# Chapitre 2 : Séries statistiques à une variable

## 1. Définition :

Une série statistique est l'ensemble des résultats d'une étude menée sur une certaine population**,** visant à mesurer la présence d'un certain caractère au sein de cette population. L'objectif de l'étude d'une série statistique est de résumer et de réunir des informations contenues dans la série statistique afin de mettre en évidence ses propriétés.

Les données d'une série statistique (appelées observations notées $O_i$) sont représentés sous forme brute (non ordonnés) ensuite on peut les disposer sous forme de tableau dont les modalités sont des aspects, des valeurs discrètes ou continues en classes.

Exemple d'une série statistique : **20, 18, 21, 18, 19, 18, 17, 17, 19**

Le nombre de données d'une série statistique est égal à N (la taille de la population).

Plusieurs outils peuvent être utilisés dans cette étude tel que : les tableaux, les graphiques et certaines caractéristiques

## 2. Représentations graphiques :

Les représentations graphiques permettent de visualiser les séries statistiques sous forme de graphes car ils ont l'avantage d'offrir une meilleure vue d'ensemble des données. De plus avec les graphes on peut mieux voir les caractéristiques essentielles de la série.

### 2.1 Diagramme en bâtons :

Le diagramme en bâtons est utilisé pour représenter des séries possédant des variables quantitatives discrètes. On représente sur l'axe des abscisses les modalités de la variable statistique ($x_i$), et pour chaque valeur on trace un bâton vertical dont la hauteur représente l'effectif ($ni_)$ correspondant à $x_i$.

**Exemple 1:**

Dans un quartier résidentiel, on a relevé le nombre de pièces par appartement :

| Modalité (nombre de pièces) $x_i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Effectif (nombre d'appartements) $n_i$ | 48 | 72 | 86 | 64 | 30 |
| $n_i$cc | 48 | 120 | 206 | 270 | 300 |

| | | | | | |
|---|---|---|---|---|---|
| $f_i$ | 0.16 | 0.24 | 0.2867 | 0.2133 | 0.1 |
| $f_icc$ | 0.16 | 0.4 | 0.6867 | 0.9 | 1 |

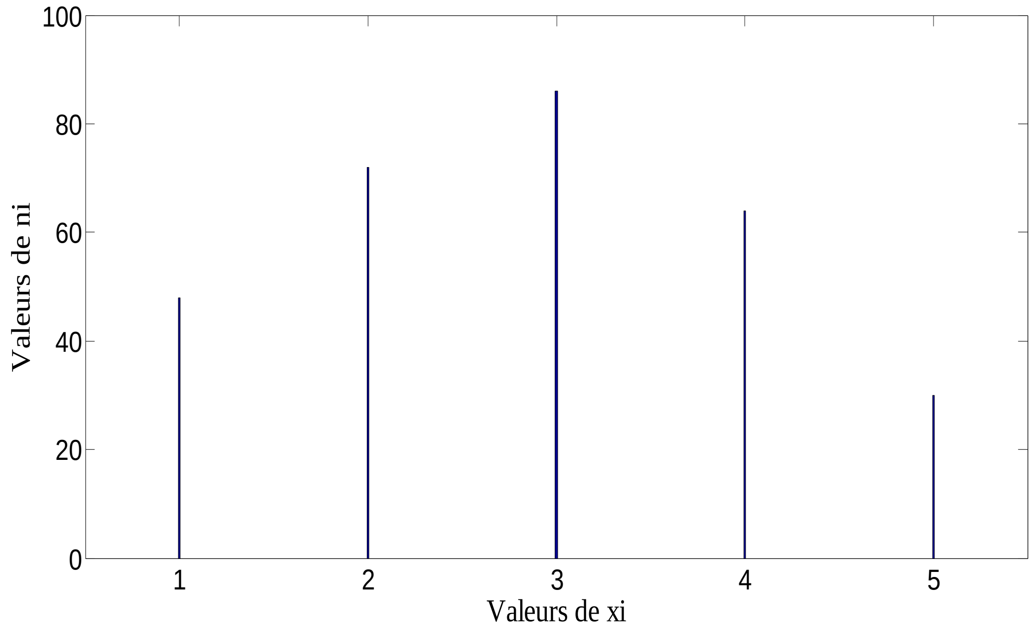


Figure 2-1 Diagramme en bâtons des effectifs pour l'exemple 1

On peut avoir aussi le diagramme pour les effectifs cumulés croissants sous forme d'escalier.

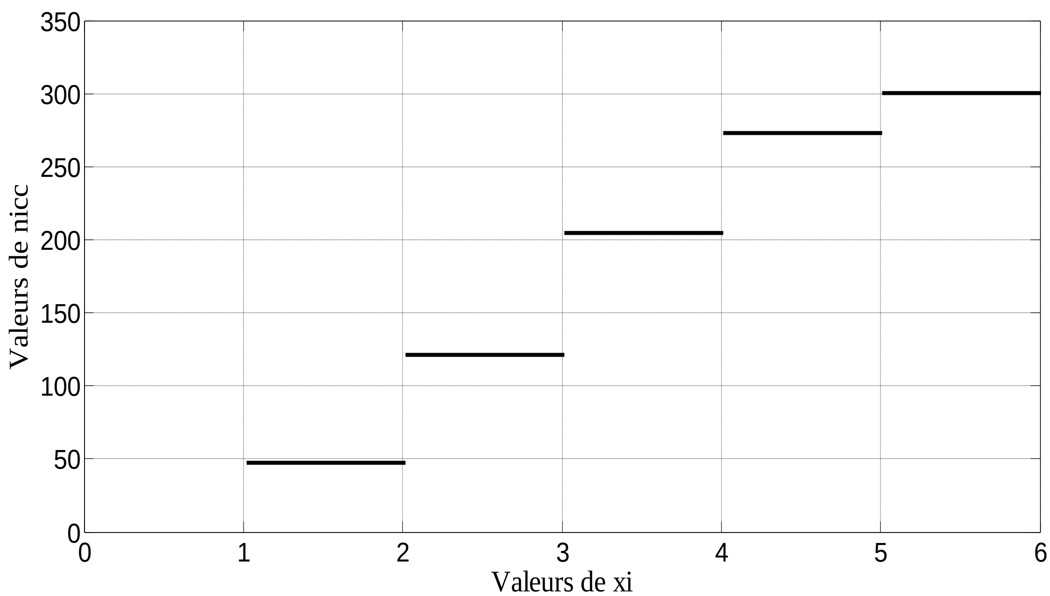


Figure 2-2 Diagramme des effectifs cumulés croissants pour l'exemple 1

De même pour les fréquences cumulées croissantes, le diagramme sous forme d'escalier est donné en figure (2-3).

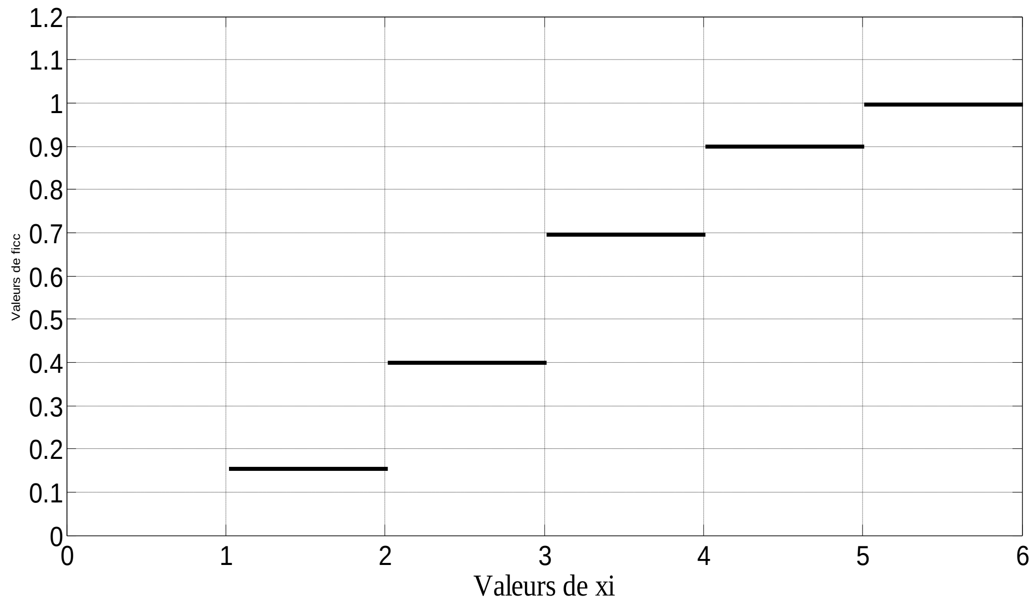


Figure 2-3 Diagramme des fréquences cumulés croissants pour l'exemple 1

### 2.2 Histogramme :

L'histogramme est utilisé lorsque les séries possèdent des variables quantitatives continues, on doit alors construire un graphe avec des rectangles représentant les classes ou les intervalles.

La largeur de chaque est égale à l'amplitude de la classe correspondante et la hauteur de chaque rectangle « i » correspond à l'effectif « $n_i$ » associé à la classe « i ».

**Exemple 2:**

Les revenus mensuels d'une centaine d'ouvriers ont été répartis selon le tableau suivant :

| | | | | | |
|---|---|---|---|---|---|
| Revenu mensuel (en milliers de DA) $x_i$ | [0, 40[ | [40, 80[ | [80, 120[ | [120, 160[ | [160, 200[ |
| Effectif (Nombre d'ouvriers) $n_i$ | 35 | 19 | 28 | 12 | 6 |
| Fréquence $\frac{n_i}{N}$ | 0.35 | 0.19 | 0.28 | 0.12 | 0.06 |

### 2.3 Polygone des effectifs :

Le polygone des effectifs est obtenu en joignant par des segments de droite les milieux des bases supérieures des rectangles de l'histogramme des effectifs.

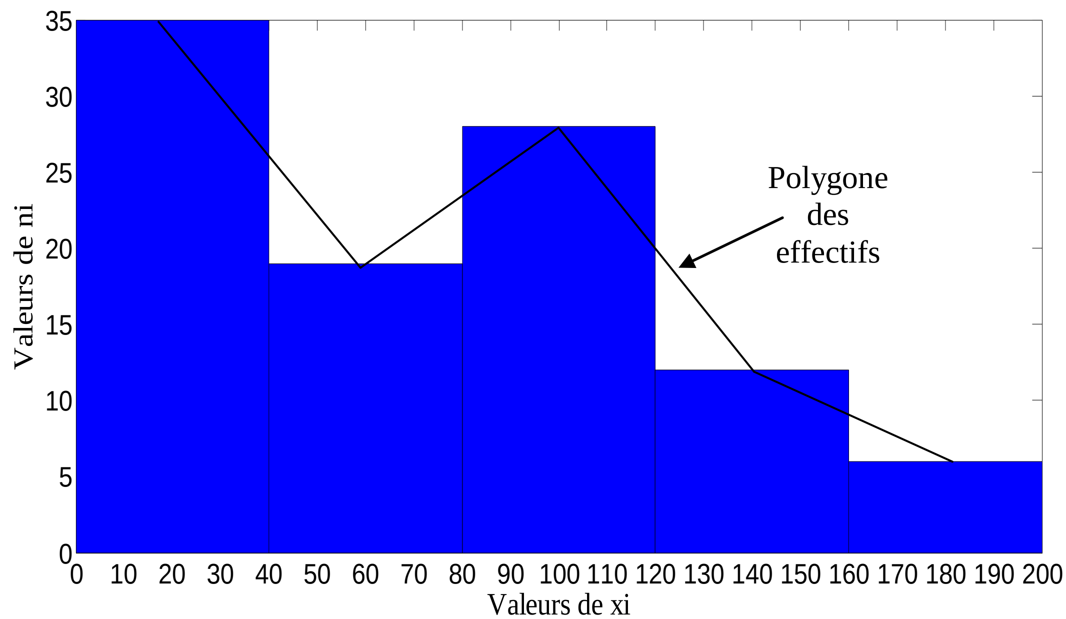


Figure 2-4 Histogramme et polygone des effectifs pour l'exemple 2

### 2.4 Polygone des fréquences :

Le polygone des fréquences est obtenu en joignant par des segments de droite les milieux des bases supérieures des rectangles de l'histogramme des fréquences.

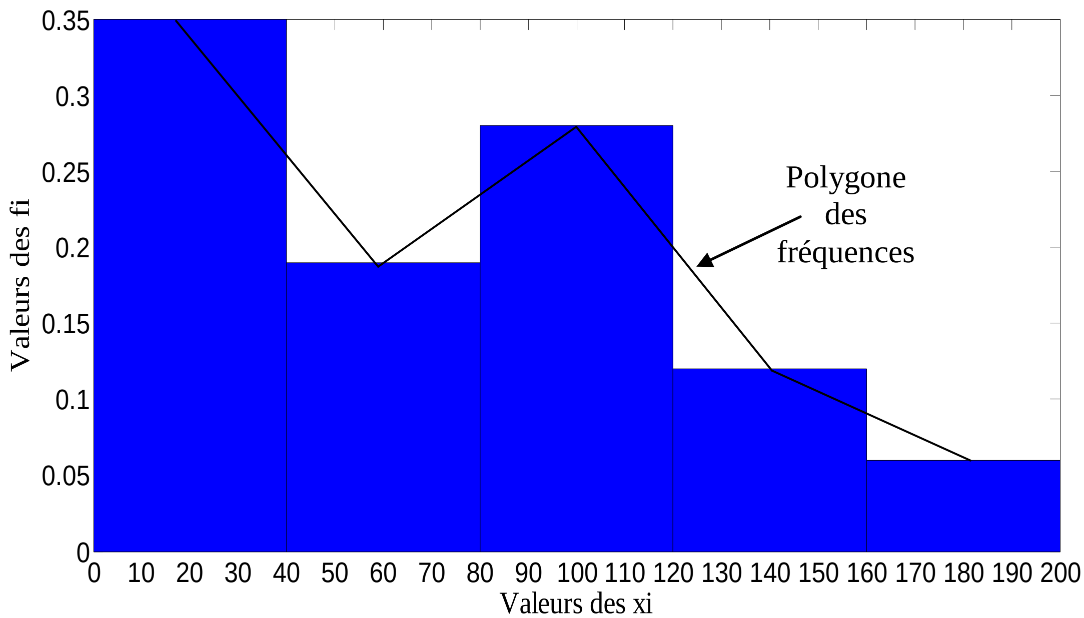


Figure 2-5 Histogramme et polygone des fréquences pour l'exemple 2

### 2.5 Diagramme en secteurs (circulaire) :

Il est représenté sur un cercle découpé en tranches issues du centre, et dont l'angle «$\theta_i$» associé à la fréquence relative « $f_i$ » est calculé suivant cette relation : $\theta_i = 360 f_i$

Généralement chaque secteur est caractérisé par son effectif donné en pourcentage suivant cette formule : $S_i = 100 f_i (\%)$. Ce type de diagramme est utilisé souvent dans le cas de variables qualitatifs.

**<u>Exemple 3:</u>**

L'analyse du sang de 100 personnes a donné les résultats suivants :

| Groupe sanguin $x_i$ | O | A | B | AB |
|---|---|---|---|---|
| Effectif (Nombre de personnes) $n_i$ | 40 | 43 | 12 | 5 |
| Fréquence $f_i = \frac{n_i}{N}$ | 0.4 | 0.43 | 0.12 | 0.05 |

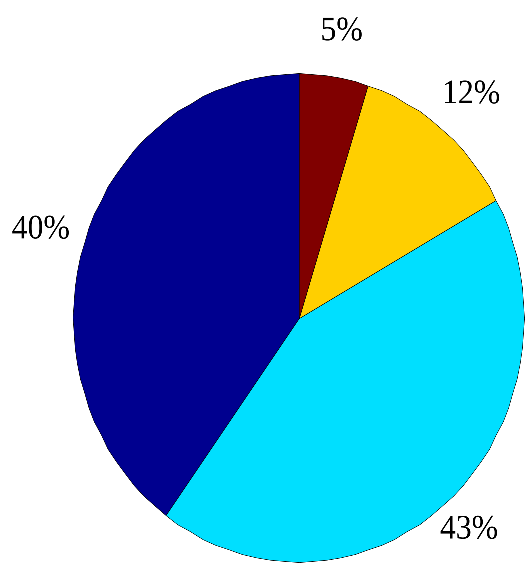


Figure 2-5 Diagramme en secteurs pour l'exemple 3

### 2.6 Diagramme en bandes :

Le digramme en bandes, utilisé pour les variables qualitatives, est construit de telle manière qu'un grand rectangle est divisé en bandes de hauteur proportionnelle à la fréquence ou l'effectif de la modalité représentée.

Reprenons le 3ème exemple, on aura les effectifs cumulés croissants donnés par :

| Groupe sanguin $x_i$ | O | A | B | AB |
|---|---|---|---|---|
| Effectif (Nombre de personnes) $n_i$ | 40 | 43 | 12 | 5 |
| $n_i cc$ | 40 | 83 | 95 | 100 |

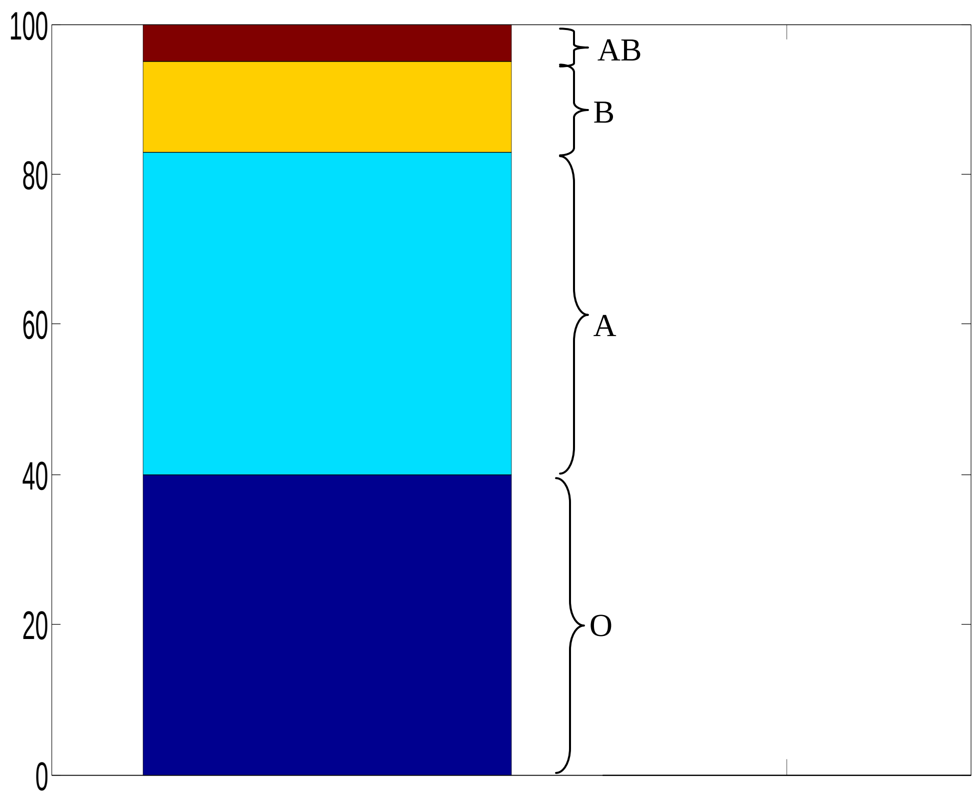


Figure 2-6 Diagramme à bandes pour l'exemple 3

## 3. Caractéristiques (indicateurs) de position :

Un indicateur de position est un nombre réel permettant de situer les valeurs d'une série statistique d'une variable quantitative. Les indicateurs de position sont le plus souvent des moyennes (arithmétique, géométrique, quadratique...) ou des quantiles comme la médiane et les quartiles

### 3.1 Le mode :

On appelle mode (ou classe modale) la valeur de la modalité ayant le plus grand effectif.

- **Cas de données brutes :**

  Dans le cas de données brutes, le mode est la valeur de l'observation la plus fréquente.

**Exemple 4 :** dans la série : 10,4, 4, 6, 1, 4, 6, le mode = 4

**Remarque 1 :**

Si nous avons deux valeurs de la variable qui apparaissent avec la même fréquence, alors on aura deux modes.

**Exemple 5 :** dans la série : 7, 6, 5, 4, 5,7 les modes sont 5 et 7.

- **Cas de tableau à variable quantitative discrète :**

Dans le cas d'un tableau à variable discrète, le mode correspond à la valeur de la modalité $x_i$ qui correspond à l'effectif $n_i$ le plus grand.

Dans l'exemple 1 des appartements, le mode = 3 car il correspond au plus grand effectif.

- **Cas de tableau à variable quantitative continue :**

Dans le cas d'une variable de type continue, on ne parle pas de mode mais de classe modale qui représente la classe la plus dense c'est à dire la classe qui contient le plus d'effectifs par amplitude.

La densité d'effectif $d_i$ de la classe « i » se calcule en divisant chaque effectif par l'amplitude correspondante de la classe : $d_i = \frac{n_i}{A_i}$

Reprenons l'exemple 2

| Revenu mensuel (en milliers de DA) $x_i$ | [0, 40[ | [40, 80[ | [80, 120[ | [120, 160[ | [160, 200[ |
|---|---|---|---|---|---|
| Effectif (Nombre d'ouvriers) $n_i$ | 35 | 19 | 28 | 12 | 6 |
| $d_i$ | $\frac{35}{40} = 0.875$ | $\frac{19}{40} = 0.475$ | $\frac{28}{40} = 0.7$ | $\frac{12}{40} = 0.3$ | $\frac{6}{40} = 0.15$ |

La plus grande densité est 0.875 alors [0; 40[est la classe modale.

### 3.2 La médiane :

La médiane (notée par Me) est la valeur de l'observation (ou de la modalité $x_i$) qui partage la distribution d'une série en deux parties égales (en deux séries aux effectifs égaux). Dans la première série on trouve les valeurs inférieures à la médiane, dans la seconde série on trouve les valeurs supérieures à la médiane.

**Remarque 2 :**

Les observations (ou les modalités) doivent être ordonnées par ordre croissant ou décroissant.

**3.2.1. Cas de données brutes:**

- Si la série possède un nombre impair d'observations (n=2k+1), alors la médiane sera la $Me = \left(\frac{n+1}{2}\right)^{\text{ème}} = (k+1)^{\text{ème}}$ observation

**Exemple 6 :** Soit les données suivantes : 23, 4, 12, 92, 67, 23, 22, 17, 29, 14, 82, 25,40, 24, 77

On ordonne ces valeurs par ordre croissant :4, 12, 14, 17, 22, 23, 23,24, 25, 29, 40, 67, 77, 82, 92

On a n = 15 =2*7+1 (impair) donc la médiane sera Me = $(7+1)^{\text{ème}} = 8^{\text{ème}}$ observation

Me = 24

- Si la série possède un nombre pair d'observations (n=2k), alors la médiane sera la moyenne des observations $(k)^{\text{ème}}$ et $(k+1)^{\text{ème}}$

$$Me = \frac{(k)^{\text{ème}}\, observation + (k+1)^{\text{ème}}\, observation}{2}$$

**Exemple 7 :** Soient les données suivantes : 8, 14, 3, 19, 24, 52, 1, 6, 10, 37

On ordonne ces valeurs par ordre croissant : 1, 3, 6, 8, 10, 14, 19, 24, 37, 52

On a n = 10 =2*5 (pair) donc la médiane sera

$$Me = \frac{(5)^{\text{ème}}\, observation + (6)^{\text{ème}}\, observation}{2} = \frac{10 + 14}{2} = 12$$

**Remarque 3 :**

La valeur de la médiane calculée doit impérativement se trouver parmi les valeurs des observations. Si ce n'est pas le cas alors on prend la valeur de l'observation immédiatement supérieure à la médiane calculée. De l'exemple précédent on remarque la valeur de Me =12 ne se trouve pas parmi les valeurs de la série, donc on prend Me = 14.

**3.2.2. Cas de tableau à variable quantitative discrète:**

Pour déterminer la médiane, on repère la valeur « 0.5 » dans la colonne de $f_i cc$ ou bien $\frac{N}{2}$ dans la colonne de $n_i cc$. On choisit ensuite la valeur de $f_i cc$ égale ou immédiatement supérieure à « 0.5 » ou bien la valeur de $n_i cc$ égale ou immédiatement supérieure à $\frac{N}{2}$ . la valeur de la médiane sera égale à la valeur de $x_i$ se trouvant sur la même ligne que $f_i cc = 0.5$ ou $n_i cc = \frac{N}{2}$

Reprenons l'exemple 1 des pièces d'appartements.

| Modalité (nombre de pièces) $x_i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Effectif (nombre d'appartements) $n_i$ | 48 | 72 | 86 | 64 | 30 |
| $n_i cc$ | 48 | 120 | 206 | 270 | 300 |
| $f_i$ | 0.16 | 0.24 | 0.2867 | 0.2133 | 0.1 |
| $f_i cc$ | 0.16 | 0.4 | 0.6867 | 0.9 | 1 |

D'après le tableau $f_i cc = 0.5$ ou $n_i cc = \frac{300}{2} = 150$ est atteint pour $x_3 = 3$ donc Me =3

**3.2.3. Cas de tableau à variable quantitative continue:**

Dans le cas continu, on ne parle pas d'une seule variable de la médiane mais plutôt de classe médiane qui possède une infinité de valeurs.

On cherche d'abord la classe dans laquelle on atteint la moitié de l'effectif total $\left(\frac{N}{2}\right)$ et cela à partir de l'effectif cumulé croissant. Ensuite on peut calculer la valeur approchée de la médiane en utilisant la relation suivante :

$$Me = a_i + A_i \left[ \frac{\left(\frac{N}{2} - n_{(i-1)}cc\right)}{n_i} \right]$$

$a_i$ : la borne inférieure de la classe médiane.

$l_2$ : la borne supérieure de la classe médiane.

$n_i$ : effectif associée à cette classe.

$n_{(i-1)}cc$ : effectif cumulé immédiatement inférieure à $a_i$

N : effectif total.

Reprenons l'exemple 2 :

| Revenu mensuel (en milliers de DA) $x_i$ | [0, 40[ | [40, 80[ | [80, 120[ | [120, 160[ | [160, 200[ |
|---|---|---|---|---|---|
| Effectif (Nombre d'ouvriers) $n_i$ | 35 | 19 | 28 | 12 | 6 |
| $n_i cc$ | 35 | 54 | 82 | 94 | 100 |

On a la moitié de l'effectif total $\frac{N}{2} = \frac{100}{2} = 50$

A partir des valeurs de $n_i cc$, la valeur « 50 » est atteinte dans la classe [40,80[. La classe médiane sera donc [40,80[.

$a_i = 40$, $n_i = 19$ et $n_{(i-1)}cc = 35$

$$Me = a_i + A_i \left[\frac{\left(\frac{N}{2} - n_{(i-1)}cc\right)}{n_i}\right] = 40 + 40\left[\frac{50-35}{19}\right]$$

$Me \approx 71.58$

### 3.3 La moyenne :

#### 3.3.1 La moyenne arithmétique :

**a) Cas de données brutes :**

La moyenne est égale à la somme des observations, divisée par leur nombre total, elle est donnée par cette relation

$$Moy = \frac{1}{n}\sum_{i=1}^{n} O_i$$

$O_i$ : la $i^{ème}$ observation

n : le nombre total d'observations

Pour l'exemple 7, on aura :

$$Moy = \frac{8+14+3+19+24+52+1+6+10+37}{10} = 17.4$$

**b) Cas de tableau à variable quantitative discrète :**

La moyenne est égale à la somme du produit des effectifs avec les modalités, divisée par l'effectif total, elle est donnée par cette relation

$$Moy = \frac{1}{N}\sum_{i=1}^{r} n_i x_i$$

r : le nombre de modalités

On reprend le $1^{er}$ exemple

$$Moy = \frac{1*48+2*72+3*86+4*64+5*30}{300} = 2.8533$$

**c) Cas de tableau à variable quantitative continue :**

Pour calculer la moyenne dans le cas où les données sont regroupées en classes, il faut calculer le centre $c_i$ de chaque classe puis on applique la formule :

$$Moy = \frac{1}{N}\sum_{i=1}^{r} n_i c_i$$

On reprend le 2ème exemple

| Revenu mensuel (en milliers de DA) $x_i$ | [0, 40[ | [40, 80[ | [80, 120[ | [120, 160[ | [160, 200[ |
|---|---|---|---|---|---|
| Effectif (Nombre d'ouvriers) $n_i$ | 35 | 19 | 28 | 12 | 6 |
| $c_i$ | 20 | 60 | 100 | 140 | 180 |

$$Moy = \frac{35 * 20 + 19 * 60 + 28 * 100 + 12 * 140 + 6 * 180}{100} = 74$$

### 3.3.2 La moyenne harmonique :

**a) Cas de données brutes :**

L'inverse de la moyenne harmonique (notée H) est la moyenne arithmétique de l'inverse des obseravations, elle est donnée par cette relation :

$$\frac{1}{H} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{O_i} \Rightarrow H = \frac{n}{\sum_{i=1}^{n}\frac{1}{O_i}}$$

Pour l'exemple 7, on aura : 8, 14, 3, 19, 24, 52, 1, 6, 10, 37

$$H = \frac{10}{\frac{1}{8}+\frac{1}{14}+\frac{1}{3}+\frac{1}{19}+\frac{1}{24}+\frac{1}{52}+\frac{1}{1}+\frac{1}{6}+\frac{1}{10}+\frac{1}{37}} = 5.1627$$

**b) Cas de tableau à variable quantitative discrète :**

L'inverse de la moyenne harmonique est donné par cette relation :

$$\frac{1}{H} = \frac{1}{N}\sum_{i=1}^{r}\frac{n_i}{x_i} \Rightarrow H = \frac{N}{\sum_{i=1}^{r}\frac{n_i}{x_i}}$$

On reprend le 1er exemple :

$$H = \frac{300}{\frac{48}{1} + \frac{72}{2} + \frac{86}{3} + \frac{64}{4} + \frac{30}{5}} = 2.2277$$

**c) Cas de tableau à variable quantitative continue :**

L'inverse de la moyenne harmonique est donné par cette relation :

$$\frac{1}{H} = \frac{1}{N}\sum_{i=1}^{r}\frac{n_i}{c_i} \Rightarrow H = \frac{N}{\sum_{i=1}^{r}\frac{n_i}{c_i}}$$

On reprend le 2ème exemple :

$$H = \frac{100}{\frac{35}{20} + \frac{19}{60} + \frac{28}{100} + \frac{12}{140} + \frac{6}{180}} = 40.5562$$

## 4. Caractéristiques (indicateurs) de dispersion :

Un indicateur de dispersion mesure la variabilité des valeurs d'une série statistique. Il est toujours positif et d'autant plus grand que les valeurs de la série sont étalées.

### 4.1.Intervalle de variation (l'étendue)

C'est la différence entre la valeur la plus grande et la valeur plus petite parmi les observations.

$e = O_{max} - O_{min}$

Pour l'exemple 7, on aura e = 52-1 = 51

- Si la variable est discrète, alors on prendra la différence entre la valeur la plus grande et la valeur plus petite parmi les modalités.

  Pour l'exemple 1 des appartements : e = 5-1=4
- Si la variable est continue, alors on prendra la différence entre la valeur la plus grande et la valeur plus petite parmi les bornes des classes.

  Pour l'exemple 2 des revenus mensuels : e = 200-0=200

### 4.2.Quartiles

On appelle quartile la valeur qui divise les données triées en quatre parts égales, de sorte que chaque partie représente 1/4 de la population.

- **Cas de données brutes :**

Soit « n » le nombre de données.

- le 1er quartile est la donnée de la série qui sépare le 1er quart (25 %) des données (notée Q1) ;

- le $2^e$ quartile est la donnée de la série qui sépare les deux quarts (50 %) (la médiane) des données (notée Q2).
- le $3^e$ quartile est la donnée de la série qui sépare les trois quarts (75 %) des données (notée Q3).
- Par extension : le $0^{ème}$ quartile est la donnée de la série la plus petite (notée $Q_0$, c'est le minimum) et le $4^{ème}$ quartile est la donnée de la série la plus grande (notée $Q_4$, c'est le maximum)

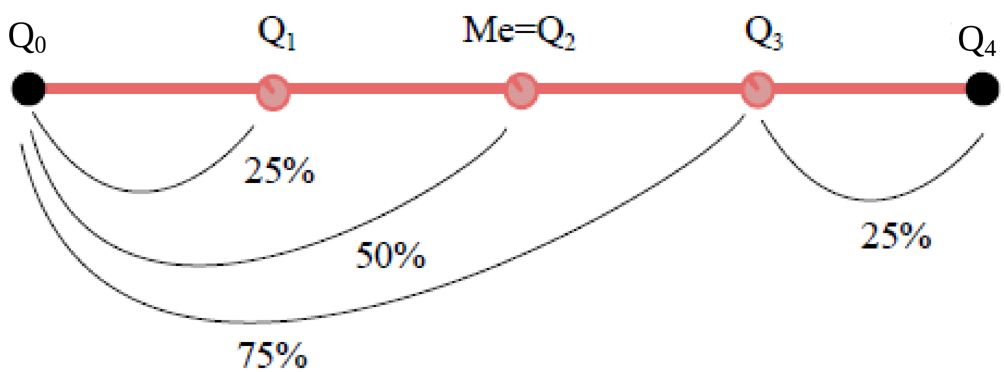


Figure 2-7 Les quartiles

- si $\frac{n}{4}, \frac{n}{2}$ et $\frac{3n}{4}$ sont des entiers naturels, le premier quartile $Q_1$ est le terme de rang $\frac{n}{4}$ , le deuxième quartile $Q_2$ est le terme de rang $\frac{n}{2}$ et le troisième quartile $Q_3$ est le terme de rang $\frac{3n}{4}$.

- si $\frac{n}{4}$ n'est pas un entier naturel on l'arrondit à la valeur entière supérieure la plus proche et la valeur de $Q_1$ sera le terme correspondant à la valeur arrondie.

On refait la même chose pour $Q_2$ et $Q_3$.

**<u>Exemple 8:</u>**

Soit la série suivante : 20, 37, 47, 68, 28, 34, 50, 19, 61, 24,1, 11, 15,

On organise les valeurs par ordre croissant 1, 11, 15, 19, 20, 24, 28, 34, 37, 47, 50, 61, 68.

- Le nombre de valeurs est n = 13, $\frac{n}{4} = 3.25$ n'est pas un entier naturel alors on l'arrondit à 4. Le $1^{er}$ quartile $Q_1$ est la $4^{ème}$ valeur, c'est-à-dire $Q_1$=19.
- $\frac{n}{2} = 6.5$ , ce n'est pas un entier naturel alors on l'arrondit à 7. Le $2^{ème}$ quartile $Q_2$ est la $7^{ème}$ valeur, c'est-à-dire $Q_2$=28.

- $\frac{3n}{4} = 9.75$, ce n'est pas un entier naturel alors on l'arrondit à 10. Le $3^{ème}$ quartile $Q_3$ est la $10^{ème}$ valeur, c'est-à-dire $Q_3$=47.

Finalement, les cinq quartiles de la série sont $Q_0$=1, $Q_1$=19, $Q_2$=28, $Q_3$=47 et $Q_4$=68.

- **Cas de tableau à variable quantitative discrète :**

Pour une variable discrète, on calcule la valeur du quart de l'effectif total ($\frac{N}{4}$ ou son arrondi), on la localise dans les effectifs cumulés croissants puis on repère la valeur de $x_i$ correspondante qui sera le $1^{er}$ quartile $Q_1$.

On fait la même chose pour $Q_2$ et $Q_3$ en prenant, respectivement, $\frac{N}{2}$ et $\frac{3N}{4}$

**<u>Exemple 9 :</u>** On a fait une étude statistique sur 50 notes obtenues à un concours. Les résultats obtenus sont donnés comme suit :

| Note du concours ($x_i$) | Nombre des concurrents ($n_i$) | $n_i cc$ |
|---|---|---|
| 0 | 1 | 1 |
| 1 | 2 | 3 |
| 2 | 2 | 5 |
| 3 | 3 | 8 |
| 4 | 2 | 10 |
| 5 | 3 | 13 |
| 6 | 2 | 15 |
| 7 | 3 | 18 |
| 8 | 4 | 22 |
| 9 | 3 | 25 |
| 10 | 2 | 27 |
| 11 | 3 | 30 |
| 12 | 4 | 34 |
| 13 | 4 | 38 |
| 14 | 3 | 41 |
| 15 | 1 | 42 |
| 16 | 2 | 44 |
| 17 | 1 | 45 |

| | | |
|---|---|---|
| 18 | 2 | 47 |
| 19 | 2 | 49 |
| 20 | 1 | 50 |

- N = 50, donc $\frac{N}{4} = 12.5$, on arrondit à 13 donc le 1er quartile $Q_1 = x_6 = 5$.
- $\frac{N}{2} = 25$ (entier), donc le 2ème quartile $Q_2 = x_{10} = 9$.

$\frac{3N}{4} = 37.5$, on arrondit à 38 donc le 3ème quartile $Q_3 = x_{14} = 13$.

Finalement, les cinq quartiles sont $Q_0=0$, $Q_1=5$, $Q_2=9$, $Q_3=13$et $Q_4=20$.

**Remarque 4 :** si les valeurs des quartiles ne se trouvent pas parmi les valeurs des $x_i$, alors on prendra la valeur immédiatement supérieure.

- **Cas de tableau à variable quantitative continue :**

Pour une variable continue en classes, on calcule les effectifs cumulés croissants puis on localise la classe où les valeurs de $\frac{N}{4}$, $\frac{N}{2}$ et $\frac{3N}{4}$ se trouvent. Enfin on utilise la formule générale suivante pour calculer les trois quartiles ($Q_1$, $Q_2$ et $Q_3$)

$$Q_j = a_i + A_i \left[ \frac{j * \frac{N}{4} - n_{i-1}cc}{n_i} \right], \qquad j \epsilon \{1, 2, 3\}$$

$a_i$ : la borne inférieure de la classe trouvée.

$n_i$ : effectif associée à cette classe.

$n_{(i-1)}cc$ : effectif cumulé immédiatement inférieur à la classe trouvée

N : effectif total.

**Exemple 10 :** Soit le tableau suivant :

| $x_i$ | $n_i$ | $n_icc$ |
|---|---|---|
| [0, 10[ | 10 | 10 |
| [10,20[ | 20 | 30 |
| [20, 30[ | 30 | 60 |
| [30,40[ | 50 | 110 |
| [40, 50[ | 40 | 150 |
| [50, 60[ | 30 | 180 |

- N = 180, donc $\frac{N}{4} = 45$, donc le 1er quartile $Q_1 \in [20, 30[$. En utilisant la formule pour le calcul de $Q_1$, on aura :

$$Q_1 = a_i + A_i \left[\frac{\frac{N}{4} - n_{i-1}cc}{n_i}\right] = 20 + 10\left[\frac{45 - 30}{30}\right] \Rightarrow Q_1 = 25$$

- $\frac{N}{2} = 90$ , donc le 2ème quartile $Q_2 \in [30, 40[$. En utilisant la formule pour le calcul de $Q_2$, on aura :

$$Q_2 = a_i + A_i \left[\frac{\frac{N}{2} - n_{i-1}cc}{n_i}\right] = 30 + 10\left[\frac{90 - 60}{50}\right] \Rightarrow Q_2 = 36$$

- $\frac{3N}{4} = 135$, donc le 3ème quartile $Q_3 \in [40, 50[$. En utilisant la formule pour le calcul de $Q_3$, on aura :

$$Q_3 = a_i + A_i \left[\frac{\frac{3N}{4} - n_{i-1}cc}{n_i}\right] = 40 + 10\left[\frac{135 - 110}{40}\right] \Rightarrow Q_3 = 46.25$$

Finalement, les cinq quartiles sont $Q_0$=0, $Q_1$=25, $Q_2$=36, $Q_3$=46.25 et $Q_4$=60.

**<u>Remarque 5 :</u>**

Pour calculer la médiane ou bien les quartiles dans le cas de tableaux à variable quantitative (discrète ou continue) on doit d'abord ordonner les modalités ar ordre croissant ou décroissant.

### 4.3.Variance

On appelle variance de la série statistique X (notée V(X)) la moyenne de la somme des carrés de la différence entre les observations et leur moyenne.

- **Donnés brutes :**

Elle est donnée par la relation suivante :

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

n : nombre d'observations

Reprenons l'exemple 7 : Soient les données suivantes : 8, 14, 3, 19, 24, 52, 1, 6, 10, 37

La moyenne $\bar{x} = 17.4$, on calcule la variance :

$$V(X) = \frac{(8-17.4)^2 + (14-17.4)^2 + (3-17.4)^2 + (19-17.4)^2 + (24-17.4)^2 + (52-17.4)^2 + (1-17.4)^2}{10}$$

$$+ \frac{(6-17.4)^2 + (10-17.4)^2 + (37-17.4)^2}{10}$$

V(X) = 238.84

- **Cas de tableau à variable quantitative discrète :**

Elle est donnée par la relation suivante :

$$V(X) = \frac{\sum_{i=1}^{r} n_i (x_i - \bar{x})^2}{N}$$

On reprend le 1er exemple

| Modalité (nombre de pièces) $x_i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Effectif (nombre d'appartements) $n_i$ | 48 | 72 | 86 | 64 | 30 |

La moyenne : $\bar{x} = 2.8533$, on calcule la variance :

$$V(X) = \frac{48(1-2.8533)^2 + 72(2-2.8533)^2 + 86(3-2.8533)^2 + 64(4-2.8533)^2 + 30(5-2.8533)^2}{300}$$

$$= 1.4718$$

- **Cas de tableau à variable quantitative continue :**

Pour une variable continue on utilise la relation suivante :

$$V(X) = \frac{\sum_{i=1}^{r} n_i c_i^2}{N} - \bar{x}^2$$

On reprend le 2ème exemple

| Revenu mensuel (en milliers de DA) $x_i$ | [0, 40[ | [40, 80[ | [80, 120[ | [120, 160[ | [160, 200[ |
|---|---|---|---|---|---|
| Effectif (Nombre d'ouvriers) $n_i$ | 35 | 19 | 28 | 12 | 6 |
| $c_i$ | 20 | 60 | 100 | 140 | 180 |

La moyenne : $\bar{x} = 74$, on calcule la variance :

$$V(X) = \frac{35 * 20^2 + 19 * 60^2 + 28 * 100^2 + 12 * 140^2 + 6 * 180^2}{100} - 74^2 = 2444$$

### 4.4.Ecart type

L'écart type d'une série statistique représente la racine carrée de la variance, elle est donnée par :

$$\sigma(X) = \sqrt{V(X)}$$

Pour les exemples précédents, on aura :

1- $V(X) = 238.84 \Rightarrow \sigma(X) = \sqrt{219.04} = 15.4544$

2- $V(X) = 1.4718 \Rightarrow \sigma(X) = \sqrt{1.4718} = 1.2132$

3- $V(X) = 2444 \Rightarrow \sigma(X) = \sqrt{2444} = 49.4368$