Part A: Statistics
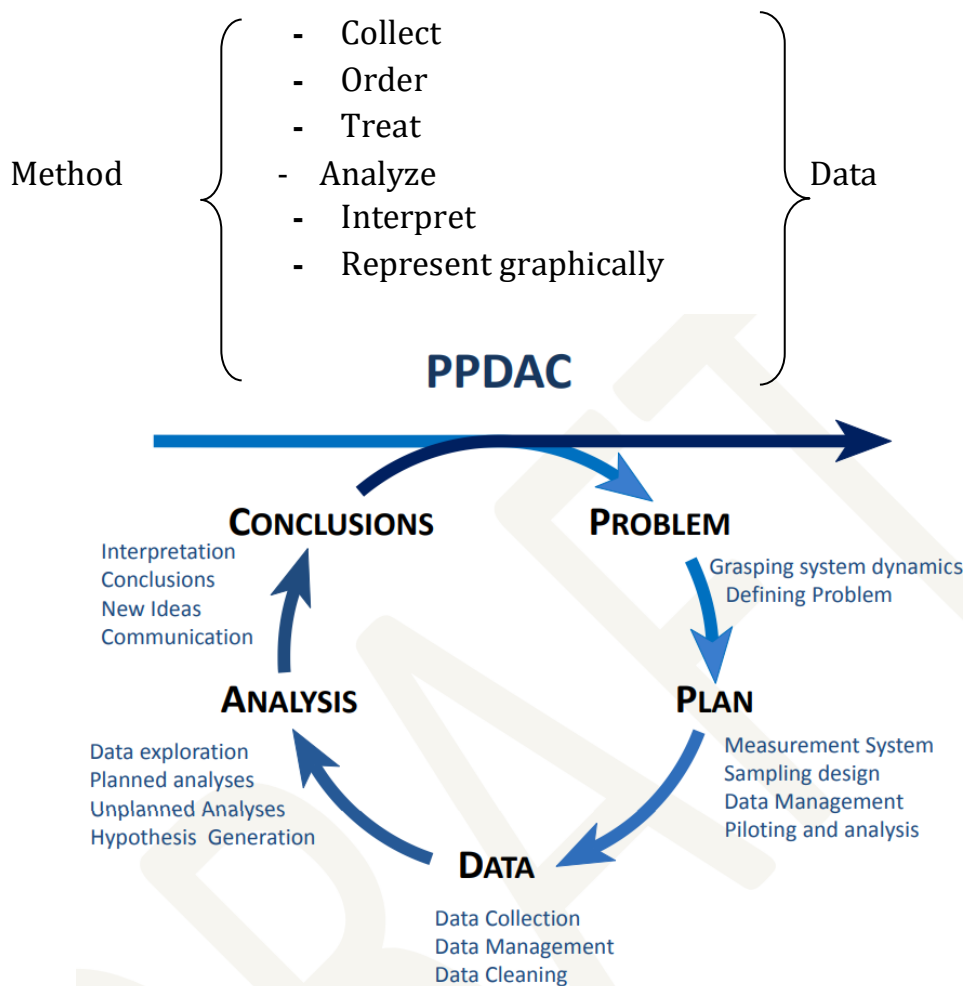
Chapter 1: Basic Definitions:

a. Statistics:

"**Statistics**" – as defined by the American Statistical Association (ASA) – "is the science of learning from data, and of measuring, controlling and communicating uncertainty". **[1]**

The study of statistics involves math and relies upon calculations of numbers. But it also relies heavily on how the numbers are chosen and how the statistics are interpreted. **[2]**

Statistics: it is a method which allows us to:

Method
- Collect
- Order
- Treat
- Analyze
- Interpret
- Represent graphically

Data

PPDAC

CONCLUSIONS
Interpretation
Conclusions
New Ideas
Communication

PROBLEM
Grasping system dynamics
Defining Problem

ANALYSIS
Data exploration
Planned analyses
Unplanned Analyses
Hypothesis Generation

PLAN
Measurement System
Sampling design
Data Management
Piloting and analysis

DATA
Data Collection
Data Management
Data Cleaning

Figure. 1.1: The statistical inquiry cycle. **[1]**

b. Descriptive statistics:

Descriptive statistics is a basic area of study in statistics. It is used to summarize data in an organized manner by describing the relationship between variables in a sample or population. It uses parameters, such as graphs, tables and data analysis methods.

Descriptive statistics involves methods of organizing, picturing and summarizing information from data.

c. Mathematical statistics or inductive (Inferential):

Inferential statistics involves methods of using information from a sample to draw conclusions about the population.

The aim of statistical induction is to bring out properties of a set of variables. It draws on the results of mathematical statistics, which applies rigorous mathematical calculations using probability theory.

d. Probability

Probability is the branch of mathematics that calculates the probability of an event, that is, the frequency of an event compared to all possible cases. This branch of mathematics originated from games of chance, more specifically from the desire to predict the unpredictable or quantify the uncertain.

- ❖ Study a large amount of data---------------- Statistics
- ❖ Random aspect of results--------------------- Probability

A.1.1 Notions of population, sample, variables, modalities:

In general, statistical methodology is based on four concepts: Population, Variables, Observations, Data.
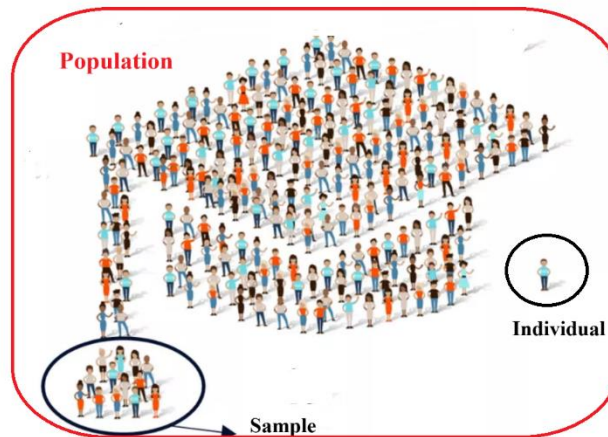
Figure. 1.2: Population, Sample, Individual

a. Population:

In statistics the notion of population represented the whole subject of the study, it is about what we want to study (individuals, objects, companies, institutions, etc.). It always brings together all individuals relating to this study. The population is generally rated: P

Example 1:

We consider all teachers at Algiers University. We are interested in number of children of each teacher. In this case: P = set of teachers

b. Individual (statistical units):

The population is the set of individuals, each person in this population represents a statistical unit (individuals), It is indivisible and can be a human, an animal, an object or a plant. The individual is rated: i

Example 2:

in the *Example 1,* the individual represents a teacher at the University of Algiers.

c. Sample:

Sometimes in statistics, we find ourselves with a very large population that we cannot study at that time, we will select someone from this population, what we call a sample.

Example 3:

We consider all Algerian students. We are interested of the eyes color of each student. In this case we are in front of a large population, we select just 500 students for the study.

500 is a sample of population.

d. Variables (Characters):

Variables are properties or characteristics of some event, object, or person that can take on different values or amounts.

> ➢ **A single character** studied represents a one-dimensional statistical series or variable
>
> ➢ **Two characters** studied represent a statistical series in two dimensions or two variables.

The characteristics studied can be: age, weight, marital status, number of children in a family, height, professional categories, eyes color, level of education, mass, list of manufacturing defects, temperature, price, professional categories of employees, year of study, nationality, etc.

Example 4:

in the *Example 1,* the variable represents the number of children of the teachers at the University of Algiers.

e. Modalities:

A modality is the value taken by a character or a statistical variable. The modalities therefore correspond to the set of possible values.

- ❖ A character can take different modalities or values.
- ❖ Each individual must present one and only one modality. (Ex: red, blond, brown}.
- ❖ A statistical variable is denoted by a capital letter X, Y, and the values it takes by lowercase letters x1, x2..., y1, y2...
- ❖ A variable must present at least two categories or values.

Example 5:

- For the variable "family situation", modalities are "single, married, divorced".
- For the variable "number of children", modalities are "1, 2, 3, 4".

f. Statistical series:

Refers to a collection of data arranged in a specific order to show the frequency, variation, and distribution of a particular phenomenon. These series are used in various research studies to represent the data in a structured and organized manner, making it easier to interpret and analyze

A.1.2 Different types of statistical variables: qualitative, quantitative, discrete, continuous:

Statistical Variables may be classified into two main categories: <u>Quantitative</u> (numerical) or <u>Qualitative</u> (categorical).

Example 6:

- Variable "Number of rooms in an apartment", their modalities are: 1, 2, 3. It is a Quantitative Variable.
- Variable "family situation", their modalities are: single, married, divorced. It is a Qualitative Variable.

A.1.2.1 Quantitative Variable:

These variables represent numbers values. It can be categorized into two types: Discrete or Continuous

a. Discrete Variables: These variables can only take on a countable and finite numbers.

Example 7:

- The number of books in the library.
- The number of patients in a hospital.

b. Continuous Variables: Continuous if it takes all the values of a finite or infinite interval (real numbers R).

Example 8:

- The weight of the students the values are: 40.8kg, 42.5kg, 45.6kg.
- Time to respond to a question.

A.1.2.2 Qualitative Variable:

It represents categories that cannot be measured numerically. It can be categorized into two types: Nominal or Ordinal.

a. **Nominal Variables:** Their modalities cannot be ordered or ranked.

Example 9:

- The variable "family situation", their modalities are: Single, Married, Divorced.
- The variable "eyes color", their modalities are: Blue, Green, Black, Brown.

b. **Ordinal Variables:** Their modalities can be ordered in a specific order or ranking.

Example 10:

- The variable "The condition of a car", their modalities are: New, Ancient, Degraded.
- The variable "strength", their modalities are: Weak, Normal, Strong, Very strong.