# 1 Paramètres de dispersion : Etendue de la série statistique, Intervalle interquartile, Ecart-type

**PRÉSENTATION D'UNE SÉRIE STATISTIQUE : :**

Cas n°1 : Les valeurs prises par le caractère sont listées

Voici les longueurs en km des vingt-et-une étapes du Tour de France 2009 :

| | | | | | | |
|---|---|---|---|---|---|---|
| 15,5 | 187 | 196,5 | 39 | 196,5 | 181,5 | 224 |
| 176,5 | 160,5 | 194,5 | 192 | 211,5 | 200 | 199 |
| 207,5 | 159 | 169,5 | 40,5 | 178 | 167 | 164 |

Cas n°2 : Les valeurs prises par le caractère sont présentées dans un tableau : qui à une valeur prise par le caractère notée $x_i$ ,appelée **modalité**, lui associe son effectif noté $n_i$ ou bien sa fréquence notée $f_i$

**9** Au large de l'Afrique du Sud, on a répertorié les tailles de 96 requins blancs.

| Taille (en m) | 1,5 | 2 | 2,5 | 3 | 3,5 | 4 | 4,5 |
|---|---|---|---|---|---|---|---|
| Effectif | 7 | 10 | 24 | 32 | 18 | 4 | 1 |

**RAPPEL , CALCUL DE LA MOYENNE D'UNE SÉRIE STATISTIQUE :**

**La moyenne $\bar{x}$ de la série statistique donnée par le tableau ci-contre est définie par :**

$$\bar{x} = \frac{n_1x_1+n_2x_2+...+n_px_p}{N} \text{ ou par } \bar{x} = f_1x_1+f_2x_2+...+f_px_p;\left(f_1=\frac{n_1}{N}\right)$$

| valeur | $x_1$ | $x_2$ | ... | $x_p$ | TOTAL |
|---|---|---|---|---|---|
| effectif | $n_1$ | $n_2$ | ... | $n_p$ | N |

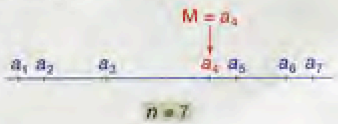
# 2 Médiane

**Définition 1**

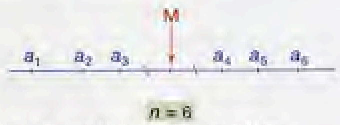
On appelle **médiane** de la série statistique $(x_1, x_2, ..., x_n)$ le nombre M obtenu de la façon suivante :

• on range d'abord les valeurs du caractère par ordre croissant, chacune figurant un nombre de fois égal à son effectif. On obtient la suite $a_1 \leqslant a_2 \leqslant ... \leqslant a_n$ ;

• si **$n$ est impair**, M est le nombre $a_i$ de cette suite situé « au milieu », c'est-à-dire tel que $i = \frac{n+1}{2}$ ;




• si **$n$ est pair**, M est le nombre situé au centre de l'intervalle $\left[a_{\frac{n}{2}} ; a_{\frac{n}{2}+1}\right]$ (intervalle médian).




**1.**

| valeur du caractère | 50 | 45 | 30 | 60 | 61 |
|---|---|---|---|---|---|
| effectif | 2 | 3 | 2 | 2 | 1 |

Commençons par ranger les valeurs du caractère par ordre croissant, **chacune figurant un nombre de fois égal à son effectif** : 30 ; 30 ; 45 ; 45 ; 45 ; 50 ; 50 ; 60 ; 60 ; 61 .
Ici $n$ est pair ($n = 10$) ; donc la médiane est le milieu de l'intervalle médian, c'est-à-dire l'intervalle délimité par la 5ᵉ et la 6ᵉ valeur, c'est-à-dire [45 ; 50] donc M = 47,5 .

**2.**

| valeur du caractère | 50 | 45 | 30 | 60 | 61 |
|---|---|---|---|---|---|
| effectif | 2 | 3 | 2 | 2 | 2 |

Commençons par ranger les valeurs du caractère par ordre croissant, **chacune figurant un nombre de fois égal à son effectif** : 30 ; 30 ; 45 ; 45 ; 45 ; 50 ; 50 ; 60 ; 60 ; 61 ; 61 .
Ici $n$ est impair ($n = 11$) ; donc la médiane est le 6ᵉ terme $(i = \frac{11+1}{2})$ ; donc M = 50 .

# 3 Quartiles

La médiane M sépare une série statistique en deux sous-séries de même effectif, l'une contenant les plus petites valeurs, l'autre les plus grandes.
*Les quartiles sont les médianes de ces sous-séries. Le premier quartile noté $Q_1$ est la médiane de la sous-série inférieure, le troisième quartile noté $Q_3$ est la médiane de la sous-série supérieure.*

**Note :** La médiane est parfois appelée deuxième quartile.

**Définition 2**

Le nombre $Q_3 - Q_1$ est appelé **écart interquartile**.
L'intervalle $]Q_1 ; Q_3[$ est appelé **intervalle interquartile**.

**Remarque 1.** Les nombres $Q_1$, M et $Q_3$ permettent de couper la population étudiée en quatre groupes contenant chacun le même nombre d'éléments.

**Remarque 2. Notion de déciles.** On peut définir de manière analogue la notion de déciles : ce sont des nombres qui permettent de couper la population étudiée en dix groupes contenant chacun le même nombre d'éléments.

**1.**

| valeur du caractère | 50 | 45 | 30 | 60 | 61 |
|---|---|---|---|---|---|
| effectif | 2 | 3 | 2 | 2 | 2 |

• Commençons par ranger les valeurs du caractère par ordre croissant, **chacune figurant un nombre de fois égal à son effectif** : 30 ; 30 ; 45 ; 45 ; 45 ; 50 ; 50 ; 60 ; 60 ; 61 ; 61 . Ici $n$ est impair ($n = 11$), donc la médiane M est la 6$^e$ valeur, c'est-à-dire 50 . Donc M = 50 .

• On a donc $\underbrace{30 ; 30 ; 45 ; 45 ; 45}_{\text{série inférieure}} ; \underset{\text{M}}{50} , \underbrace{50 ; 60 ; 60 ; 61 ; 61}_{\text{série supérieure}}$ .

• $Q_1$ est la médiane de la série inférieure (30 ; 30 ; 45 ; 45 ; 45), donc $Q_1 = 45$ .

• $Q_3$ est la médiane de la série supérieure (50 ; 60 ; 60 ; 61 ; 61), donc $Q_3 = 60$ .

**2.**

| valeur du caractère | 50 | 45 | 30 | 60 | 70 |
|---|---|---|---|---|---|
| effectif | 2 | 5 | 3 | 2 | 2 |

• Commençons par ranger les valeurs du caractère par ordre croissant, **chacune figurant un nombre de fois égal à son effectif** .

Ici $n$ est pair ($n = 14$), donc M est le centre de l'intervalle d'extrémités la 7$^e$ valeur et la 8$^e$ valeur, donc **M** = 45 .

$\underbrace{30 ; 30 ; 30 ; 45 ; 45 ; 45 ; 45}_{\text{série inférieure}} ; \underset{\text{M}}{} 45 ; 50 ; 50 ; 60 ; 60 ; 70 ; 70$ .

$$\underbrace{30 ; 30 ; 30 ; 45 ; 45 ; 45 ; 45}_{\text{série inférieure}} \; \overset{\uparrow}{\text{M}} \; \underbrace{45 ; 50 ; 50 ; 60 ; 60 ; 70 ; 70}_{\text{série supérieure}}$$

• $Q_1$ est la médiane de la série inférieure (30 ; 30 ; 30 ; 45 ; 45 ; 45 ; 45), donc $\mathbf{Q_1 = 45}$ .

• $Q_3$ est la médiane de la série supérieure (45 ; 50 ; 50 ; 60 ; 60 ; 70 ; 70), donc $\mathbf{Q_3 = 60}$ .

## 4 Diagramme en boîtes : un exemple

Reprenons l'exemple étudié dans le paragraphe précédent.

| valeur | 50 | 45 | 30 | 60 | 61 |
|---|---|---|---|---|---|
| effectif | 2 | 3 | 2 | 2 | 2 |

Nous avons trouvé : M = 50 ; $Q_1 = 45$ ; $Q_3 = 60$ .

D'autre part, la plus petite valeur de cette série est 30 et la plus grande 61 .

On peut représenter graphiquement ces résultats de la manière suivante :


**Note :** On peut aussi construire ce graphique verticalement. Les calculatrices usuelles l'affichent horizontalement.

Les valeurs $Q_1$ et $Q_3$ correspondent aux côtés verticaux délimitant la boîte, M au côté vertical intérieur à la boîte.

**Définition 3**

Un tel diagramme est appelé **diagramme en boîtes.**

## 5 Variance – Écart-type

**Définition 4**

La **variance V** de la série statistique donnée par le tableau ci-contre est définie par :

| valeur | $x_1$ | $x_2$ | ... | $x_p$ | TOTAL |
|---|---|---|---|---|---|
| effectif | $n_1$ | $n_2$ | ... | $n_p$ | N |

$$V = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \ldots + n_p(x_p - \bar{x})^2}{N}$$

où $\bar{x}$ désigne la moyenne de cette série.

L'**écart-type σ** est la racine carrée de la variance : $\sigma = \sqrt{V}$ .

**Note :** On dit que la variance et l'écart type sont des paramètres de dispersion (voir le paragraphe 1, page 72).

**Exemple :** Considérons les deux séries de notes suivantes, présentées en **1** .

Première série

| note $x_i$ | 1 | 2 | 3 | 17 | 20 |
|---|---|---|---|---|---|
| effectif $n_i$ | 3 | 1 | 1 | 1 | 4 |

Deuxième série

| note $x_i$ | 8 | 10 | 11 | 12 |
|---|---|---|---|---|
| effectif $n_i$ | 1 | 2 | 4 | 1 |

Chacune de ces séries a pour moyenne 10,5 .

Notons $V_1$ et $V_2$ les variances respectives de chacune de ces séries et $\sigma_1$ , $\sigma_2$ leurs écarts-types.

$$V_1 = \frac{(1-10{,}5)^2 \times 3 + (2-10{,}5)^2 \times 1 + (3-10{,}5)^2 \times 1 + (17-10{,}5)^2 \times 1 + (20-10{,}5)^2 \times 4}{10} = 80{,}25$$

$$V_2 = \frac{(8-10{,}5)^2 \times 1 + (10-10{,}5)^2 \times 2 + (11-10{,}5)^2 \times 4 + (12-10{,}5)^2 \times 1}{10} = 1{,}25$$

donc $\sigma_1 \approx 8{,}96$ et $\sigma_2 \approx 1{,}12$ .

**Note :** L'écart-type de la première série est nettement supérieur à celui de la seconde. Ce résultat était prévisible *a priori*, compte tenu de la plus grande dispersion des notes de la première série.