

# Statistiques

## I. Moyenne, variance et écart-type d'une série statistique

### 1. Série statistique discrète :

Exemple d'une série statistique discrète :

Prenons le cas d'une classe de 23 élèves qui réalisent un devoir noté sur 5. La série statistique discrète suivante présente les notes des élèves :

Note	0	1	2	3	4	5	Total
Nombre d'élève	1	2	8	9	2	1	23

On peut donc voir par exemple que 2 élèves ont obtenu la note de 4/5. La taille de cette série est 23, c'est-à-dire le nombre d'élèves.

On peut également calculer les fréquences :

Note	0	1	2	3	4	5	Total
Nombre d'élève	1	2	8	9	2	1	23
Fréquence	$\frac{1}{23}$	$\frac{2}{23}$	$\frac{8}{23}$	$\frac{9}{23}$	$\frac{2}{23}$	$\frac{1}{23}$	1

D'une façon générale, on note :

Valeur	$x_1$	$x_2$	...	$x_p$	Total
Effectif	$n_1$	$n_2$	...	$n_p$	$n$
Fréquence	$f_1 = \frac{n_1}{n}$	$f_2 = \frac{n_2}{n}$	...	$f_p = \frac{n_p}{n}$	1

$n$ , effectif total, est aussi appelé taille de la série.

### 2. Moyenne :

La moyenne de cette série statistique discrète, généralement notée  $\bar{x}$ , est définie par :

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_p x_p}{n}$$

soit en plus condensé :

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

Remarque :  $p$  représente le nombre de valeurs différentes

Dans l'exemple du 1., la moyenne est donc :

$$\text{moyenne} = \bar{x} = \frac{1}{23} \sum_{i=1}^6 n_i x_i$$

$$\text{moyenne} = \bar{x} = \frac{1}{23} (1 \times 0 + 2 \times 1 + 8 \times 2 + 9 \times 3 + 2 \times 4 + 1 \times 5) = \frac{58}{23} \approx 2,5$$

### 3. Variance :

La variance d'une série statistique discrète est généralement notée  $V$  et est définie par :

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

On pourrait également démontrer que :

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

Dans l'exemple du 1., la variance est donc :

$$\text{variance} = V = \left( \frac{1}{23} \sum_{i=1}^6 n_i x_i^2 \right) - \left( \frac{58}{23} \right)^2$$

$$V = \frac{1}{23} (1 \times 0^2 + 2 \times 1^2 + 8 \times 2^2 + 9 \times 3^2 + 2 \times 4^2 + 1 \times 5^2) - \left( \frac{58}{23} \right)^2$$

$$V = \frac{1}{23} (2 + 8 \times 4 + 9 \times 9 + 2 \times 16 + 1 \times 25) - \left( \frac{58}{23} \right)^2 = \frac{172}{23} - \left( \frac{58}{23} \right)^2 \approx 1,12$$

## 4. Ecart-type :

L'écart-type d'une série statistique discrète est généralement noté  $\sigma$  et est défini par :

$$\sigma = \sqrt{V}$$

L'écart-type est la racine carrée de la variance.

☝ L'écart-type représente l'écart moyen des valeurs de la série par rapport à sa moyenne. Une série statistique très dispersée (dont les valeurs sont très éloignées de la moyenne) aura un écart-type important. A l'inverse, une série dont les valeurs sont très proches de la moyenne aura un écart-type faible.

Dans l'exemple du 1., l'écart-type est :

$$\sigma \approx \sqrt{1,12} \approx 1,06$$

## II. Effet d'une transformation affine

---

Soit une série statistique  $(x_i, n_i)$  avec  $1 \leq i \leq p$  et de taille  $n$ .

On note  $\bar{x}$  la moyenne de cette série et d'écart-type  $\sigma_x$ .

Considérons  $a$  et  $b$ , deux réels. On définit la série statistique  $(y_i, n_i)$  avec  $1 \leq i \leq p$  et de taille  $n$  par la transformation affine suivante :  $y_i = ax_i + b$ .

Cette nouvelle série statistique a pour moyenne  $\bar{y} = a\bar{x} + b$  et pour écart-type  $\sigma_y = |a|\sigma_x$ .

## III. Médiane et interquartiles

---

### 1. Quartiles :

Considérons une série statistique, de taille  $n$ , et dont les termes sont rangés dans l'ordre croissant  $(x_1 \leq x_2 \leq \dots \leq x_n)$ .

Remarque : les  $x$  représentent ici les termes de la série et non les valeurs. Dans l'exemple du I.1. on aurait donc  $x_1 = 0$ ,  $x_2 = 1$ ,  $x_3 = 1$ ,  $x_4 = 2$ ,  $x_5 = 2 \dots$ ,  $x_{22} = 4$ ,  $x_{23} = 5$ .

- Le premier quartile, généralement noté  $Q_1$  est la plus petite valeur de la série pour laquelle au moins  $\frac{1}{4}$  (25%) des données sont inférieures ou égale à  $Q_1$ .

- Le troisième quartile, généralement noté  $Q_3$  est la plus petite valeur de la série pour laquelle au moins  $\frac{3}{4}$  (75%) des données sont inférieures ou égale à  $Q_3$ .

On définit également l'intervalle interquartile :  $[Q_1; Q_3]$ .

L'écart interquartile se définit par :  $Q_3 - Q_1$ .

## 2. Médiane :

La médiane d'une série est la valeur séparant les termes en deux « groupes » de même effectif.

Autrement dit, si on note  $M$  la médiane, on peut dire que au moins  $\frac{1}{2}$  (50%) des données sont inférieures ou égales à  $M$  **ET** au moins  $\frac{1}{2}$  des données sont supérieures ou égales à  $M$ .

### Méthode de calcul de la médiane :

Si  $n$  est impair (c'est le cas le plus facile), la médiane est le terme central :  $M = x_{\frac{n+1}{2}}$

Si  $n$  est pair, la médiane est la moyenne des deux termes centraux :  $M = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$

Dans l'exemple du I.1,  $n = 23$  est donc impair, on a donc  $M = x_{\frac{23+1}{2}} = x_{12} = 3$

## 3. Exercice d'application :

### Énoncé :

Des biologistes capturent 20 truites dans une rivière et les mesurent. Ils obtiennent les chiffres suivant (en cm, classés dans l'ordre croissant) :

$x_1 = 8, x_2 = 10, x_3 = 10, x_4 = 13, x_5 = 13, x_6 = 14, x_7 = 14, x_8 = 15, x_9 = 16, x_{10} = 16,$   
 $x_{11} = 16, x_{12} = 17, x_{13} = 18, x_{14} = 19, x_{15} = 19, x_{16} = 21, x_{17} = 22, x_{18} = 22, x_{19} = 25,$   
 $x_{20} = 26.$

Donner la moyenne, la variance, l'écart-type, le premier quartile, le troisième quartile, l'intervalle interquartile, l'écart interquartile et la médiane de cette série.

## Résolution de l'exercice :

La taille de cette série est 20 (il y a 20 éléments dans la série).

**La moyenne** est la somme de toutes les valeurs divisée par la taille de la série :

$$\text{moyenne} = 16,7$$

Pour calculer **la variance**, on repart de la formule :

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

$$V = \frac{1}{20} \left[ (8-16,7)^2 + 2(10-16,7)^2 + 2(13-16,7)^2 + 2(14-16,7)^2 + (15-16,7)^2 + 3(16-16,7)^2 + (17-16,7)^2 \right. \\ \left. + (17-16,7)^2 + (18-16,7)^2 + 2(19-16,7)^2 + (21-16,7)^2 + 2(22-16,7)^2 + (25-16,7)^2 + (26-16,7)^2 \right]$$

$$V = 22,71$$

**L'écart-type** est la racine carrée de la variance :

$$\sigma = \sqrt{V} = \sqrt{22,71} \approx 4,77$$

Détermination des **premier et troisième quartiles** :

Il y a 20 éléments en tout.

$\frac{1}{4} \times 20 = 5$ , donc **le premier quartile**  $Q_1 = x_5 = 13$ . Le premier quartile vaut 13 (13 est donc la plus petite valeur pour laquelle au moins  $\frac{1}{4}$  des données sont inférieures ou égale à elle-même).

$\frac{3}{4} \times 20 = 15$ , donc **le troisième quartile**  $Q_3 = x_{15} = 19$ . Le premier quartile vaut 19 (19 est donc la plus petite valeur pour laquelle au moins  $\frac{3}{4}$  des données sont inférieures ou égale à elle-même).

**L'intervalle interquartile** est donc  $[13;19]$ .

**L'écart interquartile** vaut  $Q_3 - Q_1 = 19 - 13 = 6$ .

Calculons **la médiane** : ici,  $n = 20$  est pair donc  $M = \frac{x_{\frac{20}{2}} + x_{\frac{20}{2}+1}}{2} = \frac{x_{10} + x_{11}}{2} = \frac{16+16}{2} = 16$

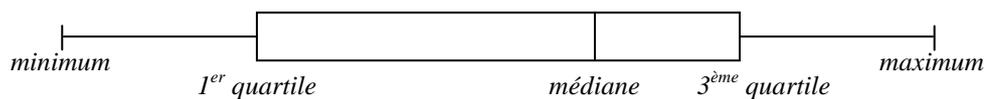
## IV. Diagrammes en boîte

---

Les diagrammes en boîte, également appelés boîtes à moustaches ou diagrammes de Tukey, permettent de représenter visuellement l'ensemble d'une série statistique.

### Méthode de construction :

- Sur un axe horizontal (ou vertical), on place les cinq valeurs suivantes : le minimum, le maximum, la médiane, le premier quartile  $Q_1$  et le troisième quartile  $Q_3$ .
- On construit un rectangle horizontal (ou vertical) allant de  $Q_1$  à  $Q_3$ . Son grand côté est donc égal à l'écart interquartile  $Q_3 - Q_1$ .
- Ce rectangle est séparé par un segment passant par la médiane.
- On ajoute ensuite deux segments : l'un allant de  $Q_1$  au minimum et l'autre allant de  $Q_3$  au maximum.



Reprendre la série statistique du **III. 3.** et construire le diagramme en boîte de cette série :

