Introduction to Statistics

1 Introduction

In the modern world of information and communication technology, the importance of statistics is very well recognized 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 of today, there is no other human walk of life, where statistics cannot be applied.

Statistics is concerned with the scientific method of collecting, organizing, summarizing, presenting and analyzing statistical information (data) as well as drawing valid conclusion on the basis of such analysis. It could be simply defined as the "science of data". Thus, statistics uses facts or numerical data, assembled, classified and tabulated so as to present significant information about a given subject. Statistic is a science of understanding data and making decisions in the face of randomness.

The study of statistics is therefore essential for sound reasoning, precise judgment and objective decision in the face of up to date accurate and reliable data. Thus many researchers, educationalists, business men and government agencies at the national, state or local levels rely on data to answer operations and programs. Statistics is usually divided into two categories namely: Descriptive statistics and inferential statistics.

1.1 Descriptive Statistics

This is the act of summarizing and given a descriptive account of numerical information in form of reports, charts and diagrams. The goal of descriptive statistics is to gain information from collected data. It begins with collection of data by either counting or measurement in an inquiry. It involves the summary of specific aspect of the data, such as averages values and measure of dispersion (spread). Suitable graphs, diagrams and charts are then used to gain understanding and clear interpretation of the phenomenon under investigation keeping firmly in mind where the data comes from. Normally, a descriptive statistics should:

- be single valued
- be algebraically tractable
- consider every observed value.

1.2 Inferential Statistics

This is the act of making deductive statement about a population from the quantities computed from its representative sample. It is a process of making inference or generalizing about the population under certain conditions and assumptions. Statistical inference involves the processes of estimation of parameters and hypothesis testing. However, this concept is not in the context of this course. We will learn those on MA 3102-Applied Statistics course.

Classification of Data

Data can be described as a mass of unprocessed information obtained from measurement of counting of a characteristics or phenomenon. They are raw facts that have to be processed in numerical form they are called quantitative data. But when data are not presented in numerical form, they are called qualitative data. E.g.: status, sex, religion, etc.

Definitions

The quantities measured on outcomes of a random phenomenon in a study are called *random variables*, and a particular outcome is called an *observation*. Several observations are collectively known as *data*. The collection of all possible outcomes is called the *population*. In practice, we cannot usually observe the whole population. Instead we observe a subset of the population, known as a *sample*. In order to ensure that the sample we take is representative of the whole population, we usually take a *random sample* in which all members of the population are *equally likely* to be selected for inclusion in the sample.

For example, if we are interested in conducting a survey of the amount of physical exercise undertaken by the general public, surveying people entering and leaving a gymnasium would provide a *biased sample* of the population, and the results obtained would not generalise to the population at large.

Variables are either *qualitative* or *quantitative*. Qualitative variables have non-numeric outcomes, with no natural ordering. For example, gender, disease status, and type of car are all qualitative variables. Quantitative variables have numeric outcomes. For example, survival time, height, age, number of children, and number of faults are all quantitative variables.

Quantitative variables can be *discrete* or *continuous*. Discrete random variables have outcomes which can take only a countable number of possible values. These possible values are usually taken to be integers, but don't have to be. For example, number of children and number of faults are discrete random variables which take only integer values, but your score in a quiz where "half" marks are awarded is a discrete quantitative random variable which can take on non-integer values. Continuous random variables can take any value over some

continuous scale. For example, survival time and height are continuous random variables. Often, continuous random variables are rounded to the nearest integer, but the are still considered to be continuous variables if there is an underlying continuous scale. Age is a good example of this.

2 Sources of Statistical Data

There are two types of data which are Primary and Secondary data.

2.1 Primary Data

These are data generated by first hand or data obtained directly from respondents by personal interview, questionnaire, measurements or observation. Statistical data can be obtained from:

1. Census – complete enumeration of all the unit of the population
2. Surveys – the study of representative part of a population
3. Experimentation – observation from experiment carried out in laboratories and research center
4. Administrative process e.g. Record of births and deaths.

Advantages

- Comprises of actual data needed
- It is more reliable with clarity
- Comprises a more detail information

Disadvantages

- Cost of data collection is high
- Time consuming
- There may larger range of non response

2.2 Secondary data

These are data obtained from publication, newspapers, and annual reports. They are usually summarized data used for purpose other than the intended one. These could be obtain from the following:

1. Publication e.g. extract from publications
2. Research/Media organization
3. Educational institutions.

Advantages

- The outcome is timely
- The information gathered more quickly
- It is less expensive to gather.

Disadvantages

- Most time information are suppressed when working with secondary data
- The information may not be reliable

3 Methods of Data Collection

There are various methods we can use to collect data. The method used depends on the problem and type of data to be collected. Some of these methods include:

1. Direct observation
2. Interviewing
3. Questionnaire
4. Abstraction from published statistics.

Direct Observation

Observational methods are used mostly in scientific inquiry where data are observed directly from controlled experiment. It is used more in the natural sciences through laboratory works than in social sciences. But this is very useful studying small communities and institutions.

Interviewing

In this method, the person collecting the data is called the interviewer goes to ask the person (interviewee) direct questions. The interviewer has to go to the interviewees personally to collect the information required verbally. This makes it different from the next method called questionnaire method.

Questionnaire

A set of questions or statement is assembled to get information on a variable (or a set of variable). The entire package of questions or statement is called a questionnaire. Human beings usually are required to respond to the questions or statements on the questionnaire. Copies of the questionnaire can be administered personally by its user or sent to people by post. Both interviewing and questionnaire methods are used in the social sciences where human population is mostly involved.

Abstractions from the Published Statistics

These are pieces of data (information) found in published materials such as figures related to population or accident figures. This method of collecting data could be useful as preliminary to other methods.

Other methods includes: Telephone method, Document/Report method, Mail or Postal questionnaire, On-line interview method, etc.