

# Chapter 1 – Introduction to Statistics

## 1.1 Introduction

**Statistics** involves the procedures associated with the data collection process, the summarizing and interpretation of data, and the drawing of inferences or conclusions based upon the analysis of the data.

**Descriptive Statistics** uses numerical and/or visual techniques to summarize or describe the data in a clear and effective manner.

The **population** is the entire collection of all individuals or objects of interest.

The **sample** is the portion of the population that is selected for study.

**Inferential Statistics** is the process of using **sample information** to draw inferences or conclusions about the **population**.

A **representative sample** is a sample that has the pertinent characteristics of the population in the same proportion, as they are included in that population.

A **statistic** is a number that describes a characteristic of a sample.

A **parameter** is a number that describes a characteristic of a population.

## 1.2 Why Sample?

**Sampling** is the process of selecting a portion, or sample, of the entire population.

A survey that includes every item or individual of the population is called a **census**.

### **Why Sample?**

Samples are taken when it is either impossible or impractical to examine the entire population. Studying an entire population is impractical when the population is extremely large.

**Sample Size** – The size of a sample, denoted by  $n$ , is the number of data values in the sample.

**Population Size** – The size of a finite population, denoted by  $N$ , is the number of data values in the population.

## Reasons for Sampling

1. A census is impossible
2. Cost
3. Time
4. Accuracy

## 1.3 Sampling Techniques

A **simple random sample** is a sample of data values selected from a population in such a way that every sample of size  $n$  has an equal probability of being selected and every data value of the population has the same chance of being selected for the sample.

Selecting a **non-representative** or **biased** sample from a population can lead to invalid inferences about the population.

## 1.4 Uses of Statistics

Statistics are used in many different fields such as economics, finance, psychology, sociology, education, and the physical sciences just to name a few.

Extensive data collection and distribution activities are performed by the federal and other governmental and private agencies in areas such as education, employment, health, crime, prices, housing, medical care, manufacturing, agriculture, construction, transportation, etc.

The development of the computer has led to a revolution in the area of data collection and analysis. Statistical analysis can now be applied to vast amounts of data accurately and quickly. Such diverse problems as weather forecasting, economic stabilization, and disease control are today being solved using statistical analysis.

## 1.5 Misuses of Statistics

The following techniques illustrate how information can be presented in a misleading fashion:

1. Misleading graphs - can be very misleading based on the scale.
2. Non-representative samples (Biased) - purposely choosing a sample that favors your desired outcome.
  - ex: Surveying people with a personal interest, such as their livelihood is dependent on a certain product
  - ex: Surveying many people but only using the ones who say what you are looking for.
3. Inappropriate comparisons
  - ex: Publishing that destination X was a more popular vacation spot for the month of August than destination Y, when destination Y is in its off-season during the month of August and destination X is not.
4. The omission of variation about an average