DEFINITIONS, SCOPE AND LIMITATIONS

In the modern world of computers and information technology, the importance of statistics is very well recogonised by all the disciplines. Statistics has originated as a science of statehood and found applications slowly and steadily in Agriculture, Economics, Commerce, Biology, Medicine, Industry, planning, education and so on. As on date there is no other human walk of life, where statistics cannot be applied.

Origin and Growth of Statistics:

The word ' Statistics' and ' Statistical' are all derived from the Latin word Status, means a political state. The theory of statistics as a distinct branch of scientific method is of comparatively recent growth. Research particularly into the mathematical theory of statistics is rapidly proceeding and fresh discoveries are being made all over the world.

Meaning of Statistics:

Statistics is concerned with scientific methods for collecting, organising, summarising, presenting and analysing data as well as deriving valid conclusions and making reasonable decisions on the basis of this analysis. Statistics is concerned with the systematic collection of numerical data and its interpretation.

The word ' statistic' is used to refer to

1. Numerical facts, such as the number of people living in particular area.

2. The study of ways of collecting, analysing and interpreting the facts.

Definition of Statistics (Croxton and Cowden)

Statistics may be defined as the science of collection, presentation analysis and interpretation of numerical data from the logical analysis. It is clear that the definition of statistics by Croxton and Cowden is the most scientific and realistic one.

Thus Statistics is the science of collecting, organizing, analyzing, and interpreting data in order to make decisions

Type of Statistics

1. *Descriptive Statistics* **consists of methods for organizing, displaying, and describing data by using tables, graphs, and summary measures.**

2. *Inferential Statistics* consists of methods that use sample results to help make decisions or predictions about a population.

Limitations of statistics

1. Statistics does not study individuals:
2. Statistical laws are not exact
3. Statistics table may be misused
4. Statistics is only, one of the methods of studying a problem

Population and Sample

A *population or Universe* consists of all elements – individuals, items, or objects – whose characteristics are being studied. The population that is being studied is also called the *target population*. **In other** words, the population is a complete set of all possible observations of the type which is to be investigated.

- The individuals constituting a population is called the unit of the population

Finite population and infinite population

A population is said to be finite if it consists of finite number of units. Number of workers in a factory, production of articles in a particular day for a company are examples of finite population.

The total number of units in a population is called population size.

A population is said to be infinite if it has infinite number of units. For example the number of stars in the sky, the number of people seeing the Television programmes etc.,

A portion of the population selected for study is referred to as a *__sample__*. In other words the sample is the representative part of the population from which information is collected.

Census and Sampling Methods

The process of collecting required information from the unit is called *__enumeration.__*

If information is collected from each and every unit of the population We call this a *__Complete enumeration, or census__*

If information is collected only from reprehensive part of the population it is called *__sampling__*.

Merits and limitations of Census method

Merits

1. The data are collected from each and every item of the population
2. The results are more accurate and reliable, because every item of the universe is required.
3. Intensive study is possible.
4. The data collected may be used for various surveys, analyses etc.

Limitations

1. It requires a large number of enumerators and it is a costly method
2. It requires more money, labour, time energy etc.
3. It is not possible in some circumstances where the universe is infinite

The following are some advantage of sampling when compared to census.

1. It is less expensive and less time consuming as information has to be collected from a lesser number of units only.
2. As the number of enumerators required is less, more efficient and better trained personal can be employed.
3. More sophisticated instruments can be used if needed for collecting data.

4. If the population is infinite or non existent sampling is the only method that can be adopted.

5. If destructive tests are involved in the collection of information, sampling alone can be adopted.

COLLECTION OF DATA

Preliminary considerations

The first step in any statistical investigation is the formulation of the problem under consideration as precisely as possible. Only then the investigator can have a clear idea of the data to be collected. If the formulation of the problem is perfect or faculty, the idea collected may be irrelevant or inadequate.

Collection of data may be done in two different ways, **primary and secondary data.** Data collected by the investigator for the purpose of the investigation at hand is called **primary data.** That is the **Primary data** is the one, which is collected by the investigator himself for the purpose of a specific inquiry or study. Such data is original in character and is generated by survey conducted by individuals or research institution or any organisation.

The data that collected by others for some other purpose and used by the investigator is called **secondary data.** Secondary data are those data which have been already collected and analysed by some earlier agency for its own use; and later the same data are used by the investigator.

Sources of secondary data

1. Government publications like censes report and bulletins of various departments.
2. Office records of municipalities, panchayaths, village offices, registration offices, employment offices etc.

3. Publications of research institutions.

4. Research journals

5. Report of enquiry commissions

6. Publications of institutions like banks, companies etc.

7. Publications of international organizations like UNO,ILO,WHO etc.

Secondary data should be accepted only after very careful scrutiny. The following are some important points to be considered.

1. The person who collected the data

Before accepting secondary data it should be verified whether it was   collected by experts in the field and people having no special bias or personal interest.

2 Purpose for which the data was collected

Difference in purpose may bring about difference in stress. So only data collected for a similar purpose can be accepted for the study at hand.

3 Definition of terms used

The same term may be used in one sense by the person who collected the data and the present investigator may be using it in an entirely different sense.

4 Degree of accuracy

A certain degree of accuracy may be required for the present purpose. So before accepting the information supplied by the secondary source, its degree of accuracy should be examined.

5 Time at which the data was collected

Time lay is an important factor which affects the acceptability of the data

6 Geographical region from which the data was collected.

The investigator may be interested in a particular geographical region and if the data available was collected from some other region it will be entirely useless for him.

7 Details contained in the data

Data supplied by a secondary source may or may not contain all the details required for the present investigation. Only after ascertaining whether it contains all the necessary details a decision regarding its acceptability can be made.

Collection of primary data

The primary data can be collected by the following five methods.
1. Direct personal interviews.
2. Mailed questionnaire method.
3. Sending enumerators to the informants.
4. Indirect investigation
5. Information from correspondents.

1) Direct personal investigation

In this method of collecting data the investigator himself collect information from the unit selected for enumeration.

Advantages

1. The informants are likely to show more interest in giving information as the person approaching them is much more respectable than an ordinary enumerators.

2. The quality of the information collected will be much better as the investigator as the better knowledge of the implication of the question and he will be a position to clear the doubts of the informants.

3. Some useful supplementary information which may be helpful at the analysis stage may also be collected.

4. The questions can be asked more tactfully.

Disadvantages

1. When information is to be collected from a very large number of units, or from a very large area, it is difficult to adopt this method.

2. The personal prejudices of the investigator are likely to influence the data.

3. The time required may be much larger.

2) Sending questionnaire through post and collecting replies also through post.

In this method questionnaire are send to the informant together with stamped covers for sending back the filled up questionnaires. A covering letter accompanying the questionnaire explains the purpose of the investigation and the importance of correct informations and requests the informants to fill in the blank spaces provided and to return the form within a specified time. This method is appropriate in those cases where the informants are literates and are spread over a wide area.

Advantages

1. This is the cheapest method when the informants are spread over a large geographical area.
2. The number of workers required for the collection of data can be minimized in this method.
3. The time required for the collection work will also be minimized.

Disadvantages

1. This method is succeeding only when the informants are sufficiently educated.
2. Unless the investigator has some compelling power like backing the response is likely to be poor.
3. The information supplied may be incomplete or incorrect.
4. It is difficult to verify the correctness of the information's furnished by the respondents.

3) Sending enumerators to the informants

In this method the necessary number of people is given intensive training and they are sending to the informants with the questionnaire. They interview the informants and fill up the forms.

Advantages

1. This method can be adopted even when the informants are illiterate.
2. Case of non response will be very small.
3. The information received will be more or less complete and correct, as it is collected by trained personnel.
4. The time schedule can be kept up.

Disadvantages

1. Of all the method this is the most costly one.
2. The success of the method depends on the training given to the enumerators and their sincerity as well as on whether their work is efficiently supervised.

This is the most commonly used method in large scale data collection.

4) Indirect investigation

In this method the investigators collects information by contacting third parties. This method is adopted when the informants are not inclined to give information or are likely to give wrong information.

5) Using the services of correspondents

In this method the investigator appoints agents in different places and they collect information and send it to the investigator. Information's to Newspapers and some departments of Government come by this method. The advantage of this method is that it is cheap and appropriate for extensive investigations. But it may not ensure accurate results because the correspondents are likely to be negligent, prejudiced and biased. This method is adopted in those cases where information's are to be collected periodically from a wide area for a long time.

Questionnaire

A questionnaire is a list of questions used for the collection of information in an investigation. Forms called schedules are usually prepared with these questions printed or written on the left side of the paper and space left for answers on the right side. Questionnaire is necessary for both census and sample studies. The only difference is that for sample studies the questionnaire can be more elaborate and complex as information is to be collected only from a small number of units and better trained personnel can be employed for enumeration purpose.

Characteristics of a questionnaire
1. The questionnaire should be capable of electing all the required information.
2. The number of questionnaire should be kept in minimum.
3. The questions should be arranged in a logical order.
4. The questions should be short, simple and unambiguous.
5. Questions which require 'Yes' or 'No' answer or one word answers should be preferred.
6. Questions which are likely to offend the feelings of the informant should be avoided.
7. Questions which require elaborate calculations or reference to records should be minimized.

8. Some very personal questions should be avoided.

9. The meaning of technical terms used in the questionnaire and explanatory notes wherever necessary should be given as foot notes.

Editing of the data

The following are some important points in editing of the data.

1) completeness

Each schedule is to be carefully examined to see whether it is complete in every respect. If some questions are left unanswered, the investigator should get those answers by contacting the informant personally or by post. If his answers are not obtained even after repeated attempts, the column should be filled by 'no' answer.

2) consistency

To examining whether the entries in the schedule are consistent or not is another important points to be remembered while editing the data.

3) Accuracy and homogeneity

The figures given in the schedules should be examined for arithmetic accuracy.

To examined whether all informants have understood the questions in the same sense

Scaling Techniques

Scaling means conversion of qualitative data into quantitative data and then measuring them. The most widely classifications of measurement scale are

1 Nominal scale

2 Ordinal scales

3 Interval scale

4 Ratio scale

1. Nominal scale

Nominal scale is a system of assigning number symbols to events in order to label them. Mode can be used as measure of central tendency, and chi square test is used for significance testing.

2 Ordinal scales

The ordinal scale places events in order. It is commonly used in research relating to qualitative characteristic. Rank order represents ordinal scale.

3 Interview scale

In interval scale the intervals are adjusted in terms of some rule that has been established as a basis for making the unit equal.

4 Ratio scale

Ratio scales have an absolute zero of measurement. Ratio scale represents the actual amount of variables.

Pilot survey

A pilot survey is a small scale study of the main study. It covers all process of research. For pilot survey small sample is taken from the proportion and questionnaire is applied on it. The main important of pilot survey is that, the researcher can know about the feasibility of the main study.