

Statistiques descriptives

Cours de maths - 2de

Une série statistique est une longue liste de nombres : impossible de la lire d'un coup d'œil ! Dans ce chapitre, tu apprends à la **résumer** par quelques nombres bien choisis : des indicateurs de **position** (moyenne, médiane, quartiles) et des indicateurs de **dispersion** (étendue, écart interquartile, écart-type). Tu apprends aussi à les **dessiner** (diagramme en boîte) et à les utiliser pour **comparer** deux séries.

1. Rappels : moyenne, médiane, étendue

Série statistique Une série statistique est la liste des valeurs $x_1, x_2, \dots, x_n$ prises par un caractère quantitatif sur une population de **n** individus (n est l'**effectif total**).

Moyenne, médiane, étendue La **moyenne** est $\bar{x} = [(x_1 + x_2 + \dots + x_n)/n]$. La **médiane** Me est une valeur qui partage la série *rangée dans l'ordre croissant* en deux groupes de même effectif : si n est impair, c'est la valeur du milieu ; si n est pair, c'est la moyenne des deux valeurs centrales. L'**étendue** est la différence entre la plus grande et la plus petite valeur.

Exemple Temps de trajet (en minutes) d'un élève sur 7 jours : 12 ; 15 ; 9 ; 20 ; 15 ; 11 ; 18. Moyenne : $100 \div 7 \approx \mathbf{14,29 \text{ min}}$. Rangée : 9 ; 11 ; 12 ; 15 ; 15 ; 18 ; 20 ; n = 7 est impair, la médiane est la 4^e valeur : **15 min**. Étendue : $20 - 9 = \mathbf{11 \text{ min}}$.

2. Effectifs, fréquences et cumuls

Quand les valeurs se répètent, on regroupe la série dans un tableau. Voici le nombre de livres lus pendant un trimestre par 30 élèves d'une classe (données inventées).

Nombre de livres lus	0	1	2	3	4	5	Total
Effectif	3	6	8	7	4	2	30
Effectif cumulé croissant	3	9	17	24	28	30	
Fréquence	10,0 %	20,0 %	26,7 %	23,3 %	13,3 %	6,7 %	100 %
Fréquence cumulée croissante	10,0 %	30,0 %	56,7 %	80,0 %	93,3 %	100,0 %	



Fréquence, effectif cumulé, fréquence cumulée La **fréquence** d'une valeur est $f_i = [n_i/n]$ (souvent écrite en pourcentage). L'**effectif cumulé croissant** d'une valeur est le nombre d'individus dont la valeur est **inférieure ou égale** à celle-ci ; la **fréquence cumulée croissante** est ce nombre divisé par n .

Lecture Dans la colonne « 2 livres », l'effectif cumulé est 17 : 17 élèves sur 30 ont lu **au plus 2 livres**, soit 56,7 % de la classe.

Calculer la moyenne avec les effectifs On multiplie chaque valeur par son effectif : $\bar{x} = [n_1x_1 + n_2x_2 + \dots + n_px_p/n]$. Ici : $\bar{x} = (3 \times 0 + 6 \times 1 + 8 \times 2 + 7 \times 3 + 4 \times 4 + 2 \times 5) \div 30 = 69 \div 30 = \mathbf{2,3 \text{ livre}}$ en moyenne. C'est la moyenne **pondérée** par les effectifs.

3. Médiane et quartiles

Quartiles On range la série dans l'ordre croissant : $x_1 \leq x_2 \leq \dots \leq x_n$.

Le **premier quartile Q_1** est la plus petite valeur de la série telle qu'**au moins 25 %** des valeurs lui sont inférieures ou égales : c'est la valeur de rang **$[n/4]$** (on arrondit $n \div 4$ à l'entier supérieur).

Le **troisième quartile Q_3** est la plus petite valeur telle qu'**au moins 75 %** des valeurs lui sont inférieures ou égales : c'est la valeur de rang **$[3n/4]$** .

La médiane Me joue le même rôle pour 50 %.

Déterminer Q_1 , Me et Q_3 sur une série brute 1) Ranger les valeurs dans l'ordre croissant. 2) Calculer $n \div 4$ et $3n \div 4$ et arrondir à l'entier supérieur : ce sont les rangs de Q_1 et Q_3 . 3) Lire ces valeurs dans la liste rangée. 4) Pour Me : valeur du milieu (n impair) ou moyenne des deux valeurs centrales (n pair).

Exemple Temps d'attente (en minutes) de 12 clients à un guichet, déjà rangés : 3 ; 5 ; 5 ; 6 ; 8 ; 9 ; 11 ; 12 ; 14 ; 15 ; 19 ; 27. $n = 12$; $12 \div 4 = 3$ donc Q_1 est la 3^e valeur : **$Q_1 = 5$** . $3 \times 12 \div 4 = 9$ donc **$Q_3 = 14$** (9^e valeur). $Me = (9 + 11) \div 2 = \mathbf{10}$.

Avec un tableau d'effectifs Pour la série des livres lus ($n = 30$) : $30 \div 4 = 7,5$ donc Q_1 est la 8^e valeur ; les effectifs cumulés valent 3 ; 9 ; 17 ; 24 ; 28 ; 30, et 8 est dépassé pour la première fois à la valeur 1 (cumul 9) : **$Q_1 = 1$** . De même $3 \times 30 \div 4 = 22,5$ donc la 23^e valeur : le cumul passe de 17 à 24 pour la valeur 3, donc **$Q_3 = 3$** . Les valeurs de rang 15 et 16 valent toutes les deux 2 : **$Me = 2$** .

Attention à la convention Certaines calculatrices ou certains tableurs utilisent d'autres méthodes d'interpolation et peuvent donner des quartiles légèrement différents. En 2de, on suit toujours la définition ci-dessus (rangs $[n/4]$ et $[3n/4]$).

Écart interquartile L'**écart interquartile** est $Q_3 - Q_1$. Il contient au moins la moitié des valeurs « du milieu » de la série et mesure leur dispersion. Pour les clients : $14 - 5 = 9$ min.

4. Diagramme en boîte

Diagramme en boîte (ou « boîte à moustaches ») Sur un axe gradué, on dessine un rectangle (la **boîte**) allant de Q_1 à Q_3 , coupé par un trait vertical à la médiane. Les **moustaches** relient la boîte au minimum et au maximum.

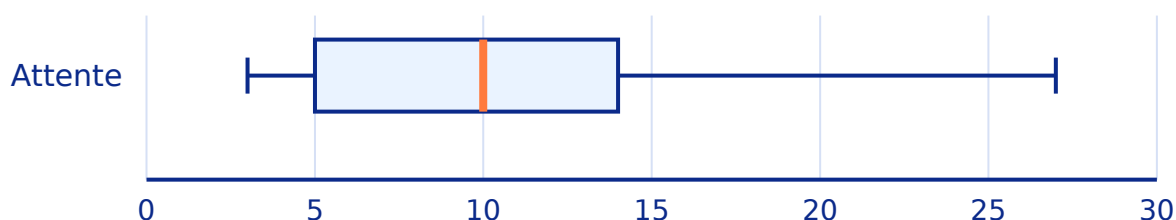


Diagramme du temps d'attente : minimum 3, $Q_1 = 5$, $Me = 10$, $Q_3 = 14$, maximum 27. Le trait orange est la médiane.

Lire une boîte Chaque « morceau » du diagramme (moustache gauche, moitié gauche de la boîte, moitié droite, moustache droite) contient **environ un quart** des valeurs. La boîte seule contient **au moins 50 %** des valeurs ; sa largeur est l'écart interquartile. Une boîte déséquilibrée (médiane non centrée) ou une longue moustache signale une série **asymétrique** : ici, quelques clients attendent très longtemps.

Astuce Mathbot « Bip ! Pour comparer deux séries, dessine leurs boîtes **sur le même axe** : tu verras en une seconde laquelle est la plus haute et laquelle est la plus étalée. »

5. Variance et écart-type

L'étendue et l'écart interquartile ne tiennent compte que de quelques valeurs. Pour mesurer à quel point **toutes** les valeurs s'écartent de la moyenne, on utilise l'écart-type.

Variance et écart-type Pour une série de moyenne $\bar{x}$, la **variance** est la moyenne des carrés des écarts à la moyenne :

$$V = [(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 / n].$$

Avec des effectifs : $V = [n_1(x_1 - \bar{x})^2 + \dots + n_p(x_p - \bar{x})^2 / n].$

L'**écart-type** est $\sigma = \sqrt{V}$. Il s'exprime dans la **même unité** que les valeurs.

Exemple Nombre de leçons apprises par un élève durant 5 semaines : 4 ; 7 ; 7 ; 9 ; 13.
Moyenne : $40 \div 5 = 8$. Écarts à la moyenne : -4 ; -1 ; -1 ; 1 ; 5. Leurs carrés : 16 ; 1 ; 1 ; 1 ; 25, de somme 44. Donc $V = 44 \div 5 = 8,8$ et $\sigma = \sqrt{8,8} \approx 2,97$.

Formule de König La variance est égale à la **moyenne des carrés moins le carré de la moyenne** : $V = [x_1^2 + \dots + x_n^2 / n] - \bar{x}^2$.

Démonstration. On développe $(x_i - \bar{x})^2 = x_i^2 - 2\bar{x}x_i + \bar{x}^2$. En faisant la moyenne des n termes : $V = (\text{moyenne des } x_i^2) - 2\bar{x} \times (\text{moyenne des } x_i) + \bar{x}^2 = (\text{moyenne des } x_i^2) - 2\bar{x}^2 + \bar{x}^2 = (\text{moyenne des } x_i^2) - \bar{x}^2$. CQFD.

Vérification Avec les leçons : moyenne des carrés = $(16 + 49 + 49 + 81 + 169) \div 5 = 72,8$; moins $8^2 = 64$: on retrouve bien **8,8**.

Calculer σ à la calculatrice ou au tableur On saisit les valeurs (et les effectifs) en mode statistiques et on lit σx (ou « σx_n »), pas « s_x » qui divise par $n - 1$. Au tableur : =ECARTYPE.PEARSON(plage) (ou ECARTYPEP selon la version). Pour la série des livres lus, on trouve $\sigma \approx 1,37$.

6. Effet d'un changement de variable

Transformation affine Si on transforme chaque valeur x_i en $y_i = a \cdot x_i + b$ (a et b réels), alors : la moyenne devient $a \cdot \bar{x} + b$; la médiane et les quartiles suivent la même transformation si $a > 0$; l'écart-type devient $|a| \cdot \sigma$; l'écart interquartile et l'étendue sont multipliés par $|a|$. Ajouter b ne change donc **aucune** mesure de dispersion.

Exemple Une série de masses en kg a pour moyenne 62 et pour écart-type 4. En grammes ($y = 1\,000x$) : moyenne 62 000 g, écart-type 4 000 g. Si on mesure l'écart à une masse de référence de 60 kg ($y = x - 60$) : moyenne 2 kg, écart-type toujours 4 kg.

Piège Si $a < 0$, les valeurs sont retournées : les quartiles Q_1 et Q_3 s'échangent. L'écart-type, lui, reste positif grâce à la valeur absolue.

7. Données regroupées en classes

Quand les données sont données par intervalles (« classes »), on ne connaît plus les valeurs
Fiche en ligne : maths-facile.eu/statistiques-descriptives-cours-2de/ · Cours, exercices et contrôles corrigés sur [maths-](https://maths-facile.eu)

exactes. On **estime** la moyenne en remplaçant chaque classe par son **centre** (milieu de l'intervalle) : on obtient une valeur approchée, pas la moyenne exacte. La classe qui contient la médiane s'appelle la **classe médiane** ; on la repère avec les effectifs cumulés. Pour estimer médiane et quartiles, on peut aussi tracer le **polygone des fréquences cumulées croissantes** : on lit l'abscisse correspondant à 25 %, 50 % et 75 %.

Lire une médiane sur un polygone des fréquences cumulées On part de 50 % sur l'axe vertical, on trace l'horizontale jusqu'à la courbe, puis on descend à la verticale sur l'axe horizontal. Le segment entre deux points de la courbe étant supposé droit, on peut aussi calculer cette valeur par proportionnalité (interpolation linéaire).

8. Choisir des indicateurs et comparer deux séries

Robustesse La **moyenne** et l'**écart-type** utilisent toutes les valeurs : ils sont **sensibles aux valeurs extrêmes**. La **médiane** et l'**écart interquartile** n'utilisent que les valeurs centrales : ils sont **peu sensibles** aux valeurs extrêmes.

Comparer deux séries 1) Comparer les **positions** : moyennes ou médianes (laquelle est la plus élevée ?). 2) Comparer les **dispersions** : écarts-types ou écarts interquartiles (laquelle est la plus régulière ?). 3) Choisir les indicateurs adaptés : (médiane ; écart interquartile) si la série a des valeurs extrêmes ou est asymétrique, (moyenne ; écart-type) sinon. 4) Rédiger une **phrase de conclusion** dans le contexte, avec les unités.

Exemple Deux joueuses ont la même moyenne de 12 points par match, mais $\sigma = 2$ pour la première et $\sigma = 7$ pour la seconde. La première est plus **régulière** ; la seconde alterne gros matchs et matchs ratés.

À retenir

- Moyenne $\bar{x} = \text{somme des valeurs}/n$; avec effectifs, $\bar{x} = [(\sum n_i x_i)/n]$. Médiane : partage la série rangée en deux moitiés.
- $Q_1 = \text{valeur de rang } [n/4]$; $Q_3 = \text{valeur de rang } [3n/4]$ (série rangée) ; écart interquartile = $Q_3 - Q_1$; étendue = $\max - \min$.
- Fréquence = **effectif/effectif total** ; les effectifs et fréquences cumulés croissants servent à repérer la médiane et les quartiles.
- Diagramme en boîte : $\min, Q_1, Me, Q_3, \max$; chaque quart de la série occupe un « morceau ».
- Variance $V = \text{moyenne des } (x_i - \bar{x})^2 = \text{moyenne des } x_i^2 - \bar{x}^2$; écart-type $\sigma = \sqrt{V}$, dans la même unité que les valeurs.
- $y = ax + b$: moyenne $a \cdot \bar{x} + b$, écart-type $|a| \cdot \sigma$.
- Moyenne et écart-type sont sensibles aux valeurs extrêmes, médiane et écart interquartile le sont peu ; on compare deux séries par une position et une dispersion,

avec une phrase de conclusion.