

Mathematics Statistics Informatics



Dr: BOUDEBANE Zineb

University of Mila

Institute of Nature and life
Sciences

Department of NLS

Email: z.boudebane@center-univ-mila.dz

March, 2026

Table of contents

Objectives	3
Introduction	4
I - Prerequisites	5
II - Descriptive Statistics	6
1. Specific objectives	6
2. Statistical Vocabulary	6
3. Position Parameters	8
4. Quartiles	11
5. Dispersion parameters	12
Bibliography	14

Objectives

At the end of this course, the learner will be able to :

- **Identify** the basic concepts related to real functions of a real variable .
- **Distinguish** between real functions of one variable and several variables.
- **Test** the nature of numerical series.
- **Analyze** the statistical data.
- **Reorganize** data into parameters and indicators.

Introduction

To study living things, biology utilizes many tools such as computer science, chemistry, statistics, etc. In particular, it uses mathematics, which is becoming increasingly important. This can be explained by the complexity of observed phenomena and the emergence of new fields of study such as ecology, molecular biology, climatology, population dynamics, and more. All of these require highly sophisticated methods for quantifying and interpreting results. Many of these methods were only introduced in the last century.

Alongside the characteristics of description or classification of living entities, contemporary biology is interested in their birth or emergence, their growth or evolution, and their extinction or disappearance. To conduct these studies, various mathematical models have been developed, which are divided into two categories: those qualified as discrete, which call upon combinatorics, sequences, etc., and those qualified as continuous, which utilize functions of one or more real variables, equations, and differential systems.

The "Statistical Mathematics" course is intended for first-year undergraduate students in the natural and life sciences.

This course is divided into four learning units (chapters). It consists of both a lecture and a tutorial session.

I Prerequisites

To successfully complete this course, you must first have:

- Basic concepts related to functions.
- Some differentiation rules .
- Knowledge of some types and properties of numerical sequences
- A general overview of statistical variable

II Descriptive Statistics

1. Specific objectives

- **Define** the basic concepts of descriptive statistics : population, modality, variable,.....
- **Classify** variables into quantitative and qualitative types
- **Illustrate** observed data using appropriate statistical tables.
- **Calculate** parameters of position and Dispersion .

2. Statistical Vocabulary

- **Population** : is the set of individuals or objects of the same nature on which the study relates.
- The **components of the population** are called individuals, elements, or statistical units.
- **Sample** : a subset of the population selected for study
- **Statistical Variable** : the character X is the subject under statistical study.
- **Statistical Modality**: The category or the different possible situations (levels) of a statistical variable.

Reminder: Types of Statistical Variables

There are two types of statistical variables

Quantitative Variables: are the variables that can be measured or give us numbers representing counts, or variables whose modalities are numbers.

Moreover the quantitative variables are divided into two main types **discrete** and **continuous**.

Discrete : a variable is discrete if it takes isolated values (usually integers).

Continuous: a variable is continuous when it can take numbers from an interval of real numbers.

Example

variable	The modalities	Type of variable
Temperature	between 15 and 56, i.e., [15, 56].	Quantitative continuous
The number of students	30, 50, 60, 80	Quantitative discrete

Qualitative Variables : are variables that are not measurable numerically.

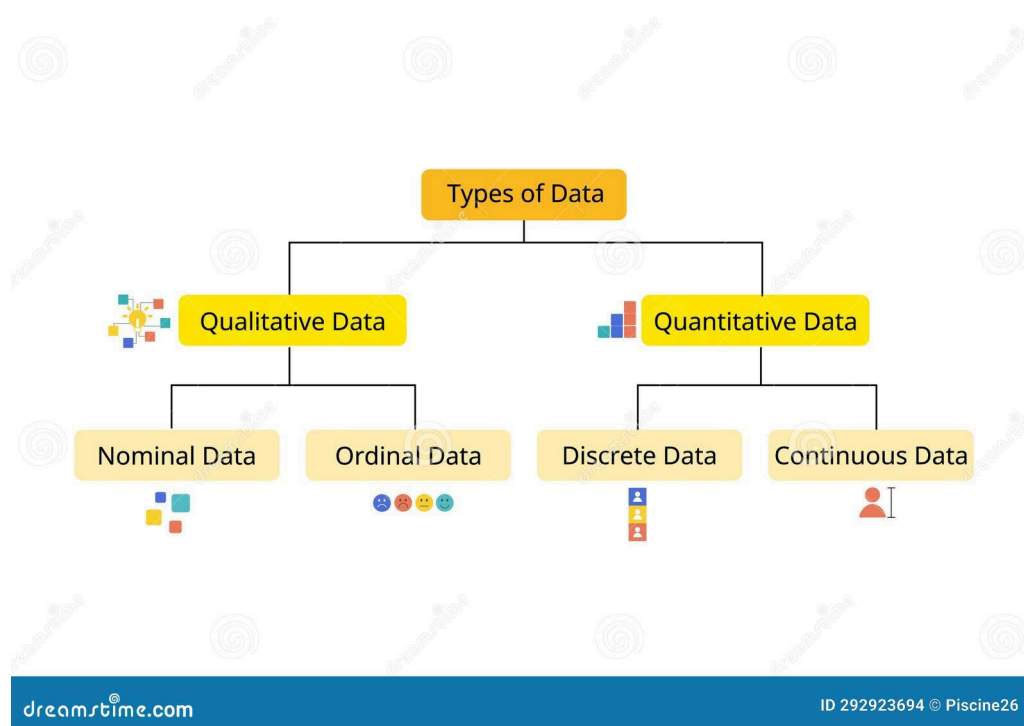
The qualitative variables can be

Nominal: Variables whose modalities cannot be ordered according to their meaning

Ordinal: Variables whose modalities are ordered according to their meaning (hierarchy).

Example

variable	The modalitie	Type of variable
Eye color	Black, Brown.	Qualitative nominal
Rating scale	bad, good, excellent	Qualitative ordinal



Organizing data

Statistical Series : a statistical series is the simplest form of presenting statistical data.

Absolute Frequency n_i : The absolute frequency n_i is the number of statistical elements corresponding to a given modality.

Cumulative Absolute Frequency $n_i^{c\uparrow}$: the cumulative absolute frequency $n_i^{c\uparrow}$ is the number of individuals corresponding to the same modality and all previous modalities.

Relative Frequency f_i : The relative frequency is defined by $f_i = \frac{n_i}{N}$, where N is the total number of observations.

Cumulative Relative Frequency $f_i^{c\uparrow}$: is defined by $f_i^{c\uparrow} = \frac{n_i^{c\uparrow}}{N}$.

Example

The table below shows the number of children per family, with absolute frequency (n_i), cumulative absolute frequency (N_i^-), relative frequency (f_i), and cumulative relative frequency (F_i^-):

Number of Children	n_i	$n_i^{c\uparrow}$	f_i	$f_i^{c\uparrow}$
0	16	16	0.250	0.250
1	18	34	0.281	0.531
2	14	48	0.218	0.749
3	11	59	0.172	0.921
4	3	62	0.047	0.968
5	2	64	0.031	1.000
Total	64	-	-	-

Continuous Statistical Variables

Class Limits:

For each class, we define:

- a lower limit,
- an upper limit.

Class Amplitude:

The amplitude of a class is $a = \text{upper limit} - \text{lower limit}$.

Class Center:

The center of a class is

$$c_i = \frac{\text{lower limit} + \text{upper limit}}{2}.$$

Example

Suppose we record the daily study hours of 30 students. We group the data into classes and calculate class limits, class amplitude, and class centers.

Class	Lower Limit	Upper Limit	Amplitude	Class Center c_i
[0, 2[	0	2	2	1
[2, 4[	2	4	2	3
[4, 6[	4	6	2	5
[6, 8[	6	8	2	7
[8, 10[	8	10	2	9

3. Position Parameters

Mean

- **Cas of discrete statistical variable:**

Let X be a discrete statistical variable and $x_1, x_2, \dots, x_k$ its valeurs with frequencies $n_1, n_2, \dots, n_k$, and $N = \sum_{i=1}^k n_i$

The mean is $\bar{x} = \frac{1}{N} \sum_{i=1}^k x_i n_i$.

Example

Calculation of the mean:

x_i 0 1 2 3

n_i 1 4 5 2

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k x_i n_i = \frac{0 \cdot 1 + 1 \cdot 4 + 2 \cdot 5 + 3 \cdot 2}{12} = \frac{0 + 4 + 10 + 6}{12} = \frac{20}{12} \approx 1.67$$

Cas of Continuous statistical variable:

For a continuous variable grouped into classes, the mean is $\bar{x} = \frac{1}{N} \sum_{i=1}^k n_i c_i$,

where c_i is the class center.

Example

Calculation of the mean:

Class	n_i	c_i
[0, 1[	0	0.5
[1, 3[	1	2
[3, 5[	3	4
[5, 7[	5	6
[7, 9[	7	8

Mode

Cas of discrete statistical variable:

The mode M_0 is the value that occurs most frequently.

Example

The highest frequency in the previous example (Cas of discrete statistical variable) is $n_3 = 4$ corresponding to $x_3 = 2$,

so the mode is: $M_0 = 2$

Cas of continuous variable: In this case the mode is calculated by the formula:

$$Mo = L_i + \left(\frac{d_1}{d_1 + d_2} \right) a$$

- L_i : the lower limit of **the modal class** (the class that has the highest frequency)
- d_1 : the absolute frequency of the modal class - the absolute frequency of the previous class ($n_i - n_{i-1}$).
- d_2 : the absolute frequency of the modal class - the absolute frequency of the next class ($n_i - n_{i+1}$).
- a : the amplitude of the modal class.

Example

class	n_i
[1,60-1,65[	3
[1,65-1,70[	8
[1,70-1,75[	2

- The modal class is : [1, 65 – 1, 70[.
- $L_i = 1.65$.
- $d_1 = 8 - 3 = 5$
- $d_2 = 8 - 2 = 6$
- $a = 1.70 - 1.65 = 0.05$ donc $Mo = 1.65 + \left(\frac{5}{5 + 6} \right) 0.05 = 1.67$

Median

Cas of discrete variable:

The median Me is the value at the center of a series of numbers arranged in ascending order.

- If n is even, then $Me = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$
- If n is odd, then $Me = x_{\frac{n+1}{2}}$

Example

- The number of children of 6 families is as follows : 7, 3, 1, 1, 5, 2

We first order the values: 1, 1, 2, 3, 5, 7

We have $n = 6$ is even so $Me = \frac{x_3 + x_4}{2} = \frac{2 + 3}{2} = 2.5$.

- The number of children of 7 families is as follows : 3, 2, 1, 0, 0, 1, 2

We first order the values: 0, 0, 1, 1, 2, 2, 3

We have $n = 7$ is odd so $Me = x_4 = 1$.

Cas of continuous variable:

In this case the median is given by

$$Me = L_i + \left(\frac{\frac{N}{2} - \sum_{i=1}^{<Me} n_i}{n_{Me}} \right) a$$

- L_i : the lower limit of the median class
- $\sum_{i=1}^{<Me} n_i$: the sum of the absolute frequencies corresponding to all classes below the median class.
- n_{Me} : the absolute frequency of the median class.
- a : the amplitude of the median class.

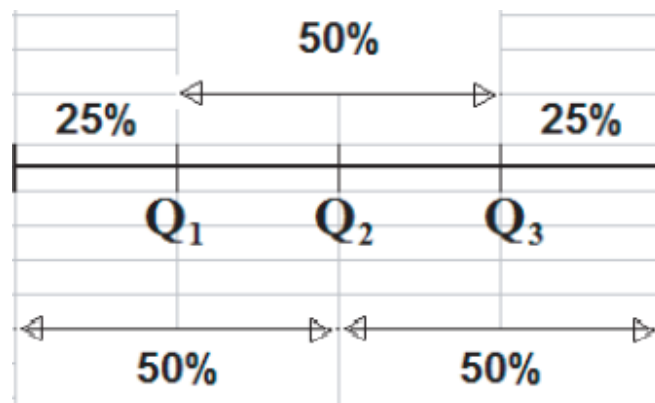
4. Quartiles

Quartiles are the three values that divide the distribution into four equal parts.

The first quartile Q_1 represents 25% of the sample i.e. Q_1 is the value x_i whose position is the smallest integer following $\frac{n}{4}$.

The second quartile Q_2 represents 50% of the sample.

The third quartile Q_3 represents 75% of the sample i.e. Q_3 is the value x_i whose position is the smallest integer following $\frac{3n}{4}$.



Interquartile Range

The interquartile range is $Q = Q_3 - Q_1$.

Example

In the example of the following observations:

x_i	1	3	5	7	9
n_i	1	2	1	2	2
n_i^c	1	3	4	6	8

- We have $n = 8$ and $\frac{n}{4} = 2$ so $Q_1 = x_2 = 3$.
- We have $n = 8$ and $\frac{3n}{4} = 6$ so $Q_3 = x_6 = 7$.

5. Dispersion parameters

Dispersion parameters are the parameters that summarize the dispersion of values around the central value.

Range:

The difference between the largest value and the smallest value observed is called the range e .

$$e = x_{\max} - x_{\min}$$

🔗 Example

The marks of 10 students are as follows: 2, 3, 10, 10, 11, 12, 15, 18, 19, 20

Then $e = x_{\max} - x_{\min} = 20 - 2 = 18$.

🔗 Variance:

A variance is the arithmetic mean of the squares of the differences between the values of a variable and the arithmetic mean.

$$V(X) = \frac{1}{N} \sum_{i=1}^k n_i (x_i - \bar{x})^2 = \frac{1}{N} \sum_{i=1}^k n_i x_i^2 - \bar{x}^2 = \sum_{i=1}^k f_i x_i^2 - \bar{x}^2$$

Standard deviation:

We call standard deviation denoted σ_X the square root of the variance.

$$\sigma_X = \sqrt{V(X)}$$

🔗 Coefficient of variation:

The coefficient of variation, CV , is defined by: $CV = \frac{\sigma_X}{\bar{x}}$

🔗 Example

x_i	0	1	2	3	4
n_i	2	3	1	1	1

$$\bar{x} = 1.5$$

The variance:

$$\begin{aligned}
 V(X) &= \frac{1}{n} \sum_{i=1}^k n_i x_i^2 - \bar{x}^2 \\
 &= \frac{1}{8} \sum_{i=1}^5 n_i x_i^2 - (1.5)^2 \\
 &= \frac{1}{8} (2 \times 0^2 + 3 \times 1^2 + 1 \times 2^2 + 1 \times 3^2 + 1 \times 4^2) - 2.25 \\
 &= \frac{2}{8} - 2.25 = 1.75
 \end{aligned}$$

The standard deviation:

$$\sigma_X = \sqrt{V(X)} = \sqrt{1.75} = 1.3$$

The coefficient of variation:

$$CV = \frac{\sigma_X}{\bar{x}} = \frac{1.3}{1.5} = 0.87$$

Bibliography

Jean Bouyer, 2000- Méthodes statistiques : médecine-biologie. Ed. Estem.

Gilles Stoltz et Vincent Rivoirard, 2012- Statistique mathématique en action. Ed. Vuibert, Paris, 448p.

Genet, J. et Pupion, G : Analyse moderne, tome 2, Vuibert, 1974.