

$$\Rightarrow Mé = b_{i-1} + a_i \left( \frac{F(Mé) - F_{i-1}}{F_i - F_{i-1}} \right) = 20 + 10 \left( \frac{0,50 - 0,36}{0,56 - 0,36} \right)$$

$$\Rightarrow Mé = 27 \text{ ans}$$

## 2.1.7 Les Quartiles

Les quartiles  $Q_1$ ,  $Q_2$ ,  $Q_3$  divisent une série statistique en 4 parties d'effectifs égaux : 25% des valeurs sont  $\leq Q_1$ , 25% comprises entre  $Q_1$  et  $Q_2$ ; 25% entre  $Q_2$  et  $Q_3$ , et 25% supérieures à  $Q_3$ .

**Remarque 2.1.10**  $Q_1$ ,  $Q_2$ ,  $Q_3$  sont respectivement l'abscisse des points d'ordonnées 0,25 ; 0,5 ; 0,75 sur la courbe cumulative croissante.  $Q_2$  est égal à la médiane.

C'est-à-dire :

1. L'ordre de  $Q_1$  est  $p = F(Q_1) = 0,25$ .
2. L'ordre de  $Q_2$  est  $p = F(Q_2) = F(Mé) = 0,50$ .
3. L'ordre de  $Q_3$  est  $p = F(Q_3) = 0,75$ .

- **Calcul des quartiles (Cas quantitatif discret)**

On n'a  $p$  l'ordre du quartile  $Q_i$ , avec  $i = 1, 2, 3$  alors :

- Si  $(N \times p)$  est un nombre entier, on a  $Q_i = \frac{v_{(N \times p)} + v_{(N \times p) + 1}}{2}$ .

- Si  $(N \times p)$  n'est pas un nombre entier, on a  $Q_i = v_{[N \times p]}$ .

Où  $[N \times p]$  représente le plus petit nombre entier supérieur ou égal à  $N \times p$  (qui est appelée **partie entière avec excès**).

### Exemples 2.1.9

1. Soit la répartition de 12 ménages selon le nombre d'enfants

$x_i$	0	1	2	3	4
$n_i$	2	2	1	5	2

Le premier quartile  $Q_1$  : Comme  $(N \times p) = 12 \times 0,25 = 3$  est un nombre entier, on a :

$$Q_1 = \frac{v_{(N \times p)} + v_{(N \times p) + 1}}{2} = \frac{v_3 + v_4}{2} = \frac{1 + 1}{2} \Rightarrow Q_1 = 1$$

Le Deuxième quartile  $Mé = Q_2$  : Comme  $(N \times p) = 12 \times 0,50 = 6$  est un nombre entier, on a :

$$Q_2 = Mé = \frac{v_{(N \times p)} + v_{(N \times p) + 1}}{2} = \frac{v_6 + v_7}{2} = \frac{3 + 3}{2} \Rightarrow Q_2 = 3$$

## Statistique descriptive univariée

Le troisième quartile  $Q_3$  : Comme  $(N \times p) = 12 \times 0,75 = 9$  est un nombre entier, on a :

$$Q_3 = \frac{v_{(N \times p)} + v_{(N \times p) + 1}}{2} = \frac{v_9 + v_{10}}{2} = \frac{3 + 3}{2} \Rightarrow Q_3 = 3$$

2. Soit la répartition de 9 ménages selon le nombre d'enfants

$x_i$	0	1	2	3	4
$n_i$	2	2	1	2	2

Le premier quartile  $Q_1$  : Comme  $(N \times p) = 9 \times 0,25 = 2,25$  n'est pas un nombre entier, on a :  $Q_1 = v_{\lceil 2,25 \rceil} = v_3 = 1$ .

Le Deuxième quartile  $Mé = Q_2$  : Comme  $(N \times p) = 9 \times 0,50 = 4,50$  n'est pas un nombre entier, on a :  $Q_2 = v_{\lceil 4,50 \rceil} = v_5 = 2$ .

Le troisième quartile  $Q_3$  : Comme  $(N \times p) = 9 \times 0,75 = 6,75$  n'est pas un nombre entier, on a :  $Q_3 = v_{\lceil 6,75 \rceil} = v_7 = 3$ .

3. Une répartition avec effectif total grand (voir le tableau ci-dessous avec  $N = 50$ )

$x_i$	$n_i$	$N_i$	$F_i$
60	3	3	0,06
65	13	16	0,32
70	9	25	0,50
75	18	43	0,86
85	7	50	1,00
Total	50		

Le premier quartile  $Q_1$   $N \times p = 50 \times \frac{1}{4} = 12,5$  qui n'est pas un nombre entier, donc :

$$Q_1 = v_{\lceil 12,5 \rceil} = v_{13} = x_2 = 65 \text{ car } N_1 = 3 \text{ et } N_2 = 16.$$

Le troisième quartile  $Q_3$   $N \times p = 50 \times \frac{3}{4} = 37,5$  qui n'est pas un nombre entier, donc :

$$Q_3 = v_{\lceil 37,5 \rceil} = v_{38} = x_4 = 75 \text{ car } N_3 = 25 \text{ et } N_4 = 43.$$

### • Calcul des quartiles (Cas quantitatif continu)

Pour le calcul de  $Q_1, Q_2, Q_3$  : On suit les étapes suivantes

1. **Détermination de la classe**  $[b_{i-1}, b_i[$  de  $Q_1 \in [b_{i-1}, b_i]$ , En cherchant la classe qui contient l'individu d'ordre  $\lceil N \times p \rceil = \lceil N/4 \rceil$ .

2. Si  $N_i$  : l'effectif cumulé croissant de la classe de  $Q_1$ ,

$N_{i-1}$  : l'effectif cumulé croissant de la classe qui précède la classe de  $Q_1$ .

$N$  : l'effectif total.

$F_i$  : la fréquence cumulée croissante de la classe de  $Q_1$ ,

$F_{i-1}$  : la fréquence cumulée croissante de la classe qui précède la classe de  $Q_1$ .

Alors on aura

$$Q_1 = b_{i-1} + a_i \left( \frac{N/4 - N_{i-1}}{N_i - N_{i-1}} \right) = b_{i-1} + a_i \left( \frac{0,25 - F_{i-1}}{F_i - F_{i-1}} \right)$$

$N_i$  : l'effectif cumulé croissant de la classe médiane,

$N_{i-1}$  : l'effectif cumulé croissant de la classe avant la classe médiane,

$N$  : l'effectif total.

**Remarque 2.1.11** Le calcul de  $Q_2$  et  $Q_3$  se fait de la même manière tel que

$$Q_2 = Mé = b_{i-1} + a_i \left( \frac{N/2 - N_{i-1}}{N_i - N_{i-1}} \right) = b_{i-1} + a_i \left( \frac{0,5 - F_{i-1}}{F_i - F_{i-1}} \right)$$

$$Q_3 = b_{i-1} + a_i \left( \frac{N(3/4) - N_{i-1}}{N_i - N_{i-1}} \right) = b_{i-1} + a_i \left( \frac{0,75 - F_{i-1}}{F_i - F_{i-1}} \right)$$

**Remarque 2.1.12** L'unité des paramètres de position est l'unité de la variable.

## 2.2 Paramètres de dispersion

**Remarque 2.2.1** Les quartiles déjà vu comme paramètres de positions peuvent être considérés comme paramètres de dispersion.

### 2.2.1 L'étendue

L'étendue noté  $E$  est simplement la différence entre la plus grande et la plus petite valeur observée  $E = x_{max} - x_{min}$ .

### 2.2.2 L'écart interquartile

Il s'agit de l'écart entre le premier et le dernier quartile. C'est-à-dire  $IQ = Q_3 - Q_1$ .

**Remarque 2.2.2** L'écart interquartile mesure l'étendue des 50% de valeurs situées au milieu d'une série de données classées.

**Exemple 2.2.1** On reprend l'exemple de la page 16 sur la répartition des 100 individus selon leur âge.

$[b_{i-1}, b_i[$	$n_i$	$N_i$	$f_i$	$F_i$
$[5, 10[$	11	11	0,11	0,11
$[10, 15[$	10	21	0,10	0,21
$[15, 20[$	15	36	0,15	0,36
$[20, 30[$	20	56	0,20	0,56
$[30, 40[$	18	74	0,18	0,74
$[40, 60[$	16	90	0,16	0,90
$[60, 80[$	10	100	0,10	1,00

Calculons les quartiles  $Q_1$ ,  $Q_3$  et l'écart interquartile.

On a :  $\left\lceil \frac{N}{4} \right\rceil = 25$ ,  $\left\lceil \frac{3N}{4} \right\rceil = 75$ , donc la classe de  $Q_1$  est  $[15, 20[$ , celle de  $Q_3$  est  $[40, 60[$  :

$$\Rightarrow Q_1 = b_{i-1} + a_i \left( \frac{0,25 - F_{i-1}}{F_i - F_{i-1}} \right) \Rightarrow Q_1 = 15 + 5 \left( \frac{0,25 - 0,21}{0,36 - 0,21} \right) = 16,33 \text{ ans}$$

Ce qui signifie que 25 % des individus sont âgés de moins de 16 ans et 4 mois ( $0,33 \times 12 = 3,96 \approx 4$ ). Et pour  $Q_3$  on aura :

$$Q_3 = b_{i-1} + a_i \left( \frac{0,75 - F_{i-1}}{F_i - F_{i-1}} \right) \Rightarrow Q_3 = 40 + 20 \left( \frac{0,75 - 0,74}{0,90 - 0,74} \right) = 41,25 \text{ ans}$$

Ce qui signifie que 75 % des individus sont âgés de moins de 41 ans et 3 mois ( $0,25 \times 12 = 3$ ). Donc l'écart interquartile est :

$$\Rightarrow IQ = Q_3 - Q_1 = 24,92 \text{ ans}$$

Ce qui signifie que la différence d'âge entre  $Q_1$  et  $Q_3$  est de 24 ans, 11 mois et 12 jours ( $0,92 \times 12 = 11,04$  et  $0,4 \times 30 = 12$ ).

**Remarque 2.2.3** Si  $N \times p = N_i$ , alors le quantile  $x_p = b_i$  malgré que  $b_i \notin [b_{i-1}, b_i[$  et la classe du quantile est  $[b_{i-1}, b_i[$ .

**Exemple 2.2.2** Soit la répartition de 100 personnes selon leur âge

Classes	$n_i$	$N_i$	$f_i$	$F_i$
$[20, 30[$	25	25	0,25	0,25
$[30, 40[$	20	45	0,20	0,45
$[40, 60[$	35	80	0,35	0,80
$[60, 80[$	20	100	1,00	1,00

Calcul du 1<sup>er</sup> quartile  $Q_1$  :

On a l'ordre du 1<sup>er</sup> quartile  $Q_1$  est  $p = 0,25$ .

Comme  $\lceil N \times p \rceil = \lceil 100 \times 0,25 \rceil = \lceil 25 \rceil = 25$  et  $N_1 = 25$ , alors :

La casse de  $Q_1$  est  $[20, 30[$ , c'est à dire  $Q_1 \in [20, 30]$ . Donc :

$$Q_1 = b_0 + a_1 \left( \frac{N(1/4) - N_0}{N_1 - N_0} \right) = 20 + 10 \left( \frac{25 - 0}{25 - 0} \right) = 30 \in [20, 30]$$

**Remarque 2.2.4** On peut obtenir des valeurs approximatives des quartiles graphiquement à partir de la **courbe cumulative**.

### 2.2.3 La variance

La variance d'une variable  $X$  notée  $V(x)$  est la somme des carrés des écarts à la moyenne divisée par le nombre d'observations (Effectif total  $N$ ).

- **Le cas d'une variable discrète**

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

- **Le cas d'une variable continue**

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

avec  $c_i = \frac{b_{i-1} + b_i}{2}$ , le centre de la classe  $[b_{i-1}, b_i]$ .

**Remarque 2.2.5** La variance peut être écrite sous une autre forme dite « formule développée »

- **La formule développée de la variance pour le cas d'une variable discrète**

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

- **La formule développée de la variance pour le cas d'une variable continue**

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

### Preuve de la formule développée

On a :

$$\begin{aligned}V(x) &= \frac{1}{N} \sum_{i=1}^{i=r} n_i (x_i - \bar{x})^2 = \frac{1}{N} \sum_{i=1}^{i=r} n_i (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\ \Rightarrow V(x) &= \frac{1}{N} \sum_{i=1}^{i=r} n_i x_i^2 - 2 \frac{\bar{x}}{N} \sum_{i=1}^{i=r} n_i x_i + \frac{\bar{x}^2}{N} \sum_{i=1}^{i=r} n_i \\ \Rightarrow V(x) &= \left( \frac{1}{N} \sum_{i=1}^{i=r} n_i x_i^2 \right) - 2\bar{x} \cdot \bar{x} + \bar{x}^2\end{aligned}$$

### Remarque 2.2.6

1. Cette formule développée de la variance est plus facile à retenir et plus rapide à calculer.
2. La variance est exprimée dans le carré de l'unité de la variable. Par exemple, la variance de la variable âge est exprimée en « années au carré (année<sup>2</sup>) » car :

## 2.2.4 L'écart type

On appelle écart type que l'on le note par  $\sigma(x)$ , la racine carrée de la variance :

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

### Remarque 2.2.7

1. L'écart type est exprimé dans la même unité de mesure que la variable.
2. Il est utilisé comme un indicateur de la dispersion de la série statistique, de façon que dans un rangement croissant la moyenne  $\bar{x}$  partage la population en deux parties tel que les individus ayant la valeur de la variable inférieure à  $\bar{x}$  auront approximativement  $\bar{x} - \sigma(x)$ , les autres ( $X > \bar{x}$ ) auront  $\bar{x} + \sigma(x)$ .
3. Plus l'écart type est grand, plus la dispersion des observations autour de la moyenne de la variable est forte.
4. Une distribution aura un écart-type proche de 0 si ces valeurs seront ramassées autour de la moyenne.

**Exemple 2.2.3** Considérons les notes suivantes en statistique d'un groupe de 20 étudiants :

$x_i$	$n_i$	$n_i \cdot x_i$	$n_i \cdot x_i^2$
2	2	4	8
3	2	6	18
7	4	28	196
8	2	16	128
12	3	36	432
17	2	34	578
18	5	90	1620
Total	20	214	2980

$$\text{Donc } \bar{x} = \sum_{i=1}^{i=r} \frac{n_i x_i}{N} = \frac{214}{20} = 10,7 \text{ et } V(x) = \left( \frac{1}{N} \sum_{i=1}^{i=r} n_i x_i^2 \right) - \bar{x}^2$$

$$\Rightarrow V(x) = \frac{2980}{20} - (10,7)^2 \Rightarrow V(x) = 149 - 114,49 = 34,51$$

$$\Rightarrow \sigma(x) = 5,87.$$

Donc, certains étudiants (les bons) auront approximativement la note moyenne (10,7) plus (+) 5,87 (=16,57) les autres (les mauvais) auront la note moyenne (10,7) moins (-) 5,87 (= 4,83).

### 2.2.5 L'écart absolu moyen

Pour une série statistique dont la variable peut prendre les valeurs  $x_1, x_2, \dots, x_r$ , les effectifs correspondants étant  $n_1, n_2, \dots, n_r$ , l'écart-absolu moyen de la variable  $X$  est la moyenne arithmétique des valeurs absolues des écarts à la moyenne arithmétique est défini par :

$$e_{\bar{x}} = \frac{n_1|x_1 - \bar{x}| + n_2|x_2 - \bar{x}| + \dots + n_r|x_r - \bar{x}|}{N} = \frac{1}{N} \sum_{i=1}^{i=r} n_i |x_i - \bar{x}|$$

#### Remarque 2.2.8

1. L'écart absolu moyen permet de mesurer la dispersion d'une série. Par exemple, si un premier (1) étudiant a eu pour notes 5, 10, 15 et un deuxième (2) étudiant 9, 10, 11, ils ont même moyenne alors que clairement leurs cas sont très différents. L'écart absolu moyen du premier étudiant est  $10/3 = 3,333$ , tandis que celui du deuxième étudiant est  $2/3 = 0,6666$ . On lit donc bien cette différence sur l'écart absolu moyen.

2. Plus l'écart absolu moyen par rapport à la moyenne est élevé, plus il y a de valeurs éloignées de la moyenne : l'écart absolu moyen est donc bien un paramètre de dispersion, c'est-à-dire un indicateur de l'étalement des valeurs recueillies.

3. On peut, de manière similaire, définir un écart moyen par rapport à la médiane quand on juge que la médiane est un paramètre de position plus approprié à l'étude d'une série statistique que la moyenne arithmétique.

### 2.2.6 Le coefficient de variation

L'écart type, et la moyenne, s'exprime dans la même unité que la variable statistique, mais dans certain cas on peut être ramené à comparer les dispersions qui ne sont pas exprimées dans la même unité. Donc le coefficient de variation se calcule comme le **rapport de l'écart type à la moyenne**, et s'exprime en pourcentage.

$$CV(x) = \frac{\sigma(x)}{\bar{x}}$$

#### Remarque 2.2.9

1. Le coefficient de variation permet de comparer le degré de variation d'un échantillon à un autre, même si les moyennes sont différentes.
2. Plus la valeur du coefficient de variation est élevée, plus la dispersion autour de la moyenne est grande. Il est généralement exprimé en pourcentage. Sans unité, il permet la comparaison de distributions de valeurs dont les échelles de mesure ne sont pas comparables.
3. Plus le coefficient de variation est faible, plus les données statistiques sont regroupées autour de la moyenne et plus il est grand, plus les données sont dispersées.

**Exemple 2.2.3** Dans une maternité on a relevé le poids (en kilogramme) à la naissance de 47 nouveaux nés. Les données collectées sont résumées dans le tableau suivant :

Classes	$n_i$	$c_i$	$n_i c_i$	$n_i c_i^2$
[ 2,5 ; 3,0 [	08	2,75	22,00	60,50
[ 3,0 ; 3,5 [	15	3,25	48,75	158,438
[ 3,5 ; 4,0 [	20	3,75	75,00	281,25
[ 4,0 ; 4,5 [	04	4,25	17,00	72,25
<b>Total</b>	47		162,75	572,438

$$\text{Donc } \bar{x} = \sum_{i=1}^{i=4} \frac{n_i c_i}{N} = \frac{162,75}{47} = 3,463 \text{ et } V(x) = \left( \frac{1}{N} \sum_{i=1}^{i=r} n_i x_i^2 \right) - \bar{x}^2$$

$$\Rightarrow V(x) = \frac{572438}{47} - (3,463)^2 \Rightarrow V(x) = 0,189$$

$$\Rightarrow \sigma(x) = \sqrt{V(x)} = 0,434 \Rightarrow CV(x) = \frac{\sigma(x)}{\bar{x}} = \frac{0,434}{3,463} = 0,125.$$