LES STATISTIQUES EN PREMIÈRE

Dans tout ce chapitre, on considère une série statistique représentée par le tableau :

Valeurs	x_1	x_2	...	x_p	**Total**
Effectifs	n_1	n_2	...	n_p	N

1. PARAMÈTRES DE POSITION

DÉFINITION

La **moyenne** d'une série statistique est le nombre :

$$\overline{x} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_p x_p}{N} = \frac{1}{N} \sum_{k=1}^{p} n_k x_k$$

EXEMPLE

Les âges des élèves d'un lycée sont donnés par le tableau :

Ages	14	15	16	17	18	19	20	**Total**
Effectifs	2	52	78	75	81	25	2	315

La moyenne des âges vaut :

$\overline{x} = \frac{1}{315}(2 \times 14 + 52 \times 15 \ \ +78 \times 16 + 75 \times 17 + 81 \times 18 + 25 \times 19 + 2 \times 20)$

$\overline{x} = \frac{5304}{315} \approx 16,84$ à 10^{-2} près.

DÉFINITION

La **médiane** d'une série statistique est la valeur du caractère qui partage la population en deux classes de même effectif.

REMARQUE

En pratique pour trouver la médiane d'une série statistique d'effectif global N :

- On ordonne les valeurs du caractère dans l'ordre croissant.
- Si N est pair, la médiane sera la moyenne des valeurs du terme de rang $\frac{N}{2}$ et du terme de rang $\frac{N}{2}+1$.
- Si N est impair, la médiane sera la valeur du terme de rang $\frac{N+1}{2}$.
- Lorsque l'effectif global est élevé, il est souvent utile de calculer les effectifs cumulés pour trouver cette valeur.

EXEMPLE

On lance 10 fois un dé à six faces. Les résultats obtenus sont : 1 ; 5 ; 6 ; 6 ; 3 ; 2 ; 3 ; 1 ; 4 ; 1

On trie ces valeurs par ordre croissant : 1 ; 1 ; 1 ; 2 ; 3 ; 3 ; 4 ; 5 ; 6 ; 6

N=10 étant pair on effectue la moyenne du cinquième et du sixième terme (3 et 3) et on obtient donc 3.

REMARQUE

Voir la fiche de Statistiques en seconde pour un exemple plus détaillé.

2. PARAMÈTRES DE DISPERSION

DÉFINITIONS

La **variance** d'une série statistique est le nombre :

$$\begin{aligned} V &= \frac{1}{N}\left(n_1\left(x_1-\overline{x}\right)^2+n_2\left(x_2-\overline{x}\right)^2+\ldots\ +n_p\left(x_p-\overline{x}\right)^2\right) \\ &= \frac{1}{N}\sum_{k=1}^{p} n_k\left(x_k-\overline{x}\right)^2 \end{aligned}$$

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

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

PROPRIÉTÉ

La **variance** d'une série statistique est égale à :

$$V=\frac{n_1x_1^2+n_2x_2^2+\ldots+n_px_p^2}{N}-\overline{x}^2=\overline{x^2}-\overline{x}^2$$

DÉFINITIONS

- Le **premier quartile** Q1 d'une série statistique est la plus petite valeur des termes de la série pour laquelle au moins un quart des données sont inférieures ou égales à Q1.
- Le **troisième quartile** Q3 d'une série statistique est la plus petite valeur des termes de la série pour laquelle au moins trois quarts des données sont inférieures ou égales à Q3.
- Le **premier décile** D1 d'une série statistique est la plus petite valeur des termes de la série pour laquelle au moins 10% des données sont inférieures ou égales à D1.
- Le **neuvième décile** D9 d'une série statistique est la plus petite valeur des termes de la série pour laquelle au moins 90% des données sont inférieures ou égales à D9

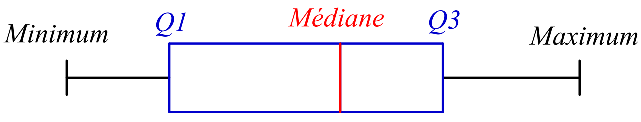
DÉFINITION

L'**écart interquartile** est la différence entre le troisième et le premier quartile $Q_3 - Q_1$.

REMARQUE

L'écart interquartile mesure la dispersion autour de la médiane.

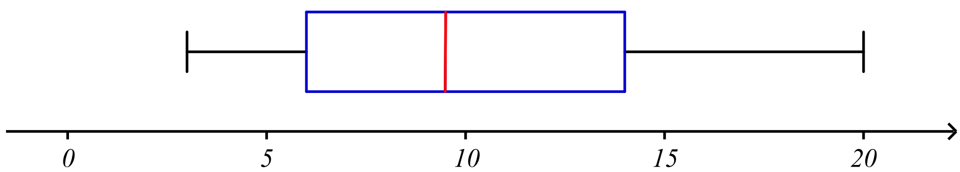
3. DIAGRAMME EN BOÎTE



On peut résumer un certain nombre d'informations relatives à une série statistique grâce à un **diagramme en boîte** (aussi appelé *boîte à moustache*) qui fait apparaître (voir figure ci-dessus) :

- les valeurs minimum et maximum
- le premier et le troisième quartile (Q1 et Q3)
- la médiane

EXEMPLE



Le figure ci-dessus représente une série statistique de valeurs extrêmes 3 et 20, de premier quartile 6, de troisième quartile 14 et de médiane 9,5.

REMARQUE

Parfois, notamment lorsqu'on étudie des séries dont certaines valeurs peuvent être erronées, on remplace les valeurs minimum et maximum par les premier et neuvième déciles afin d'éliminer les valeurs aberrantes.