1. INTRODUCTION

1.1 What is Statistics and what is its importance?

Statistics is the science of collecting, organizing, presenting, analyzing and interpreting data for the purpose of assisting in making more effective decision under uncertainty.

Today, statistics has become an important tool in the work of many academic disciplines such as medicine, psychology, education, sociology, geography, engineering and physics, just to name a few. Statistics is also important in many aspects of society such as business, industry and government. Statisticians use quantitative abilities, statistical knowledge, and communication skills to work on many challenging problems such as;

- Human Brain Mapping and Neuroimaging Researches which are new multidisciplinary fields that demand novel statistical models and analytic tools for quantifying disease, temporal, age, genotypic and neurophysiological variation in the brain anatomy and function.
- Designing and analyzing studies to determine if **new drugs** and medical devices are safe and effective (at a pharmaceutical company, medical research center, or the Food and Drug Administration)
- Analyzing **consumer demand** for products and services (at a consumer marketing firm, corporation, or consulting firm)
- Estimating the **unemployment rate** in Sri Lanka (at the Department of Census & Statistics)
- Statistical Modeling of Network load, bandwidth, saturation and information flow (e.g., the Internet)

Because of the increasing use of statistics in so many areas of our lives, it has become very desirable to understand and practice statistical thinking. This is important even if you do not use statistical methods directly.

Some statistical aspects include regression analysis, analysis of variance, hypothesis testing and probability. Statistics has a wide variety of applications. Business analytics, econometrics, actuarial science, biostatistics, geostatistics, environmental statistics etc- are some of the fields that require the application of statistics.

1.2 Types of statistics

The two main branches of statistics are:

Descriptive statistics- As the name suggests, the descriptive statistics merely describe the data and consist of methods and techniques used in collection, organization, presentation and analysis of data in order to describe the various features and characteristics of such data; Graphical techniques, tabulated techniques, measures of central tendency and dispersion are there.

Inferential statistics- This branch of statistics involves drawing conclusions from limited information taken on sample basis and testing the reliability of the estimates. It deals with analysis of data, making estimates and drawing conclusions. The analysis usually starts with a hypothesis and the consistency of the data in accordance to the hypothesis is checked. Parameter estimation, Data prediction and Model comparison are also there.

The two general "philosophies" in inferential statistics are **frequentist inference** and **Bayesian inference**. The main difference of these approaches is Frequentists consider parameters are constants while Bayesians consider those as random variables.

Parameter estimation **-** For example, the parameters of a normal distribution are its *mean* and its *standard deviation***.** Suppose, you are interested in studying the height of women in this university**.** A sample can be taken and after measuring the heights of females in a sample, you can estimate the mean and standard deviation of the distribution for all adult females**.**

Model comparison - For example, after building a model to illustrate the relationship between someone's height with genetic components and physical exercises, it can be found that genetic components influence height more than physical exercise.

1.3 Types of variables

There are different ways that variables can be described according to the ways they can be studied, measured, and presented.

1.3.1 Numerical Variables

Numerical variables have values that describe a measurable quantity as a number, like 'how many' or 'how much'**.** Therefore numerical variables are **quantitative variables**. Numerical variables may be further described as either continuous or discrete:

- **Continuous variables** can take a value based on a measurement at any point along a continuum range**.** The value given to an observation for a continuous variable can include values as small as the instrument of measurement allows. Examples of continuous variables include height, time, age, and temperature.

- **Discrete variables** can take a value based on a count from a set of distinct whole values**.** A discrete variable cannot take the value of a fraction between one value and the next closest value. Examples of discrete variables include the number of registered cars, number of business locations, and number of children in a family, all of which measured as whole units (i.e. 1, 2, 3 cars).

1.3.2 Categorical Variables

Categorical variables have values that describe a 'quality' or 'characteristic' of a data unit, like 'what type' or 'which category'**.** Categorical variables fall into mutually exclusive (in one category or in another) and exhaustive (include all possible options) categories**.** Therefore, categorical variables are **qualitative variables** and tend to be represented by a non-numeric value. The data collected for a categorical variable are qualitative data. Sometimes numerical data can be categorized according to the need.

1.4 Levels of Measurement of Variables

Levels of measurement describe how much information the numbers associated with the variable contain. Mathematical operations depend on levels of measurements.
Four levels of measurements are recognized as follows;

- Nominal
- Ordinal
- Interval
- Ratio

1.4.1 Nominal Measures

It is a designation of an observation with a value which is merely a name or label. They cannot be organized in a logical sequence.
Examples: Marital status, Gender

1.4.2 Ordinal Measures

Observations have all the features of nominal measures. They also represent rank order. The categories associated with ordinal variables can be ranked higher or lower than another, but do not necessarily establish a numeric difference between the each category.
Examples: Quality of a product (Excellent, Good, Fair or Poor), Highest education qualification

1.4.3 Interval Measures

Observations have all the features of ordinal measurement while the differences between arbitrary pairs of numbers can be meaningfully compared
Examples: Shoe size. The difference between size 3 and size 4 shoe is equal to the distance between size 7 and size 8 shoe, but size 6 shoe is not equal to 2 * size 3 shoe. Also, size 0 shoe does not mean that there is no shoe, it's simply a shoe with zero size i.e an arbitrary zero point.

1.4.4 Ratio Measures

The numbers have all the features of interval measurements and they also have meaningful ratios between arbitrary pairs of numbers. A ratio scale has an absolute zero (a point where none of the quality being measured exists).
Examples: Weight of a person. Zero weight means that the person has no weight. Also, we can add, subtract, multiply and divide weights at the real scale for comparisons.