

## Chapter One: What is Statistics?

When you completed this chapter, you will be able to:

- ✓ understand the board overview of the subject of statistics;
- ✓ distinguish descriptive and inferential statistics; and
- ✓ classify types of data and data level.

*Reference(s): Mason Chapter 1*

*Exercise(s): Seminars 1 and 2, Mason Chapter1 Exercises 1-7, 15-18*

---

### What is Statistics?

Statistics is the science that deals with collecting, organizing, presenting, analyzing, and interpreting data to assist in making effective decisions.

### Reasons for studying statistics

- Numerical information is everywhere;
- Statistical techniques help in making decision, and understanding others' decision; and
- Decision making always involves data.

### Types of possible uses of Statistics

#### 1. Descriptive Statistics

Methods of arranging, organizing, summarizing, and presenting data. Data presentation could be in the form of tables, charts and statistical measures.

#### 2. Inferential Statistics

Problems involving statistical inference arise when a statistician takes a *sample* from a *population* and wishes to make statements about the population's characteristics from the information contained the sample

SAMPLE    $\hat{P}$    STATISTICS    $\hat{P}$    POPULATION

Prediction, or estimation of the relationships between two or more variables for deciding among alternative courses of action when it is impossible to predict exactly what the consequence of each of these will be.

**Population :** Population is a set or collection of *items of interest*. Numerical characteristics of populations are referred to as *Population Parameters*, or simply *Parameters*.

**Sample :** Sample is a *subset* of a population. Numerical characteristics of samples are referred to as *Sample Statistics*, or simply *Statistics*.

e.g.1 Suppose we are interested in the mean age of the students in the class. The age of all students is the *population*, and the mean age is the *population parameters*. If 10 students are selected from the class, the age of the 10 students is the *sample*, and the mean age of the 10 students is the *sample statistics*.



YOU DON'T HAVE TO EAT THE WHOLE  
OX TO KNOW THAT THE MEAT IS TOUGH

By Samuel Johnson

## Data, Data Types, and Data Sources

Data Set is a collection of data. Data Sets or Data Files are often collections of data that are organized to facilitate data analysis. Data are collected about entities known as *items*. An item may be a person, a company, a product, or any other entity of interest. Items are variously referred to in the practice of statistics as elements, objects, or units.

**Variable :** Variable is a characteristics of interest for a set of items. A variable takes on different values for different items.

e.g.2 if  $X$  is used to denoted to represent the age of students, then  $X$  is equal to 17 for representing John's age, or  $X$  is equal to 16 for representing Mary's age, and etc.

**Data Value :** Data value is a measurement for a variable, and many data sets include several variables. The data values for a single item from all of the variables make up an observation. An observation is variously referred to in the practice of statistics as a *case*, a *row*, or a *record*.

e.g.3 A record, called CHARACTER, which includes several characteristics of an item, such as  $X$  denotes the age of the item,  $Y$  denotes the weight of the item, and  $Z$  denotes the height of the item. All these  $X$ ,  $Y$  and  $Z$  together form a record for the item. Different Records collect the information of age, weight and height for different items.

## Data Types

### 1. Qualitative Data

Qualitative or Categorical Data are data values, each of which can be classified into single category that belongs to a set of categories. It is **nonnumeric**.

e.g. 4 Comment on a movie: [*Excellent Good Fair Bad Very Bad*]

Selection of transportation means: [*Bus MTR Train Mini-bus Taxi*]

### 2. Quantitative Data

Quantitative data are data values that are measured on a **numerical** scale.

e.g. 5 Age of IVE students (in years old): [*0-9 10-19 20-29 30-39*]

Life of a battery (in months): [*0 to less than 2 2 to less than 4 4 or above*]

## Levels of Measurement

### 1. Nominal Level Data

It is the lowest level of data measurement that classifies item into one category taken from a set of categories. The observations can only be **counted** or **classified**. There is **no natural order**.

e.g.6 [*Man Woman Boy Girl*].

### 2. Ordinal Level Data

It is higher level than Nominal. Data are classified into categories that can be **rank** or **ordered** for the variable.

e.g.7 [*Superior Good Average Poor Inferior*]

### 3. Interval Level Data

It is the next highest level. It includes all characteristics of the ordinal level. It is **quantitative** data and the difference between measurements is a **meaningful quantity** and **constant**.

e.g.8 Suppose that the date of 3 consecutive days, Mon, Tue and Wed are 5<sup>th</sup> Jan 99, 6<sup>th</sup> Jan 99 and 7<sup>th</sup> Jan 99 respectively. It is quantitative data and can be ranked such as Wed is one day after Tue, and Tue is one day after Mon. The one-day differences are the same regardless of their position. Date is interval level data.

#### 4. Ratio Level Data

It is the highest level. It includes all characteristics of the interval level. Moreover, data has a *meaningful zero point* and the ratio of two data values is *meaningful*.

e.g.9 A person who is 76 inches tall is twice as tall as someone who is 38 inches tall. Zero inch means nothing, so the zero point of the height measure is meaningful. Height is ratio level data.

What is the level of temperature measurement?

Temperature is a *trickier case. Temperature scale in degree Celsius is interval but not ratio. Since the "0" demarcation is arbitrary, not real. No one should say that 76 degrees Celsius is twice as hot as 38 degrees Celsius. But when measured from absolute zero, in Kelvin scale, temperature is on ratio scale.*

#### Statistical Studies

Statistical Studies can be classified into two broad categories as *Observational Studies* and *Experimental Studies*.

In an observational study, the investigator examines variables of interest by using *observed* or *historical* data. The investigator *does not directly control*, or determine which subjects or items receive treatments that are thought to affect the variables of interest in the study.

In an experimental study, the investigator *directly controls* or determines which subjects or experimental items or materials receive treatments that are thought to affect variables of interest.