

# Chapitre 1 : Statistiques descriptive

## 1)- Définitions et vocabulaires:

### Définition :

1. La statistique est une méthode scientifique qui consiste à réunir des données chiffrées sur des ensembles nombreux, puis à analyser, à commenter et à critiquer ces données.
2. La statistique est une méthode qui vise à la description quantitative des ensembles nombreux. (définition donnée par Gerard-Carlot). C'est une méthode et non une théorie.

Le but de la **statistique descriptive** est de structurer et de représenter l'information contenue dans les données

### Définition : La biostatistique

La biostatistique c'est la statistique appliquée à la biologie.

### Exemple :

Etude descriptive des poids des étudiants inscrits en deuxième année de biologie à l'université de Jijel.

Comme toute science, la statistique fait appel à un vocabulaire spécialisé :

- **POPULATION** La collection d'objets ou de personnes étudiées (élèves, habitants, voitures...).
- **INDIVIDU** élément de la population étudiée. (un élève, un habitant, une voiture,...).
- **ECHANTILLON** partie de la population étudiée. Nombre d'individus dans un échantillon noté  $n$  est appelé taille de l'échantillon
- **LA TAILLE** représente le nombre d'individus d'un échantillon ou d'une population. Elle est symbolisée par «  $n$  » dans le cas d'un échantillon et par «  $N$  » dans le cas d'une population.
- **VARIABLE (CARACTERE)** propriété commune aux individus de la population, que l'on veut étudier.
- **LES MODALITE** les différentes manières d'être que peut présenter un caractère.

### Exemple :

Nous résumons les différents concepts dans cet exemple :

Population : l'ensemble des tous les employés d'une usine.

Individu : chaque employé de l'usine.

Caractère : le salaire, l'état matrimonial, le nombre d'enfants,... etc.

Les modalités du caractère : marié, célibataire, divorcé et veuf sont les modalités de l'état matrimonial, par exemple.

## 2) Type de caractères

**Caractère quantitatif** : Modalités mesurables (tout ce qu'on peut mesurer par un instrument)

<b>Discret</b>	<b>Continu</b>
<ul style="list-style-type: none"> <li>Les modalités prennent des valeurs isolées (ne peut prendre qu'un nombre fini ou dénombrable de valeurs)</li> </ul> <p><b>Exemple:</b> Nombre d'enfants dans une famille</p> <p>0,1,2,.....etc</p>	<ul style="list-style-type: none"> <li>Les modalités prennent des valeurs très rapprochées (peut prendre toutes les valeurs d'un intervalle de l'ensemble des nombres réels)</li> </ul> <p><b>Exemple:</b> Taux du glucose dans le sang</p> <p>0.88, 0.89, 1.01, 1.03, 1.27...etc</p>

**Caractère qualitatif** : modalités non mesurables (on ne peut associer ni valeur numérique ni un ordre naturel (type de voiture, couleur des cheveux, ...)).

<b>Ordinale</b>	<b>Nominale</b>
<p>Les modalités sont des qualités peuvent être ordonnées</p> <p><b>Exemples:</b></p> <p>Taille de vêtements (S, M,L, XL...)</p>	<p>Pas d'ordre</p> <p><b>Exemples:</b></p> <ul style="list-style-type: none"> <li>le sexe</li> <li>La couleur des yeux (Bleu, vert, Marron...)</li> </ul>

## 3) Représentation des données :

Après la réalisation d'une étude statistique, les données recueillies peuvent être résumées et représentées sous forme de

- Tableaux statistiques
- Graphes
- Paramètres statistiques

### Notations

- **Effectif ou fréquence absolue** : (noté  $n_i$ ) nombre d'apparitions de la valeur associée à un caractère dans un échantillon.
- **Fréquence relative** : (noté  $f_i$ )

$$f_i = \frac{n_i}{n}$$

- **Série statistique** : l'ensemble des valeurs du caractère avec en regard, les fréquences absolues ou relative correspondante.
- **Les effectifs (la fréquence) cumulés croissants** : la fréquence (effectif) cumulé croissante d'une valeur  $X_i$  est égale à la somme des effectifs ou fréquences des valeurs inférieure ou égale  $X_i$ . On note  $n_i^\uparrow$  ( $f_i^\uparrow$ ).
- **Les effectifs (la fréquence) cumulés décroissants** : la fréquence (effectif) cumulé décroissante d'une valeur  $X_i$  est égale à la somme des effectifs ou fréquences des valeurs supérieures ou égale  $X_i$

**Remarque:** On a toujours

$$\sum_{i=1}^k n_i = n$$

$$0 \leq f_i \leq 1, \quad \sum_{i=1}^k f_i = 1$$

### 1. Représentation sous forme des tableaux statistiques

#### 1) Série statistique dans le cas d'une variable quantitative discrète

**Exemple 1 :**

Nombre d'enfant	0	1	2	3	4	5	total
Nombre de famille effectif : $n_i$	16	18	14	11	3	2	64
Fréquence relative : $f_i$	0,25	0,281	0,21	0,172	0,047	0,031	1
	0		8				

- **Population étudiée** : les familles
- **L'échantillon sur lequel porte l'étude** : familles d'un immeuble ;  $n=64$ .
- **Le caractère étudié** est le nombre d'enfants par famille. C'est un caractère quantitatif discret.

## 2) Série statistique dans le cas d'une variable quantitative continue

**Exemple :** poids de 161 nouveau-nés.

Les classes	Centre de classe $x_i$	Effectif $n_i$	Fréquence relative $f_i$	$n_i^\uparrow$
2.2-2.5	2.35	5	0.031	5
2.5-2.8	2.65	11	0.068	16
2.8-3.1	2.95	24	0.148	40
3.1-3.4	3.25	40	0.248	80
3.4-3.7	3.55	42	0.259	122
3.7-4.0	3.85	20	0.124	142
4.0-4.3	4.15	13	0.08	155
4.3-4.6	4.45	6	0.037	161

## 3) Série statistique dans le cas d'une variable qualitative

**Exemple :** L'analyse du sang de 100 individus à donner les résultats suivants

Groupe SANGUIN	Effectif $n_i$	Fréquence relative $f_i$
O	40	0.40
A	43	0.43
B	12	0.12
AB	5	0.05

**Remarque :** On peut découper une série de valeurs en classe c.-à-d. passer de caractère discret au caractère continue en appliquant l'une des formules suivantes :

**1. Si  $n < 50$**

$$\text{Nombre de classes} = k \cong \sqrt{n}$$

**2. Formule de Sturge**

$$k = 1 + 3.332 \log_{10}(n)$$

**3. Formule de Yule**

$$k = 2.5\sqrt[4]{n}$$

**Amplitude des classes :**

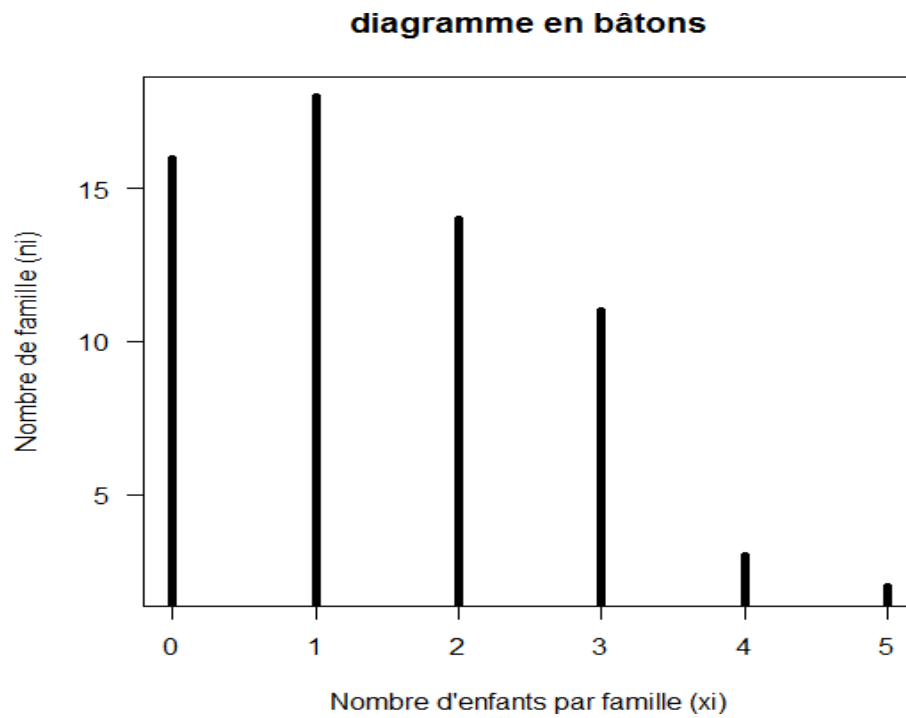
Il faut mieux d'utiliser des classes de même amplitude  $a$  donnée par

$$a = \frac{X_{max} - X_{min}}{k}$$

## 2. Représentation graphique d'une série statistique

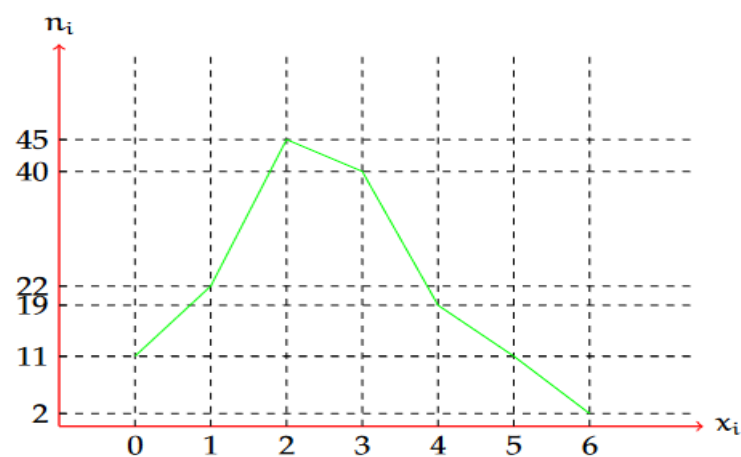
### 1) Caractère discret (discontinu)

- **Diagramme en Bâtons:** C'est un ensemble de bâtons ayant pour abscisses les valeurs  $X_1, X_2, \dots, X_n$  du caractère et en chacun des points d'abscisses  $x_i$  une ordonnée proportionnelle à l'effectif  $n_i$  de  $X_i$



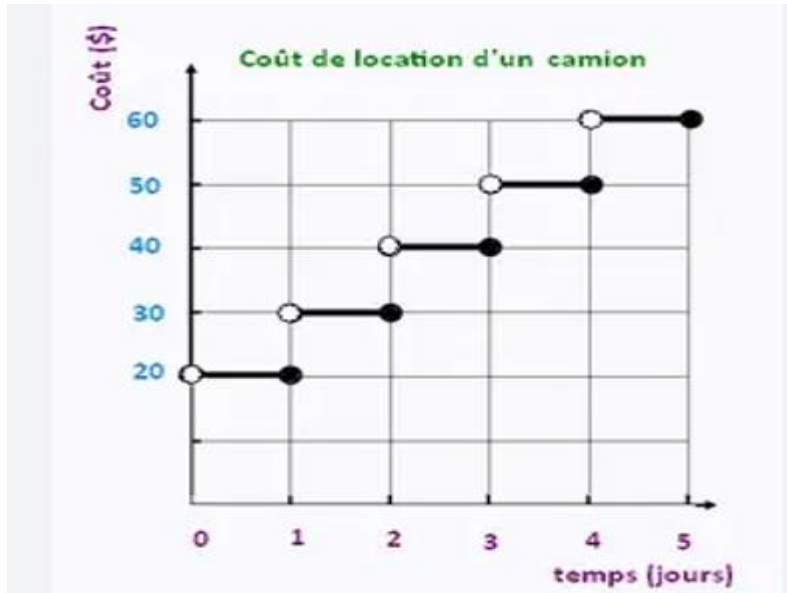
- **Polygone des Fréquences**

On l'obtient en joignant par des segments de droite les extrémités des bâtons



- **Diagramme cumulatif (diagramme en escaliers)**

En abscisse figurent, encore une fois, les observations de la variable considérée, tandis qu'en ordonnée figurent maintenant les effectifs cumulés, les fréquences cumulées ou les pourcentages cumulés. Dans le diagramme cumulatif les bâtons ont des longueurs proportionnelles aux effectifs cumulés (ou aux fréquences cumulées)



**Exemple 2**

Tracer le diagramme en escaliers de tableau suivant

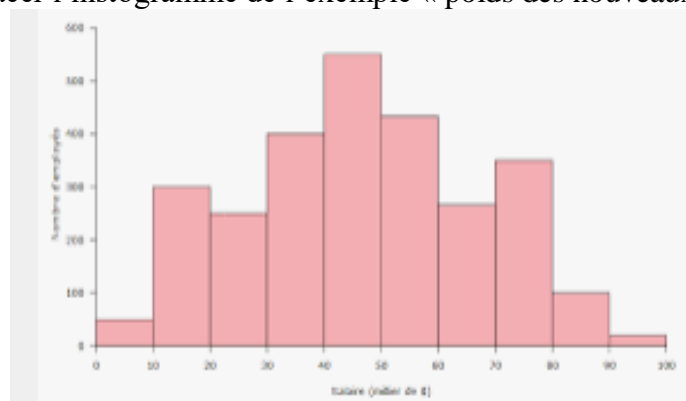
Valeurs	Effectif	Fréquence
0	10	0.2
1	5	0.1
3	10	0.2
4	15	0.3
5	10	0.2

**2) Caractère continue**

- **Histogramme** : c'est un ensemble de rectangles ayant pour largeur l'amplitude de la classe et pour hauteur l'effectif de la classe

**Exemple:**

Tracer l'histogramme de l'exemple « poids des nouveaux nés »



- **Polygone des effectifs ou polygone des fréquences**

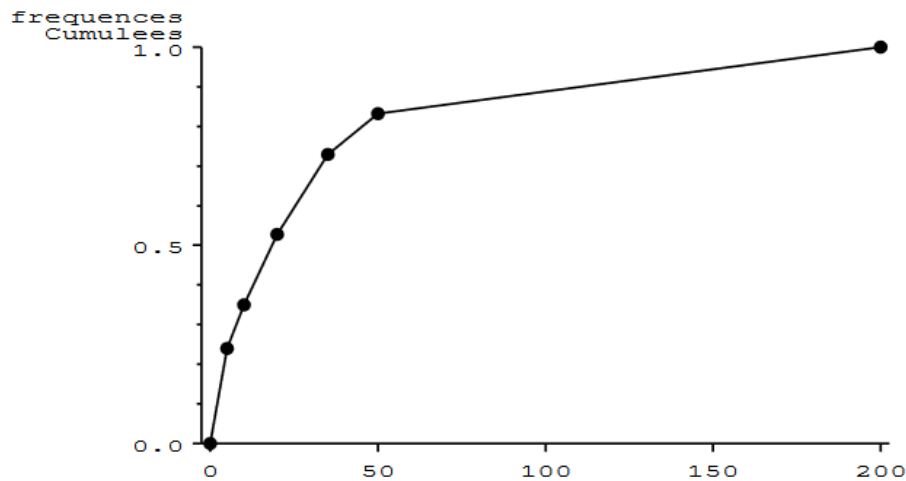
Est la ligne brisée joignant les milieux des bases supérieures des différents rectangles adjacents.

**Exemple:**

Tracer le polygone des effectifs pour le même exemple

La courbe cumulative croissante et décroissante (poids des nouveaux nés)

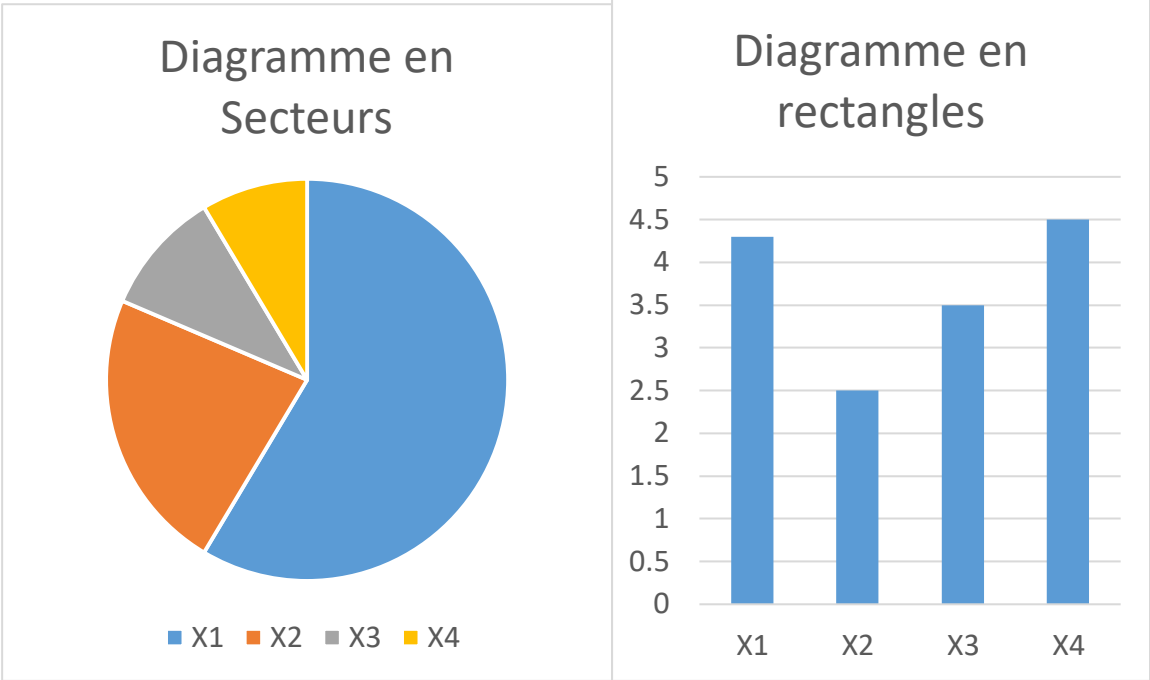
Pour la variable quantitative continue étudiée, chaque classe considérée doit d'abord être représentée par un point unique dont l'abscisse est la borne supérieure/inférieure de la classe et l'ordonnée est l'effectif (ou la fréquence, ou le pourcentage) cumulé de cette classe. La courbe cumulative est alors la courbe joignant les points en question



### 3) Dans le cas d'une variable qualitatives

Les représentations graphiques que l'on rencontre avec les variables qualitatives sont assez nombreuses. Les deux plus courantes, qui sont aussi les plus appropriées, sont :

- Le diagramme en colonnes (en rectangles).
- Le diagramme en secteurs (diagramme en camembert)



### 3. Les Paramètres (Description Numérique)

La réduction des données permet de condenser les données sous forme des paramètres typiques qui sont les suivants :

- Paramètre de position
- Paramètre de dispersion
- Paramètre de dissymétrie et d'aplatissement

#### 1) Paramètre de position

Ce sont des valeurs moyennes qui servent à caractériser l'ordre de grandeur des observations. Ce sont principalement :

- la moyenne - Le mode La médiane- La médiane

#### 1.1. Moyenne arithmétique:

Soit l'ensemble de mesure d'une même variable  $X : X_1, X_2, \dots, X_n$ , la moyenne arithmétique notée  $\bar{X}$  est définie par :

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

**Exemple 1:** La moyenne arithmétique des valeurs 8,5,3,6,2

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{8 + 5 + 3 + 6 + 2}{5} = 4.8$$

- Lorsque les valeurs  $X_1, X_2, \dots, X_n$  se répètent respectivement 1,2,...,n fois, on obtient la **moyenne arithmétique pondérée** en comptant chaque valeur  $X_i$  autant de fois qu'elle se présente: ceci revient à pondérer la valeur  $X_i$  par l'effectif  $n_i$  qui lui correspond. On aura

$$\bar{X} = \frac{\sum_{i=1}^n n_i X_i}{n}$$

**Exemple 2 :** Si les valeurs 8,5,3,6, et 2 se reproduisent respectivement 1,4,2,2,1 fois, la moyenne arithmétique est

$$\bar{X} = \frac{8 \times 1 + 5 \times 4 + 3 \times 2 + 6 \times 2 + 2 \times 1}{10} = 4.8$$

- Si les valeurs  $X_i$  sont groupé dans des classes dont les centres  $x_i$ , la moyenne arithmétique est donnée par :

$$\bar{X} = \frac{\sum_{i=1}^n n_i x_i}{n}$$

**Exemple 3 :**

Classe	$x_i$	$n_i$
[1, 2[	1.5	3
[2, 3[	2.5	1
[3, 4[	3.5	2

$$\bar{X} = \frac{1}{n} \sum_{i=1}^3 n_i x_i = \frac{1}{6} (3 \times 1.5 + 1 \times 2.5 + 2 \times 3.5) = \frac{14}{6} = 2.33.$$

#### 1.2. La médiane (the Mediane)

### ❖ Cas d'une variable statistique discrète

La médiane est le quantile d'ordre 1/2. Elle partage donc la série des observations en deux ensembles d'effectifs égaux.

Si la série possède un nombre impair de valeurs, la médiane sera  $\left(\frac{n+1}{2}\right)^{\text{ième}}$  valeur

Si la série compte un nombre pair de valeurs, la médiane sera la demi somme de la  $\left(\frac{n}{2}\right)^{\text{ième}}$  et la  $\left(\frac{n}{2} + 1\right)^{\text{ième}}$  valeurs.

**Exemple 1 :** On a la série de 15 valeurs suivantes :

12, 10, 11, 4, 5, 4, 2, 1, 6, 8, 7, 8, 9, 9, 4.

On ordonne d'abord les valeurs

1, 2 \_\_\_\_\_ , 4

On a  $n = 15$  impair donc  $me = \left(\frac{15+1}{2}\right)^{\text{ième}} = 7$ .

**Exemple 2:** 4,5,8,8,9,11,12,14,17,19. la médiane  $me = \frac{9+11}{2} = 10$ .

### ❖ Cas d'une variable statistique continue (médiane par interpolation)

$$me = l_1 + \frac{\left(\frac{n}{2} - n_{l_1}^{\uparrow}\right)}{n_0} a$$

Avec:

- $l_1$  : la borne inférieure de la classe médiane (classe médiane: la classe correspondante à  $n_i^{\uparrow}$  égale ou supérieure à  $\frac{n}{2}$ )
- $a$ : l'amplitude de la classe médiane
- $n_0$ : l'effectif de la classe médiane
- $n_{l_1}^{\uparrow}$ : l'effectif cumulé jusqu'à  $l_1$  ( avant  $n/2$  ) .

**Remarque :** La médiane Graphiquement (le cas continue)

la médiane est le point d'intersection de deux graphiques cumulés croissant et décroissant

## 1. 3. Le mode

### ❖ Cas d'une variable statistique discrète

Le mode d'une série est la valeur la plus fréquente de la série

**Exemple 1 :** le mode de la série {4 , 2, 4, 3, 2, 2} est 2.

### Cas d'une variable statistique continue (