

STATISTICS

The word 'statistics' comes from the Italian word '*statista*' or the German word '*statistik*' each of which means a Political State. It was first used by Professor **Gottfried Achenwell** (1917-1772). **Biostatistics** is the term used when tools of statistics are applied to the data derived from biological sciences. Biostatistics is also called **biometry**.

Prof.A.L.Bowley defines, "Statistics may rightly be called as science of averages. According to **Croxton and Cowden**, "Statistics may be defined as the collection, presentation, analysis and interpretation of numerical data. According to this definition, there are four stages:

1. **Collection of data:** Collection of data is the first step in a statistical investigation. There are different methods of collection of data such as census, sampling, primary, secondary etc. The investigator should make use of the correct method.
2. **Presentation of data:** The collected data are ready for presentation. Data is presented in an orderly manner to facilitate statistical analysis. The collected data may be presented in tabular or diagrammatic or graphic form.
3. **Analysis of data:** The purpose of analysing data is to find out information useful for decision-making. The most commonly used methods of statistical analysis are measure of central tendency, measure of variation, correlation, regression etc.
4. **Interpretation of data:** The last stage in statistical investigation is interpretation, i.e., concluding the data collected and analysed. The interpretation for data is a difficult task and necessitates a high degree of skill and experience. Correct interpretation will lead to a valid conclusion of the study.

Functions of Statistics:

1. To simplify complex facts.
2. To provide comparative study.
3. To study the relationship between different facts.
4. To formulate policies in different fields.
5. To test a hypothesis.
6. To forecast.

Collection of Data

The collection of data is the first and most important step in the statistical study. Since all calculations and results depend upon the data collected. If there is any defect or deficiency in the collection of data then it will affect the whole investigation. Thus, it is most important for the one who collects the data to be *alert, conscious, thoughtful, firm* and *patient* so that the data collected by him is faultless and dependable. So before collecting data, the investigator should know the plan of his work and the method of collecting data.

Types of data:

Statistical data are of two types (i) Primary data and (ii) Secondary data

- i) **Primary data:** The data collected by the investigator first time from his experimental studies is called primary data.
- ii) **Secondary data:** The data obtained from some secondary sources such as journals, magazines, newspapers etc. are known as secondary data.

Difference between primary and secondary data:

1. The primary data are new and are **collected afresh**. The secondary data are those which were previously collected and published by some investigator and are being simply used by the present investigator.
2. The primary data is collected according to the **plan of the investigator**, so it requires much money, labour and time to collect them whereas the secondary data is always readily available and can be borrowed easily by the present investigator for his purpose.
3. The primary data is collected according to the **requirement of the problem** under consideration while the secondary data was collected for another purpose and is used according to the requirement of the present investigator.

Secondary Data:

Secondary data are those data, which have been already collected and analysed by some earlier agency for its use, and later the same data are used by a different agency. The sources of secondary data can broadly be classified under two heads: (1) Published sources and (2) Unpublished sources.

1. Published sources:

- a) Reports and official publications of International bodies such as the World Bank, the United Nations, Official publications of Central and State Governments and Reports of various committees and commissions appointed by the Government.
- b) Semi-official publications of various local bodies such as Municipal Corporations and District Boards.
- c) Publications of Research institutions like the Indian Statistical Institution, and Indian Council of Agricultural Research (ICAR).
- d) Journals and newspapers.

2. Unpublished sources:

These are various sources of unpublished data such as records maintained by various Government and private offices, studies made by research institutions, scholars, etc., such sources can be used where necessary.

Methods of Collecting Primary Data

There are five methods of collecting primary data as follows:

- a. Direct Personal interviews
- b. Indirect oral (Investigation) interviews
- c. Information from correspondents
- d. Mailed questionnaire method
- e. Schedule sent through enumerators

a) Direct Personal interviews:

In this method, the investigator collects the data according to his requirements. In this method, the investigator must be present on the **spot and collect** the data himself directly. This method is very important as the investigator himself inspects the situation. So, the investigator avoids unnecessary collection of data too.

Merits:

1. There is original and accurate data.
2. This is most suitable if the area of investigation is limited.
3. Response is more encouraging as most people are willing to supply information when approached personally.
4. Checking of data is automatically done along with the collection.

Demerits:

1. It is not applicable in an extensive enquiry.
2. It requires too much labour and time.
3. The prejudice of the investigator influences the data to a great extent.
4. It is not scientific.

b) Indirect oral (investigation) interviews:

This method is happening when the person related to the facts feels reluctant (i.e., not willing) to give information. In this method, the investigator makes enquiries from other persons who are in touch with the facts. Commission and enquiry committees generally use this method. This method is adopted only if: (i) It is difficult for the investigator to contact the related persons or on making contact he refuses to give information (ii) it is most suitable when the area of investigation is vast (iii) when it is felt that the investigation should be kept secret from the related person.

Merits:

1. It is very useful in extensive enquiry.
2. It is most inexpensive since very little time, labour and money is required.
3. It is free from the prejudices of the investigator.

Demerits:

1. The data collected depends on information received indirectly, so the results are likely to be inaccurate and unreliable.
2. The information received is not free from the ignorance of the informers.

c) Information from correspondents:

In this method the investigator instead of contacting the related persons, directly and indirectly, engages local people or correspondents. The correspondents obtain information in their way and communicate it to the investigator. Newspaper and magazines where the accuracy in investigations is not so important generally adopt this method.

Merits:

1. This method is applicable where the field of extensive enquiry and places of enquiry are scattered.
2. This method is economical as it saves time, labour and money.
3. Speedy information is possible.

Demerits:

1. The information may be biased.
2. Uniformity cannot be maintained.
3. The delay in sending information often makes the investigator ineffective.

d) Mailed questionnaire method:

Under this method, a list of questions about the survey is prepared and sent to the various informants by post. The questionnaire contains questions and provides space for answers. A request is made to the informants through a cover to fill up the questionnaire and send it back within a specified time.

Merits:

1. The mailed questionnaire method is the most economical.
2. It can be widely used when the area of investigation is large.
3. Error in the investigation is very small because the information is obtained directly from the respondents.

Demerits:

1. The greatest disadvantage of this method is the willingness of persons to reply to all or in full the questionnaire due to carelessness or fear of exposure.
2. The method is suitable only for literate people.
3. Asking supplementary questions is not possible.

e) Schedule sent through enumerators:

It is the most widely used method of collection of primary data. Several enumerators are selected and trained. The enumerators contact the persons to get replies to the questions contained in a schedule and fill them in their handwriting in the questionnaire form. The essential difference between the mailed questionnaire method and this method is that whereas in the former the questionnaire is sent to the informants by post, in the latter the enumerators carry the schedule personally to the informants.

Merits:

1. This method is very useful in extensive enquiries.
2. It yields reliable and accurate results because the enumerators are educated and trained.
3. Even if the respondents are illiterate, this technique can be widely used.

Demerits:

1. This is a very costly method.
2. Personal bias of the enumerators may lead to false conclusions.