

# Module : Introduction aux probabilités et statistique descriptive

## Introduction :

La statistique descriptive est une Méthode d'analyse numérique que les scientifiques utilisent pour recueillir, classer, présenter et analyser des informations provenant des ensembles comportant un grand nombre d'éléments.

Son utilisation intensive dans de nombreux domaines tels ~~les sciences~~ expérimentales (Biologie, physique, agronomie, ... Mathématique ...)

## Chapitre 01: Notions de base et vocabulaire statistique

- Population <sup>مجتمع إحصائي</sup>: c'est l'ensemble de tous les éléments concernés par l'étude.
- L'individu <sup>الأفراد</sup>: c'est chaque élément de la population.
- Caractère <sup>صية</sup> (variable) <sup>متغير</sup> statistique: c'est la propriété caractéristique des éléments de la population, les différents caractères étudiés habituellement sont: l'âge, la taille, le poids, la nationalité, le groupe sanguin,...

~~L'observation: les différents résultats d'un caractère associé aux individus d'une population sont dits des observations.~~

- L'échantillon: est un sous ensemble de la population souvent le nombre des individus d'une population est assez grand alors le traitement des résultats sera très délicat, dans ce cas on doit prendre un sous ensemble de la population choisi aléatoirement pour avoir toutes les propriétés qui existent dans la population.
- Taille: le nombre des éléments de l'échantillon.

Les caractères statistique se décomposent en deux types:

- Caractère quantitatif: c'est un caractère mesurable autrement dit on peut associer à chaque individu de la population une valeur réelle, par exemple: la taille, le poids, la moyenne des étudiants.
- Caractère qualitatif: Tout caractère qu'on ne peut pas mesurer, par exemple: la nationalité, la couleur des yeux, la profession, le groupe sanguin.

Remarque: Par fois, pour des raisons de traitement ou essaye de traduire à l'aide d'un codage les caractères qualitatifs en nombres réels.

- Modalité: Le même caractère peut changer l'état d'un individu à un autre, habituellement lorsque le caractère est qualitatif chaque état est dite une modalité, lorsque le caractère est quantitatif chaque valeur réelle est une modalité.

Exemple: Un échantillon composé de 6 étudiants parmi les nouveaux bacheliers, on s'intéresse à leur distribution selon les spécialités choisies.

Nbr d'étudiants	1	2	3	4	5	6
Spécialités	SM	MI	MI	ST	MI	SM

Population: les nouveaux bacheliers.

Individu: les nouveaux bacheliers

Caractère étudié: la spécialité choisie

Nature: qualitative.

Modalités: 3 (SM, MI, ST)

## Tableaux statistiques et représentation graphiques.

Modalités $x_i$	$x_1$	$x_2$	$x_3 \dots x_i$	$x_h$
Nbr. d'individu $n_i$	$n_1$	$n_2$	$n_3 \dots n_i$	$n_h$

Effectif: (fréquence absolue) est le nbr  $n_i$  d'individu ayant pris la modalité  $x_i$  du caractère  $X$ .

Les observations ordonnées forment une série statistique (ou distribution statistique) qui est constituée de l'ensemble des données et les effectifs correspondants.

$\{(x_i, n_i), i=1, k\}$  série statistique

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

• Fréquence (fréquence relative) de la modalité  $x_i$

$f_i = \frac{n_i}{n}$  : c'est la proportion d'individus ayant pris la modalité  $x_i$ .

$$\sum_{i=1}^k f_i = f_1 + f_2 + \dots + f_k = 1$$

Spécialité	Effectif $n_i$	Fréquence $f_i$
SM	2	$\frac{2}{6} = 0,33$
MI	3	$\frac{3}{6} = \frac{1}{2}$
ST	1	$\frac{1}{6} \approx 0,17$
$\Sigma$	6	1

• Représentation graphique d'une variable qualitative:

(a) Représentation circulaire par des secteurs:

$n$  (l'effectif total)  $\rightarrow 360^\circ$

$n_i$   $\rightarrow d_i$  (dégéné)

$$d_i = \frac{n_i \times 360^\circ}{n}$$

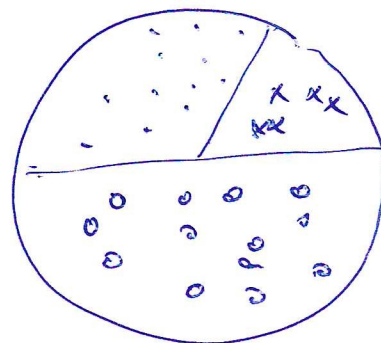
exemple

Spécialité	SM	MI	ST
effectif $n_i$	2	3	1

$$6 \rightarrow 360^\circ \Rightarrow d_1 = \frac{2 \times 360^\circ}{6} = 120^\circ$$

$$d_2 = 180^\circ, \quad d_3 = 60^\circ$$

 SM  
 MI  
 ST



(b) Représentation en tige et feuille d'origine:

exemple

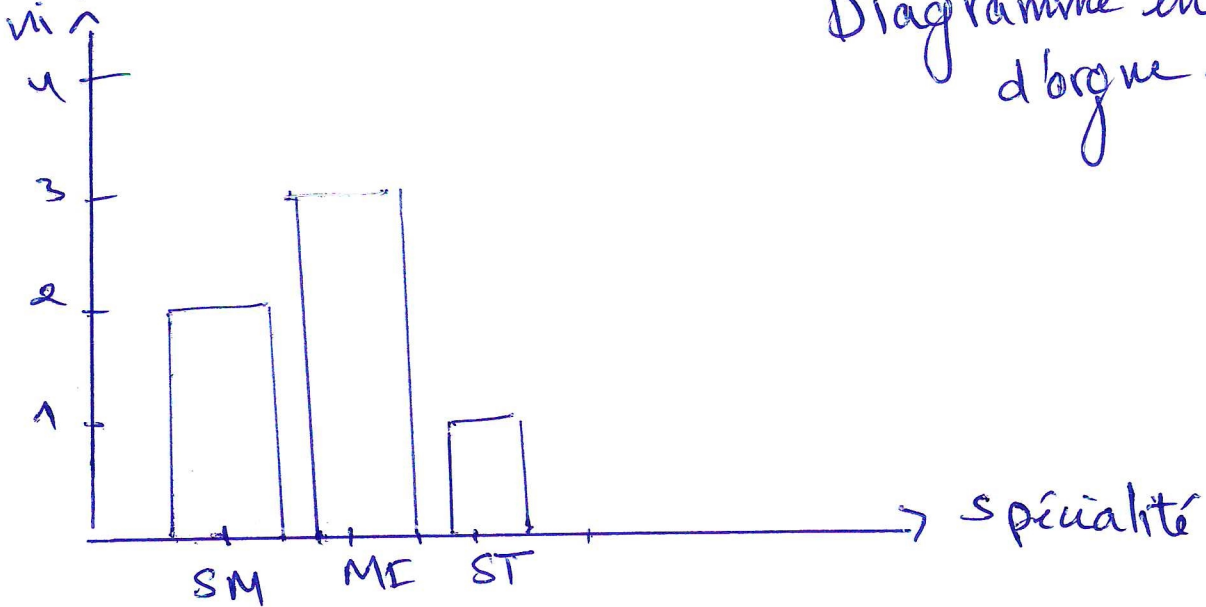
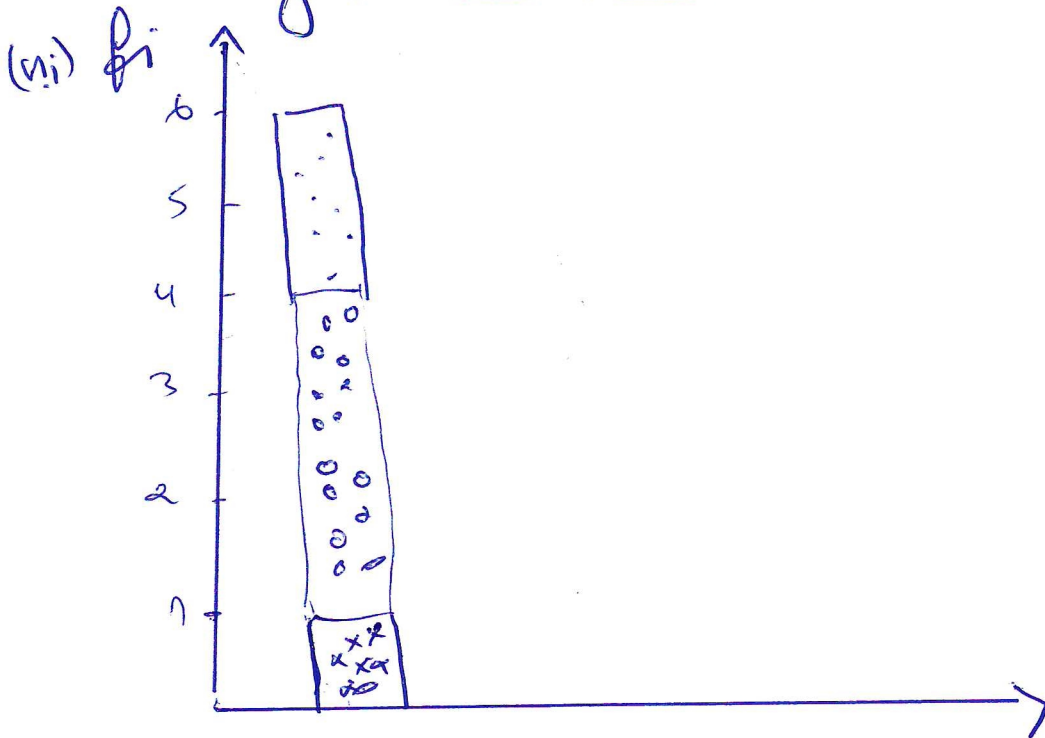


Diagramme en tige et feuille d'origine.

(c) Diagramme en bandes



SM	2
ME	3
ST	1

\* Variable quantitative : Il y a deux types de caractères (variables) quantitatives.

(a) Un caractère discret (ou discontinu) : Un caractère quantitatif est dit discret lorsqu'il ne peut prendre que des valeurs isolées dans son intervalle de variation, par exemple : le nbr d'enfants dans chaque famille dans un ensemble de familles, les modalités : 0, 1, 2, 3, ...

~~(b) Un caractère continu :~~

Exemple : Dans une cité résidée par 64 familles on s'intéresse au nombre d'enfants chez chaque famille parmi les 64, les résultats sont donnés par : (0, 16), (1, 18), (2, 14), (3, 11), (4, 3), (5, 2)

$\begin{matrix} \nearrow & \searrow \\ \text{nbr d'enfant} & \text{nbr de famille} \end{matrix}$

$x_i$	0	1	2	3	4	5	
$n_i$	16	18	14	11	3	2	64
$f_i$	$\frac{16}{64} = 0,25$	0,291	0,218	0,172	0,047	0,031	1

Population : les familles

Individu : chaque famille

Taille : 64

Caractère étudié : nbr d'enfant chez chaque famille

Nature : Quantitatif discret

Modalités :  $x_i \in \{0, 1, 2, 3, 4, 5\}$



• La représentation graphique d'une variable quantitative discrète :

• Le diagramme en bâtons : Il consiste à représenter les diverses valeurs  $x_i$  prises par une variable discrète en fonction des fréquences  $f_i$  ou des effectifs  $n_i$ , de la manière suivante :

On représente en abscisse les observations  $x_i$  et au dessus de chaque valeur on trace un bâton dont la hauteur est proportionnelle à  $f_i$  ou  $n_i$ .

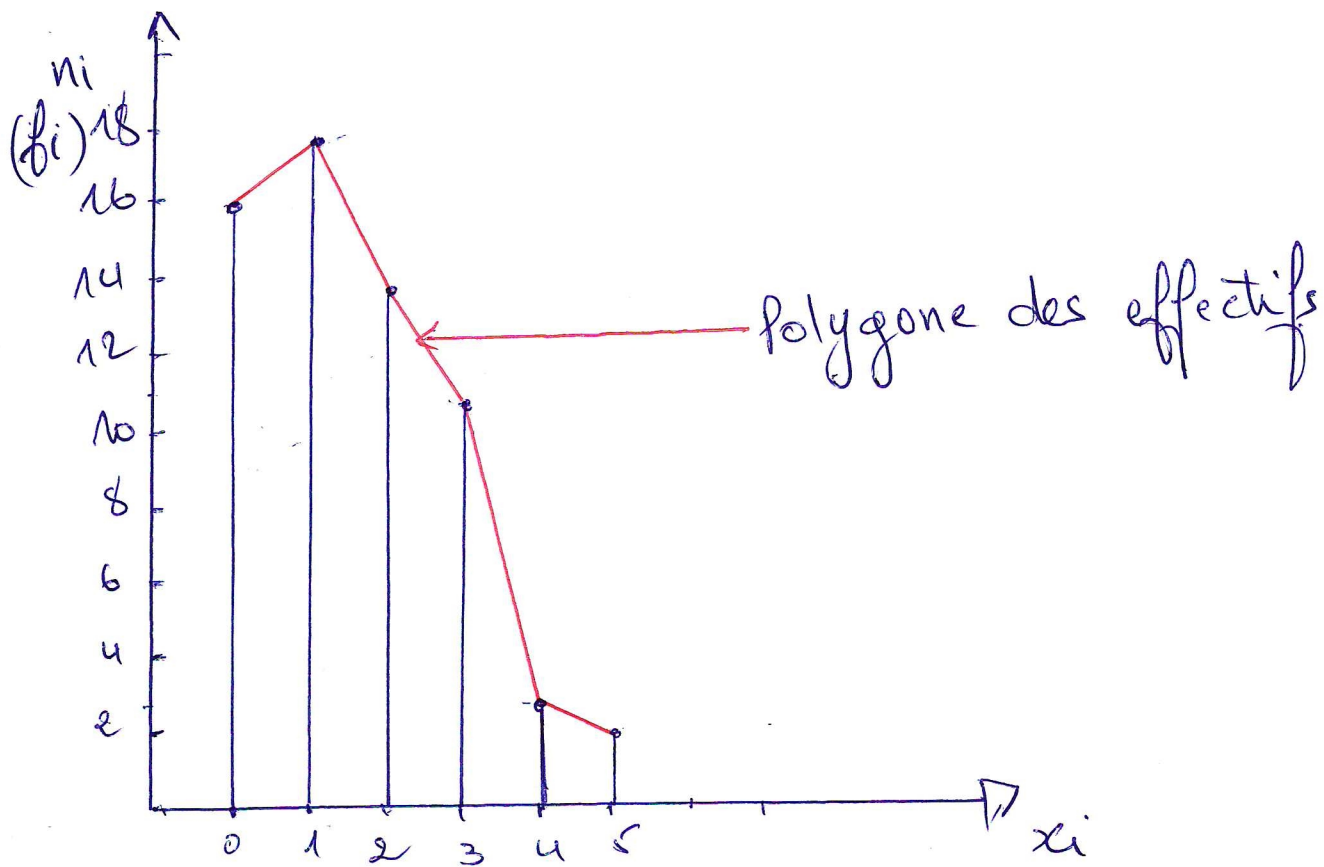


Diagramme en bâtons.

(b) Un caractère continu : Un caractère quantitatif est dit continu lorsqu'il peut prendre toutes les valeurs dans son intervalle de variation, par exemple :  
 Le poids d'un ensemble de personnes [50,3 kg - 80 kg]  
 La taille de 100 personnes [155,26 cm - 176,24 cm]

- La classe : pour étudier un caractère quantitatif continu, on divise son intervalle de variation en des sous intervalles de valeurs ayant une amplitude (la longueur de l'intervalle) cte (constante) ou variable.

Si  $[a_i, b_i[$  est la  $i^{\text{ème}}$  classe, alors l'amplitude de

$$k_i = b_i - a_i$$

Exemple : On considère l'âge de 1000 personnes, la répartition de leur âge est donnée par le tableau suivants :

Tranche d'âge	[0-20[	[20,40[	[40,60[	[60-80[
nbr de personne	360	280	160	100

Population : Les personnes

Individus : la personne et sa taille : 1000

le caractère : Quantitatif continu.

les modalités : 4 [0-20[, [20,40[, [40,60[, [60-80[.

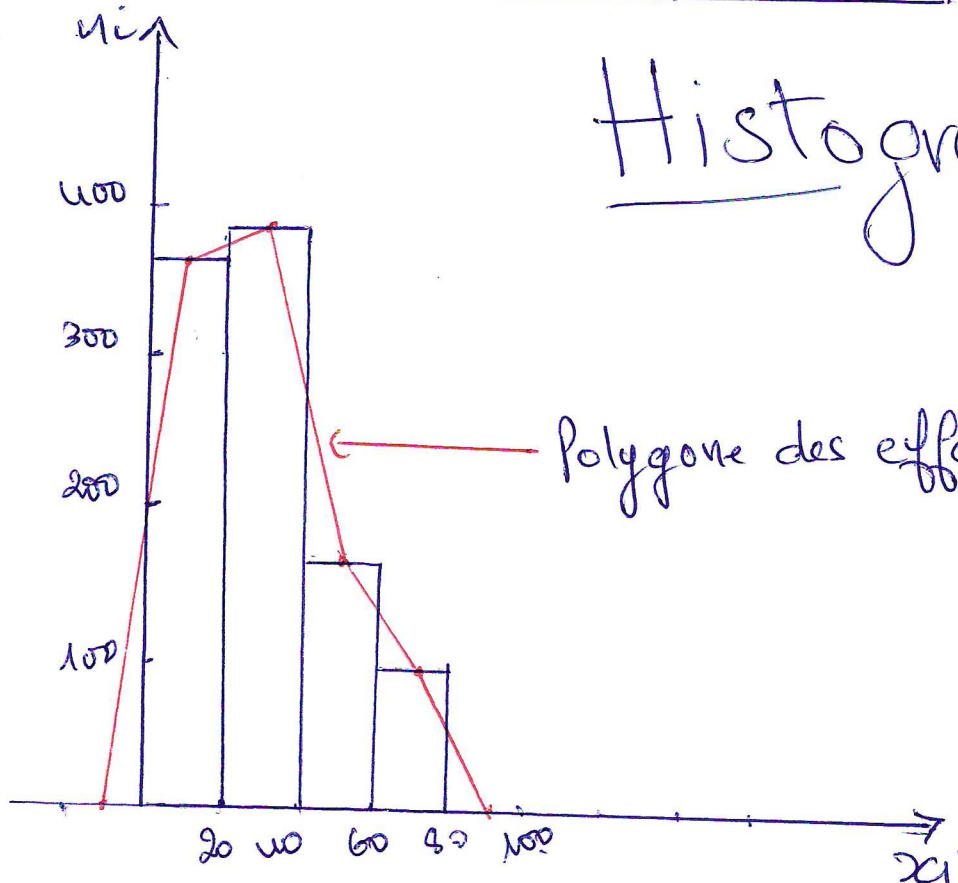
La représentation graphique d'une v. quantitative continue :

• Histogramme : L'histogramme est utilisé pour représenter une variable continue.

On représente sur l'axe des abscisses les différentes classes (supposées de même amplitude) et au dessus de chaque d'elle on trace des rectangles, dont les hauteurs sont directement proportionnelles aux fréquences ou aux effectifs.

Exemple :

Tranche d'âge	[0-20[	[20-40[	[40-60[	[60-80[
nbr. de personne	360	380	160	100



Histogramme

Exemple (Cas où les amplitudes des classes sont différentes)

• La méthode des effectifs rectifiés.

$$n_i' = n_i \times \frac{\text{la plus petite amplitude} = \text{PGCD}(a_i)}{\text{amplitude de la classe } i}$$

$$Tq: a_i' = \frac{a_i}{\text{PGCD}}$$

classe	$a_i$ amplitude	$n_i$	$a_i' = \frac{a_i}{\text{PGCD}}$	$n_i' = \frac{n_i}{a_i}$
[20-40[	20	8	2	4
[40-50[	10	8	1	8
[50-60[	10	12	1	12
[60-80[	20	8	2	4
[80-100[	20	4	2	2

$$\text{PGCD}(20, 10) = 10$$

Histogramme

