

Statistiques

Cours de maths - 3ème

Les **statistiques** servent à résumer et à comparer des données : temps de trajet, notes, tailles, prix... Dans ce chapitre, tu apprends à calculer des **indicateurs de position** (moyenne, médiane, quartiles) et de **dispersion** (étendue, écart interquartile), à regrouper des données en classes, à lire ou construire des graphiques (barres, histogramme, boîte à moustaches) et à comparer deux séries en justifiant ta conclusion.

1. Rappels : vocabulaire, effectifs, fréquences

Vocabulaire La **population** est l'ensemble des individus étudiés ; le **caractère** est ce que l'on observe chez chacun. Il est **qualitatif** s'il ne se mesure pas (un sport, une couleur) et **quantitatif** s'il se mesure ou se compte (une taille, un nombre de frères et sœurs). L'**effectif** d'une valeur est le nombre d'individus qui la présentent ; l'**effectif total** est le nombre d'individus de la population.

Fréquence La fréquence d'une valeur est le quotient (effectif de la valeur) ÷ (effectif total). Elle s'exprime par une fraction, un nombre décimal ou un pourcentage. La somme de toutes les fréquences vaut 1 (soit 100 %).

Exemple On interroge 30 élèves sur le sport qu'ils pratiquent. Le caractère « sport » est qualitatif.

Sport	Foot	Danse	Natation	Judo	Aucun	Total
Effectif	9	6	7	3	5	30
Fréquence (%)	30,0	20,0	23,3	10,0	16,7	100

Pour le foot : $9 \div 30 = 0,3 = 30 \%$. Les fréquences sont arrondies au dixième, leur somme reste 100.

2. La moyenne

Moyenne d'une série La moyenne d'une série de valeurs est le quotient de la **somme des valeurs** par le **nombre de valeurs**.

Exemple Léa a obtenu 12 ; 15 ; 9 ; 14 et 11 en maths. Somme : $12 + 15 + 9 + 14 + 11 = 61$. Moyenne : $61 \div 5 = 12,2$.

Moyenne pondérée (série avec effectifs) Quand une valeur apparaît plusieurs fois, on multiplie chaque valeur par son effectif, on additionne les produits, puis on divise par l'effectif total.

Exemple Nombre de frères et sœurs dans une classe de 25 élèves :

Nombre de frères et sœurs	0	1	2	3	4	Total
Effectif	6	9	6	3	1	25

Somme : $0 \times 6 + 1 \times 9 + 2 \times 6 + 3 \times 3 + 4 \times 1 = 0 + 9 + 12 + 9 + 4 = 34$. Moyenne : $34 \div 25 = 1,36$ frère ou sœur par élève. Cette moyenne n'est pas un nombre entier : ce n'est pas gênant, c'est un résumé de la série.

Piège La moyenne est sensible aux valeurs extrêmes : un seul très grand ou très petit nombre peut la déplacer beaucoup. C'est pourquoi on la complète par la médiane.

3. La médiane

Médiane La médiane d'une série est une valeur qui partage la série **rangée dans l'ordre croissant** en deux groupes de même effectif : au moins la moitié des valeurs lui sont inférieures ou égales, et au moins la moitié lui sont supérieures ou égales.

Calculer une médiane 1) Range les valeurs dans l'ordre croissant. 2) Si l'effectif total N est **impair**, la médiane est la valeur centrale, de rang $(N + 1) \div 2$. 3) Si N est **pair**, la médiane est la moyenne des deux valeurs centrales, de rangs $N \div 2$ et $N \div 2 + 1$.

Deux exemples (temps de trajet en minutes) Série de 7 valeurs : 12 ; 5 ; 20 ; 8 ; 15 ; 30 ; 10. Rangée : 5 ; 8 ; 10 ; 12 ; 15 ; 20 ; 30. $N = 7$ est impair, le rang central est 4 : la médiane est **12 min**.

On ajoute un 8ème trajet de 25 min : 5 ; 8 ; 10 ; 12 ; 15 ; 20 ; 25 ; 30. $N = 8$ est pair, on prend les valeurs de rang 4 et 5 : $(12 + 15) \div 2 = 13,5$ min. Ici la médiane n'est pas une valeur de la série.

Médiane et valeurs extrêmes Si on remplace 30 min par 90 min dans la première série, la moyenne passe de 14,29 min à 22,86 min mais la médiane reste égale à 12 min : la médiane est **peu sensible** aux valeurs extrêmes.

4. Étendue, quartiles, écart interquartile

Étendue L'étendue est la différence entre la plus grande et la plus petite valeur : étendue = maximum – minimum. Elle mesure la dispersion totale de la série.

Quartiles Le **premier quartile** Q_1 est la plus petite valeur de la série rangée telle qu'**au moins 25 %** des valeurs lui soient inférieures ou égales. Le **troisième quartile** Q_3 est la plus petite valeur telle qu'**au moins 75 %** des valeurs lui soient inférieures ou égales. L'**écart interquartile** est $Q_3 - Q_1$: il décrit la dispersion de la moitié centrale des données.

Trouver Q_1 et Q_3 Avec N valeurs rangées dans l'ordre croissant : Q_1 est la valeur de rang $N \div 4$ (arrondi **au rang entier supérieur** si ce n'est pas un entier) ; Q_3 est la valeur de rang $3N \div 4$ (même arrondi).

Exemple ($N = 15$) Temps de trajet de 15 élèves, rangés : 5 ; 8 ; 10 ; 10 ; 12 ; 15 ; 15 ; 18 ; 20 ; 20 ; 25 ; 25 ; 30 ; 35 ; 45.

Médiane : rang 8, donc **18 min**. Q_1 : $15 \div 4 = 3,75$, on prend le rang 4, donc **$Q_1 = 10$ min**. Q_3 : $3 \times 15 \div 4 = 11,25$, on prend le rang 12, donc **$Q_3 = 25$ min**.

Étendue : $45 - 5 = 40$ min. Écart interquartile : $25 - 10 = 15$ min.

Contrôle : 4 valeurs sur 15 ($\approx 26,7\%$) sont ≤ 10 , mais seulement 3 sur 15 (20 %) sont ≤ 8 : c'est bien 10 le premier quartile.

À savoir Au moins 25 % des valeurs sont $\leq Q_1$, au moins 50 % sont $\leq$ la médiane, au moins 75 % sont $\leq Q_3$. Entre Q_1 et Q_3 se trouvent environ la moitié des valeurs. Si on remplace le maximum 45 par 145, l'étendue passe à 140 mais l'écart interquartile reste 15 : il ignore les valeurs extrêmes.

5. Regrouper en classes

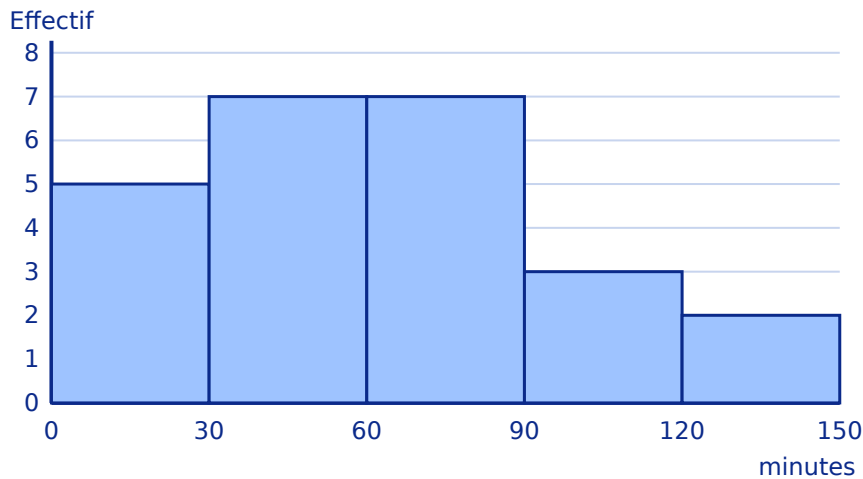
Classe Quand les valeurs sont nombreuses ou très variées, on les regroupe par **classes**, c'est-à-dire des intervalles de même amplitude. La classe $[30 ; 60[$ contient les valeurs x avec $30 \leq x < 60$ (30 inclus, 60 exclu). On donne alors l'effectif de chaque classe.

Exemple : temps passé sur un téléphone 24 élèves ont noté leur temps d'écran quotidien (en minutes) : 10 ; 25 ; 40 ; 15 ; 55 ; 70 ; 95 ; 20 ; 65 ; 80 ; 35 ; 110 ; 45 ; 60 ; 125 ; 30 ; 50 ; 85 ; 75 ; 100 ; 18 ; 140 ; 58 ; 68. En classes de 30 minutes :

Classe (min)	[0 ; 30[	[30 ; 60[	[60 ; 90[	[90 ; 120[	[120 ; 150[	Total
Effectif	5	7	7	3	2	24

Histogramme

Pour des classes de même amplitude, on représente les effectifs par un **histogramme** : chaque classe est un rectangle, collé aux autres, dont la hauteur est proportionnelle à l'effectif.



Les classes $[30 ; 60[$ et $[60 ; 90[$ ont le plus grand effectif (7) : ce sont les classes modales.

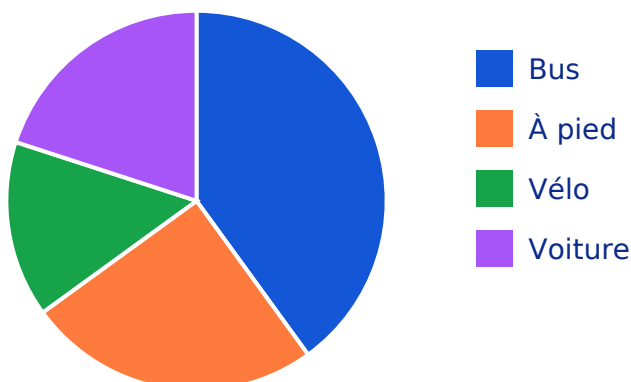
Moyenne approchée avec des classes Quand on ne connaît que les classes, on remplace chaque classe par son **centre** (milieu de l'intervalle), puis on calcule une moyenne pondérée. Le résultat est une **valeur approchée**.

Ici, centres : 15 ; 45 ; 75 ; 105 ; 135. Moyenne approchée = $(15 \times 5 + 45 \times 7 + 75 \times 7 + 105 \times 3 + 135 \times 2) \div 24 = 1500 \div 24 \approx \mathbf{62,5 \text{ min}}$. La vraie moyenne des 24 valeurs brutes est 61,4 min : l'approximation est bonne.

6. Lire et construire des graphiques

Choisir le bon graphique Diagramme en **barres** : comparer des effectifs de modalités. Diagramme **circulaire** : montrer une répartition (parts d'un tout) ; l'angle d'un secteur est proportionnel à l'effectif (360° pour le total). **Courbe** : suivre une évolution dans le temps. **Histogramme** : données regroupées en classes.

Diagramme circulaire 120 élèves se rendent au collège ainsi : bus 48, à pied 30, vélo 18, voiture 24. L'angle d'un secteur vaut $360^\circ \times \text{effectif} \div 120 = 3^\circ \times \text{effectif}$. Bus : $3 \times 48 = 144^\circ$; à pied : 90° ; vélo : 54° ; voiture : 72° . Vérification : $144 + 90 + 54 + 72 = 360$.

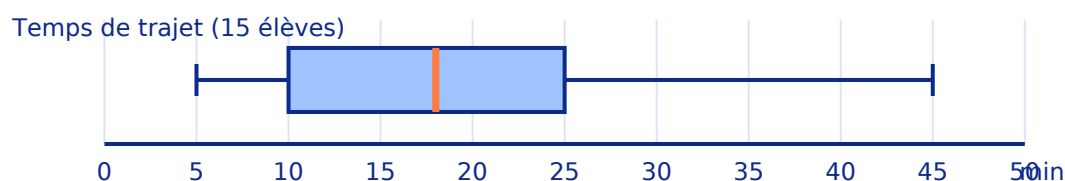


Le secteur « Bus » mesure 144° : il représente 40 % des élèves ($144 \div 360$).

Piège de lecture Regarde toujours l'origine de l'axe vertical. Un axe qui ne commence pas à 0 exagère les différences entre les barres.

7. Le diagramme en boîte

Diagramme en boîte (ou boîte à moustaches) Il résume une série par cinq nombres : minimum, Q_1 , médiane, Q_3 , maximum. La **boîte** va de Q_1 à Q_3 , un trait marque la médiane, et deux « moustaches » vont jusqu'au minimum et au maximum. Les graduations sont régulières.



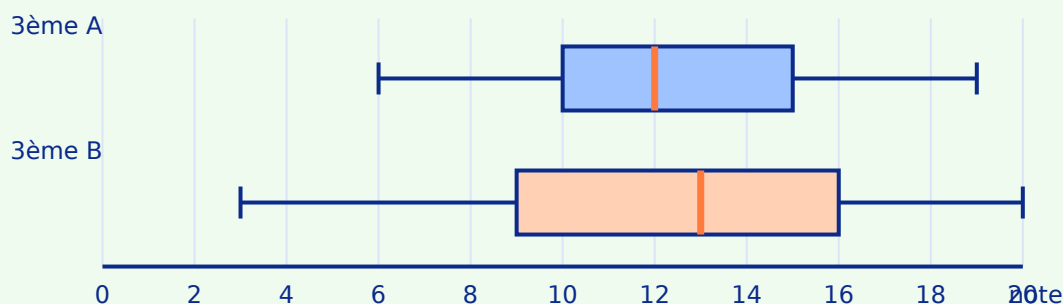
Longueur totale = étendue (40) ; longueur de la boîte = écart interquartile (15). Chaque quart de la série occupe une portion du dessin, même si ces portions sont de longueurs différentes.

8. Comparer deux séries

Méthode 1) Compare un indicateur de **position** : médiane (ou moyenne). 2) Compare un indicateur de **dispersion** : étendue ou écart interquartile. 3) Rédige une phrase avec les nombres et le contexte.

Deux classes de 3ème au même devoir (15 élèves chacune)

	Minimum	Q_1	Médiane	Q_3	Maximum	Moyenne
3ème A	6	10	12	15	19	12,3
3ème B	3	9	13	16	20	12,4



Les médianes sont voisines (12 pour A, 13 pour B). La classe A est plus homogène : son écart interquartile est 5 contre 7, et son étendue est 13 contre 17. La classe B est plus dispersée, avec la note la plus basse mais aussi la plus haute.

Mathbot conseille « Une moyenne seule ne dit pas tout : donne toujours un indicateur de position ET un indicateur de dispersion avant de conclure ! »

9. Le tableur et l'interprétation

Un tableur calcule vite sur de grandes séries. Si les valeurs sont dans les cellules B2 à B16 :

- =SOMME(B2:B16) donne la somme ; =MOYENNE(B2:B16) la moyenne ;
- =MEDIANE(B2:B16) la médiane ; =MIN(B2:B16) et =MAX(B2:B16) les extrêmes ;
- =MAX(B2:B16) - MIN(B2:B16) l'étendue ; =B2/B\$18*100 une fréquence en pourcentage (le total est en B18, le signe \$ bloque la ligne quand on recopie vers le bas).

Attention aux quartiles du tableur Les fonctions de quartile d'un tableur utilisent souvent une méthode d'interpolation : elles peuvent donner un résultat différent de celui de la définition du cours. À l'écrit, on applique la méthode du cours (rang $N \div 4$ arrondi au supérieur).

Interpréter une série Une conclusion statistique se rédige avec une phrase complète et des unités : « La moitié des élèves mettent au plus 18 minutes pour venir au collège ; un quart met plus de 25 minutes. » Rappelle-toi que « au moins 50 % » ne veut pas dire « exactement 50 % ».

À retenir

- **Moyenne** = somme des valeurs $\div$ nombre de valeurs ; avec des effectifs, on pondère par les effectifs.
- **Médiane** : on range, puis valeur centrale (N impair) ou moyenne des deux valeurs centrales (N pair).
- **Étendue** = maximum – minimum ; **écart interquartile** = $Q_3 - Q_1$.
- Q_1 : rang $N \div 4$, Q_3 : rang $3N \div 4$, toujours arrondi au rang supérieur.
- La moyenne et l'étendue sont sensibles aux valeurs extrêmes ; la médiane et l'écart interquartile beaucoup moins.
- Classes : même amplitude, histogramme, moyenne approchée avec les centres des classes.
- Pour comparer deux séries : une mesure de position, une mesure de dispersion, une phrase de conclusion.