Chapter 1: Basic Definitions

Introduction

Statistics (statistique) is a set of methods used to collect, organize, present, analyze, and interpret data (données) obtained from observations, experiments, or measurements.

Statistical analysis is generally divided into two main parts:

1. Descriptive statistics (statistique descriptive)
2. Inferential statistics (statistique inférentielle)

1. Descriptive statistics

Descriptive statistics (statistique descriptive) aims to describe, summarize, and represent a set of data (données) in a simple and meaningful way, without making conclusions beyond the data that have been observed.

It uses several types of tools:

- Tables (tableaux statistiques) to organize the observations;
- Graphical representations (représentations graphiques) such as bar charts, histograms, pie charts, and cumulative-frequency curves;
- Measures of position (paramètres de position) such as the mean, median, and mode;
- Measures of dispersion (paramètres de dispersion) such as the range, variance, and standard deviation;
- Measures of relationship (paramètres de relation) such as covariance and correlation, which describe the relationship between two or more variables.

For example, suppose that the examination marks of five students are: 10,12,12,14,17.

Descriptive statistics can be used to calculate the mean, identify the median and mode, determine the range and standard deviation, and represent the data graphically.

The objective is simply to describe the observed data.

2. Inferential statistics (الإحصاء الاستدلالي)

Inferential statistics (statistique inférentielle) uses information obtained from a sample (échantillon) to draw conclusions about a larger population (population statistique).

It includes methods such as:

- Estimation (estimation) of unknown population parameters;

- Confidence intervals (intervalles de confiance);
- Hypothesis testing (tests d'hypothèses);
- Statistical models (modèles statistiques);
- Regression and prediction (régression et prédiction).

For example, suppose we want to study the average examination mark of all students in a university. It may be difficult to collect the marks of every student. We can therefore select a sample (un échantillon) of students, calculate the average mark of the sample, and use this information to estimate the average mark of the entire population.

Thus, while descriptive statistics summarizes and describes the data that have been observed, inferential statistics uses observed data to make conclusions or estimates about a larger population.

I. statistics language

1. Population and sample

In statistics, two important concepts are the population (population statistique) and the sample (échantillon).

A population is the complete set of individuals, objects, measurements, or observations that we want to study.

A sample is a subset of the population selected for observation or analysis. The number of individuals in a sample, denoted by (n), is called the **sample size (taille de l'échantillon)**.

An **individual (individu)** is one element of the population being studied.

A **variable (variable)**, also called a **characteristic (caractère)** and usually denoted by **X**, is a property or characteristic of the individuals in a **population (population statistique)** that we want to study. The value taken by the variable may be different from one individual to another.

Example:
If the population consists of all second-year Physics students at a university,

- **one student** is an individual of this population.

a group of 10 selected students constitutes **a sample**; The **sample size (taille de l'échantillon)** is [n=10.].

- **the measured** height of each student is an **observation** (observation);
- **height** is a **statistical variable** (variable statistique).

The distinction between population and sample is particularly important in inferential statistics because conclusions about a population are generally obtained from information collected from a sample.

Remark: The role of statistics in physics

Statistics is particularly important in physics, because physical experiments generally involve measurements affected by experimental variability and measurement uncertainty.

For example, when measuring the period of a pendulum several times, we may obtain:
1.98,2.01,2.00,1.97,2.03 s

These measurements are not exactly identical. Statistical methods allow us to:

- summarize the measurements;
- calculate their mean value;
- quantify their dispersion;
- study relationships between physical quantities;
- estimate unknown physical parameters;
- evaluate experimental results and their uncertainty.

Therefore, statistics provides useful tools for analyzing experimental data and extracting meaningful information from physical measurements.

Example:
Suppose that we want to study the heights of 30 students. We can organize the collected data in a table, represent them using a graph, and calculate the mean height. These are examples of descriptive statistics.

2. Random Variable (variable aléatoire)

A **random variable (variable aléatoire)** is a variable whose value depends on the outcome of a random experiment or on the individual selected from a population. Its value may therefore vary from one observation to another.

For example, if we select a student at random and record their height, the observed height depends on the student selected. Therefore, the height can be considered a **random variable (variable aléatoire)**.

Similarly, if we select a student at random and record their examination mark, the obtained mark is a value of a random variable.

In statistics, variables are mainly divided into two types:

1. **Qualitative variables (variables qualitatives)**

2. **Quantitative variables (variables quantitatives)**

2.1. Qualitative Variable (variable qualitative)

A **qualitative variable (variable qualitative)** describes a characteristic that is not expressed as a numerical measurement. Its possible values are called **categories (modalités)**.

Qualitative variables describe qualities, attributes, or categories of individuals.

Examples:

- Eye color: blue, brown, green, etc.
- Gender: male, female, etc.
- Blood group: A, B, AB, O.
- Type of material: metal, plastic, glass, etc.
- Type of transport: car, bus, train, etc..

A **qualitative variable (variable qualitative)** describes a characteristic that is not expressed by a numerical measurement.

For example:

Eye color is the variable we want to study.

Different students can have different eye colors. The possible values of this variable are: Blue,Brown,Green,Grey . Each of these possible values is called a **category (modalité)**.

So:

- **Variable:** eye color
- **Categories (modalités):** blue, brown, green, grey
- **One observation:** if a particular student has blue eyes, the observed value is **blue**.

We can represent this as: **Variable**=Eye color **and Categories**={Blue, Brown, Green, Grey

The categories are not numerical values, even if numbers may sometimes be used as codes to represent them

The important idea is to understand **what "order" means** in an ordinal qualitative variable.

What is an ordinal qualitative variable?

An **ordinal qualitative variable (variable qualitative ordinale)** is a qualitative variable whose categories can be placed in a **natural and meaningful order**.

For example, consider the variable:

Level of satisfaction (niveau de satisfaction)

We could have the following categories:

Very dissatisfied < Dissatisfied < Neutral < Satisfied < Very satisfied

There is a logical order from the lowest level of satisfaction to the highest. This ordering has a real meaning. Therefore:

Satisfaction level is an ordinal qualitative variable**. What does "natural order" mean?**

It means that we can say that one category is **higher or lower** than another according to the characteristic being studied.

However, the categories themselves are **not numerical measurements**.

For example, we cannot say that: Very satisfied=5 in the same sense as 55 metres or 55 kilograms.

We could **code** the categories as 1,2,3,4,5 but these numbers are only labels representing the order.

Compare it with a non-ordinal qualitative variable

This distinction is very important.

Consider: **Eye color (couleur des yeux)**

The categories might be: Blue,Brown,Green,Grey

We cannot naturally say: Blue<Brown<Green<Grey.\

There is **no natural statistical order** among these categories.

Therefore:

- **Eye color → nominal qualitative variable (variable qualitative nominale)**
- **Satisfaction level → ordinal qualitative variable (variable qualitative ordinale)**

If the categories can be naturally ordered → ordinal qualitative variable.

If the categories cannot be naturally ordered → nominal qualitative variable.

For example:

Variable	**Categories**	**Natural order? Type**

Variable	Categories	Natural order?	Type
Eye color	Blue, Brown, Green	No	Nominal qualitative
Blood group	A, B, AB, O	No	Nominal qualitative
Satisfaction	Very dissatisfied → Very satisfied	Yes	Ordinal qualitative
Education level	Primary → PhD	Yes	Ordinal qualitative
Performance	Poor → Excellent	Yes	Ordinal qualitative

One important point

An ordinal variable tells us **which category comes before or after another**, but it does **not necessarily tell us how large the difference is** between two categories.

For example: Dissatisfied < Neutral < Satisfied

tells us the order, but we cannot conclude that the difference between "Dissatisfied" and "Neutral" is exactly the same as the difference between "Neutral" and "Satisfied."

That is the key difference between an **ordinal qualitative variable** and a **quantitative variable**.

2.2. Quantitative Variable (variable quantitative)

A **quantitative variable (variable quantitative)** is a variable whose values are numerical and represent quantities that can be measured or counted. in other words, a discrete variable (variable discrète) is a quantitative variable whose possible values form a finite or countable set.

Examples include:

- height;
- age;
- weight;
- temperature;
- examination mark;
- number of children in a family;
- number of particles detected by a detector.

Quantitative variables are divided into two types:

a. Discrete quantitative variables (variables quantitatives discrètes)

b. Continuous quantitative variables (variables quantitatives continues)

a. Discrete Variable (variable discrète)

A **discrete variable (variable discrète)** is a quantitative variable that can take a finite or countable number of distinct values.

It usually results from **counting (dénombrement)**. A discrete variable generally takes separated values.

Examples:

- Number of students in a classroom: 20,21,22,23,…
- Number of children in a family: 0,1,2,3,…
- Number of particles detected by a detector: 0,1,2,3,…

2.2. Continuous Variable (variable continue)

A **continuous variable (variable continue)** is a quantitative variable that can take any value within an interval. It usually results from **measurement (mesure)**.

In practice, a continuous variable is recorded with a limited number of decimal places because measuring instruments have limited precision. However, theoretically, it can take any value within a given interval.

Remark: The distinction between discrete and continuous variables is important when constructing statistical tables and choosing an appropriate graphical representation.

Très bien. Je vous propose de reprendre la section **One-Variable Statistical Series** dans une version entièrement adaptée à Word, sans notation LaTeX.

III. One-Variable Statistical Series (Séries statistiques à une variable)

A **statistical series (série statistique)** is a set of observations collected from a group of individuals belonging to the same **population (population statistique)**.

When only one **statistical variable (variable statistique)** is studied for each individual, we obtain a **one-variable statistical series (série statistique à une variable)**.

For example, suppose that we want to study the examination marks of a group of students. For each student, we record only one characteristic: the examination mark.

The observations may be:

10, 12, 15, 10, 14, 12, 8, 15, 10, 12

This is a one-variable statistical series because only one variable, the **examination mark (note d'examen)**, is studied.

Chapter 1: Basic Definitions

1. Raw Data (Données brutes)

The observations collected during a statistical study are initially called **raw data (données brutes)**.

For example, the examination marks of 10 students may be:

12, 15, 10, 12, 8, 15, 14, 10, 12, 16

These data are called raw data because they are presented without any particular organization.

We can reorganize them in ascending order:

8, 10, 10, 12, 12, 12, 14, 15, 15, 16

This is called an **ordered statistical series (série statistique ordonnée)**.

Ordering the observations makes it easier to identify the smallest value, the largest value, repeated values, and the different statistical values.

2. Statistical Values (Valeurs statistiques)

The different values taken by a variable are called **statistical values (valeurs statistiques)**.

For a qualitative variable, these values are generally called **categories or modalities (modalités)**.

For a quantitative variable, they are numerical values.

Example:

Consider the following examination marks:

10, 12, 10, 14, 12, 10

The different statistical values are:

10, 12, 14

Notice that 10 appears three times and 12 appears twice. When we list the different statistical values, each value is written only once.

---

3. Count or Frequency of Occurrence (Effectif)

The **count (effectif)** of a statistical value is the number of times that value occurs in the data set.

The count associated with a value x_i is denoted by n_i.

Example:

Consider:

10, 12, 10, 14, 12, 10

The value 10 appears three times.

Therefore:

$n_1 = 3$

The value 12 appears two times:

$n_2 = 2$

The value 14 appears one time:

$n_3 = 1$

The **total count (effectif total)** is the total number of observations.

It is denoted by N.

Here:

$N = 3 + 2 + 1 = 6$

Therefore, there are 6 observations in the series.

---

4. Frequency (Fréquence)

The **frequency (fréquence)** of a statistical value indicates the proportion of observations having that value.

The frequency is calculated as:

Frequency = Count / Total number of observations

Using symbols: $f_i = \frac{n_i}{N}$

where:

- f_i is the frequency of the value x_i;
- n_i is the count of x_i;
- N is the total number of observations.

The frequency can be expressed as a decimal number or as a percentage.

Example:

Suppose that 3 students out of 10 obtained the mark 12.

Then:

f = 3 / 10 = 0.3 = 30%

Therefore, the frequency of the mark 12 is 0.3, or 30%.

Important property

The sum of all frequencies is equal to 1:

$f_1 + f_2 + ... + f_n = 1$

When frequencies are expressed as percentages, their sum is equal to 100%.

---

5. Statistical Frequency Table (Tableau statistique)

A statistical table allows us to organize the data in a clear way.

Consider the following examination marks:

10, 12, 10, 14, 12, 10

The corresponding statistical table is:

Mark x_i	**Count n_i**	**Frequency f_i**
10	3	0.50 = 50%
12	2	0.333 = 33.3%
14	1	0.167 = 16.7%
Total	**6**	**1 = 100%**

This table gives us a much clearer representation of the data than the original list.

---

6. Ascending Cumulative Count (Effectif cumulé croissant)

When the statistical values are ordered from the smallest to the largest, we can calculate the **ascending cumulative count (effectif cumulé croissant)**.

The ascending cumulative count corresponding to a value x_i is the number of observations whose value is less than or equal to x_i.

Example:

Consider the following table:

Mark x_i	**Count n_i**	**Ascending cumulative count**
10	3	3
12	2	5
14	1	6

For the value 10:

Ascending cumulative count = 3

This means that 3 students obtained a mark less than or equal to 10.

For the value 12:

Ascending cumulative count = 3 + 2 = 5

This means that 5 students obtained a mark less than or equal to 12.

For the value 14:

Ascending cumulative count = 3 + 2 + 1 = 6

Therefore, all 6 students obtained a mark less than or equal to 14.

Important property

The final ascending cumulative count is always equal to the total number of observations:

Final cumulative count = N

---

7. Ascending Cumulative Frequency (Fréquence cumulée croissante)

The **ascending cumulative frequency (fréquence cumulée croissante)** represents the proportion of observations whose value is less than or equal to a given value.

It can be calculated using:

Cumulative frequency = Cumulative count / Total number of observations

Example:

Mark x_i	**Count n_i**	**Cumulative count**	**Cumulative frequency**
10	3	3	3/6 = 0.50 = 50%
12	2	5	5/6 = 0.833 = 83.3%
14	1	6	6/6 = 1 = 100%

For the mark 12, the cumulative frequency is 83.3%.

This means that:

83.3% of the students obtained a mark less than or equal to 12.

Important property

The final cumulative frequency is always:

1 or 100%

---

8. Complete Statistical Table

The different quantities can be combined in one table.

Statistical value x_i	Count n_i	Frequency f_i	Cumulative count	Cumulative frequency
10	3	0.50	3	0.50
12	2	0.333	5	0.833
14	1	0.167	6	1.00
Total	**6**	**1.00**	**6**	**1.00**

This type of table is one of the most important tools in the analysis of a one-variable statistical series.

---

9. Graphical Representation

Statistical data can also be represented graphically.

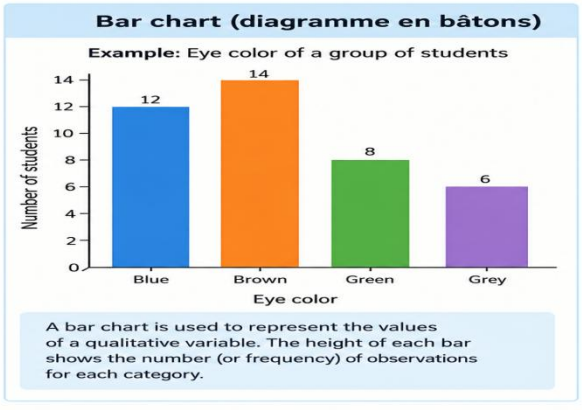
The appropriate graphical representation depends on the type of variable.

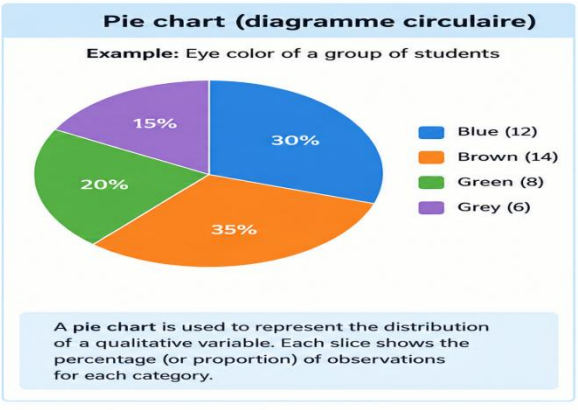
9.1. Qualitative Variable

For a qualitative variable, we can use:

- **Bar chart (diagramme en bâtons)**
- **Pie chart (diagramme circulaire)**

For example, if we study the eye colors of a group of students, we can represent the number of students with blue, brown, green, and grey eyes using a bar chart.

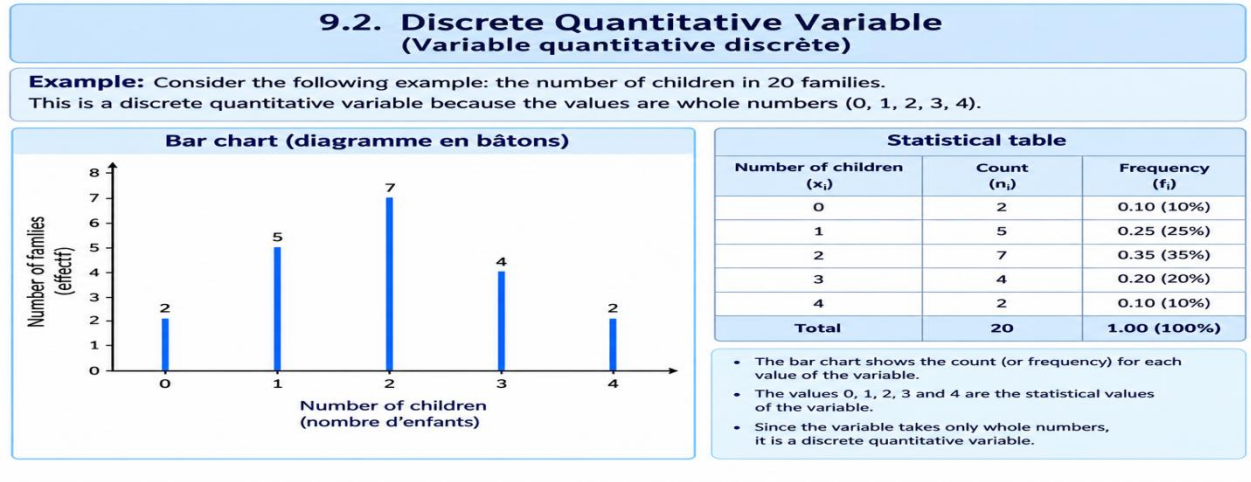




9.2. Discrete Quantitative Variable

For a discrete quantitative variable, a **bar chart (diagramme en bâtons)** can be used.

The statistical values are placed on the horizontal axis, and their counts or frequencies are represented on the vertical axis.



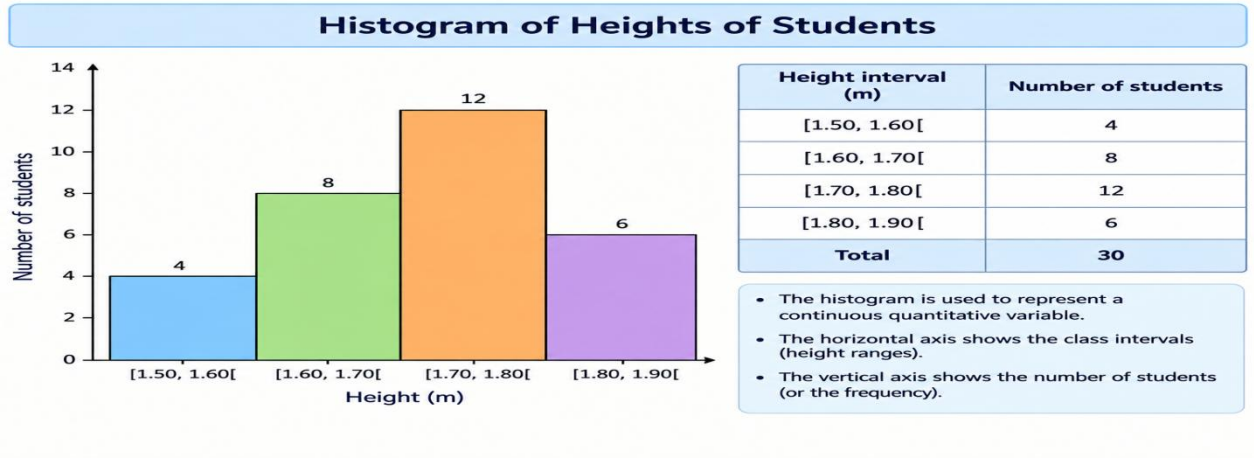
9.3. Continuous Quantitative Variable

For a continuous quantitative variable, the observations are usually grouped into intervals.

For example:

Height interval	**Number of students**
[1.50, 1.60[	4
[1.60, 1.70[	8
[1.70, 1.80[	12
[1.80, 1.90[	6

A **histogram (histogramme)** is generally used to represent such data.



Height interval (m)	Number of students
[1.50, 1.60[	4
[1.60, 1.70[	8
[1.70, 1.80[	12
[1.80, 1.90[	6
Total	**30**

- The histogram is used to represent a continuous quantitative variable.
- The horizontal axis shows the class intervals (height ranges).
- The vertical axis shows the number of students (or the frequency).

9.4. Ascending Cumulative Curve

Cumulative counts or cumulative frequencies can be represented by a **cumulative curve (courbe cumulative)**.

For an ascending cumulative frequency, the curve represents the proportion of observations satisfying:

$X \leq x$

In words, it represents the proportion of observations whose value is less than or equal to x.

---

12. Complete Example

Consider the following examination marks obtained by 20 students:

10, 12, 8, 10, 14, 12, 10, 15, 8, 12,

10, 14, 12, 15, 10, 8, 12, 14, 10, 12

Step 1: Identify the population

The population consists of 20 students.

Step 2: Identify the variable

The variable is the **examination mark (note d'examen)**.

It is a **quantitative variable (variable quantitative)**.

Step 3: Identify the different statistical values

The different marks are:

8, 10, 12, 14, 15

Step 4: Count the occurrences

Mark x_i	**Count n_i**
8	3
10	6
12	6
14	3
15	2
Total	**20**

Step 5: Calculate the frequencies

For the mark 8:

f = 3 / 20 = 0.15 = 15%

For the mark 10:

f = 6 / 20 = 0.30 = 30%

The complete table is:

Mark x_i	**Count n_i**	**Frequency f_i**	**Cumulative count**	**Cumulative frequency**
8	3	15%	3	15%
10	6	30%	9	45%
12	6	30%	15	75%
14	3	15%	18	90%
15	2	10%	20	100%
Total	**20**	**100%**	**20**	**100%**

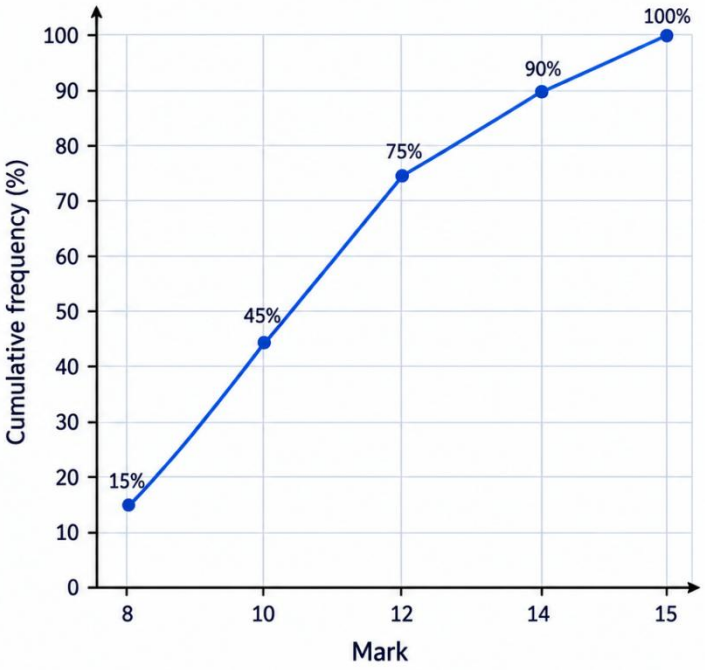
Cumulative Curve (Ogive)

Examination marks of 20 students

Mark x_i	Count n_i	Cumulative count N_i	Cumulative frequency F_i (%)
8	3	3	15%
10	6	9	45%
12	6	15	75%
14	3	18	90%
15	2	20	100%
Total	**20**	**20**	**100%**

The cumulative curve represents the cumulative frequency (or cumulative count) for each value of the variable.

In this example, the curve shows the proportion of students whose mark is less than or equal to a given value ($x \leq x_i$).



Interpretation

From the table, we can make several observations.

- 3 students obtained a mark of 8.
- 6 students obtained a mark of 10.
- 6 students obtained a mark of 12.
- 3 students obtained a mark of 14.
- 2 students obtained a mark of 15.

The cumulative count corresponding to 12 is 15.

Therefore:

15 students obtained a mark less than or equal to 12.

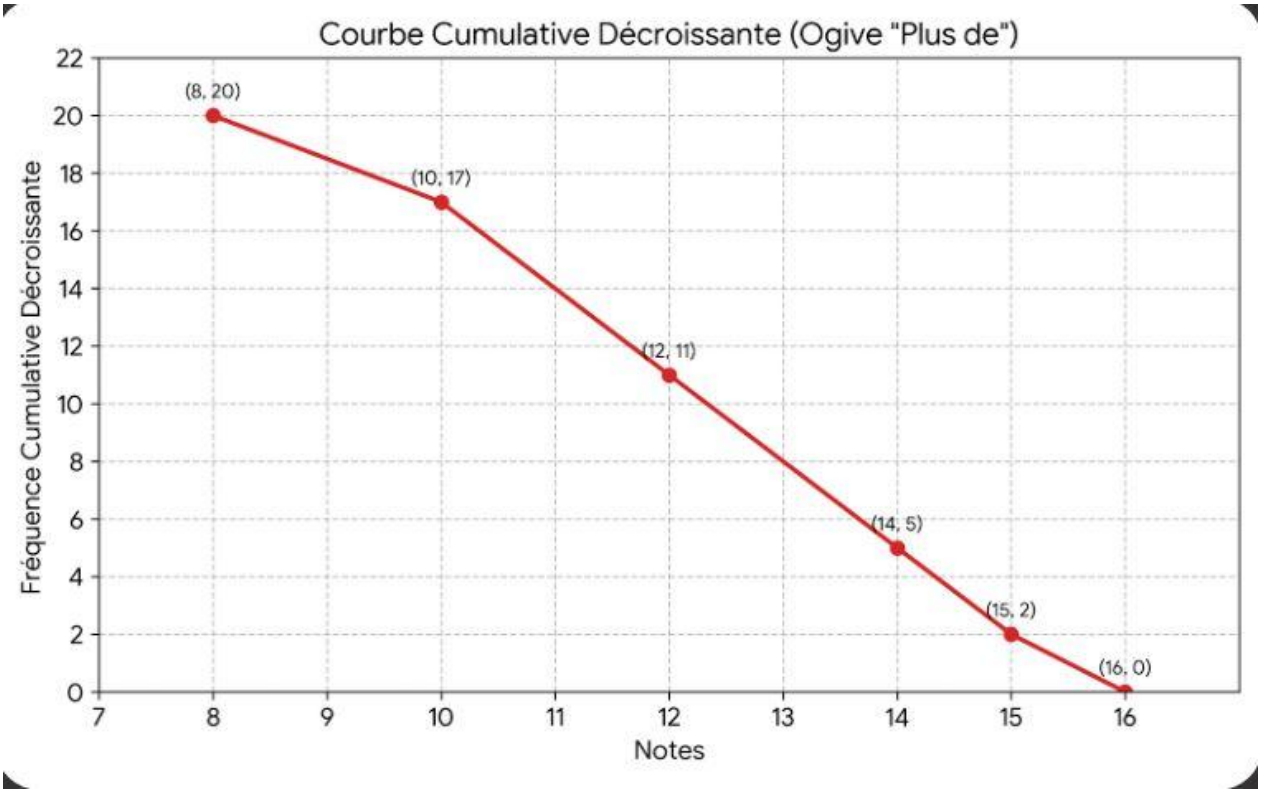
The cumulative frequency corresponding to 12 is 75%.

Therefore:

75% of the students obtained a mark less than or equal to 12.

This is one of the main advantages of a statistical table: it allows us not only to organize the data, but also to interpret them easily.

9.5. Ascending Cumulative curve



Descending Cumulative Curve (More-Than Ogive)

Data Overview

The following is the distribution of examination marks obtained by 20 students:

- **Dataset:** 10, 12, 8, 10, 14, 12, 10, 15, 8, 12, 10, 14, 12, 15, 10, 8, 12, 14, 10, 12

- **Total Number of Students (N):** 20

1. Frequency & Descending Cumulative Frequency Table

The descending cumulative frequency indicates the number of students who scored **equal to or greater than** a specific mark.

Marks (X)	**Frequency (f)**	**Calculation**	**Descending Cumulative Frequency (DCF)**	**Coordinates to Plot (X,DCF)**
8	3	All students scored >=8	**20**	(8, 20)
10	6	$20 - 3$	**17**	(10, 17)
12	6	$17 - 6$	**11**	(12, 11)
14	3	$11 - 6$	**5**	(14, 5)
15	2	$5 - 3$	**2**	(15, 2)
16	0	$2 - 2$ *(baseline anchor)*	**0**	(16, 0)

2. Steps to Draw the Curve manually in Word / Graph Paper

1. **X-Axis (Horizontal):** Plot the **Marks** (range from 7 to 17).
2. **Y-Axis (Vertical):** Plot the **Cumulative Frequency** (range from 0 to 22).
3. **Plot Points:**

3. Key Observations & Interpretation

- **Point (8, 20):** All 20 students (100%) scored 8 or higher.

- **Point (10, 17):** 17 students scored 10 or higher (excluding the 3 students who scored 8).

- **Point (12, 11):** 11 students scored 12 or higher.

- **Point (15, 2):** Only 2 students scored the maximum mark of 15.