

Les statistiques descriptives

cours et exercices corrigés de maths 2nde

Deux classes ont la même moyenne de 12. Dans l'une, tout le monde est entre 11 et 13 ; dans l'autre, la moitié est à 5 et l'autre à 19. Un seul nombre ne résume donc jamais une série : il en faut DEUX — un qui dit où elle se tient, un qui dit de combien elle s'étale.

MOTS CLÉS

Effectif, fréquence, moyenne, médiane, quartiles, dispersion

LE SECRET

Un indicateur de position ne vaut rien sans son indicateur de dispersion

OUTIL

Le diagramme en boîte : cinq nombres, un seul coup d'œil

À quoi ça sert : les statistiques descriptives

En France, le salaire MOYEN à temps plein tourne autour de 2 600 € nets par mois, mais le salaire MÉDIAN autour de 2 100 €. L'écart n'est pas une erreur : les très hauts salaires tirent la moyenne vers le haut sans déplacer la médiane. Dire « le salaire médian est de 2 100 € » signifie qu'une personne sur deux gagne moins que ça — c'est une information que la moyenne ne donne pas. ★ C'est pour cette raison que l'INSEE publie les deux, et que les journalistes qui ne citent que la moyenne racontent une autre histoire que celle des données.

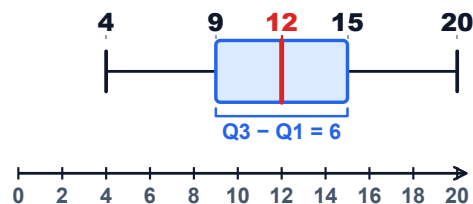
Un peu d'histoire : les statistiques descriptives

Le mot « statistique » vient du latin status, l'État : pendant des siècles, compter, c'était compter les sujets et les impôts. Au XIX^e siècle, le Belge Adolphe Quetelet invente « l'homme moyen » et fait de la moyenne le résumé roi — au point qu'on a longtemps cru qu'elle suffisait. Le diagramme en boîte est bien plus récent : l'Américain John Tukey le publie en 1977, précisément pour qu'on VOIE la dispersion au lieu de la calculer, et pour qu'on puisse comparer plusieurs séries d'un seul regard.

Définition : les statistiques descriptives

Une **SÉRIE STATISTIQUE**, ce sont des valeurs relevées sur un ensemble d'individus, avec leurs **EFFECTIFS** — combien de fois chaque valeur apparaît. La **FRÉQUENCE** d'une valeur est son effectif divisé par l'effectif total : c'est une proportion, donc un nombre entre 0 et 1, souvent donné en pourcentage. Résumer la série, c'est en tirer deux sortes de nombres : des indicateurs de **POSITION** (moyenne, médiane), qui disent autour de quoi elle se tient, et des indicateurs de **DISPERSION** (étendue, écart interquartile, écart type), qui disent de combien elle s'étale.

Les notes d'une classe



★ Le diagramme en boîte porte **CINQ** nombres : les deux bouts des moustaches sont le minimum et le maximum ; les deux bords de la boîte sont Q_1 et Q_3 ; le trait rouge à l'intérieur est la médiane. La boîte contient donc la **MOITIÉ CENTRALE** de la classe — ici, la moitié des élèves ont entre 9 et 15.

Propriétés : les statistiques descriptives

La moyenne : on partage tout également

On additionne toutes les valeurs et on divise par l'effectif. Quand les valeurs se répètent, on pondère par les effectifs. ➖ La moyenne est **SENSIBLE** aux valeurs extrêmes : une seule valeur très grande la tire vers le haut, même si elle ne concerne qu'un individu.

$$\bar{x} = \frac{\text{somme des valeurs}}{\text{effectif total}}$$
$$\bar{x} = \frac{3 \times 8 + 5 \times 12}{8} = 10,5$$

➖ On divise par l'effectif **TOTAL** (8 ici), pas par le nombre de valeurs différentes (2).

La médiane : elle coupe l'effectif en deux

La médiane partage la série **ORDONNÉE** en deux groupes de même effectif : au moins la moitié des valeurs lui sont inférieures ou égales, au moins la moitié lui sont supérieures ou égales. ★ Elle **RÉSISTE** aux valeurs extrêmes — changer la plus grande valeur ne la déplace pas.

7 9 12 15 19
7 9 12 15 1000

★ Le maximum passe de 19 à 1000 : la médiane ne bouge pas. La moyenne, elle, passe de 12,4 à 205,6.

Les quartiles : on coupe en quatre

Q_1 est la plus petite valeur telle qu'AU MOINS un quart de la série lui soit inférieur ou égal ; Q_3 , celle telle qu'au moins trois quarts le soient. Sur une série ordonnée de N valeurs, on lit Q_1 au rang $N \div 4$ et Q_3 au rang $3N \div 4$, chaque rang arrondi à l'entier SUPÉRIEUR.

Pour $N = 14$:

$$\frac{14}{4} = 3,5 \rightarrow Q_1 \text{ au rang } 4$$

$$\frac{42}{4} = 10,5 \rightarrow Q_3 \text{ au rang } 11$$

⊖ On arrondit toujours AU-DESSUS, jamais au plus proche : il faut AU MOINS un quart des valeurs en dessous.

★ La dispersion : deux couples, jamais mélangés

L'ÉTENDUE ($\max - \min$) se calcule vite mais dépend entièrement des deux valeurs extrêmes. L'ÉCART INTERQUARTILE ($Q_3 - Q_1$) mesure l'étalement de la moitié centrale et ignore les extrêmes. L'ÉCART TYPE mesure l'écart moyen à la moyenne. ★ Chaque indicateur de position va avec le sien.

Moyenne



Écart type

série sans valeur
extrême

Médiane



**Écart
interquartile**

série avec valeurs
extrêmes

Méthode : les statistiques descriptives

1. 1. Je range la série

➖ Avant toute médiane et tout quartile, j'ordonne les valeurs dans l'ordre croissant. C'est l'erreur numéro un : lire la valeur du milieu de la liste telle qu'elle arrive.

12 7 19 9 15

7 9 12 15 19

Sans ranger, on aurait lu 19 : la 3^e valeur de la liste d'origine, qui ne veut rien dire.

2. 2. Je choisis mon couple

Y a-t-il des valeurs très à l'écart des autres ? Si oui, je prends médiane et quartiles, qui y résistent. Sinon, moyenne et écart type disent tout, et se calculent plus vite.

valeurs extrêm

MÉDIANE

série régulière

MOYENNE

La forme de la série décide, pas l'habitude.

3. 3. Je vérifie l'ordre de grandeur

Une moyenne et une médiane sont toujours comprises entre le minimum et le maximum. Un résultat qui sort de la série est faux, sans avoir besoin de refaire le calcul.

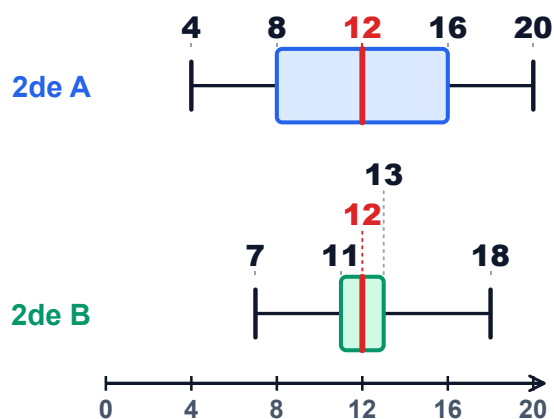
$$\min \leq \bar{x} \text{ et } Me \leq \max$$

Le contrôle de vraisemblance qui coûte trois secondes.

Selon ce que l'on cherche : les statistiques descriptives

★ Comparer deux séries

C'est l'usage numéro un du diagramme en boîte : on empile les deux boîtes sur le même axe et la comparaison se lit sans calcul. Une boîte plus à droite, c'est une médiane plus élevée. Une boîte plus étroite, c'est une série plus régulière.



Repérer une valeur qui détonne

Quand la moyenne et la médiane s'écartent beaucoup, c'est le signe qu'une poignée de valeurs extrêmes tire la série. La médiane dit alors mieux « le cas courant » que la moyenne.

$$\bar{x} = 205,6 \quad \text{Me} = 12$$

Un tel écart ne se discute pas : il faut regarder la série avant de conclure.

Changer d'unité sans tout recalculer

Si l'on ajoute la même valeur à toutes les données, la moyenne augmente d'autant. Si on les multiplie toutes par un même nombre, la moyenne est multipliée par ce nombre. C'est la LINÉARITÉ de la moyenne, et elle évite de refaire la somme.

$$y_i = ax_i + b \longrightarrow \bar{y} = a\bar{x} + b$$

$$\bar{x} = 12 \longrightarrow \overline{2x + 3} = 27$$

⚠ L'écart type, lui, est multiplié par $|a|$ — mais le $+3$ ne le change pas.

Exemples corrigés : les statistiques descriptives

Une moyenne pondérée

Dans une classe de 25 élèves : 5 ont eu 8, 12 ont eu 12, 8 ont eu 16.

Calculer la moyenne de la classe.

$$\frac{5 \times 8 + 12 \times 12 + 8 \times 16}{25} = \frac{40 + 144 + 128}{25} = \frac{312}{25} = 12,48$$

On multiplie chaque note par son effectif, on additionne, et on divise par l'effectif TOTAL : 25 et non 3. La moyenne vaut 12,48. ➖ Diviser par 3 — le nombre de notes différentes — donnerait 104, un résultat hors de la série : la vérification de l'étape 3 l'aurait attrapé.

Une médiane sur une série en désordre

Les temps (en min) de 7 coureurs : 32, 28, 41, 25, 36, 30, 45.

Déterminer la médiane.

25 28 30 **32** 36 41 45

On range d'abord. Il y a 7 valeurs, donc la médiane est la 4^e : 32 min. La moitié des coureurs ont mis moins de 32 min. ➖ Sans ranger, la 4^e valeur de la liste d'origine était 25 — le minimum, pris pour la médiane.

Les quartiles d'une série de 10 valeurs

Série ordonnée : 3, 5, 6, 8, 9, 11, 12, 14, 17, 20.

Déterminer Q_1 , la médiane et Q_3 .

$$Q_1 : \frac{10}{4} = 2,5 \rightarrow \text{rang } 3 \rightarrow 6$$

$$\text{Me} : \frac{9 + 11}{2} = 10$$

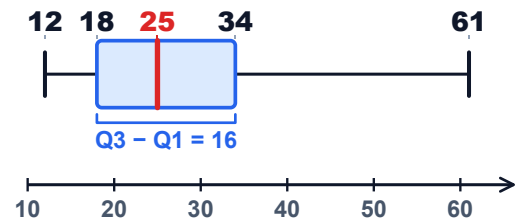
$$Q_3 : \frac{30}{4} = 7,5 \rightarrow \text{rang } 8 \rightarrow 14$$

$Q_1 = 6$, médiane = 10, $Q_3 = 14$. ⚠ L'effectif est PAIR : la médiane est la moyenne des 5^e et 6^e valeurs, et elle ne fait pas partie de la série. Les quartiles, eux, sont toujours des valeurs de la série.

Lire un diagramme en boîte

Le diagramme résume les âges des adhérents d'un club.

Quel est l'écart interquartile, et que dit-il ?



$Q_3 - Q_1 = 34 - 18 = 16$ ans : c'est la largeur de la boîte, donc l'étalement de la MOITIÉ CENTRALE des adhérents. L'étendue, elle, vaut $61 - 12 = 49$ ans — trois fois plus, parce qu'elle dépend entièrement de l'adhérent le plus jeune et du plus âgé. ★ L'écart interquartile décrit le club ; l'étendue décrit ses deux cas isolés.

Pièges à éviter : les statistiques descriptives

- Lire la médiane sans RANGER la série : c'est l'erreur numéro un, et elle donne une valeur qui n'a aucun sens.
- Diviser par le nombre de valeurs DIFFÉRENTES au lieu de l'effectif total dans une moyenne pondérée.
- Arrondir le rang d'un quartile au plus proche : on arrondit toujours à l'entier SUPÉRIEUR.
- Confondre Q_1 (bord de la boîte) et le minimum (bout de la moustache) sur un diagramme en boîte.
- Croire qu'une même moyenne signifie deux séries semblables : sans la dispersion, la moyenne ne dit presque rien.
- ⚠ La moyenne n'est pas toujours une valeur de la série (12,48 n'est la note de personne) ; les quartiles, eux, en sont toujours.

À retenir : les statistiques descriptives

★ Un indicateur de POSITION ne va jamais sans un indicateur de DISPERSION.

Moyenne ↔ écart type. Médiane ↔ quartiles. On ne mélange pas les deux couples.

La médiane résiste aux valeurs extrêmes, la moyenne non.

Rang de Q_1 : $N \div 4$ arrondi au-dessus. Rang de Q_3 : $3N \div 4$ arrondi au-dessus.

Écart interquartile = $Q_3 - Q_1$: c'est la largeur de la boîte.

Le diagramme en boîte porte cinq nombres : min, Q_1 , médiane, Q_3 , max.

Exercices corrigés : les statistiques descriptives

1. Série : 4, 7, 7, 9, 13. Calculer la moyenne et l'étendue.

Correction :

$$\text{Moyenne} = \frac{4 + 7 + 7 + 9 + 13}{5} = \frac{40}{5} = 8. \text{ Étendue} = 13 - 4 = 9.$$

2. Dans une classe, 8 élèves ont 10, 10 élèves ont 14 et 2 élèves ont 18. Quelle est la moyenne ?

Correction :

$$\frac{8 \times 10 + 10 \times 14 + 2 \times 18}{20} = \frac{80 + 140 + 36}{20} = \frac{256}{20} = 12,8. \text{ } \ominus \text{ On divise par 20, l'effectif total.}$$

3. Quelle est la médiane de la série : 15, 9, 22, 11, 17 ?

Correction :

On range : 9, 11, 15, 17, 22. Il y a 5 valeurs, la médiane est la 3^e : 15.

4. Série ordonnée de 6 valeurs : 2, 5, 8, 12, 14, 19. Quelle est la médiane ?

Correction :

L'effectif est pair : c'est la moyenne des 3^e et 4^e valeurs, $\frac{8 + 12}{2} = 10$. La médiane n'est pas une valeur de la série.

5. Série ordonnée de 11 valeurs. À quels rangs lit-on Q_1 et Q_3 ?

Correction :

$\frac{11}{4} = 2,75 \rightarrow$ rang 3 pour Q_1 . $\frac{33}{4} = 8,25 \rightarrow$ rang 9 pour Q_3 . On arrondit au-dessus dans les deux cas.

6. Une série a $Q_1 = 14$ et $Q_3 = 23$. Quel est l'écart interquartile ?

Correction :

$23 - 14 = 9$. La moitié centrale de la série s'étale sur 9 unités.

7. Le tableau donne le nombre de frères et sœurs de 21 élèves : 0 pour 4 élèves, 1 pour 7, 2 pour 6, 3 pour 4. Quelle est la médiane ?

Correction :

Effectif total 21, impair : la médiane est la 11^e valeur. Effectifs cumulés : 4, 11, 17, 21. Le premier qui atteint 11 correspond à la valeur 1. La médiane vaut 1.

8. Sur 200 personnes interrogées, 46 font du vélo. Quelle est la fréquence, en pourcentage ?

Correction :

$\frac{46}{200} = 0,23$, soit 23 %.

9. Sur un diagramme en boîte, la boîte va de 18 à 34 et les moustaches de 12 à 61. Quelle est l'étendue ?

Correction :

$61 - 12 = 49$.  Ce n'est pas $34 - 18 = 16$, qui est l'écart interquartile : l'étendue se lit d'un bout de moustache à l'autre.

10. Une série a pour moyenne 12. On ajoute 3 à toutes les valeurs, puis on double le tout. Quelle est la nouvelle moyenne ?

Correction :

$(12 + 3) \times 2 = 30$. La moyenne suit la même transformation que les données : c'est la linéarité de la moyenne.