

# Statistiques et information chiffrée

## Cours de maths - 1ère

Une série statistique, ce sont des données brutes. Pour **en tirer une information**, il faut les résumer par quelques nombres (position et dispersion), les dessiner (boîtes à moustaches), les comparer à d'autres séries, mesurer des évolutions en pourcentage et, surtout, **garder son esprit critique** : un chiffre peut être exact et pourtant trompeur. Ce chapitre te donne les outils pour lire, comparer et juger des données chiffrées.

## 1. Vocabulaire et tableaux de données

**Population, individu, caractère** Une **population** est l'ensemble des individus étudiés (des élèves, des magasins, des jours...). Un **caractère** est la propriété observée sur chaque individu. Il est **quantitatif** s'il se mesure par un nombre (temps de trajet, note) et **qualitatif** sinon (mode de transport, couleur). **L'effectif** d'une valeur est le nombre d'individus qui la prennent ; sa **fréquence** est l'effectif divisé par l'effectif total, souvent donnée en pourcentage.

**Exemple** On interroge 40 élèves sur le nombre de livres lus pendant un trimestre. Les résultats sont regroupés dans un tableau d'effectifs.

| Livres lus    | 0    | 1    | 2    | 3    | 4    | 5   | 6   | Total |
|---------------|------|------|------|------|------|-----|-----|-------|
| Effectif      | 4    | 8    | 10   | 9    | 5    | 3   | 1   | 40    |
| Fréquence (%) | 10,0 | 20,0 | 25,0 | 22,5 | 12,5 | 7,5 | 2,5 | 100   |

Par exemple, la fréquence de la valeur 3 est  $9/40 = 0,225 = 22,5 \%$ .

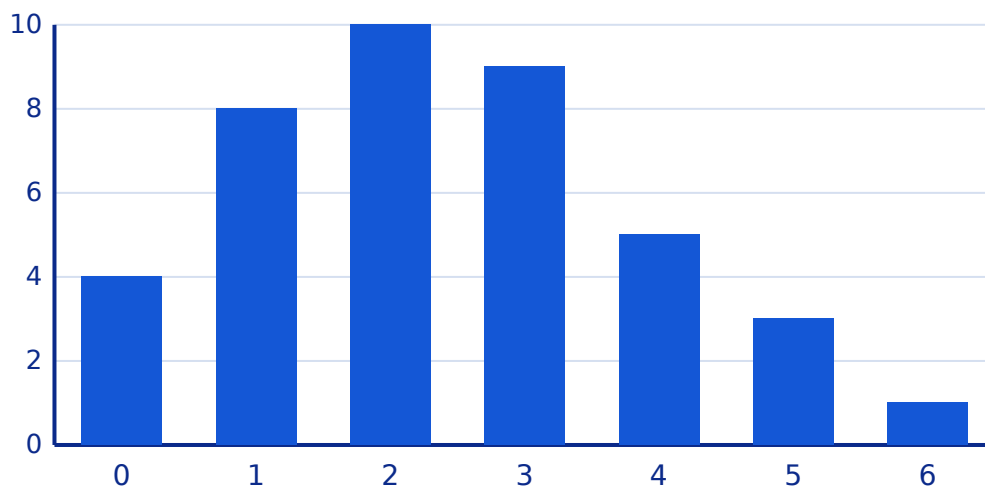


Diagramme en barres : la hauteur de chaque barre est l'effectif de la valeur.

**Piège** Un pourcentage n'a de sens que si l'on connaît la population de référence : « 50 % » de 4 personnes, ce n'est pas la même information que 50 % de 4 000 personnes. Précise toujours : « pourcentage de quoi ? ».

## 2. Indicateurs de position : moyenne, médiane, quartiles

**Moyenne** La **moyenne** d'une série de valeurs  $x_1, x_2, \dots, x_n$  est  $m = \frac{x_1 + x_2 + \dots + x_n}{n}$ . Si la valeur  $x_i$  a l'effectif  $n_i$  (effectif total  $N$ ), alors  $m = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{N}$  : c'est une **moyenne pondérée**.

**Médiane** La **médiane**  $Me$  partage la série *rangée dans l'ordre croissant* en deux groupes de même effectif : au moins la moitié des valeurs sont  $\leq Me$  et au moins la moitié sont  $\geq Me$ . Si l'effectif  $n$  est impair, c'est la valeur centrale ; s'il est pair, c'est la moyenne des deux valeurs centrales.

**Quartiles** Le **premier quartile**  $Q_1$  est la plus petite valeur de la série 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.

### Calculer les quartiles d'une série de $n$ valeurs rangées

1. Range les valeurs dans l'ordre croissant.
2.  $Q_1$  est la valeur de rang  $n \times 0,25$ , arrondi au rang entier supérieur.
3.  $Q_3$  est la valeur de rang  $n \times 0,75$ , arrondi au rang entier supérieur.

**Exemple : les temps de trajet** Temps de trajet (en minutes) de 12 élèves, déjà rangés : 5, 8, 10, 12, 15, 15, 18, 20, 25, 30, 35, 50.

- Moyenne :  $243 \div 12 = 20,25$  min.
- Médiane :  $n = 12$  est pair ; les 6<sup>e</sup> et 7<sup>e</sup> valeurs sont 15 et 18, donc  $Me = \frac{15 + 18}{2} = 16,5$  min.
- $Q_1$  :  $12 \times 0,25 = 3$ , c'est un entier, donc rang 3 :  $Q_1 = 10$  min.
- $Q_3$  :  $12 \times 0,75 = 9$ , donc rang 9 :  $Q_3 = 25$  min.

On lit : au moins 25 % des élèves mettent 10 min ou moins, au moins 75 % mettent 25 min ou moins.

### Étendue, écart interquartile, déciles

- L'**étendue** est la différence entre la plus grande et la plus petite valeur : ici  $50 - 5 = 45$  min.
- L'**écart interquartile** est  $Q_3 - Q_1$  : ici  $25 - 10 = 15$  min. Il mesure la largeur de l'intervalle qui contient la moitié « centrale » des données.

- Les **déciles**  $D_1$  et  $D_9$  se calculent comme les quartiles avec 10 % et 90 % :  $D_1$  est la plus petite valeur telle qu'au moins 10 % des valeurs lui soient inférieures ou égales.

**Moyenne ou médiane ?** La moyenne tient compte de *toutes* les valeurs, y compris les valeurs extrêmes : elle est sensible. La médiane, elle, ne bouge presque pas si une valeur extrême change : elle est **robuste**. Ici la moyenne (20,25 min) est supérieure à la médiane (16,5 min) à cause du trajet de 50 min.

### 3. Indicateurs de dispersion : variance et écart-type

Deux séries peuvent avoir la même moyenne et pourtant être très différentes : l'une regroupée autour de la moyenne, l'autre très étalée. L'étendue et l'écart interquartile mesurent la dispersion à partir de valeurs particulières ; la variance et l'écart-type utilisent **toutes** les valeurs.

**Variance et écart-type** Pour une série de moyenne  $m$ , la **variance** est la moyenne des carrés des écarts à la moyenne :  $V = (x_1 - m)^2 + (x_2 - m)^2 + \dots + (x_n - m)^2 / n = n_1(x_1 - m)^2 + \dots + n_p(x_p - m)^2 / N$  (avec effectifs). L'**écart-type** est  $\sigma = \sqrt{V}$ . Il s'exprime dans la même unité que les valeurs.

**Exemple détaillé** Série : 3, 5, 6, 10. Moyenne :  $24/4 = 6$ . Écarts à la moyenne : -3, -1, 0, 4. Carrés : 9, 1, 0, 16 (somme 26). Variance :  $V = 26/4 = 6,5$ . Écart-type :  $\sigma = \sqrt{6,5} \approx 2,55$ .

**Avec des effectifs : les livres lus** La moyenne vaut  $m = 96/40 = 2,4$ . La variance vaut  $V = 89,60/40 \approx 2,240$ , donc  $\sigma \approx 1,50$  livre. À la calculatrice, on saisit les valeurs et les effectifs en mode statistiques et on lit directement la moyenne et l'écart-type ( $\sigma$  ou «  $\sigma x$  »).

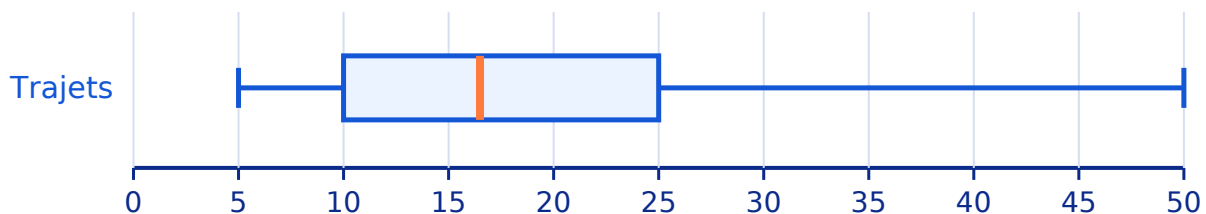
**Effet d'un changement affine** Si l'on remplace chaque valeur  $x$  par  $ax + b$  ( $a$  et  $b$  sont des constantes), alors la **moyenne** devient  $a \times m + b$  et l'**écart-type** devient  $|a| \times \sigma$ . Ajouter  $b$  décale la série mais ne change pas sa dispersion ; multiplier par  $a$  étire la dispersion dans la même proportion.

**Exemple** Des températures ont pour moyenne 10 °C et pour écart-type 3 °C. En degrés Fahrenheit ( $F = 1,8 C + 32$ ), la moyenne est  $1,8 \times 10 + 32 = 50$  °F et l'écart-type est  $1,8 \times 3 = 5,4$  °F.

**Interprétation** Un écart-type n'est « grand » ou « petit » que par rapport à la moyenne et au contexte. Une dispersion de 5 minutes est énorme pour un temps de trajet de 8 minutes et négligeable pour un voyage de 6 heures.

## 4. Les boîtes à moustaches

**Boîte à moustaches (diagramme en boîte)** On la construit à partir de **cinq nombres** : le minimum,  $Q_1$ , la médiane,  $Q_3$  et le maximum. La **boîte** va de  $Q_1$  à  $Q_3$  avec un trait à la médiane ; les **moustaches** relient la boîte au minimum et au maximum. Tout est tracé sur *une graduation régulière*.



Boîte à moustaches des temps de trajet : 5 ; 10 ; 16,5 ; 25 ; 50 (en minutes).

### Ce que raconte une boîte

- Environ **un quart** des valeurs est dans chaque « morceau » : moustache gauche, moitié gauche de la boîte, moitié droite de la boîte, moustache droite.
- La **largeur de la boîte** est l'écart interquartile : elle contient au moins la moitié centrale des données.
- Si la médiane est décentrée dans la boîte, ou une moustache beaucoup plus longue que l'autre, la série est **asymétrique** (ici, étalée vers les grandes valeurs).

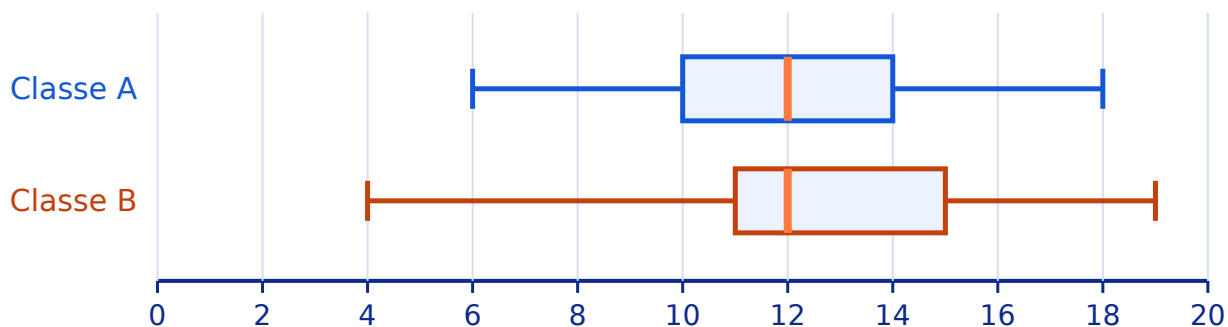
### Tracer une boîte à moustaches

1. Range la série et calcule min,  $Q_1$ , Me,  $Q_3$ , max.
2. Choisis une graduation qui contient toutes les valeurs.
3. Trace le rectangle de  $Q_1$  à  $Q_3$ , le trait de la médiane, puis les moustaches jusqu'au min et au max.

## 5. Comparer le même caractère sur plusieurs populations

Pour comparer, par exemple, les notes de deux classes à un même devoir, on dessine les deux boîtes sur **la même graduation**, l'une au-dessus de l'autre.

| Série    | Minimum | $Q_1$ | Médiane | $Q_3$ | Maximum |
|----------|---------|-------|---------|-------|---------|
| Classe A | 6       | 10    | 12      | 14    | 18      |
| Classe B | 4       | 11    | 12      | 15    | 19      |



### Rédiger une comparaison

1. **Position** : compare les médianes (et/ou les moyennes). « La classe A a une médiane de 12, la classe B de 12. »
2. **Dispersion** : compare les écarts interquartiles (et/ou les écarts-types) et les étendues. « Écart interquartile : 4 pour A, 4 pour B. »
3. **Conclusion en contexte** : phrase qui répond à la question posée (classe plus homogène, plus forte, plus étalée...).

**Application** Ici, les médianes sont identiques (12 et 12), mais l'écart interquartile est 4 pour A contre 4 pour B et l'étendue est 12 contre 15 : les deux classes ont un niveau comparable, mais la classe B est **plus dispersée**, donc moins homogène.

**Attention aux effectifs** Comparer des populations d'effectifs très différents oblige à raisonner en **fréquences** (pourcentages) et non en effectifs bruts.

**Moyenne de plusieurs sous-populations** Si une population est réunie à partir de groupes d'effectifs  $n_1, n_2, \dots$  et de moyennes  $m_1, m_2, \dots$ , la moyenne globale est la **moyenne pondérée**  $m = \frac{n_1 m_1 + n_2 m_2 + \dots}{n_1 + n_2 + \dots}$ . On ne fait jamais la moyenne simple des moyennes sauf si les effectifs sont égaux.

## 6. Comparer deux caractères d'une même série

Quand chaque individu est décrit par **deux caractères**, on range les données dans un **tableau à double entrée** (tableau croisé). Exemple : 120 abonnés d'une médiathèque, selon leur âge et leur emprunt préféré.

|         | Livres | Films | Total |
|---------|--------|-------|-------|
| Adultes | 42     | 28    | 70    |
| Jeunes  | 20     | 30    | 50    |
| Total   | 62     | 58    | 120   |

### Fréquences marginales et conditionnelles Les fréquences marginales se

calculent par rapport à l'effectif total, à partir des totaux en marge du tableau. Une **fréquence conditionnelle** se calcule par rapport à un sous-groupe : on divise par l'effectif de la ligne (ou de la colonne) choisie.

### Exemple

- Fréquence marginale des adultes : **70/120**  $\approx$  58,3 %.
- Parmi les adultes, la fréquence de ceux qui préfèrent les livres : **42/70** = 60,0 %.
- Parmi les jeunes : **20/50** = 40,0 %.

Les adultes préfèrent donc nettement plus souvent les livres que les jeunes : les deux caractères semblent **liés**.

**Quel dénominateur ?** « Parmi les adultes, 60 % préfèrent les livres » n'a rien à voir avec « parmi ceux qui préfèrent les livres, 67,7 % sont des adultes ». Le mot « parmi » désigne le groupe qui sert de dénominateur.

**Un caractère quantitatif selon un caractère qualitatif** On peut aussi comparer une grandeur (temps de trajet, note...) selon une catégorie (mode de transport, classe) : on calcule alors la moyenne, la médiane et l'écart interquartile **par catégorie**, puis on compare comme dans la partie 5.

## 7. Évolutions et pourcentages

**Taux d'évolution et coefficient multiplicateur** Si une valeur passe d'une valeur de départ  $V_{\text{départ}}$  à une valeur d'arrivée  $V_{\text{arrivée}}$ , son **taux d'évolution** est  $t = \frac{V_{\text{arrivée}} - V_{\text{départ}}}{V_{\text{départ}}}$ . Le **coefficient multiplicateur** est  $CM = \frac{V_{\text{arrivée}}}{V_{\text{départ}}} = 1 + t$ . Ainsi  $V_{\text{arrivée}} = V_{\text{départ}} \times CM$ .

### Exemples

- Une hausse de 8 % :  $CM = 1 + 0,08 = 1,08$ . Une baisse de 15 % :  $CM = 1 - 0,15 = 0,85$ .
- Un prix passe de 40 € à 46 € :  $t = \frac{6}{40} = 0,15$  : hausse de 15 %.

**Évolutions successives** Les coefficients multiplicateurs **se multiplient** :  $CM_{\text{global}} = CM_1 \times CM_2$ . **On n'additionne jamais** les taux.  
Exemple : +10 % puis +20 % donne  $1,10 \times 1,20 = 1,32$ , soit +32 % (et non +30 %).

**Évolution réciproque** L'évolution qui annule une évolution de coefficient  $CM$  a pour coefficient **1/CM**.  
Exemple : après +25 % ( $CM = 1,25$ ), il faut  $CM' = \frac{1}{1,25} = 0,8$  : une baisse de **20 %** (et non 25 %).

**Indice base 100** Choisir une valeur de référence et lui donner l'indice 100 permet de suivre une évolution. Si une valeur est passée de 250 à 275, son indice est  $100 \times \frac{275}{250} = 110$  : elle a augmenté de 10 %.

**Points de pourcentage** Lorsqu'un taux (par exemple un taux de chômage) passe de 8 % à 6 %, on parle d'une baisse de **2 points de pourcentage**. En valeur relative, la baisse est de  $\frac{2}{8} = 25$  %. Ce sont deux informations différentes.

**Taux d'évolution moyen** Si une grandeur évolue d'un coefficient global CM sur n périodes identiques, le coefficient par période est  $CM^{1/n}$ , c'est-à-dire la racine n-ième de CM. Exemple : une population multipliée par 1,44 en 2 ans a un coefficient annuel  $\sqrt{1,44} = 1,2$ , soit +20 % par an en moyenne.

## 8. Lire des données avec esprit critique

Un chiffre est une information, pas une preuve. Voici les questions à se poser devant une statistique.

- **Qui a été interrogé ?** Un échantillon doit représenter la population visée. Sonder 30 personnes à la sortie d'une salle de sport ne dit rien sur « les Français ».
- **Combien de personnes ?** Un pourcentage sans effectif est fragile : 100 % de réussite sur 3 essais n'est pas très informatif.
- **Moyenne ou médiane ?** Les revenus, les prix de l'immobilier ou les temps d'attente comportent des valeurs extrêmes : la moyenne peut être très éloignée de la situation « typique ».
- **Le graphique est-il honnête ?** Un axe vertical qui ne commence pas à zéro, des graduations irrégulières ou des échelles différentes sur deux graphiques exagèrent ou masquent un écart.
- **Corrélation n'est pas causalité.** Deux phénomènes qui varient ensemble ne sont pas forcément liés par une cause : une troisième cause peut agir sur les deux.
- **Les sous-groupes racontent parfois l'inverse du total** (paradoxe de Simpson).

**Le paradoxe de Simpson** Deux lycées présentent leur taux d'admission selon deux filières. Lycée Alpha : filière générale 19 admis sur 20 (95 %), filière technologique 30 sur 80 (37,5 %). Lycée Bêta : générale 72 sur 80 (90 %), technologique 5 sur 20 (25 %). Alpha fait mieux dans *chaque* filière, mais globalement Alpha admet 49 candidats sur 100 (49 %) contre 77 sur 100 (77 %) pour Bêta ! La cause : les deux lycées n'ont pas la même répartition entre les filières. Moralité : pour comparer, regarde aussi les sous-groupes.

**Mathbot te conseille** Devant un chiffre spectaculaire, pose-toi toujours trois questions : de quoi parle-t-on exactement, sur combien de personnes, et par rapport à quoi ?

## À retenir

---

- **Position** : moyenne, médiane, quartiles  $Q_1$  et  $Q_3$  (au moins 25 % et 75 % des valeurs sont  $\leq$ ). **Dispersion** : étendue, écart interquartile  $Q_3 - Q_1$ , variance et écart-type  $\sigma$ .
- La médiane et l'écart interquartile sont robustes ; la moyenne et l'écart-type sont sensibles aux valeurs extrêmes.
- Avec  $ax + b$  : moyenne  $\rightarrow a \cdot m + b$  ; écart-type  $\rightarrow |a| \cdot \sigma$ .
- Boîte à moustaches : min,  $Q_1$ , Me,  $Q_3$ , max sur une même graduation. Pour comparer : position, dispersion, conclusion en contexte.
- Tableau croisé : fréquences marginales (total) et conditionnelles (sous-groupe) ; toujours préciser « parmi qui ? ».
- Évolution :  $t = (V_{\text{arrivée}} - V_{\text{départ}}) \div V_{\text{départ}}$  ;  $CM = 1 + t$  ; évolutions successives : on multiplie les CM ; réciproque :  $1 \div CM$ .
- Esprit critique : échantillon, effectifs, moyenne ou médiane, axes du graphique, causalité, sous-groupes.