Statistics

Definitions & Terminologies

What is **Statistics**?

Statistics is obtaining data, organizing, summarizing, analyzing, making interpretation of the data, and drawing conclusions based on our findings.
We can split statistics into two branches, descriptive and inferential.

What is **Descriptive Statistics**?

Descriptive Statistics is the term given to the analysis of data that helps describe, show or summarize data in a meaningful way.

What is **Inferential Statistics**?

Inferential Statistics are techniques and methods that allow us to use samples to make generalizations about the populations from which the samples were drawn.

What is a **Population**?

Population is a complete collection of all elements under study.

What is a **Parameter**?

Parameter is any numerical measurement related to a population.

Example:

The average math score of all SAT participants in year 2010 was 516. Identify the population/parameter.

Solution:

Population: All SAT participants
Parameter: Average math score: 516

What is a **Sample**?

Sample is randomly selected elements from the population.

What is a **Statistic**?

Statistic is any numerical measurement related to a sample.

Example:

The average age of 25 randomly selected nurses in Southern California was reported to be 34.5 years.
Identify the sample/statistic.

Solution:

Sample: Randomly selected nurses
Statistic: Average age: 34.5, Number of nurses selected: 25

What are **Data**?

Data are the values or information that we have obtained.

How do we use the collected **Data**?

Data are usually used for two types of studies:
- **Observational Study**
- **Experimental Study**

What is an **Observational Study**?

Observational Study is when we only observe and measure specific characteristics but make no attempt to modify.

What is an **Experimental Study**?

Experimental Study is when we apply some treatment and observe its effect.

Example:

CNN's Report: The study published online in the journal Atherosclerosis found eating egg yolks regularly increases plaque buildup about two-thirds as much as smoking does. Specifically, patients who ate three or more yolks a week showed significantly more plaque than those who ate two or less yolks per week.

Identify the observational or experimental studies.

Solution:

<u>Observational</u>: There were no sign to modify the result, it was solely an observation.

Example:

The math department assigned a supplemental instruction leader to sit in a class, and help those students who chose to attend the extra sessions on regular basis. The progress of these students were compared to those that did not attend these sessions.

Identify the observational or experimental studies.

Solution:

Experimental: The extra sessions modify the progress of those attending. These were compared to those that did not attend the extra sessions.
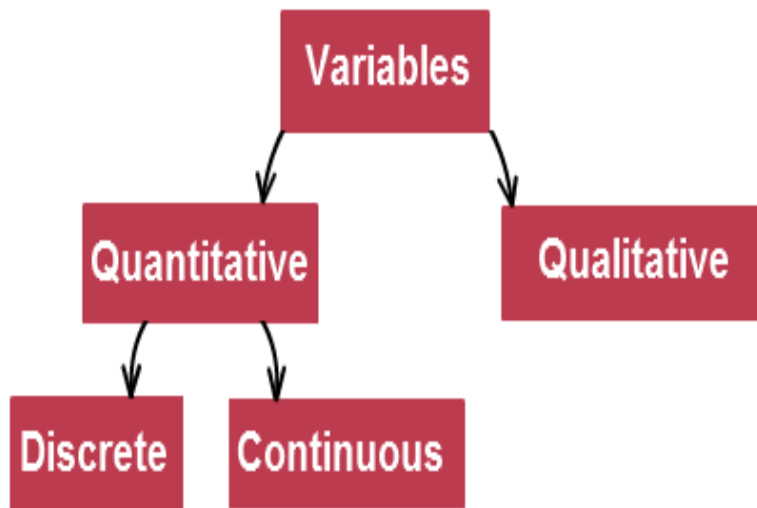
What are **Variables**?

Variables are the characteristics of the data that we have collected.

Are there any **Special Variables**?

Variables can be divided into two types, **Qualitative** or **Quantitative** .

What are **Qualitative Variables**?

Qualitative Variables deal with quality, and cannot be measured. These variables arrange individuals into categories according to shared qualities or characteristics.

What are **Quantitative Variables**?

Quantitative Variables deal with numbers and quantity. These variables provide some sort of measurement like how much or how many.

What are **Discrete Variables**?

Discrete Variables are quantitative variables that will assume only values in a discrete set like the set of integers, such as the number of chairs in the classroom.

What are **Continuous Variables**?

Continuous Variables are quantitative variables that will assume an infinite number of possible values in any given interval such as room temperature.

Example:

Identify the qualitative/quantitative and discrete/continuous variables.

1. Ron Steere sold 7 black cars and 8 red cars last month.

2. Sara took her time when taking her final exam in a sociology class and got 85 correct answers in a multiple–choice test with 100 questions.

Solution:

1. Qualitative: Color of cars he sold.
   Quantitative: The number of cars he sold.

2. Discrete: Number of correct answers she got on the exam
   Continuous: Length of time she took to finish the exam.
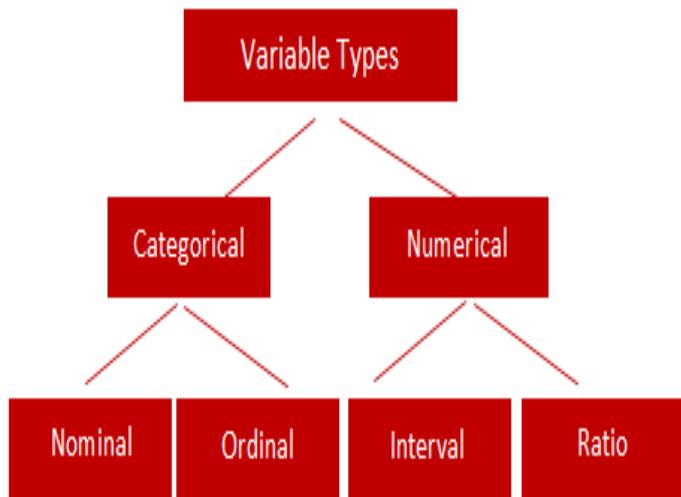
What are **Levels of Measurement**?

The process of collecting any data involves measurement of some characteristics such as height, weight, favorite color, level of pain, and so on. A good measurement scale should be both reliable and valid, therefore there are different level of measurements.

What are some **Levels of Measurement**?

Here is a list of common level of measurements:

- ▶ **Nominal**
- ▶ **Ordinal**
- ▶ **Interval**
- ▶ **Ratio**

What is a **Nominal Scale**?

Nominal Data are qualitative that deal with names, categories, or labels. Nominal data cannot be ordered in a meaningful way, and no calculations can be performed with them.

Example:

❶ Eye color, yes or no responses to a survey.

❷ Jersey's number for athlete, or license plate number on cars.

What is a **Ordinal Scale**?

Ordinal Scale is for nominal data type that can be ordered.

Example:

1. Grades that are assigned to students at the end of the semester.
2. Travel.com recently revealed the top 10 cities to visit in U.S.
3. Movie.com releases movie ratings for new movies every Monday morning.

What is a **Interval Scale**?

Interval Scale is for ordinal data type with meaningful addition or subtraction operations.

Example:

❶ Room temperature.

❷ Range for grading policy.

❸ Race time on sporting events.

What is a **Ratio Scale**?

Ratio Scale is for interval scale but we can divide one value by another value and that ratio is meaningful and has a natural starting point.

Example:

❶ Income salary of doctors versus nurses.

❷ Years of working experience.

❸ Heights of athletes in different sports.

What is a **Sampling Method**?

Sampling Method is a strategy on how to select a sample from within a general population while eliminating bias in the selection process.

What are some common **Sampling Methods**?

Here is a list of common sampling methods:

- ▶ **Simple Random sampling**
- ▶ **Convenience Sampling**
- ▶ **Systematic Sampling**
- ▶ **Stratified Sampling**
- ▶ **Cluster Sampling**

What is **Simple Random Sampling**?

Simple Random is when every sample of size n taken from the population being studied has an equal chance of being selected.

Example:

❶ Toss a coin 1000 times, and record the number of tails.

❷ Suppose there are 50 students in this class, assign one number ranging from 1 to 50 to each student in the class, use random number generator from your calculator and randomly generate 10 numbers ranging from 1 to 50, now select only those students that have these random numbers.

What is **Convenience Sampling**?

Convenience is when a sample is not drawn by a well-defined random method and it is very easy access.

Example:

❶ Asking a relative, friend or neighbor for their opinion on certain political issue.

❷ If I were to ask my class whether or not they have some sort of smart phone.

What is **Systematic Sampling**?

Systematic is when we choose a starting point at random, then select every Kth item after that.

Example:

❶ Recoding every 10th caller to the customer service line, and use it to improve quality of customer service.

❷ Inspecting every 25th item on the assembly line and check for quality control.

What is **Stratified Sampling**?

Stratified is when we divide our population of interest into groups that share some characteristics, then we randomly select individuals from each of these groups.
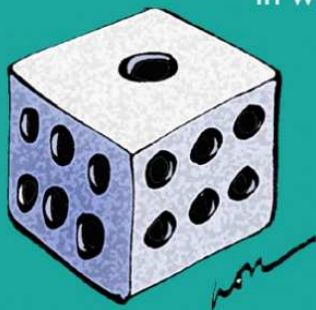

Example:

1. Hospital administration divided all patients in two groups, males and females, then randomly selected 10 patients from each group to conduct a survey.

2. All students at a four–year university are identified as freshmen, sophomore, junior, or senior. University officials randomly selected 25 students from each group to conduct a survey on campus services.

What is **Cluster Sampling**?

Cluster is when we divide our population of interest into groups that share some characteristics, then we randomly select some of these groups, and we choose all members of these selected groups.

Example:

❶ Math department used each section of math classes as clusters, selected 10 sections(clusters) randomly, and then distributed a survey to every students on these sections.

❷ A well-known bank with many branches randomly selected 25 of these branches, then interviewed all employees to see if there were any cause of concerns for bullying in the workplace.