

# What is Statistics?

---

- ❑ Different authors have defined statistics differently. The best definition of statistics is given by Croxton and Cowden according to whom **statistics may be defined as the science, which deals with collection, presentation, analysis and interpretation of numerical data.**
- ❑ The science and art of dealing with variation in data through collection, classification, and analysis in such a way as to obtain reliable results. —(John M. Last, **A Dictionary of Epidemiology** )
- ❑ Branch of mathematics that deals with the collection, organization, and analysis of numerical data and with such problems as experiment design and decision making. —(Microsoft **Encarta Premium 2009**)

- 
- ❑ A branch of mathematic staking and transforming numbers into useful information for decision makers.
  - ❑ Methods for processing & analyzing numbers
  - ❑ Methods for helping reduce the uncertainty inherent indecision making

## What is biostatistics?

---

- It is the science which deals with development and application of the most appropriate methods for the:
  - Collection of data.
  - Presentation of the collected data.
  - Analysis and interpretation of the results.
  - Making decisions on the basis of such analysis
- The methods used in dealing with statistics in the fields of medicine, biology and public health.

## Why study statistics?

---

### □ Decision Makers Use Statistics To:

- Present and describe data and information properly
- Draw conclusions about large groups of individuals or information collected from subsets of the individuals or items.
- Improve processes.

# Statistics

---

Descriptive Statistics

Experimental Statistics

Inferential Statistics

Methods for processing,  
summarizing, presenting  
and describing data

Techniques for planning  
and conducting  
experiments

Drawing conclusions and  
/ or making decisions  
concerning a population  
based only on sample  
data

# DATA

---

## ➤ Definition:-

- ❑ A set of values recorded on one or more **observational units**. Data are raw materials of statistics.
- ❑ **Data set** : A collection of data is data set
- ❑ **Data point** : A single observation
- ❑ **Raw data** : Information before it arranged and analysed

## ➤ Sources of data:-

- ❑ Experiments
- ❑ Surveys
- ❑ Records

---

□ Example of Raw data:

<b>Blood Pressure</b>	<b>Systolic BP</b>	<b>Diastolic BP</b>
	120	80
	135	90
	125	85
	140	95
	138	86

## Elements, Variables, and Observations

---

- ❑ The **elements** are the entities on which data are collected.
- ❑ A **variable** is a characteristic of interest for the elements.
- ❑ The set of measurements collected for a particular element is called an **observation**.
- ❑ The total number of data values in a data set is the number of elements multiplied by the number of variables.



## Data, Data Sets, Elements, Variables, and Observations

---

The diagram shows a data table with callouts identifying its components. A large blue callout labeled 'Observation' points to the entire table. A smaller callout labeled 'Variables' points to the columns. A callout labeled 'Element Names' points to the 'Company' header. A callout labeled 'Company' points to the first column. A callout labeled 'Data Set' points to the bottom of the table.

Company	Stock Exchange	Annual Sales(\$M)	Earn/Share(\$)
Dataram	AMEX	73.10	0.86
EnergySouth	OTC	74.00	1.67
Keystone	NYSE	365.70	0.86
LandCare	NYSE	111.40	0.33
Psychemedics	AMEX	17.60	0.13

# Descriptive statistics

---

- Summarizing and describing the data
- Uses numerical and graphical summaries to characterize sample data

# Inferential Statistics

---

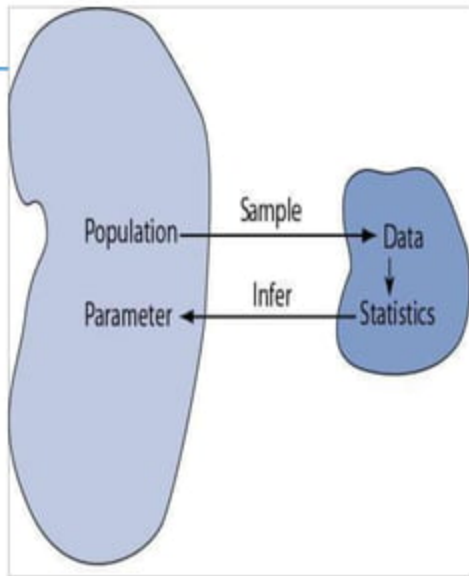
- Estimation
  - e.g., Estimate the population mean weight using the sample mean weight
- Hypothesis testing
  - e.g., Test the claim that the population mean weight is 120 pounds



Drawing conclusions about a large group of individuals based on a subset of the large group.

# Inferential statistics

- It refers to the process of selecting and using a sample to draw inference about population from which sample is drawn.
- Two forms of statistical inference
  - **Hypothesis testing**
  - **Estimation**



## Basic Vocabulary of Statistics

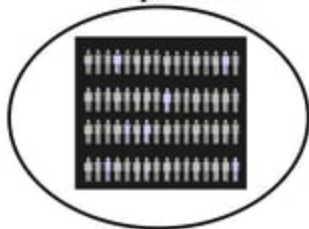
---

- ❑ **POPULATION** : A **population** consists of all the items or individuals about which you want to draw a conclusion. Ex: People who live within 25 kms of radius from centre of the city.
- ❑ **SAMPLE** : A sample is the portion of a population selected for analysis. It has to be **representative**.
- ❑ **PARAMETER** : A **parameter** is a numerical measure that describes a **characteristic of a population**.
- ❑ **STATISTIC** : A **statistic** is a numerical measure that describes a **characteristic of a sample**.

# Population vs. Sample

---

**Population**



Measures used to describe the population are called **parameters**

**Sample**



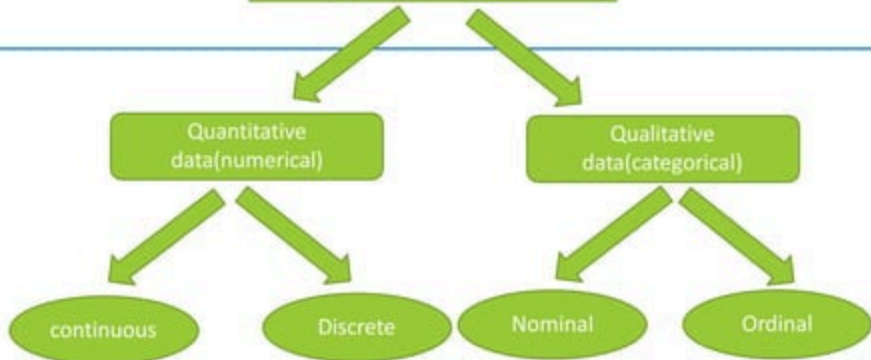
Measures computed from sample data are called **statistics**

---

	Population	Sample
Definition	Complete enumeration of items is considered	Part of the population chosen for study
Characteristics	Parameters	Statistics
Symbols	Population size = $N$ Population mean = $\mu$ Population S.d = $\sigma$	Sample size = $n$ Sample mean = $\bar{x}$ Sample S.d = $s$

# Types of data

---



take forever to count  
Ex: time

countable in a finite  
amount of time  
Ex: count change of  
money in your pocket



## Type of variables

---

- **Categorical** (qualitative) variables have values that can only be placed into categories, such as “yes” and “no.”
  
- **Numerical** (quantitative) variables have values that represent quantities.

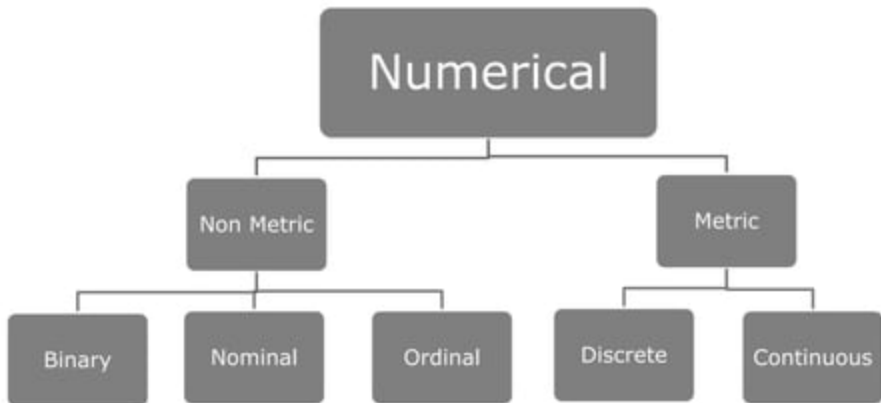
## Qualitative Data

---

- ❑ Non Numerical
- ❑ Categorical
- ❑ No numbers are use to describe it
- ❑ Word, picture, image
- ❑ Ex. Do you smoke? Yes No

# Quantitative Data

---



## REASONS FOR ASSIGNING NUMBERS

---

Numbers are usually assigned for two reasons:

- numbers permit statistical analysis of the resulting data
- numbers facilitate the communication of measurement rules and results

## Primary Scales of Measurement

### Scale Nominal

Numbers  
Assigned  
to Runners



Finish

### Ordinal

Rank Order  
of Winners



Third  
place



Second  
place



First  
place

Finish

### Interval

Performance  
Rating on a  
0 to 10 Scale

8.2

9.1

9.6

### Ratio

Time to Finish  
in Seconds

15.2

14.1

13.4