Statistics: is the study of how best to

a) Collect data
b) Summarize or describe data
c) Draw conclusions or inference based on data

When we have a research problem (project), the science of statistics help us to answer questions like

- How to collect the data
- How much data should we collect
- How to summarize the data
- How to draw conclusions based on the data
- How to assess the strength of our conclusions.

Statistical methods are classified into two broad categories:

- Descriptive Statistics
- Inferential Statistics

   - Estimation
   - Point Estimation
   - Confidence Interval Estimation
   - Hypothesis testing

You will learn about each of these methods as we proceed with the course.

Note that statistics is not just the study of collecting and summarizing data and making conclusions based on data. It is also a science that deals with designing experiments. This will not be covered in this course.
What is Data

Data refers to numerical or non-numerical facts or information.

**Example 1:** Height of your family members

<table>
<thead>
<tr>
<th>Member</th>
<th>Height (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.5</td>
</tr>
<tr>
<td>2</td>
<td>2.0</td>
</tr>
<tr>
<td>3</td>
<td>1.6</td>
</tr>
<tr>
<td>4</td>
<td>1.4</td>
</tr>
<tr>
<td>5</td>
<td>1.7</td>
</tr>
</tbody>
</table>

These numbers collectively are called **data**.

**Example 2:** Gender of your family members

<table>
<thead>
<tr>
<th>Member</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Male</td>
</tr>
<tr>
<td>2</td>
<td>Female</td>
</tr>
<tr>
<td>3</td>
<td>Male</td>
</tr>
<tr>
<td>4</td>
<td>Female</td>
</tr>
<tr>
<td>5</td>
<td>Female</td>
</tr>
</tbody>
</table>

These are collectively referred to as **data**.

**A data point (An Observation)**

A data point or an observation is the observed value of a variable.

**Example 1:** In Example 1 and 2 above there are five observations or five data points.

**Example 2:** Weight of your family members

<table>
<thead>
<tr>
<th>Member</th>
<th>Weight (Lbm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Julia</td>
<td>120</td>
</tr>
<tr>
<td>Steve</td>
<td>140</td>
</tr>
<tr>
<td>Chris</td>
<td>150</td>
</tr>
<tr>
<td>Pat</td>
<td>135</td>
</tr>
<tr>
<td>Randi</td>
<td>105</td>
</tr>
<tr>
<td>Gino</td>
<td>210</td>
</tr>
<tr>
<td>Jennifer</td>
<td>185</td>
</tr>
</tbody>
</table>

There are seven observations or seven data points.

The first data point is the weight for Julia = 120 Lbm

The 3rd observation is the weight for Chris = 150 Lbm
**Variable**

A Variable is an observable or measurable characteristic of a subject. The subject (unit) could be a place, a person, a plant, an animal, a thing.

**Examples of Variable:**
- Height
- Weight
- Age
- Gender
- Eye color
- GPA

* Values of a Variable varies from subject to subject.

**Example:** Suppose I measure the weight of new born babies (lbm). The data are given below:

<table>
<thead>
<tr>
<th>baby</th>
<th>weight (lbm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.1</td>
</tr>
<tr>
<td>2</td>
<td>8.3</td>
</tr>
<tr>
<td>3</td>
<td>9.2</td>
</tr>
<tr>
<td>4</td>
<td>7.6</td>
</tr>
<tr>
<td>5</td>
<td>5.4</td>
</tr>
</tbody>
</table>

What is the variable being measured?  
What is the subject?  
What is the first observation (data point)?  
How many observations do we have?

**Note:** In order to find out what the subject is - ask yourself "what is the thing I get the observations from?" The answer to this question is the subject.

**Example:** Suppose I record the gender of few students from this class. The data are given below:

<table>
<thead>
<tr>
<th>Student</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Male</td>
<td>Female</td>
</tr>
</tbody>
</table>

What is the variable?  
What is the subject?  
Do the data contain (a) numerical or (b) non-numerical information?  
What is the 2nd data point?
Variable types

In statistics, variables are classified into two broad categories:

**Numerical Variables**
- Discrete Variables
- Continuous Variables

**Categorical Variables**
- Ordinal Variables (ranked, ordered)
- Nominal Variables (unranked, categorical)

**1. Numeric Variable:** is a variable whose measurement represents some quantity (amount). A numeric variable is a variable whose possible values are numbers, but the numbers should measure some meaningful quantity.

**Example**

- **Weight** (200, 150, 185) in kg
- **Height** (2m, 1.5m, 1.8m)
- **The number of people in a household** (2, 5, 3, 6)
- **The number of insects per square meter**

**Note (think about these)**

a) Social security number of a student (SSN) is a variable. It satisfies the defn. of a variable. Do you think it is a numeric variable? Why or why not?

b) Phone number of a student. Is this a numeric variable? Why or why not?
Tip: An easy way of telling whether a variable is numeric or not is "try to add two possible values of the variable - if the sum makes sense -> then it is a numeric variable".

2. Categorical Variable: is a variable whose possible values are categories; the possible values of a categorical variable do not represent meaningful quantity.

Example:
- Gender (Male, Female)
- Eye color (brown, black, etc)
- Blood type (A, B, O, AB)
- Your favorite football team
- Phone number
- SSN
- Student ID

Note: Some categorical variables are labelled with numbers instead of categories (words). Example: phone number, SSN, student ID. What is an easy way to tell that these variables (although labelled with numbers) are not numeric variables?

Types of Numeric Variables

Numeric variables are further divided into

a. Discrete numeric variables
b. Continuous numeric variables

A. Discrete variables: are numeric variables that can only take a finite number of values. These are numeric variables whose possible values are integers.

Example:
- Number of Students in Stat 211 class: (40, 65, 80, 105)
- Number of cars in a parking lot: (2, 10, 15, 20)
- Number of trees per hectare: (150, 180, 200)
- Number of traffic accidents: (10, 12, 30)
b. **Continuous Variables**: are numeric variables that can take any possible value within a range.

*Example*: Height, weight, time you wait for PRT

---

**Types of Categorical Variables**

Categorical variables are further divided into two groups:

a. **Ordinal Variables**

b. **Nominal Variables**

---

**Ordinal Variable**: is a categorical variable whose categories imply some order or relative position.

*Example*: Letter grade (A, B, C, D, F), smoking status (heavy smoker, mild smoker, none smoker)

**Ordinal Variables** = are also called **Ranked Categorical Variables** or **Ordered Categorical Variables**.

**Nominal Variable**: is a categorical variable whose categories do not have a natural order.

*Example*: Gender (male, female), hair color, political affiliation, religion
**Population and Sample**

*These are the two most important terminologies in this course.*

**Definition:**

The **population** is the entire collection of individuals or measurements about which information is desired.

The **sample** is a part of the population that we actually examine in order to gather information.

An experiment, observational study, or a survey selects a sample from the **population** of all individuals about which we desire information. We base conclusions about the population on data from the sample.

**Example 1:** Suppose you are interested to find out the percentage of WVU students who enjoy playing basketball. If you randomly select 300 students for your study, these 300 students would be the **sample**. The **population** would be all of the students who attend WVU.

**Example 2:** Suppose you are interested to find the average household income of residents in Morgantown. You randomly select 500 residents.

Identify the population and sample in this study.

The **population**

The **sample**

For more information on population and sample, read the materials on "Practice Exam II" part B. => This located on my web page.
Summary

From Chapter 1, you should know the following concepts and terminologies:

1. Data
2. Data point = Observation
3. Variable
4. Numerical Variables
5. Categorical Variables
6. Discrete Variables
7. Continuous Variables
8. Ordinal Variables
9. Nominal Variables
10. Quantitative Variables
11. Qualitative Variables
12. Population
13. Sample