2. Collection of data.
3. Editing of data.
4. Presentation of data.
5. Analysis of data.
6. Interpretation of data.
7. Preparation of the report.
PLANNING OF STATISTICAL INVESTIGATION

A proper system is essential for conducting a statistical investigation. Planning must precede the execution. Careful planning is essential to get the best results at the minimum cost and time. It is essential to consider the following points while planning a statistical investigation.

1. Objective of the enquiry should be fully known.
2. Scope of the enquiry should be determined.
3. Nature of information to be collected should be decided.
4. Unit of data collection should be defined.
5. Source of data collection or type of data to be used, that is, primary or secondary should be decided.
6. Method of data collection, that is, census or sampling method, should be decided beforehand.
7. Choice of frame should be made.
8. Reasonable standard of accuracy should be fixed.

COLLECTION OF DATA

Collection data is the first step in a statistical investigation. The person who conducts the enquiry is known as ‘investigator’. The persons who help the investigator in collecting the information are called ‘enumerators’. The persons from whom the information is collected are known as ‘respondents’.

The primary task in any statistical enquiry is to determine its aims and objectives. Once these objectives have been determined, the next task is to collect the data. The data to be used can be of two types, namely (1) Primary data and (2) Secondary data.

PRIMARY DATA

It is the data collected by a particular person or organisation for his own use from the primary source. According to Wessel, “Data originally collected in the process of investigation are known as Primary data.”

Methods of Collecting Primary Data

1. Direct personal investigation: Under this method, the investigator collects the data personally. He has to go to the spot for conducting enquiry and has to meet the persons concerned. It is essential that the investigator should be polite, tactful and has a sense of observation.

   Merits
   1. Original data is collected.
   2. Reliable information is obtained.
   3. There is uniformity in the collection of information.
4. The method yields more accurate results.
5. It is most useful when the area of investigation is very small.

**Demerits**

1. It is a very costly method.
2. It is very time-consuming.
3. The chances of personal bias are greater.
4. It can give wrong results.
5. It can be used only when the area of investigation is small.

2. **Indirect oral investigation:** Here the investigator gets the information indirectly. In this method a third person is contacted who is expected to know the necessary details about the persons for whom the enquiry is meant. This method is generally suitable for commissions and enquiry committees appointed by the government.

**Merits**

1. It is a simple and convenient method of investigation.
2. Intensive and extensive investigation is possible.
3. It is economical.
4. The investigator can take the help of expert enumerators to collect the data.

**Demerits**

1. The results may not always be true and accurate; these are open to suspicion and doubt.
2. The informants may not be serious in furnishing the proper replies and hence, degree of accuracy can be affected.

3. **Information through correspondents:** In this method, the investigator appoints local agents and correspondents to collect the data. This method is open to personal prejudices and likings of the reporters and agents. The data collected by this method is not strictly comparable and the person who does not deal with this data may know the exact nature of the variation in the reported data.

**Merits**

1. It is the most economical and extensive method of investigation.
2. It is useful where the information is needed from time to time.

**Demerits**

1. There is lack of originality in data.
2. The information can be biased.
3. Degree of accuracy is generally very low.
4. There may not be uniformity in the collected data.

4. **Questionnaire method:** In this method, the necessary information is collected from the respondents through a questionnaire. A questionnaire is a set of questions relating to the enquiry. The information can be collected through questionnaires in two ways.
(a) **Questionnaire sent through post:** In this method, the questionnaire is mailed to the informant. It also carries a request for the quick response and lays down the aims and objectives of collecting the information. In this method, the return postage expenses are generally borne by the investigator. This method is suitable when the area under investigation is very large and when the informants are educated and are likely to cooperate.

**Merits**

1. It can conveniently be used when the area of investigation is very large.
2. It is economical and involves less labour and time.
3. Collected information is more reliable.
4. Original data can be obtained directly from the informants.

**Demerits**

1. It involves uncertainty about the response.
2. Information can be collected from the literates only.
3. Supplementary questions cannot be asked.

(b) **Questionnaires sent through enumerators:** In this method, the enumerators are appointed by the investigators and they contact the informants, get replies and fill them in the questionnaire form. This method is suitable when the informant is illiterate.

**Merits**

1. There is no chance of non-response since the enumerators go to all the informants.
2. Personal contact makes it possible to collect reliable data.
3. Personal bias is minimised.
4. Illiterate informants can also be approached.

**Demerits**

1. It is costly because the enumerator has to be paid.
2. It is time-consuming.
3. It requires trained enumerators.
4. It can be employed only by big organisations.

5. **Technique of drafting of a good questionnaire:** The following points should be kept in mind while drafting a questionnaire:

    1. The size of the questionnaire should be small.
    2. The questions should be simple.
    3. Possible answers could be suggested.
    4. The questions asked should be in proper sequence.
    5. Irrelevant questions should be avoided.
    6. Questions should be free from personal bias and should not injure the feelings of the respondents.
    7. Questions involving the mathematical calculations should be avoided.
8. Necessary instructions and definitions should be given.
9. Proper words should be used in the questionnaire.
10. The covering letter should contain the following steps:
   (i) The letter should state the purpose of the survey, and how the informant would benefit from it.
   (ii) A self-addressed and stamped envelope should be enclosed for the convenience of respondents for returning the questionnaire.
   (iii) The respondent should be assured that his answer will kept in strict confidence.
   (iv) The respondent should be promised that he will not be solicited after he fills up the questionnaire.
   (v) The respondent may be offered special inducements to return the filled in questionnaire.
   (vi) If the respondent is interested, the investigator should promise to send him a copy of the results of the investigation.

SECONDARY DATA

It is the data collected by some other person or organisation for their own use but the investigator also gets it for his use. According to M.M. Blair, “Secondary data are those already in existence for some other purpose than the answering of the question in hand.”

Sources of Secondary Data

The chief sources of secondary data can be classified in two types. These are: (i) Published sources and (ii) Unpublished sources.

Published sources: The published data may be obtained from various international, national and local publications. Following are the main sources of published data.

1. International Publications: Foreign governments and various national agencies publish regular and occasional reports of Economics and other related matters of significance, i.e., International Monetary Fund, International Bank for Reconstruction and Development, Statistical Year Book, annual reports of International Labour Organisation, etc.

2. Official Publications of Central and State Governments: The Central Government and the State Governments and the different departments publish reports on current issues. The information is generally more reliable and authentic. Some of the important publications are Census of India, Reserve Bank of India Bulletins, Statistical Abstracts of the States, Report of Currency and Finance, etc.

3. Committee Reports: Sometimes the government appoints survey and enquiry commissions to get the expert views on matters of great importance, for example, Reports of Public Accounts Committee of Lok Sabha.

4. Newspapers and Magazines: Newspapers and magazines also maintain their research bureaus and publish data on important issues.
5. **Individual Research Scholars:** Individual research scholars and other allied agencies also supply material on matters of importance.

6. **Semi-Official Publications:** Semi-government institutions like municipalities, corporations, district boards, panchayat samities, publish reports and data regarding birth and death, health and other related subjects.

**Unpublished data:** There are various sources of unpublished data such as the records maintained by the various government and private offices, studies made by the research scholars in the universities and other research institutions.

**Precautions in the Use of Secondary Data**

Before making use of the secondary data, the investigator must satisfy himself about the following.

1. Whether the data is reliable? Whether it is collected by qualified persons or not?
2. Whether the data is suitable for the purpose?
3. Whether the data is adequate?

<table>
<thead>
<tr>
<th><strong>Comparison Between Primary and Secondary Data</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary data</strong></td>
</tr>
<tr>
<td>2. Investigating agency collects the data.</td>
</tr>
<tr>
<td>3. It requires longer time for collection.</td>
</tr>
<tr>
<td>4. The investigator himself collects it hence, it is more reliable and suitable to the enquiry.</td>
</tr>
<tr>
<td>5. It requires elaborate organisation.</td>
</tr>
<tr>
<td>6. No extra precautions are required.</td>
</tr>
</tbody>
</table>

**METHODS OF DATA COLLECTION**

The word ‘**population**’ has a different meaning in Statistics than in ordinary usage. Ordinarily, the word population is used to mean the number of the people living in an area, a region or a country. In Statistics the word ‘population’ or ‘universe’ refers to the number of observations connected with the enquiry.

**Sample** refers to a part of the population selected for analysis. This particular part is representative of the whole population.

Two methods are generally used to collect the data, namely: (1) Census Method and (2) Sample Method.

When the entire population is taken into account for study, it is known as Census Method. When a small group is taken into account as representative of the whole, it is known as Sample Method.
Census Method

Census method means the complete enumeration of the population. In the world, every country conducts the population census at a regular interval of time. In our country the population census is conducted every 10 years.

Merits

1. The advantage of the consistency of large numbers is also available.
2. The data collected by this method gives correct result.
3. The analysis of data becomes more representative and true.

Demerits

1. This method requires a large amount of time, energy and money.
2. Qualified enumerators are required to conduct the study.
3. Sometimes information regarding some units is not available; in that case, this method may not be used successfully.

Sample Method

In the sampling technique instead of every unit of the universe only a part of it is studied and the conclusions are drawn on that basis for the entire universe. This technique was first used by A.L. Bowley in 1912 to study the extent of the poverty of labourers. In 1934, the Indian government also adopted this method.

(a) Random sampling: Random sampling refers to the sampling technique in which each and every item of the population is given an equal chance of being included in the sample. U.M. Harper observes “A random sample is a sample selected in such a way that every item in the population has an equal chance of being included”.

(i) Simple Random Sampling: In this method, a sample is selected by which each item of the population has an equal and independent chance of being included in the sample. Under this method certain number of items are chosen at random without any predetermined basis. These items are supposed to contain the characteristics of the population.

• Simple Random Sampling with Replacement: In this technique units are drawn from the population one after the other. Before drawing any unit, the unit drawn in the previous drawn is replaced into the population. Therefore, in SRSWR there is the possibility for the units to be repeated into the sample. Through this technique \(N^n\) samples of size ‘\(n\)’ can be drawn from a population containing ‘\(N\)’ units and therefore the probability of selecting any one of them is equal to \(\frac{1}{N^n}\).

• Simple Random Sampling without Replacement: In this technique, the unit drawn at any given draw is not replaced into the population before drawing the next unit. The number of samples of size ‘\(n\)’ that can be drawn from a population containing ‘\(N\)’ units is equal to \(Nc_n\) and the probability of selecting any one of them is equal to \(\frac{1}{Nc_n}\).
Stratified Random Sampling: This sampling technique is generally recommended when the population is heterogeneous. So in this method the population which is to be sampled, is divided into homogeneous groups and these groups are named strata possessing the similar characteristics. Samples are selected by taking equal proportion of items from each group.

Systematic Random Sampling: This method is popularly used in those cases where a complete list of the population from which the sample is to be drawn is available. Under this method, units to be drawn into the sample are selected at evenly spaced intervals. This means that the units in the population are to be arranged in some order. The first unit is selected by following any of the simple random techniques. Subsequent units are drawn at an equal sampling interval.

(b) Non-random sampling: A non-random sample is selected on the basis other than the probability. Considerations such as expert judgements, convenience, etc., determine the choice of sample. In other words the choice of sample is not free, but is determined by a number of considerations. Some of the important methods of non-random sampling are as follows.

(i) Judgement Sampling: In this method, the selection of items to be included in a sample is left to the discretion of the investigator. The items which are considered to be representative of a given population are selected for studying the characteristics of a population. Here the probability that an individual item will be selected is not known because the inclusion of items in a sample depends exclusively upon the preference or nature of the investigator. In judgement sampling, the accuracy of results depends upon the investigator. If he is not biased and has full knowledge of the population to be sampled, then his findings would be accurate.

(ii) Quota Sampling: Quota sampling is a type of judgement sampling. In this method quotas are set up for each specified characteristic, such as age, religion, rural or urban salary groups, etc. Each interviewer is asked to collect information accordingly. But the choice of the respondents is left to the discretion of the investigator. The reliability of this method mainly depends upon the efficiency, integrity and skill of the investigator.

(iii) Convenience Sampling: This method is also known as ‘Chunk’ which refers to the fraction of the population being investigated. This selection is neither based on random nor on judgement but on convenience. The convenience may be in regard to place, time and availability or resources, etc.

Census Vs Sampling: After studying the census method and the sampling method, one has to make a choice between the two. So, when the population is small and precise information is needed, the census method will be most appropriate; whereas, when the population is very large (or) the field of investigation is very wide and quick results are needed with minimum cost, sampling method is the most appropriate.