MOHAMED KHIDER UNIVERSITY OF BISKRA.
FACULTY OF EXACT SCIENCES AND NATURAL AND LIFE SCIENCES
DEPARTMENT OF BIOLOGY

COURSE TITLE

Mathematic and statistic Level 1^{st} year LMD

Dr. AFROUN Faïrouz

Biskra university, 2025/2026

Contents

List of Figures

List of Tables

Chapter 1

Introduction to descriptive statistical analysis

Introduction

Statistics is a scientific method that consists of: reducing data on large sets, then analyzing, commenting on, interpreting, and finally critiquing this data.

1.1 Basic concepts

In this section we will present some basic concepts and definitions associated with statistical language.

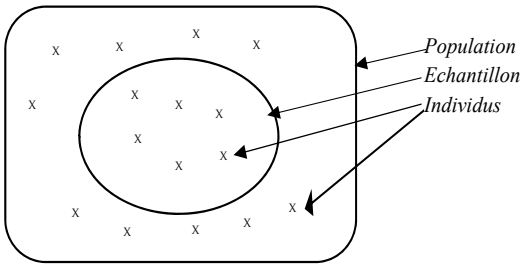
1.1.1 Definitions

Population: The *population*, also called the *universe*, is a well-defined set of homogeneous elements on which a statistical study is to be carried out.

Sample: A *sample* is the subset of the population on which the observations are made.

Individuals: The *individuals*, also called *statistical units*, may be human beings, objects, or animals.

These three concepts (population, sample, and individuals) can be illustrated as follows:



Characteristic: A *characteristic* is a feature or property that allows us to identify individuals and classify them into subsets. Note that each individual may be described by one or several characteristics. Moreover, the modalities must be mutually exclusive, meaning that an individual cannot belong to more than one modality at the same time.

Categories: The categories (modalities) of a characteristic are the different possible situations of that characteristic. For example:

1. The characteristic "Sex" has two categories: {Female, Male}.
2. The characteristic "Marital status" has the following categories: {Married, Single, Widowed, Divorced}.

1.1.2 Character type

We distinguish two types of characteristics:

Qualitative characteristic: A characteristic is said to be qualitative when its categories are not measurable. They are identified by words describing a state. **Example:** Sex, occupation, nationality, ... This type of variable can in turn be classified into two categories:

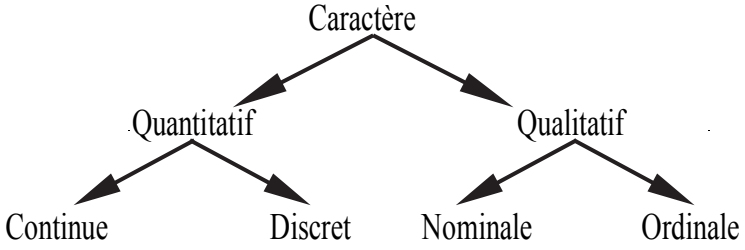
1. **Nominal:** where the categories are measured on a nominal scale, meaning they are expressed by names. Example: eye color, types of plants, ...
2. **Ordinal:** where the categories can be presented on an ordinal scale; they express the degree or level of a state characterizing an individual.

   Example 1: The characteristic "metal resistance to heat" has the following categories: slightly resistant, moderately resistant, highly resistant.
   Example 2: The characteristic "high school diploma grade" has the categories: excellent, very good, good, fair.

Quantitative characteristic: A characteristic is said to be quantitative when its categories are measurable, that is, expressed as numbers. The characteristic is then called a statistical variable, and the different categories are the possible values of the variable.

Example: Age, height, weight, and number of children are quantitative characteristics, statistical variables whose categories are measurable in various specific units. Quantitative variables are of two different types:

1. **Discrete variables (or discontinuous):** A quantitative variable is said to be discrete if it can take only isolated values. A discrete variable that takes only integer values is called discrete. For example, the number of children per household can only be 0, 1, 2, 3, ...; it can never take a value strictly between 0 and 1, or between 1 and 2, or between 2 and 3, ...
2. **Continuous variables:** A quantitative variable is said to be continuous when it can take any value in a finite or infinite interval. For example, the diameter of a tree, its height, or the average grade of a semester, ...



1.2 Statistical tables and graphical representations

The statistical information collected in its *raw* form is practically unusable. In order to give it meaning and usefulness, it must be organized, classified, and processed, mainly by using tables and graphs.

Presenting qualitative or quantitative statistical data in the form of statistical tables is a very important and essential step for subsequent statistical procedures. Afterwards, the statistical tables are represented by graphs in order to visualize the behavior of the statistical variable.

In what follows, after introducing the notion of frequency (absolute and relative), we will highlight the difference between graphs specific to qualitative characteristics, discrete quantitative characteristics, and continuous quantitative characteristics.

1.2.1 Statistical tables: Count and frequency

Let us consider a population composed of N individuals described by a characteristic X, which consists of the categories $x_1, x_2, \ldots, x_k$. Presenting this information in a table consists of counting the number of individuals corresponding to each category and then organizing them.

The theoretical table may be presented as follows:

Categories x_i of the characteristic	x_1	x_2	...	x_i	...	x_k
Number of individuals n_i	n_1	n_2	...	n_i	...	n_k

The number n_i of individuals having category x_i of the characteristic X is called the **count** or **absolute frequency**. Thus, we have the notion of the **total count**, denoted by n, defined as

$$n = \sum_{i=1}^{k} n_i = n_1 + n_2 + \cdots + n_k,$$

which represents the size of the sample taken for statistical analysis.

The observations organized in the table form a statistical series (or statistical distribution), which consists of all the data and their corresponding counts, denoted by $\left\{(x_i, n_i)\,,\ i = \overline{1,k}\right\}$.

The **frequency** or **relative frequency** of the category x_i is the number $f_i = \frac{n_i}{n}$, which measures the proportion of individuals having category x_i in the sample. Note that the frequency f_i satisfies the following two properties:

1. For all $i \in \{1, \ldots, k\}$, we have $0 \leq f_i \leq 1$.

2. $\sum_{i=1}^{k} f_i = f_1 + f_2 + \cdots + f_k = 1$.

Exemple 1 *Let the sample below of size $n = 50$, taken from a discrete quantitative variable:*

1	4	3	5	1	6	3	1	6	1	5	2	1	4	6	1	6	6	4	2	1	5	1	3	5
2	4	2	6	1	3	2	4	3	1	4	5	6	1	5	2	2	4	4	2	5	1	1	3	2

The grouping of the observations (counting of frequencies) provides us with the following table:

X_i	*1*	*2*	*3*	*4*	*5*	*6*
n_i	*13*	*9*	*6*	*8*	*7*	*7*
$f_i = n_i/n$	*0.26*	*0.18*	*0.12*	*0.16*	*0.14*	*0.14*

1.2.2 Graphical representation of a qualitative variable

To graphically present a qualitative variable, several types of diagrams are available. Below are the most commonly used ones in practice.

- **Bar chart**: This type of representation is obtained by constructing as many columns as there are categories of the characteristic. These columns are rectangles with a constant base and height proportional to n_i (or f_i).
- **Pie chart**: The pie chart allows us to visualize the relative share of each category of the characteristic. The base of this representation is a circle divided into as many sectors as there are categories, such that the angle θ_i representing the share of x_i is given by:

$$\begin{array}{ccc} 360^\circ & \rightarrow & n \\ \theta_i = ? & \rightarrow & n_i \end{array} \Rightarrow \theta_i = 360^\circ \times \frac{n_i}{n} = 360^\circ \times f_i$$

Exemple 2 The distribution of workers in a company according to their qualification is summarized as follows:

Qualification	Workers	Employees	Technicians	Engineers	Total
Number of workers n_i	140	30	20	10	200
Relative frequency f_i	0.7	0.15	0.10	0.05	1

The angles corresponding to the distribution of workers according to qualification are:

$$\theta_1 = 360^\circ \times f_1 = 360^\circ \times 0.7 = 252^\circ \qquad \theta_2 = 360^\circ \times f_2 = 360^\circ \times 0.15 = 54^\circ$$
$$\theta_3 = 360^\circ \times f_3 = 360^\circ \times 0.10 = 36^\circ \qquad \theta_4 = 360^\circ \times f_4 = 360^\circ \times 0.05 = 18^\circ$$

The graphical presentation of the distribution of workers can be done using one of the following diagrams:

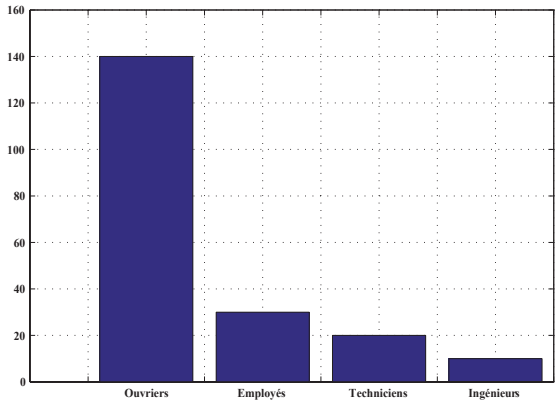


Diagramme en Tuyaux des effectifs

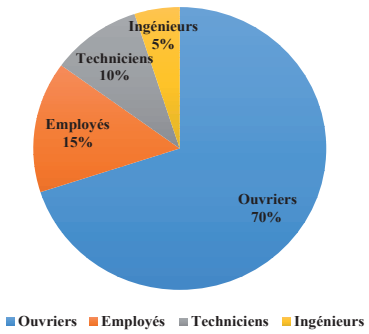


Diagramme circulaire des fréquences

1.2.3 Graphical representation of a discrete quantitative variable

The appropriate graph to represent a statistical series from a discrete quantitative variable is the **bar chart**, where each value x_i of the variable corresponds to a bar whose height is proportional to n_i or f_i.

Suppose that the statistical distribution of the number of rooms per dwelling in a certain locality is given as follows:

Number of rooms x_i	1	2	3	4	5	6	Total
Number of dwellings n_i	5	10	20	30	25	10	100
f_i	0.05	0.1	0.20	0.30	0.25	0.10	1

The corresponding diagram of this series is shown in Figure 1.6 (left). If we connect the tops of the bars, we obtain the **frequency polygon** (see Figure 1.6 (right)).

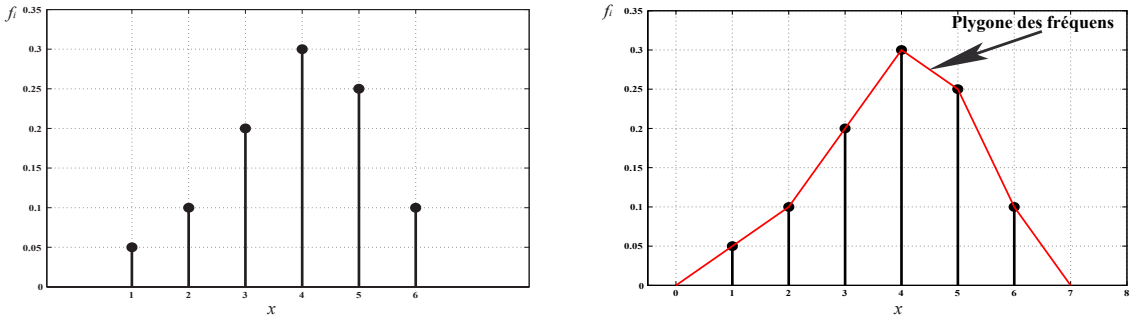


Figure 1.1: *Bar chart of room frequencies per dwelling.*

1.2.4 Cumulative Counts and Frequencies

Définition 1.1 *Cumulative counts (resp. cumulative frequencies) are defined as the number N_i (resp. F_i) such that:*

$$N_i = \sum_{j=1}^{i} n_j \quad (resp.\ F_i = \sum_{j=1}^{i} f_j),$$

The cumulative frequency F_i answers the question: what proportion of individuals have a value less than x_{i+1} or greater than or equal to x_i?

Définition 1.2 *The cumulative curve (or cumulative frequency polygon) (absolute or relative) is the graphical representation of these cumulative frequencies.*

For a discrete statistical variable, the cumulative curve is the representation of a step function whose horizontal steps have coordinates (x_i, F_i). This function is called the empirical distribution function, defined by:

$$\begin{array}{rcl} F & : & \mathbb{R} \longrightarrow [0,1] \\ x & \longrightarrow & F(x),\ such\ that \end{array}$$

$$F(x) = \begin{cases} 0 & if\ x < x_1 \\ f_1 & if\ x_1 \leq x < x_2 \\ f_1 + f_2 & if\ x_2 \leq x < x_3 \\ \sum_{j=1}^{i} f_j & if\ x_i \leq x < x_{i+1} \\ 1 & if\ x \geq x_k \end{cases}$$

Example: distribution of dwellings according to the number of rooms

$$F(x) = \begin{cases} 0 & if\ x < 1 \\ 0.05 & if\ 1 \leq x < 2 \\ 0.15 & if\ 2 \leq x < 3 \\ 0.35 & if\ 3 \leq x < 4 \\ 0.65 & if\ 4 \leq x < 5 \\ 0.90 & if\ 5 \leq x < 6 \\ 1 & if\ x \geq 6 \end{cases}$$

x_i	n_i	f_i	$F_i^{\nearrow}$	$F_i^{\searrow}$
1	5	0.05	0.05	0.95
2	10	0.10	0.15	0.85
3	20	0.20	0.35	0.65
4	30	0.30	0.65	0.35
5	25	0.25	0.90	0.10
6	10	0.10	1	0
Total	100	1		

Statistical table of increasing and decreasing cumulative frequencies for the number of rooms per dwelling.

Remarque 1.1 *The function F is discontinuous at each point of the statistical variable.*

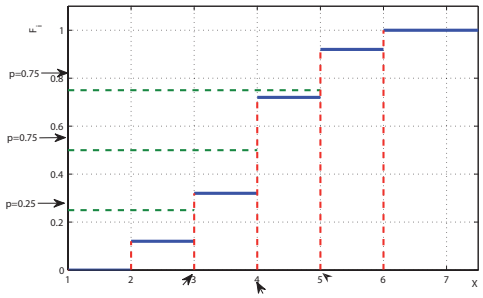


Figure 1.2: *Increasing cumulative curve for the number of rooms per dwelling.*

1.2.5 Graphical representation of a continuous quantitative variable

For a continuous variable, to establish the statistical table, it is necessary first to group the data into **classes**. A **class** is defined by its lower and upper bounds: by convention $[a_{i-1}, a_i[$. It is clear that this requires defining the number of classes and the amplitude associated with each.

Let the i^{th} class be given by $[a_{i-1}, a_i[$, which is fully defined by:

$\trianglerighteq$ The lower bound of the class: a_{i-1}.

$\trianglerighteq$ The upper bound of the class: a_i.

$\trianglerighteq$ The **amplitude** of the class: $A_i = a_i - a_{i-1}$.

$\trianglerighteq$ The class midpoint: $x_i = \frac{a_i + a_{i-1}}{2}$.

Remarque 1.2 *In practice:*

1. *Generally, classes of equal amplitude are chosen.*
2. *The choice of the number of classes and their amplitude depends on the total number of observations n.*
3. *Any reduction in the number of classes and any increase in amplitude leads to a loss of information.*

4. *Sturges' rule: Number of classes* $\simeq 1 + 3.3 \log N$, *and the amplitude is given by*

$$A = \frac{\max x_i - \min x_i}{number\ of\ classes}.$$

5. *The values of a continuous statistical variable are the class midpoints.*

The statistical table of a continuous variable series is generally presented as follows:

X	$[a_0, a_1[$	$[a_1, a_2[$	$\cdots$	$[a_{i-1}, a_i[$	$\cdots$	$[a_{m-1}, a_m[$
n_i	n_1	n_2	$\cdots$	n_i	$\cdots$	n_m

where n_i represents the number of observations in the i^{th} class, while the frequency is defined as in Section 1.2.1, i.e., $f_i = \frac{n_i}{n}$ with n the total sample size.

To graphically represent a continuous variable, we use the **histogram**, which is a generalization of the bar chart to the notion of classes. Two situations are possible: series with equal class widths and series with unequal class widths, as illustrated below.

Each class is represented by a rectangle whose **base** is the class amplitude and whose **height** is proportional to the **frequency** or **number of observations** (see Figure 1.3).

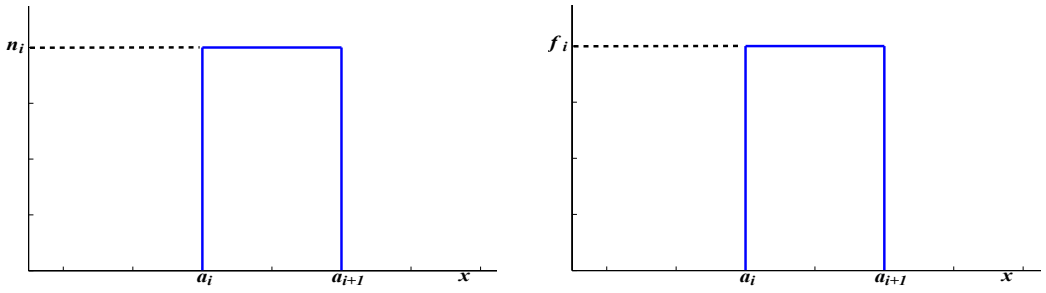


Figure 1.3: *Illustration of graphical representation of a class.*

The frequency polygon in this case is the line joining the midpoints of the top sides of the rectangles (see Figure 1.6).

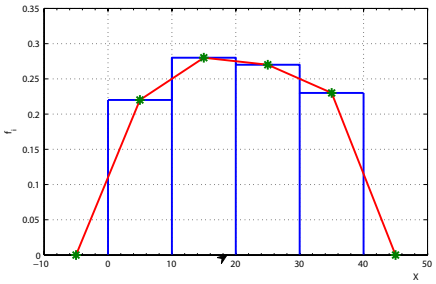


Figure 1.4: *The frequency polygon.*

Exemple 3 *The data below come from an experiment in which the plasma calcium concentration was measured in 40 individuals who received a hormonal treatment.*

X	*[10,16[*	*[16,22[*	*[22,28[*	*[28,34[*	*[34,40[*	*[40,46[*
n_i	*4*	*6*	*17*	*8*	*4*	*1*
f_i	*0.100*	*0.150*	*0.425*	*0.200*	*0.100*	*0.025*

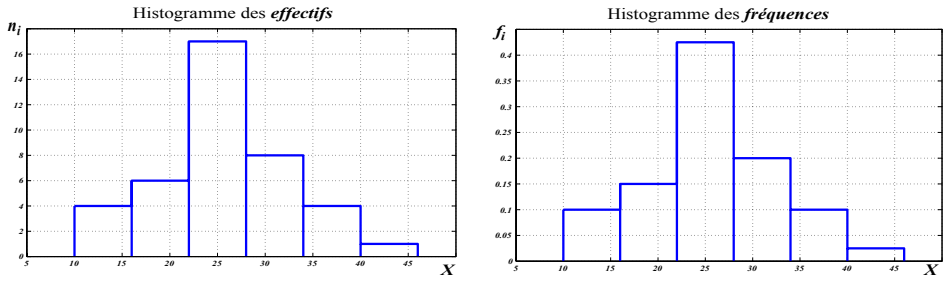


Figure 1.5: *Histogram of counts and Histogram of frequencies of the plasma calcium concentration.*

Exemple 4 *The distribution of a group of individuals according to their height (cm) is given in the following table:*

Height (cm)	n_i	f_i	$F_i^{\nearrow}$	$F_i^{\searrow}$
$[149.5, 159.5[$	*14*	*0.08*	*0.08*	*0.92*
$[159.5, 169.5[$	*32*	*0.18*	*0.26*	*0.74*
$[169.5, 179.5[$	*65*	*0.37*	*0.63*	*0.37*
$[179.5, 189.5[$	*47*	*0.27*	*0.9*	*0.1*
$[189.5, 199.5[$	*17*	*0.10*	*1*	*0*
Total	*175*	*1*		

Statistical table of increasing and decreasing cumulative frequencies for the distribution of the height(cm) of a group of individuals.

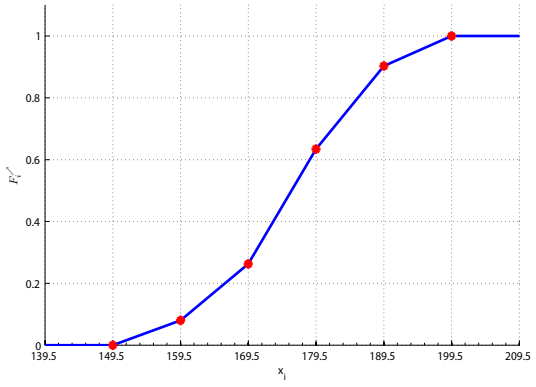


Figure 1.6: *Increasing cumulative curve.*

1.3 Numerical Characteristics of a Statistical Variable

When observing a graphical representation of a statistical series, two impressions may be noted:

1. The order of magnitude of the statistical variable, characterized by the values located at the center of the distribution; this is called the "central tendency characteristic" (or position measure).

2. The fluctuations of the observations around the central value; this is called the "measure of dispersion".

1.3.1 Measures of Central Tendency

They are intended to define the central values of the statistical series. These are: the mode, the median, and the arithmetic mean.

1.3.1.1 The Mode

a) Discrete Case

The mode, denoted by M_o, of a discrete statistical variable is the value with the highest frequency.

Exemple 5 *In Example 2 (Number of rooms per dwelling), $M_o = 4$ (because value 4 has the highest count).*

Remarque 1.3 *On a bar chart, the mode corresponds to the bar with the greatest height.*

b) Continuous Case

In this case we refer to the "modal class", which is the class with the highest frequency per unit width.

In Example 3, the calcium concentration in plasma, the modal class is $[22, 28[$.

Remarque 1.4 *In the continuous case, the mode may be taken as the midpoint of the modal class.*

For an approximate calculation, use the formula: for the modal class $[e_{i-1}, e_i[$,

$$M_o \in [e_{i-1}, e_i[$$

$$M_o = e_{i-1} + a_i \frac{\Delta_1}{\Delta_1 + \Delta_2},$$

where a_i is the class width,
Δ_1: excess of the modal class over the preceding class,
Δ_2: excess of the modal class over the following class.
Graphically: $M_o = e_{i-1} + d$ (d calculated using the scale).

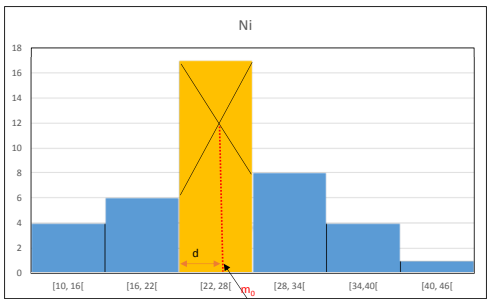


Figure 1.7: *The mode graphically*

1.3.1.2 The median

The median M_e is the value of the statistical variable that divides the individuals, arranged in ascending (or descending) order, into two equal groups.

Exemple 6 *Consider the dataset:* $\{12; 26; 6; 3; 32; 15; 21\}$.
Ordered: $\{3; 6; 12; 15; 21; 26; 32\}$.
The median is $M_e = 15$.

In general, the median of a statistical variable is the value M_e such that $F(M_e) = \frac{1}{2} = 0.5$, where F is the empirical distribution function.

Calculation of the median

a) Discrete case:

Let $x_{(1)}, x_{(2)}, ...x_{(n)}$ be the n observations arranged in increasing order:

$$x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}.$$

1st case: n odd, $n = 2p + 1$

$$\underbrace{x_{(1)} \leq x_{(2)} \leq ... \leq x_{(p)}}_{\text{p observations}} \leq \boxed{x_{(p+1)}} \leq \underbrace{x_{(p+2)} \leq ... \leq x_{(2p+1)}}_{\text{p observations}}.$$

$M_e = x_{(p+1)}$: the value at position $(p + 1)$.

2nd case: n even, $n = 2p$

$$\underbrace{x_{(1)} \leq x_{(2)} \leq ... \leq x_{(p)}}_{p \text{ observations}} \leq \boxed{x_{(p+1)}} \leq \underbrace{x_{(p+2)} \leq ... \leq x_{(2p)}}_{\text{p observations}}.$$

$M_e = \frac{x_{(p)} + x_{(p+1)}}{2}$, with $\left[x_{(p)}, x_{(p+1)}\right]$ the median interval.

In Example 2: number of children per household.

$n = 10 = 2 \times 5 \Longrightarrow M_e = \frac{x_{(5)} + x_{(6)}}{2} = \frac{1+1}{2} = 1.$

$$\begin{array}{cccccccccc} 0 & 0 & 1 & 1 & 1 & & 1 & 1 & ... \\ & & & & \uparrow & & \uparrow & & \\ & & & & X(5) & & X(6) & & \end{array}$$

Graphical determination of the median:

1. No horizontal plateau of the cumulative curve has ordinate (0.5).

Figure.

$$\begin{array}{rcl} 0.2 & < & 0.5 < 0.7. \\ F(0) & < & 0.5 < F(1). \\ & \Rightarrow & M_e = 1 \text{ child}. \end{array}$$

2. If a horizontal plateau of the cumulative curve has ordinate 0.5, the median is undetermined between two consecutive possible values:

$M_e = \frac{x_{i-1} + x_i}{2}$

b) Continuous case:

Since the function F is continuous and monotonic ($\nearrow$) between 0 and 1, the median is the unique solution of the equation $F(x) = \frac{1}{2}$.

Graphical determination of the median (interpolation method):

Let $[e_{i-1}, e_i[$ be the median class.

Figure.

$$\begin{aligned} F(e_{i-1}) &< 0.5 < F(e_i) \\ \tan\alpha &= \frac{F(e_i) - F(e_{i-1})}{e_i - e_{i-1}} = \frac{0.5 - F(e_{i-1})}{M_e - e_{i-1}} \\ &\Rightarrow \frac{f_i}{a_i} = \frac{0.5 - F(e_{i-1})}{M_e - e_{i-1}} \end{aligned}$$

Hence,

$$f_i(M_e - e_{i-1}) = a_i\,(0.5 - F(e_{i-1}))$$

And therefore,

$$M_e = e_{i-1} + a_i \frac{0.5 - F(e_{i-1})}{f_i}$$

a_i : class width
f_i : frequency of the median class.
Graphically:
$M_e = e_{i-1} + d$ (where d is calculated from the scale).

1.3.1.3 Arithmetic Mean

It is the sum of all observations divided by the total number of observations.

a) Discrete case:

Let X be a discrete variable taking the values $x_1, x_2, \ldots, x_k$ with corresponding frequencies $n_1, n_2, \ldots, n_k$ such that $\sum_{i=1}^{k} n_i = n$.

The arithmetic mean is denoted by

$$\begin{aligned} \overline{x} &= \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \frac{n_1x_1 + n_2x_2 + \ldots + n_kx_k}{n} \\ \overline{x} &= \sum_{i=1}^{k} f_i x_i \quad \text{(Weighted mean).} \end{aligned}$$

b) Continuous case:

The values are grouped into classes; by convention, we choose x_i as the midpoints of each class, and we use the same formula:

$$\begin{aligned} \overline{x} &= \sum_{i=1}^{k} f_i x_i \quad (x_i = \text{midpoint of the } i^{\text{th}} \text{ class}). \\ x_i &= \frac{e_{i-1} + e_i}{2} \end{aligned}$$

Algebraic properties of the arithmetic mean:
Property 1: (Change of origin)
Let $x_1, x_2, \ldots, x_k$ be the observed values of a statistical variable X and $n_1, n_2, \ldots, n_k$ their frequencies.

Let x_0 be a new origin.

Define the new variable
$y_i = x_i - x_0,\ \forall i = \overline{1,n}$, then $\overline{x} = \overline{y} + x_0$ with $\overline{y} = \sum_{i=1}^{k} f_i y_i$.
Indeed, since $x_i = y_i + x_0$,

$$\begin{aligned} \overline{x} &= \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \frac{1}{n}\sum_{i=1}^{k} n_i (y_i + x_0) \\ &= \frac{1}{n}\sum_{i=1}^{k} n_i y_i + \frac{1}{n}\sum_{i=1}^{k} n_i x_0 = \overline{y} + x_0. \end{aligned}$$

Remarque 1.5 *This change of variable is used to simplify calculations; in practice, one often chooses $x_0 = M_o$ or M_e.*

Property 2: (Change of scale and origin)
If we choose $y_i = \frac{x_i - x_0}{a}$, where x_0 and a are constants,
then $\overline{x} = a\overline{y} + x_0$.

Remarque 1.6 *In practice, we choose x_0 as: the median, the mode, or the class center.*
$a = gcd(a_i)$ (continuous case),
and the spacing between x_i (discrete case).

Proposition 1.1 *The sum of deviations from the arithmetic mean is zero:*
$\sum_{i=1}^{k} n_i(x_i - \overline{x}) = 0$.
Indeed: $\sum_{i=1}^{k} n_i(x_i - \overline{x}) = \sum_{i=1}^{k} n_i x_i - \sum_{i=1}^{k} n_i \overline{x} = n\overline{x} - n\overline{x} = 0$.

Proposition 1.2 *The sum of squared deviations from the arithmetic mean is minimal:*
$\varphi(a) = \sum_{i=1}^{k} n_i(x_i - a)^2$ *is minimal for* $a = \overline{x}$.

Relative position of the mode, median, and mean
Consider a unimodal statistical distribution.
1. When the distribution is symmetric, the three measures of central tendency coincide.

Figure

2. When the distribution is asymmetric, the median is generally between the mode and the mean, and is usually closer to the latter.

Figure

1.3.2 Measures of Dispersion

Exemple 7 *Consider the two statistical series:*
$X = \{6; 6; 7; 7; \boxed{8}; 9; 9; 10; 10\}$.
$Y = \{1; 2; 4; 6; \boxed{8}; 10; 12; 14; 15\}$.
We notice that X and Y have the same mean and the same median $\overline{x} = \overline{y} = M_e = 8$, but they are different: the first series X is less dispersed than the second.

Exemple 8 *Consider the statistical series X of k values arranged in increasing order:*
$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(k)}$.

1.3.2.1 Range

The range, denoted by "e", is the difference between the two extreme values: the smallest and the largest observed value.

$$e = \max_{1\leq i\leq k}(x_i) - \min_{1\leq i\leq k}(x_i) = x_{(k)} - x_{(1)}.$$

1.3.2.2 Variance and Standard Deviation

The empirical variance of the statistical variable X taking values x_i, $1 \leq i \leq k$, with frequencies n_i, $1 \leq i \leq k$, and $\sum_{i=1}^{k} n_i = n$, is:

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

The empirical standard deviation, denoted by σ_X, is:

$$\sigma_X = \sqrt{V(X)} = \sqrt{\frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \overline{x})^2} = \sqrt{\sum_{i=1}^{k} f_i(x_i - \overline{x})^2}.$$

Properties of the variance:

1. $V(X) \geq 0$.
2. $V(X) = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - \overline{x}^2 = \sum_{i=1}^{k} f_i x_i^2 - \overline{x}^2$.
3. Let $\overline{x}$ and $V(X)$ be the mean and variance of the statistical variable X.

   Define a new variable X' with mean $\overline{x}'$ and variance $V(X')$ such that:

$$x'_i = \frac{x_i - x_0}{a}, \quad i = \overline{1,k}$$

   where x_0 and a are constants. Then:

$$V(X) = a^2 V(X') \quad \text{and} \quad \sigma_X = a\,\sigma_{X'}.$$

Remarque 1.7 *When comparing two statistical series of the same nature, the one with the larger σ_X is the more dispersed.*

Coefficient of Variation:
It is a measure of relative dispersion defined by:

$$Cv = \frac{\sigma_X}{|\overline{x}|}.$$

Properties:

1. Cv is a dimensionless quantity.
2. Cv does not depend on the units used.
3. Cv makes it possible to compare two series expressed in different units.

1.3.2.3 Interquartile Range

a) Quantiles: these generalize the median.

· A quantile of order α $(0 \leq \alpha \leq 1)$, denoted by x_α, is the solution of the equation $F(x) = \alpha$.
That is, a proportion α of the individuals have the characteristic X less than x_α.

· Quartiles are commonly used.

These are the values of the variable x_i that divide the series into 4 equal parts. There are 3 quartiles, denoted Q_1, Q_2, Q_3, with Q_1 being the quantile of order $\frac{1}{4}$, Q_2 the quantile of order $\frac{1}{2}$, and Q_3 the quantile of order $\frac{3}{4}$.

That is, $F(Q_1) = \frac{1}{4}, F(Q_2) = \frac{1}{2} = F(M_e)$, and $F(Q_3) = \frac{3}{4}$.

· Interquartile Range:

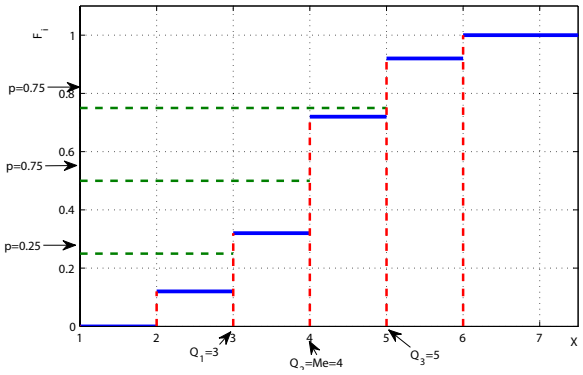
It contains 50% of the population, leaving 25% on each side.

The interquartile range is given by: $Q_3 - Q_1$.

• Practical Determination:

To determine $Q_3 - Q_1$, first calculate Q_1 and Q_3 using the same method used for finding the median.

• Discrete Case (graphically)



If $F(x_{i-1}) < 0.5 < F(x_i) \Rightarrow M_e = x_i$.
If $F(x_{i-1}) < 0.25 < F(x_i) \Rightarrow Q_1 = x_i$.
If $F(x_{i-1}) < 0.75 < F(x_i) \Rightarrow Q_3 = x_i$.
If $\forall x \in]x_{i-1}, x_i[, F(x) = 0.5 \Rightarrow M_e = \frac{x_i + x_{i-1}}{2}$.

Continuous Case:
Same method as for the median.
If $F(e_i) = 0.5$ (resp. 0.25, 0.75), then $e_i = M_e$ (resp. Q_1, Q_3).
If $F(e_{i-1}) \leq 0.5 \leq F(e_i)$, then:

$$M_e = e_{i-1} + a_i \frac{0.5 - F(e_{i-1})}{f_i}, \quad \text{with } [e_{i-1}, e_i[\text{ being the median class.}$$

$$Q_1 = e_{i-1} + a_i \frac{0.25 - F(e_{i-1})}{f_i}, \quad Q_1 \in [e_{i-1}, e_i[.$$

$$Q_3 = e_{i-1} + a_i \frac{0.75 - F(e_{i-1})}{f_i}, \quad Q_3 \in [e_{i-1}, e_i[.$$