# AN INTRODUCTION TO STATISTICS

LECTURE 1
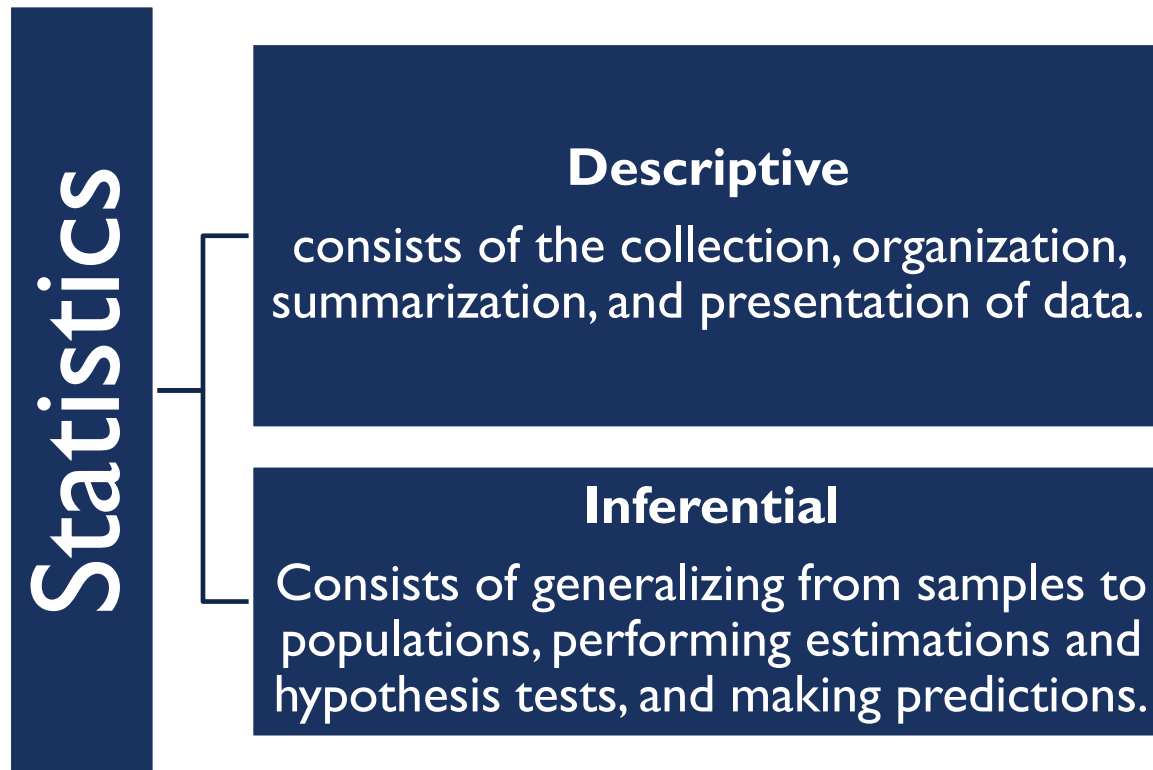
## BY/ ALY MAHER ABDELFATTAH

# DEFINITIONS

- **Statistics** a collection of methods for planning experiments, obtaining data, and then then organizing, summarizing, presenting, analyzing, interpreting, and drawing conclusions based on the data.

- **Variable** is a characteristic or attribute that can assume different values.

- **Data** is observations (such as measurements, genders, survey responses) that have been collected and the variables can assume.

- Variables whose values are determined by chance are called **random variables**.

# STATISTICS TYPES

## Statistics

### Descriptive
consists of the collection, organization, summarization, and presentation of data.

### Inferential
Consists of generalizing from samples to populations, performing estimations and hypothesis tests, and making predictions.

# DEFINITIONS

- **Population** is the complete collection of all elements (scores, people, measurements, and so on) to be studied.

- The collection is **complete** in the sense that it includes all subjects to be studied.

- The population is **usually too big** to be studied directly, then statistics is used

- **Census** is the collection of data from every member of the population.

- **Sample** is a sub-collection of elements drawn from a population.

# DATA TYPES

**Data**

**Qualitative**
can be separated into different categories that are distinguished by some nonnumeric characteristics. Example: genders (male/female).

**Quantitative**
numbers representing counts or measurements. Example: weights of students.

**Continuous**
(numerical) data result from infinitely many possible values that correspond to some continuous scale that covers a range of values without gaps, interruptions, or jumps.

**Discrete**
data result when the number of possible values is either a finite number or a 'countable' number of possible values.

# LEVELS OF MEASUREMENT

This is another way to classify data using four levels of measurement (Nominal, Ordinal, Interval, Ratio )

- **Nominal level of measurement:** names of things without a mathematical interpretation (e.g., hair color, marital status).

- **Ordinal level of measurement:** Attributes with a meaningful order or rank but without a fixed magnitude difference (e.g., size, grades).

- **Interval level of measurement:** Attributes where the difference between values is meaningful, with an equal interval scale, but no true zero point (e.g., temperature in Celsius or Fahrenheit).

- **Ratio level of measurement:** Attributes with a meaningful order, equal intervals, and a true zero point, allowing for statements about how many times greater one value is than another (e.g., weight, height, income).

# SAMPLING TYPES

- **Random Sampling:** selection so that each has an equal chance of being selected

- **Stratified Sampling:** subdivide the population into at least two different subgroups that share the same characteristics, then draw a sample from each subgroup.

- **Systematic Sampling:** Select some starting point and then select every $K\,th$ element in the population

- **Cluster Sampling:** divide the population into sections (or clusters); randomly select some of those clusters; choose all members from selected clusters

# PARAMETER VS STATISTIC

## Parameter

- a numerical measurement describing some characteristic of a population Example Mean $\mu$
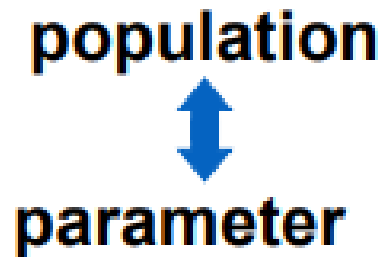


## Statistic

- a numerical measurement describing some characteristic of a sample. Example sample Mean $\chi$.