

الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et populaire

وزارة التعليم العالي و البحث العلمي

Ministère de l'enseignement supérieur et de la recherche scientifique



**Centre Universitaire AbdeLhafid boussouf Mila**

**Institut des lettres et des langues**

**Département des langues Etrangère**

**Destiné aux étudiants de première année Master .**

**Filière : langue anglaise**

**Matière : statistique**

**Réaliser par :**

**- Boumezbeur Meroua**

**- Bouzaraa Abir**

**Année universitaire : 2022/2023**

# SOMMAIRE

## Les Statistiques descriptives

### 01 Introduction générale

Eléments de statistique  
- Définition et Objectif  
- Notion de base

### 02 Tableaux statistiques

Tableaux statistique  
- cas d'une seule variable  
- cas de deux variables

### 03 Caractéristiques de tendance centrale et de position

- Mode  
- Médiane  
- Moyenne ( arithmétique , géométrique, harmonique, quadratique )  
- Quantiles

### 04 Caractéristiques de dispersion

- Etendue  
- Intervalle interquartile  
- Coefficient de variation

### 05 La concentration

- valeurs globales  
- Médiale  
- Courbe de concentration ou de LORENZ  
- Indice de GINI

# Plan du cours

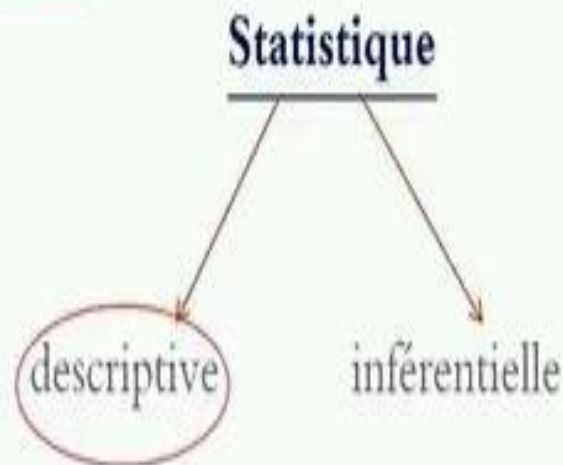
- 1) Les termes de la statistique.
- 2) Étude d'une variable quantitative discrète (tendance centrale, dispersion, concentration, forme)
- 3) Étude d'une variable quantitative continue (tendance centrale, dispersion, concentration, forme)
- 4) Caractéristiques de concentration.
- 5) Caractéristiques de forme.
- 6) Les indices et taux de croissance.
- 7) Les caractéristiques des distributions à deux caractères.

## À la fin de ce cours vous serez capables de:

- Tirer à partir d'une série brute de données les différents paramètres statistiques.
- Déterminer les valeurs centrales d'une série et mesurer sa dispersion.
- Représenter et analyser graphiquement une série ou une variable statistique.
- Mesurer l'évolution dans le temps d'une grandeur (prix, quantité d'un produit.....).
- Mesurer la corrélation entre deux variables statistiques étudiées simultanément.

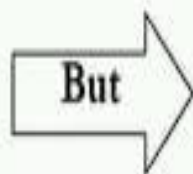
# 1) Les termes de la statistique descriptive:

## 1-1) Introduction :



## Statistique descriptive:

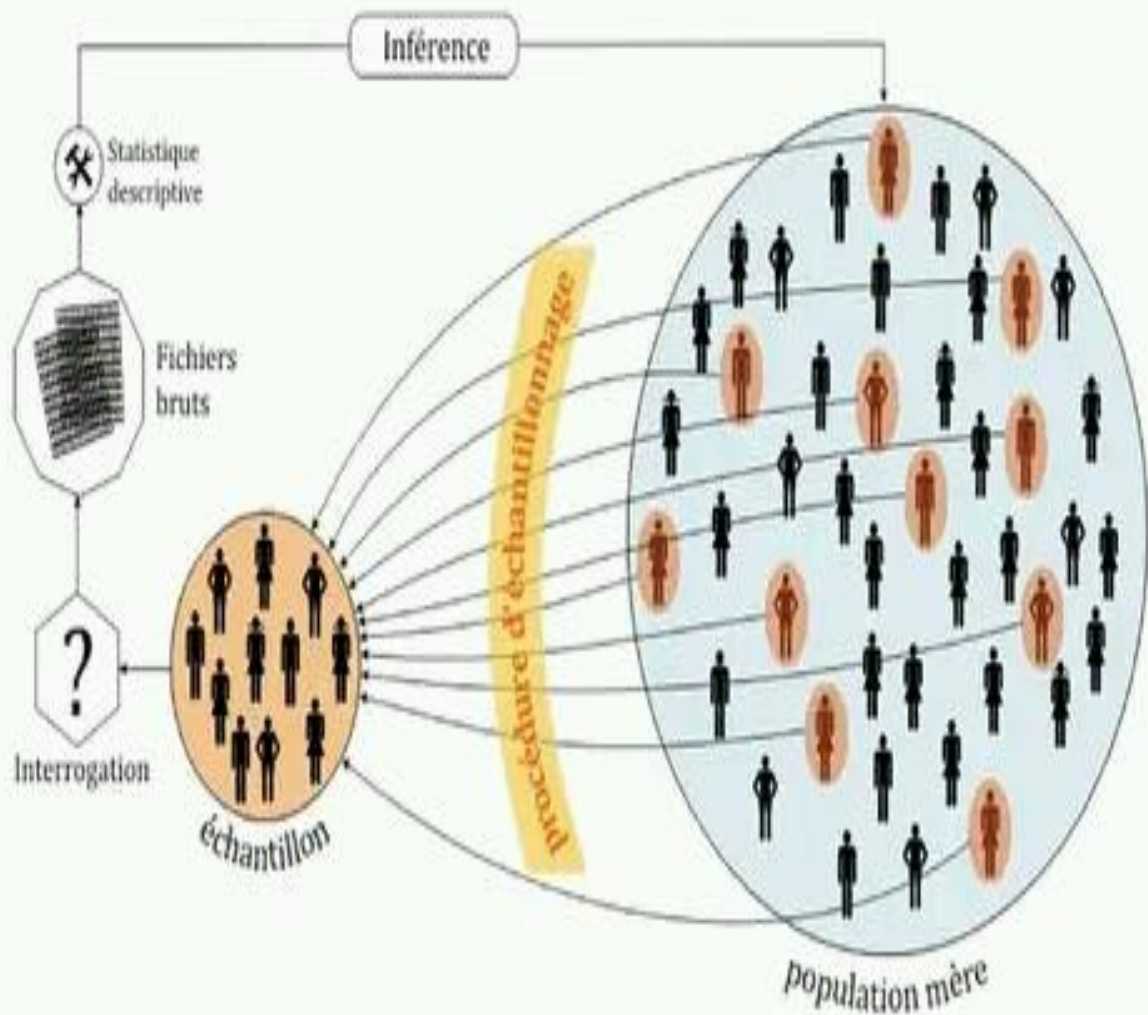
Analyse et synthèse **numérique** et **graphique** d'un ensemble des données.



décrire l'information contenue dans les données.



# Le lien de complémentarité entre statistique inférentielle et statistique descriptive



## 1-2) Définitions:

**Population:** Ensemble des éléments sur lesquelles porte l'étude.



Les étudiants, les villes marocaines, les voitures produites entre 2005 et 2010....

**Remarque:** La population doit être définie d'une manière précise, il diffère donc de considérer:

- ☞ Les étudiants.
- ☞ Les étudiants de 15 à 23 ans.

**Individu (unité statistique):** élément sur lequel porte l'étude (=un élément de la population).

Voiture, pays, ville, personne.....

**Caractère (variable statistique):** Ce qu'on observe sur chacun des individus de la population.

Couleur des yeux, nombre d'habitants, date de production...

## Il existe deux types de caractères :

### Caractères qualitatifs:

Les caractères qualitatifs sont tous les caractères qui ne sont pas représentés par des nombres. (-non mesurables).

Il existe deux type de caractères qualitatifs:

Couleur des yeux, nom et prénom, appréciations dans un examen...

- ☞ **Caractère qualitatif ordinal**- on peut les ordonner ou les hiérarchiser
- ☞ **Caractère qualitatif nominal**- on peut pas les ordonner.

### Caractères quantitatifs:

Les caractères quantitatifs sont représentés par des nombres et sur lesquels les opérations arithmétiques de base ont un sens.(-mesurable)

Age, taille, notes d'examen,.....

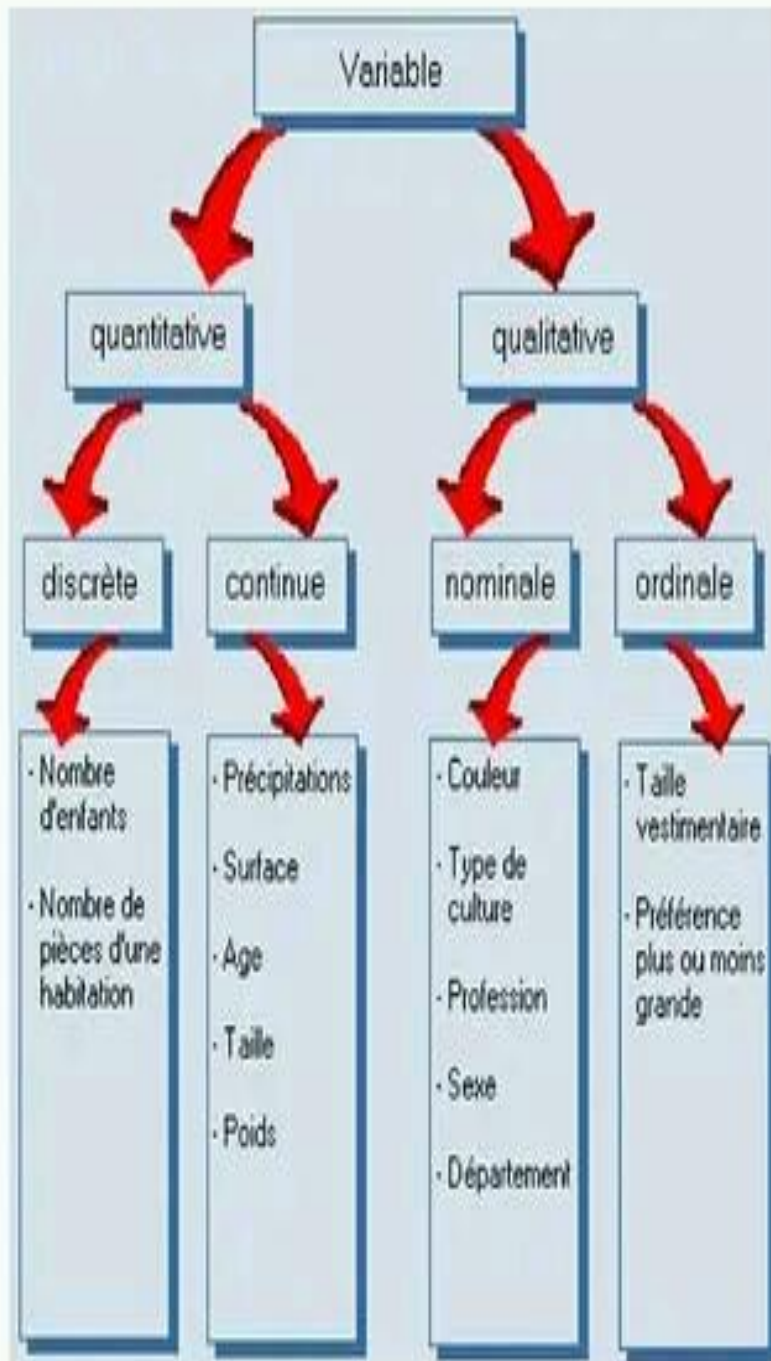
## Il existe deux types de caractères quantitatifs:

Caractères quantitatifs discrets: qui peuvent prendre un nombre faible et fini de valeurs (Exp: nombre d'enfants d'une famille, nombre de pièce d'une immeuble...).

Caractères quantitatifs continus: qui peuvent prendre un nombre théoriquement infini de valeurs dans un intervalle donné (Exp: Age et taille d'une personne, nombre de diamètres d'une pièce...).



## Types de caractères



## Rappel:

### (2) Un outil : L 'OPERATEUR SOMME S

PROPRIETE 1:  $\sum_i ka_i = k \sum_i a_i$

PROPRIETE 2:  $\sum_i (a_i + b_i) = \sum_i a_i + \sum_i b_i$

PROPRIETE 3:  $\sum_i k(a_i + b_i) = k \sum_i a_i + k \sum_i b_i$

PROPRIETE 4:  $\sum_{i=1}^n k = nk$

PROPRIETE 5:  $\sum_{i=p}^q k = (q - p + 1)k$

## 2) Étude d'une variable quantitative discrète.

### 2-1) Définitions:

#### Série statistique:

- ❑ On appelle **série statistique** les différentes valeurs du caractère étudié (noté **xi**).
- ❑ On appelle **effectif d'une population** le nombre d'individus de cette population.
- ❑ On appelle également **fréquence** la quantité  $f_i = \frac{n_i}{N}$  (où **ni** l'effectif de la valeur **xi** et **N** représente l'effectif total)

#### Exemple 1:

Les notes sur 20 obtenues lors d'un devoir de mathématiques dans une classe de seconde sont les suivantes :

10, 8, 11, 9, 12, 10, 8, 10, 7, 9, 10, 11, 12, 10, 8, 9, 10, 9, 10, 11

**Population étudiée:** Classe des élèves dans une classe de seconde.

**Individus:** Les élèves.

**Effectif total:** **N=20** élèves.

**Caractère étudié:** la note obtenue au devoir.

#### Distribution statistique:

la liste (dans une ligne) des valeurs possibles de la variable est associée à une deuxième ligne dans laquelle sont repris les fréquences ou l'effectifs. Elle est représentée sous forme d'une tableau ordonné de données.

### La distribution statistique définie par les effectifs:

Exemple: 10, 8, 11, 9, 12, 10, 8, 10, 7, 9, 10, 11, 12, 10, 8, 9, 10, 9, 10, 11 (N=20)

Valeurs du caractère (note obtenue)	7	8	9	10	11	12
Effectif (nb d'élève ayant cette note)	1	3	4	7	3	2

### La distribution statistique définie par les fréquences:

Valeurs du caractère (note obtenue)	7	8	9	10	11	12
Fréquence en % : $f_i = \frac{n_i}{N} \times 100$	5	15	20	35	15	10

### Effectif et fréquence cumulés:

L'**effectif cumulé croissant** (resp. **fréquence cumulée croissante**) d'une valeur  $x$  est la somme des effectifs (resp. fréquences) des valeurs  $y$  tels que  $y \leq x$ .

L'**effectif cumulé décroissant** (resp. **fréquence cumulée décroissante**) d'une valeur  $x$  est la somme des effectifs (resp. fréquences) des valeurs  $y$  tels que  $x \leq y$ .

Valeurs du caractère (note obtenue)	7	8	9	10	11	12
Effectif	1	3	4	7	3	2
Fréquence en %	5	15	20	35	15	10
Effectif cumulé croissant	1	4	8	15	18	20
Fréquence cumulée croissante en %	5	20	40	75	90	100
Fréquence cumulée décroissante en %	100	95	80	60	25	10

### 2-2) Propriété:

on a toujours :  $\sum_{i=1}^k n_i = N$ ,  $0 \leq f_i \leq 1$ ,  $\sum_{i=1}^k f_i = 1$  Où:  $N$  est l'effectif total  
Et  $k$  est le nombre des valeurs du caractère



### 2-3) Représentation graphique:

Pour les caractères quantitatifs discrets, on utilise le diagramme en bâton, le diagramme à secteur et le diagramme cumulatif:

#### a) Diagramme en bâton:

Dans un repère orthogonal, pour chaque valeur de la série statistique on trace un trait vertical dont la hauteur est **proportionnelle** à l'effectif.

#### b) Diagramme à secteur:

On appelle un diagramme à secteur un graphique qui divise un disque en secteurs angulaires dont les aires sont **proportionnels** au effectifs de chaque modalité.

Pour une modalité  $M_i$  d'effectif  $n_i$ , l'angle  $\alpha_i$  correspondante est:  $\alpha_i = \frac{n_i}{n} \times 360 = f_i \times 360$  (°)

Exemple: ( précédent)

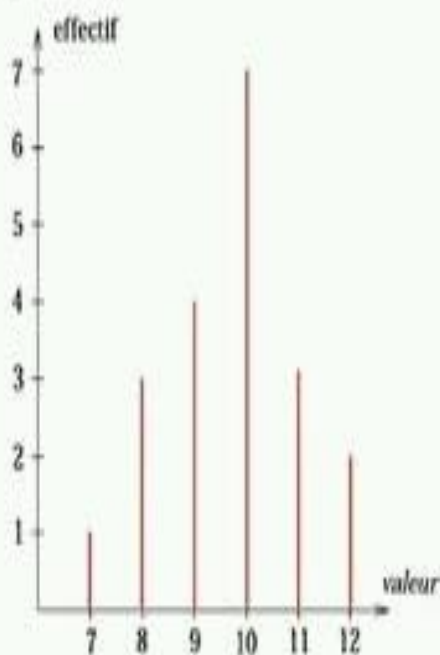


Diagramme en bâton

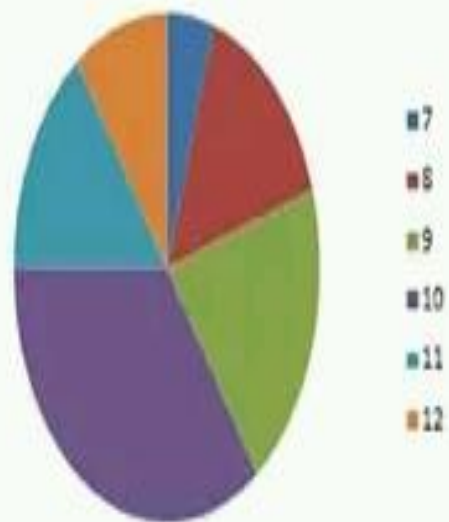


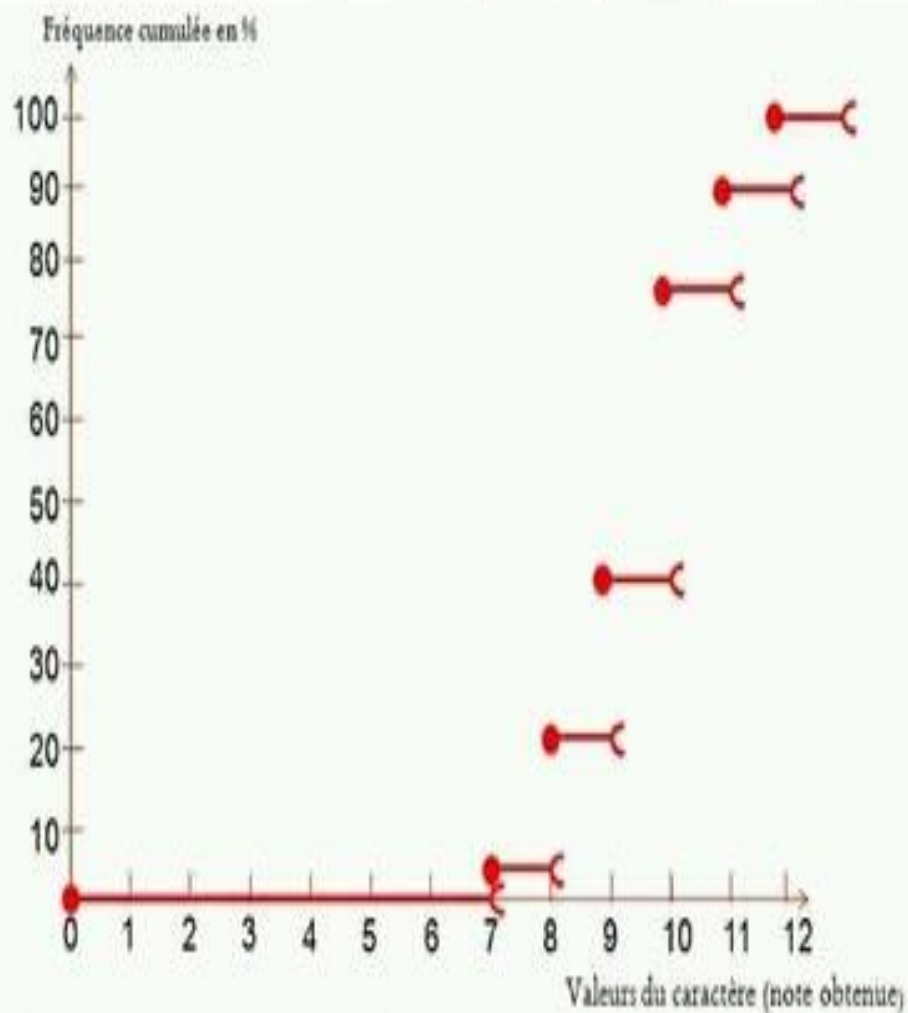
Diagramme à secteur

### c) Diagramme cumulatif (ou fonction de répartition)

Ce diagramme représente les fréquences cumulées en fonction des valeurs du caractère

#### Exemple:

Valeurs du caractère (note obtenue)	7	8	9	10	11	12
Fréquence en % : $f_i = \frac{n_i}{N} \times 100$	5	15	20	35	15	10
Fréquence cumulée croissantes en %	5	20	40	75	90	100



## 2-4) Les caractères de la tendance centrale (paramètres de position)

### A) La moyenne

#### a-1) La moyenne arithmétique.

On définit la moyenne arithmétique des valeurs  $x_i$  ( $i=1, \dots, n$ ) d'un caractère pour les  $n$  individus d'une population par :

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

#### Remarque:

Si le caractère est groupé dans un tableau de données contenant les effectifs ou les fréquences de chaque valeur du caractère alors :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i x_i = f_1 x_1 + f_2 x_2 + \dots + f_k x_k \quad \text{avec} \quad n = \sum_{i=1}^k n_i \quad \text{et} \quad f_i = \frac{n_i}{N} \quad \text{Et} \quad 0 \leq k \leq n$$

#### Exemple:

Valeurs du caractère (note obtenue)	7	8	9	10	11	12
Effectif	1	3	4	7	3	2
Fréquence cumulée en %	5 %	20 %	40 %	75 %	90 %	100 %

15

$$\bar{x} = \frac{1 \times 7 + 3 \times 8 + 4 \times 9 + 7 \times 10 + 3 \times 11 + 2 \times 12}{20} = 9,7$$

### a-2) Propriétés de la moyenne

- 1)  $\sum_{i=1}^k n_i (x_i - \bar{x}) = 0$  Où **k** est le nombre de valeurs que prend la variable **X** ( $0 \leq k \leq n$ )
- 2) si **a** et **b** désignent deux constantes telles que pour toutes les valeurs observées ou tous les centres de classes **x<sub>i</sub>** on a **y<sub>i</sub> = ax<sub>i</sub> + b**, alors:  $\bar{y} = a\bar{x} + b$

### a-3) Moyenne géométrique

On définit la moyenne géométrique des valeurs **x<sub>i</sub>** ( $i=1, \dots, n$ ) d'effectifs **n<sub>i</sub>** d'un caractère pour les **n** individus d'une population par :

$$G = \sqrt[n]{x_1^{n_1} x_2^{n_2} \dots x_k^{n_k}}$$

Utilisée dans le cas de phénomènes multiplicatifs (taux de croissance moyen)

### a-4) Moyenne harmonique

On définit la moyenne harmonique des valeurs **x<sub>i</sub>** ( $i=1, \dots, n$ ) d'effectifs **n<sub>i</sub>** d'un caractère pour les **n** individus d'une population par :

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

16

Utilisée dans le cas où l'on combine 2 variables sous forme de rapport (km/heure,.....)



## b) La médiane:

La médiane **Mé**, correspond au centre de la série statistique classée par ordre croissant, ou à la valeur pour laquelle 50% des valeurs observées sont supérieures et 50% sont inférieures.

### Pratiquement:

On classe la série  $X_i$  ( $i=1 \dots N$ ), par ordre croissant:

Si  $N$  est impair  $\rightarrow N=2m+1$  alors  $Mé=X_{m+1}$ .  $m \in \mathbb{N}$

Si  $N$  est pair  $\rightarrow N=2m$  alors  $Mé = \frac{X_m + X_{m+1}}{2}$

Exemple 1: 6 6 6 8 9 9 12 13 13 14 17 17 18



Mé= 12

Exemple 2: 4 5 5 6 6 12 13 13 14 14 18 18



Mé= 12,5

### c) Le mode:

Le mode d'une série est la valeur ou la modalité qui revient le plus fréquemment dans la série ou la distribution.

Exemple : Soit la série  $\{8, 4, 4, 3, 4, 3, 8, 2, 5\}$

$\Rightarrow$   $Mo = 4$

Remarque 1: Une série peut avoir plusieurs modes (série multimodale).

Elle est **bimodale**

Soit la série  $S = \{4, 0, 1, 1, 2, 2, 2, 3, 3, 4, 2, 3, 4, 5, 2, 1, 3, 3, 4, 5\}$

$\Rightarrow$   $Mo = 2$  ,  $Mo' = 3$

Remarque 2 : Le mode n'existe pas forcément

$S = \{8, 6, 5, 7, 3, 1\}$



Remarque 3: Mettre la série sous forme d'une distribution pour repérer le mode.

## 2-5) Caractéristiques de dispersion

Ces paramètres permettent de mesurer la façon dont les valeurs du caractère sont réparties autour de la moyenne et de la médiane. On distingue 3 principaux paramètres: **variance**, **écart-type** et **l'écart interquartile**.

### a) La variance:

La variance d'une série statistique  $X_i, i= 1, \dots, n$  est donnée par :

$$V = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

la variance est la moyenne des carrés des écarts à la moyenne

### b) Écart-type:

l'écart-type d'un caractère quantitatif discret est défini alors par :  $S = \sqrt{V}$

### Propriété

- Si on ajoute à toutes les valeurs d'une série statistique le même nombre **b**, la variance et l'écart-type restent inchangés.
- Si les valeurs d'une série statistique sont multipliées ou divisées par un même nombre **a**, l'écart-type est multiplié ou divisé par  $|a|$ .

### Remarque:

On est amené à considérer la racine carrée de la variance pour avoir un résultat exprimé dans la même unité que le caractère étudié.

### c) Ecart interquartile les quartiles:

Notés **Q1**, **Q2**, **Q3** les quartiles partagent la population en quatre parties de même effectif.

Etant donné une série statistique de médiane **Mé**, dont la liste des valeurs est rangée dans l'ordre croissant.

#### **On appelle:**

1<sup>er</sup> quartile Q1 est la valeur du caractère tel que: 25% des valeurs observées lui sont **inférieurs** ou **égale**.

3<sup>ème</sup> quartile Q3 est la valeur du caractère tel que: 75% des valeurs observées lui sont **inférieurs** ou **égale**.

On note que **Q2=Mé**.



### Dans la pratique :

- Si  $\frac{N}{4}$  est un entier, le premier quartile **Q1** est la valeur qui dans cette liste occupe le rang  $\frac{N}{4}$  et le troisième quartile **Q3** est la valeur qui dans cette liste occupe le rang  $\frac{3N}{4}$ .
- Si  $\frac{N}{4}$  n'est pas un entier, le premier quartile **Q1** est la valeur qui dans cette liste occupe le rang immédiatement supérieur à  $\frac{N}{4}$  et le troisième quartile **Q3** est la valeur qui dans cette liste occupe le rang immédiatement supérieur à  $\frac{3N}{4}$ .

#### Exemple 1 :

6 6 7 8 11 12 12 13 15 17 17 18      N=12      Q1=7 Q2=12 Q3=15

#### Exemple 2 :

6 6 7 8 8 11 12 12 13 13 15 17 17 18      N=14      Q1=8 Q2=12 Q3=15

### L'écart interquartile

L'écart interquartile d'une série statistique est le nombre  $EII=Q3 - Q1$

#### Remarque:

L'écart interquartile mesure la dispersion des valeurs autour de la médiane ; plus l'écart est petit, plus les valeurs de la série appartenant à l'intervalle interquartile sont concentrées autour de la médiane.

### Remarques:

- 1) On peut également définir l'**intervalle interquartile** qui est égale:  $[Q1 \ Q3]$
- 2) De même l'**étendue** ou amplitude d'une distribution est égale à la différence entre la plus grande et la plus petite valeur de la distribution :  $\text{Etendue de } X = X_{\max} - X_{\min}$

### Exercice d'application

Les étudiants d'un groupe de TD forment une population statistique  $P$  dont le caractère étudié est la note obtenue à un devoir de mathématiques. Ces notes étant entières, la série statistique  $S1$  ainsi recueillie est discrète :

$$S1 = \{11; 12; 10; 9; 4; 18; 19; 12; 16; 12; 13; 14; 18; 9; 8; 8; 12; 5; 11; 10\}$$

### Questions:

- Déterminer la distribution des effectifs de la série  $S1$  ainsi que la fréquence de ses valeurs et les fréquences cumulées.
- Tracer les diagrammes en bâton et cumulatif.
- Déterminer les paramètres de tendance centrale (médiane, moyenne, mode)
- Déterminer les paramètres de dispersion ( variance, écart-type, intervalle interquartile)

