

# Probability and Statistics

Dr. ARRAR - Prof. REMITA  
National Higher School of Artificial Intelligence

2022-2023

## **Part 1 : Descriptive statistics**

Chapter 1 Generalities

Chapter 2 Statistical series of one-dimension

Chapter 3 Bivariate statistical series

## **Part 2 : Introduction to probability calculus**

Chapter 4 Events and set algebra

Chapter 5 Probability calculus

# Chapter 1: Generalities

## Introduction

The essential purpose of statistics is to facilitate decision-making under conditions of uncertainty.

There is no precise definition of statistics, but for us we will use the following definition:

### Definition

We call statistics a set of methods or techniques used to analyze or process sets of observations that we will call data.

The methods used are based on mathematics and make extensive use of computer tools for their implementation.

### Remark

*We must not confuse statistics, which is the science that has just been defined, with a statistic, which is a set of numbers on a specific subject.*

# Chapter 1: Generalities

## History

We distinguish three important phases in the evolution of statistics:

- From antiquity until the end of the 19th century, statistics remained mainly a set of enumeration.
- From the 19th century to the 1960s, mathematical statistics was built up school (K. Pearson, W. Gosset (Student), R. Fisher, J. Neyman ...).
- Since the 1960's, and with the development of computer and graphic tools, statistics graphics, statistics has undergone a considerable development.

# Chapter 1: Generalities

## Basic terminology

- **Population:** set concerned by a statistical study (noted  $\Omega$ ).
- **Individual element or statistical unit (individu):** is one member of a set of entities being studied (noted  $\omega \in \Omega$ ).
- **Sample (échantillon):** subset of the population on which the observations are made.
- **Survey (enquête):** operation consisting of observing (measuring, questioning, etc.) all the individuals in a sample.
- **Census (recensement):** survey in which the sample observed is the entire population (exhaustive survey).
- **Poll (sondage):** survey in which the observed sample is a strict subset of the population (non-exhaustive survey).

# Chapter 1: Generalities

## Basic terminology

- **Character:** It is a characteristic defined on the population and observed on the sample. The characteristic can be:
- **Modality or observations:** the different values taken by each character.

# Chapter 1: Generalities

## Type of characters

- ① **Qualitative:** it is a character which is not measurable and it can be :
  - **nominal data** (gender, profession, family situations, ...)
  - **ordinal data** (military rank, grade in university, ...)
- ② **Quantitative:** it is a characteristic that can be measured and we distinguish two cases :
  - **discrete** (number of children, number of rooms in an apartment, ...) also called **discrete statistical variable**.
  - **continuous** (height, age, speed, weight, rate, ...) also called **continuous statistical variable**.

# Chapter 1: Generalities

## Statistical approach

To carry out a statistical study, it is advisable to follow the following procedure:

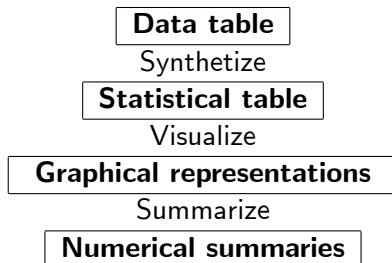
- First of all, we start by collecting the raw observations that we group in a table called data table.
- In many situations, the observations may be different and in large numbers, and the data table is not easily exploitable, so the observations are synthesized by grouping them in the statistical table.
- We complete the study by making graphical representations and numerical calculations in order to give a good interpretation of the observed data.



# Chapter 1: Generalities

## Statistical approach

Schematically the statistical approach is summarized as follows:



# Chapter 1: Generalities

## Data tables - Qualitative data

### Example (1)

The study of the family situation of 30 employees of a company is summarized in the following table:

<b>01</b>	<b>02</b>	<b>03</b>	<b>04</b>	<b>05</b>	<b>06</b>	<b>07</b>	<b>08</b>	<b>09</b>	<b>10</b>
S	M	S	S	D	W	M	M	S	D
<b>11</b>	<b>12</b>	<b>13</b>	<b>14</b>	<b>15</b>	<b>16</b>	<b>17</b>	<b>18</b>	<b>19</b>	<b>20</b>
S	M	S	S	M	D	S	S	M	M
<b>21</b>	<b>22</b>	<b>23</b>	<b>24</b>	<b>25</b>	<b>26</b>	<b>27</b>	<b>28</b>	<b>29</b>	<b>30</b>
S	S	M	S	W	S	S	M	S	S

# Chapter 1: Generalities

## Data tables - Discrete statistical variable

### Example (2)

We study the number of children in a family in a city inhabited by 100 families. The following results were obtained:

<b>01</b>	<b>02</b>	<b>03</b>	<b>04</b>	<b>...</b>	<b>98</b>	<b>99</b>	<b>100</b>
2	0	4	2	...	6	3	1

# Chapter 1: Generalities

## Data tables - Continuous statistical variable

### Example (3)

We study the time taken by 30 workers to manufacture a given part. The following results were obtained:

<b>01</b>	<b>02</b>	<b>03</b>	<b>04</b>	<b>05</b>	<b>06</b>	<b>07</b>	<b>08</b>	<b>09</b>	<b>10</b>
56	60	60	65	67	70	71	73	73	74
<b>11</b>	<b>12</b>	<b>13</b>	<b>14</b>	<b>15</b>	<b>16</b>	<b>17</b>	<b>18</b>	<b>19</b>	<b>20</b>
75	76	77	77	77	78	78	78	78	78
<b>21</b>	<b>22</b>	<b>23</b>	<b>24</b>	<b>25</b>	<b>26</b>	<b>27</b>	<b>28</b>	<b>29</b>	<b>30</b>
79	81	81	82	82	83	84	85	87	90