BURHAN

Statistique

Tronc Commun · Sciences · Mathématiques

Version en ligne : burhan.ma/tc/sciences/mathematiques/statistique/cours

Chapitre	Semestre
Statistique	Semestre 2

Deux villes, même moyenne ?

Voici les températures maximales relevées pendant une même semaine dans deux villes : une ville côtière, notée A, et une ville de l'intérieur, notée B. Les valeurs sont inventées pour l'exemple.

Jour	lun.	mar.	mer.	jeu.	ven.	sam.	dim.
Ville A (en °C)	22	24	25	23	25	24	25
Ville B (en °C)	18	24	30	21	26	22	27

Dans chaque ville, la somme des sept températures vaut 168. Les deux moyennes sont donc égales : $\frac{168}{7} = 24$ °C.

Devine : Une prédiction

Les deux villes ont la même moyenne. Laquelle a eu le climat le plus régulier cette semaine ?

A. Aucune : avec la même moyenne, les deux climats se valent.

Faux.

La moyenne ne dit rien des écarts d'un jour à l'autre.

B. La ville B, car elle a eu le jour le plus chaud, 30 °C.

Faux.

Un jour très chaud ne dit rien de la régularité : la ville B passe de 18 °C à 30 °C.

C. La ville A.

Juste.

Juste : ses températures restent entre 22 °C et 25 °C.

Correction

Les deux moyennes valent 24 °C.

Pourtant, la ville A reste entre 22 °C et 25 °C, alors que la ville B passe de 18 °C à 30 °C. Le climat de la ville A a donc été le plus régulier.

À la fin du chapitre, tu mesureras cette régularité par un seul nombre.

Dans ce chapitre, tu ranges des données dans des tableaux, tu calcules des cumuls et tu traces un histogramme.

Tu calcules ensuite des **paramètres**. La moyenne, la médiane et le mode disent où se placent les valeurs. L'étendue, la variance et l'écart type disent si elles sont groupées ou dispersées.

Tu retrouves au passage ce que tu as vu au collège : effectifs, fréquences, moyenne et médiane.

Population, caractère, effectifs et fréquences

Population, individus, caractère

En clair. Pour un travail de SVT, une classe de 40 élèves a relevé le groupe sanguin de chacun. Les 40 élèves forment l'ensemble étudié. Pour chacun, on note une seule information : son groupe sanguin.

Avant de compter, il faut savoir **sur qui** porte l'étude et **ce que** l'on observe.

Devine : Un groupe sanguin moyen ?

Pour chaque élève, on note son groupe sanguin : O, A, B ou AB. Peut-on calculer « le groupe sanguin moyen » de la classe ?

A. Oui : on fait la moyenne des effectifs des quatre groupes.

Faux.
Ce calcul donne un nombre d'élèves, pas un groupe sanguin.

B. Oui : on fait la moyenne des quatre groupes.

Faux.
On ne peut pas additionner O, A, B et AB.

C. Non : la classe compte trop d'élèves.

Faux.
Le nombre d'élèves n'y change rien : une moyenne de 40 nombres se calcule sans peine.

D. Non : les groupes sanguins ne sont pas des nombres.

Juste.
Juste : une moyenne se calcule sur des nombres.

Correction
Les modalités O, A, B et AB ne sont pas des nombres : on ne peut ni les additionner, ni en faire la moyenne. On dira que ce caractère est **qualitatif**.

On peut en revanche compter les élèves de chaque groupe.

Définition : Population, individu, caractère

Une **population** est l'ensemble des personnes ou des objets sur lesquels porte une étude statistique ; chacun d'eux est un **individu**. Le **caractère** (ou variable statistique) est la propriété étudiée sur chaque individu.

- Le caractère est **qualitatif** lorsque ses modalités ne sont pas des nombres (groupe sanguin, couleur, moyen de transport).
- Il est **quantitatif** lorsque ses valeurs sont des nombres. Il est **discret** s'il ne prend que des valeurs isolées (nombre de frères et sœurs, pointure) ; il est **continu** s'il peut prendre toutes les valeurs d'un intervalle (taille, masse, durée).

La **série statistique** est la liste des valeurs, ou des modalités, du caractère relevées sur les individus de la population.

Exemple : quatre caractères

Dans ce chapitre, tu rencontres plusieurs études.

- Le groupe sanguin des élèves d'une classe : chaque élève est un individu. Le caractère est **qualitatif** : ses modalités sont O, A, B et AB.
- Le nombre d'enfants des familles d'un quartier : chaque famille est un individu. Le caractère est **quantitatif discret**.
- La taille des élèves, en centimètres : le caractère est **quantitatif continu**.
- La température maximale d'une journée : elle aussi est quantitative continue.

Effectifs et fréquences

Vérifie : Quelle nature ?

Un élève affirme : « Le groupe sanguin est un caractère quantitatif, puisqu'on compte les élèves de chaque groupe. » Quelle est la nature de ce caractère ?

A. Quantitatif discret.
Faux.
On compte bien des élèves, mais les modalités O, A, B et AB ne sont pas des nombres.

B. Quantitatif continu.
Faux.
Un caractère continu peut prendre toutes les valeurs d'un intervalle : ce n'est pas le cas ici.

C. Ce n'est pas un caractère.
Faux.
C'est bien la propriété étudiée sur chaque élève.

D. Qualitatif.
Juste.
Juste : ses modalités ne sont pas des nombres.

Correction
Le caractère étudié est le groupe sanguin. Ses modalités O, A, B et AB ne sont pas des nombres : d'après la définition, il est **qualitatif**.

Compter les élèves de chaque groupe donne des effectifs, qui sont des nombres. Mais ce sont les modalités qui décident de la nature du caractère.

Vérifie : Un rappel du collège

Tu l'as vu au collège : la fréquence d'une valeur est son effectif divisé par l'effectif total. Dans une classe de 20 élèves, 8 sont du groupe O. Quelle est la fréquence du groupe O ?

A. 8

Faux.
C'est l'effectif : le nombre d'élèves du groupe O.

B. $0{,}4$

Juste.
Juste : $\dfrac{8}{20} = 0{,}4$.

C. $2{,}5$

Faux.
Tu as divisé 20 par 8 : c'est l'inverse.

D. $0{,}08$

Faux.
Tu as divisé par 100 au lieu de diviser par l'effectif total, 20.

Correction
La fréquence est l'effectif divisé par l'effectif total :

$$\frac{8}{20} = 0{,}4.$$

Contrôle : une fréquence est comprise entre 0 et 1. Ici, $0{,}4$ signifie « 4 élèves sur 10 ».

Définition : Effectif et fréquence

On considère une série statistique dont le caractère prend les valeurs (ou modalités) distinctes x_1, x_2, …, x_p.

- L'**effectif** n_i de la valeur x_i est le nombre d'individus pour lesquels le caractère prend la valeur x_i.
- L'**effectif total** est le nombre d'individus de la population : $N = n_1 + n_2 + \cdots + n_p$.
- La **fréquence** de la valeur x_i est le nombre $f_i = \dfrac{n_i}{N}$. On l'écrit sous forme de fraction ou de nombre décimal ; multipliée par 100, elle donne un pourcentage.

Exemple : les groupes sanguins de la classe

Voici le relevé de la classe de 40 élèves.

Groupe sanguin	O	A	B	AB
Effectif n_i	18	15	5	2
Fréquence f_i	0,45	0,375	0,125	0,05

L'effectif total vaut $N = 18 + 15 + 5 + 2 = 40$.

D'après la définition de la fréquence, celle du groupe A vaut $\dfrac{15}{40} = 0{,}375$. Une fréquence se lit comme une part : $0{,}375$, c'est 3 élèves sur 8.

La somme des fréquences

Additionne les fréquences du tableau :

$$\begin{aligned} &0{,}45 + 0{,}375 \\ &+ 0{,}125 + 0{,}05 \\ &= 1. \end{aligned}$$

Ce n'est pas un hasard.

Pourquoi est-ce vrai ? Chaque fréquence est une part de l'effectif total. Réunies, toutes les parts redonnent la classe entière.

Propriété : Somme des fréquences

Avec les notations de la définition de l'effectif et de la fréquence, chaque fréquence est comprise entre 0 et 1 :

$$0 \leq f_i \leq 1 \quad \text{pour tout } i \text{ de } 1 \text{ à } p,$$

et la somme des fréquences vaut 1 :

$$f_1 + f_2 + \cdots + f_p = 1.$$

Démonstration

Étape 1. Chaque effectif n_i compte une partie des N individus, donc $0 \leq n_i \leq N$. Comme $N > 0$, on peut diviser par N sans changer le sens des inégalités : $0 \leq \dfrac{n_i}{N} \leq 1$, c'est-à-dire $0 \leq f_i \leq 1$.

Étape 2. Par définition des fréquences,

$$f_1 + f_2 + \cdots + f_p = \frac{n_1}{N} + \frac{n_2}{N} + \cdots + \frac{n_p}{N} = \frac{n_1 + n_2 + \cdots + n_p}{N}.$$

Étape 3. Par définition de l'effectif total, le numérateur vaut N. Donc la somme des fréquences vaut $\frac{N}{N} = 1$.

À retenir : Ce qu'il faut retenir de la preuve

Le geste de la preuve : **écrire chaque fréquence** sous la forme $f_i = \frac{n_i}{N}$, puis tout mettre sur le **même dénominateur** N. Le numérateur devient $n_1 + n_2 + \cdots + n_p$, c'est-à-dire N.

Pour l'encadrement, on divise $0 \leq n_i \leq N$ par N, qui est strictement positif.

Vérifie : Le numérateur de la preuve

Dans la preuve, quelle définition permet de remplacer $n_1 + n_2 + \cdots + n_p$ par N ?

A. La définition de l'effectif total.

Juste.
Juste : N est, par définition, la somme des effectifs.

B. La définition de la fréquence.

Faux.
Elle sert à l'étape d'avant, pour écrire $f_i = \frac{n_i}{N}$.

C. La définition du caractère.

Faux.
Elle dit ce que l'on observe, pas combien d'individus on compte.

D. La définition de la moyenne.

Faux.
La preuve ne calcule aucune moyenne.

Correction
Après la mise au même dénominateur, la somme des fréquences vaut :

$$\frac{n_1 + n_2 + \cdots + n_p}{N}.$$

Par la définition de l'effectif total, le numérateur vaut N. La somme vaut donc $\frac{N}{N} = 1$.

Exemple : la somme sert de contrôle

- Un élève oublie le groupe AB dans son tableau. Il trouve $0{,}45 + 0{,}375 + 0{,}125 = 0{,}95$. La somme n'est pas 1 : une modalité manque. L'énoncé exige en effet que **toutes** les valeurs prises par le caractère figurent dans le tableau.

- Un autre élève écrit $\frac{40}{5} = 8$ pour la fréquence du groupe B. Le contrôle $0 \leq f_i \leq 1$ échoue : il a renversé la fraction. La bonne fréquence est $\frac{5}{40} = 0{,}125$.

La méthode, un piège, puis à toi

Méthode : dresser un tableau d'effectifs et de fréquences

1. **Nomme** la population, l'individu et le caractère, puis donne sa nature.
2. **Range** les modalités ; si le caractère est quantitatif, range les valeurs dans l'ordre croissant.
3. **Compte** l'effectif de chaque modalité, puis vérifie que leur somme est N.
4. **Divise** chaque effectif par N et vérifie que les fréquences ont pour somme 1.

Attention : effectif ou fréquence ?

✗ « Dans la classe, la fréquence du groupe B est 5. »

✓ 5 est l'**effectif** du groupe B. Sa fréquence est $\frac{5}{40} = 0{,}125$: on divise toujours par l'effectif total.

Si tu arrondis les fréquences, leur somme peut donner $0{,}99$ ou $1{,}01$. C'est l'effet des arrondis, pas une faute.

À toi : Une autre classe

Dans une autre classe de 25 élèves, on a relevé 12 élèves du groupe O, 9 du groupe A, 3 du groupe B et 1 du groupe AB. Dresse le tableau des effectifs et des fréquences, puis contrôle-le. Écris ta réponse avant de dévoiler la correction.

Indice 1
Suis la méthode : l'effectif total d'abord, puis une fréquence par groupe.
Indice 2
Chaque fréquence s'obtient en divisant par 25 : par exemple, $\frac{12}{25} = 0{,}48$.
Indice 3
Point délicat : on divise par l'effectif total, 25, et non par le nombre de groupes, 4.

Réponse
La population est la classe ; le caractère, le groupe sanguin, est qualitatif.

L'effectif total vaut $12 + 9 + 3 + 1 = 25$. Chaque fréquence est l'effectif divisé par 25 : par exemple, $\frac{12}{25} = 0{,}48$.

Groupe sanguin	O	A	B	AB
Effectif	12	9	3	1
Fréquence	0,48	0,36	0,12	0,04

Contrôle, d'après la somme des fréquences :

$$\begin{aligned} &0{,}48 + 0{,}36 \\ &+ 0{,}12 + 0{,}04 \\ &= 1. \end{aligned}$$

Effectifs et fréquences cumulés

Compter « au plus »

En clair. Une enquête a porté sur les 50 familles d'un quartier. On a noté le nombre d'enfants de chaque famille.

Nombre d'enfants x_i	0	1	2	3	4	5
Effectif n_i	7	11	15	9	5	3

Le quartier veut savoir combien de familles ont **au plus** deux enfants. L'effectif d'une seule colonne ne suffit pas.

Devine : Au plus deux enfants

Combien de familles du quartier ont au plus deux enfants ?

A. 15

Faux.
C'est le nombre de familles qui ont exactement deux enfants.

B. 18

Faux.
C'est le nombre de familles qui ont au plus un enfant : il manque celles qui en ont deux.

C. 32

Faux.
C'est le nombre de familles qui ont au moins deux enfants.

D. 33

Juste.
Juste : $7 + 11 + 15 = 33$.

Correction
« Au plus deux enfants » signifie 0, 1 ou 2 enfants. On additionne les effectifs de ces trois valeurs :

$$7 + 11 + 15 = 33.$$

Donc 33 familles ont au plus deux enfants. Cette somme de proche en proche s'appelle un effectif cumulé.

Définition : Effectifs et fréquences cumulés

On considère une série statistique d'un caractère quantitatif de valeurs x_1, x_2, ..., x_p, d'effectifs n_1, n_2, ..., n_p et d'effectif total N, les valeurs étant rangées dans l'ordre croissant : $x_1 < x_2 < \cdots < x_p$. On note f_1, f_2, ..., f_p les fréquences.

- L'**effectif cumulé croissant** de la valeur x_i est $n_1 + n_2 + \cdots + n_i$: c'est le nombre d'individus pour lesquels le caractère est **inférieur ou égal** à x_i.
- L'**effectif cumulé décroissant** de la valeur x_i est $n_i + n_{i+1} + \cdots + n_p$: c'est le nombre d'individus pour lesquels le caractère est **supérieur ou égal** à x_i.
- On définit de même la **fréquence cumulée croissante** $f_1 + f_2 + \cdots + f_i$ et la **fréquence cumulée décroissante** $f_i + f_{i+1} + \cdots + f_p$ de la valeur x_i.

Exemple : les cumuls des familles

On complète le tableau par les effectifs cumulés.

Nombre d'enfants x_i	**0**	**1**	**2**	**3**	**4**	**5**
Effectif n_i	7	11	15	9	5	3
Cumulé croissant	7	18	33	42	47	50
Cumulé décroissant	50	43	32	17	8	3

D'après la définition des effectifs cumulés, 42 familles ont au plus trois enfants.

De même, 8 familles ont au moins quatre enfants.

Lire un cumul

Vérifie : Un cumul décroissant

Dans le tableau, l'effectif cumulé décroissant de la valeur 2 est 32. Que signifie ce nombre ?

A. 32 familles ont plus de deux enfants.
Faux.
« Plus de deux » exclut les familles de deux enfants : elles sont 17.

B. 32 familles ont au moins deux enfants.
Juste.

Juste : $15 + 9 + 5 + 3 = 32$.

C. 32 familles ont au plus deux enfants.

Faux.

Celles-là sont comptées par le cumul croissant : elles sont 33.

D. 32 familles ont exactement deux enfants.

Faux.

Exactement deux : c'est l'effectif simple, 15.

Correction

Par la définition, le cumul décroissant de la valeur 2 additionne les effectifs des valeurs 2, 3, 4 et 5 :

$$15 + 9 + 5 + 3 = 32.$$

Donc 32 familles ont **au moins** deux enfants.

Contrôle : les 18 familles qui ont au plus un enfant complètent ce groupe, car $18 + 32 = 50$.

Méthode : calculer et lire des effectifs cumulés

1. **Range** les valeurs dans l'ordre croissant.
2. **Cumule** de gauche à droite : chaque cumul croissant est le précédent plus l'effectif de la colonne.
3. **Vérifie** que le dernier cumul croissant vaut N, tout comme le premier cumul décroissant.
4. **Traduis** la question : pour « au plus », prends un cumul croissant ; pour « au moins », un cumul décroissant.

Exemple : les fréquences cumulées

Les fréquences s'obtiennent en divisant chaque effectif par $N = 50$. On les cumule de la même façon.

Nombre d'enfants x_i	**0**	**1**	**2**	**3**	**4**	**5**
Fréquence f_i	0,14	0,22	0,3	0,18	0,1	0,06
Fréquence cumulée croissante	0,14	0,36	0,66	0,84	0,94	1

Ainsi, $0{,}66$ des familles ont au plus deux enfants : c'est 33 familles sur 50.

Un contrôle, puis à toi

Remarque : un contrôle gratuit

Le dernier effectif cumulé croissant vaut toujours N : il compte tous les individus. De même, la dernière fréquence cumulée croissante vaut 1. Si tu trouves autre chose, une addition est fausse ou une valeur manque.

À toi : Plus de trois enfants

Avec le tableau des familles du quartier, calcule la fréquence cumulée décroissante de la valeur 3. Déduis-en la part des familles qui ont plus de trois enfants. Écris ta réponse avant de dévoiler la correction.

Indice 1
Additionne les fréquences des valeurs 3, 4 et 5.
Indice 2
Contrôle ton résultat avec la fréquence cumulée croissante de la valeur 2 : les deux parts se complètent.
Indice 3
Point délicat : « plus de trois » exclut les familles de trois enfants ; retire leur fréquence.

Réponse
Par la définition, la fréquence cumulée décroissante de la valeur 3 vaut :

$$0{,}18 + 0{,}1 + 0{,}06 = 0{,}34.$$

Ainsi, $0{,}34$ des familles ont au moins trois enfants. Contrôle : $0{,}66 + 0{,}34 = 1$.

Plus de trois enfants, c'est au moins trois, sauf exactement trois :

$$0{,}34 - 0{,}18 = 0{,}16.$$

Contrôle : $0{,}1 + 0{,}06 = 0{,}16$. Donc $0{,}16$ des familles, soit 8 familles sur 50, ont plus de trois enfants.

Classes et histogramme

Regrouper en classes

En clair. On a mesuré 40 élèves de tronc commun. Les tailles vont de 141 cm à 188 cm, et presque toutes sont différentes. Un tableau valeur par valeur serait illisible.

On regroupe donc les tailles par tranches de 10 cm. La tranche qui va de 150 cm compris à 160 cm non compris s'écrit $[150; 160[$.

Vérifie : Dans quelle classe ?

Un élève mesure exactement 160 cm. Dans quelle classe le ranges-tu ?

A. Dans $[150; 160[$.
Faux.

Le crochet tourné vers l'extérieur exclut 160 de cette classe.

B. Dans $[150; 160[$ et dans $[160; 170[$.

Faux.

Les classes n'ont aucune valeur commune : chaque élève est compté une seule fois.

C. Dans $[160; 170[$.

Juste.

Juste : 160 est compris dans $[160; 170[$.

D. Dans aucune : 160 est une borne.

Faux.

Toute taille de 140 cm à moins de 190 cm appartient à une classe, et à une seule.

Correction

La classe $[160; 170[$ contient les tailles x telles que $160 \leq x < 170$: elle contient 160.

La classe $[150; 160[$ s'arrête juste avant 160. Ainsi, chaque élève est compté dans une seule classe.

Définition : Classes et histogramme

Lorsque le caractère est continu, ou prend de nombreuses valeurs, on regroupe ses valeurs en **classes** : des intervalles consécutifs $[a_0; a_1[$, $[a_1; a_2[$, ..., $[a_{p-1}; a_p[$, sans valeur commune. L'effectif d'une classe est le nombre d'individus dont la valeur appartient à cette classe.

- L'**amplitude** de la classe $[a; b[$ est le nombre $b - a$; son **centre** est le nombre $\dfrac{a+b}{2}$.
- L'**histogramme** de la série est formé, pour chaque classe $[a; b[$, d'un rectangle qui a pour base le segment $[a; b]$ de l'axe des abscisses et dont l'**aire est proportionnelle à l'effectif** de la classe.

Lorsque toutes les classes ont la même amplitude, on peut prendre pour hauteurs des rectangles les effectifs eux-mêmes.

Exemple : le tableau des tailles

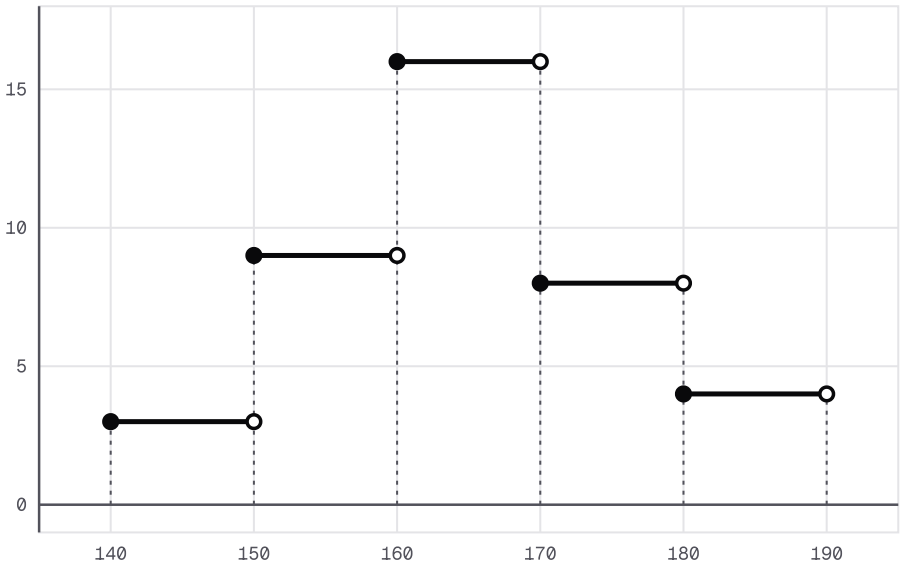
Voici les 40 tailles, regroupées en cinq classes de même amplitude.

Taille (en cm)	$[140; 150[$	$[150; 160[$	$[160; 170[$	$[170; 180[$	$[180; 190[$
Effectif	3	9	16	8	4
Centre	145	155	165	175	185

Chaque classe a pour amplitude 10 cm. Par exemple, la classe $[150; 160[$ a pour amplitude $160 - 150 = 10$ et pour centre la moyenne de ses bornes, $\dfrac{150 + 160}{2} = 155$.

L'effectif total vaut $3 + 9 + 16 + 8 + 4 = 40$.

L'histogramme



Histogramme des tailles des 40 élèves, en classes de 10 cm. Les rectangles ne sont pas remplis ; sur chaque palier, le point plein marque la borne comprise de la classe et le point creux la borne exclue.

Borne exclue en x = 150 : point creux à l'ordonnée estimée ≈ 3.

Borne exclue en x = 160 : point creux à l'ordonnée estimée ≈ 9.

Borne exclue en x = 170 : point creux à l'ordonnée estimée ≈ 16.

Borne exclue en x = 180 : point creux à l'ordonnée estimée ≈ 8.

Borne exclue en x = 190 : point creux à l'ordonnée estimée ≈ 4.

Méthode : construire un histogramme

1. **Calcule** l'amplitude de chaque classe.
2. **Choisis** une **amplitude de référence**, ici 10 cm : un rectangle de base 10 cm et de hauteur 1 représente 1 individu. La hauteur d'un rectangle est l'effectif de sa classe divisé par le nombre d'amplitudes de référence qu'elle contient. Si toutes les classes ont l'amplitude de référence, c'est l'effectif.
3. **Dessine** les rectangles sans espace entre eux, chacun posé sur sa classe.
4. **Indique** l'unité de chaque axe.

Attention : des classes de largeurs différentes

X « On regroupe les grands élèves dans la classe $[170; 190[$, d'effectif $8 + 4 = 12$, et on dessine un rectangle de hauteur 12. »

✓ Prends l'amplitude de référence 10 cm : un rectangle de base 10 cm et de hauteur 1 représente 1 élève. La classe $[170;190[$ contient deux amplitudes de référence.

C'est l'**aire** qui représente l'effectif : la hauteur est $\frac{12}{2}=6$. Avec la hauteur 12, le dessin ferait croire à 24 grands élèves.

Vérifie : Une classe plus large

On garde l'amplitude de référence 10 cm. Une classe $[180;200[$ compte 6 élèves. Quelle hauteur donnes-tu à son rectangle ?

A. 6

Faux.
C'est l'effectif : un rectangle deux fois plus large et de hauteur 6 représenterait 12 élèves.

B. 3

Juste.
Juste : la classe contient deux amplitudes de référence, et $\frac{6}{2}=3$.

C. 12

Faux.
Tu as multiplié l'effectif par 2 au lieu de le diviser.

D. $0{,}3$

Faux.
Tu as divisé par 20 : on compte en amplitudes de référence, pas en centimètres.

Correction
L'amplitude de la classe vaut $200-180=20$ cm, soit deux amplitudes de référence. La hauteur est donc $\frac{6}{2}=3$.

Contrôle : l'aire du rectangle vaut $2\times 3=6$ unités d'aire, c'est-à-dire 6 élèves.

Cumuler des classes

Devine : Moins de 160 cm

Combien des 40 élèves mesurent strictement moins de 160 cm ?

A. 9

Faux.
C'est l'effectif de la seule classe $[150;160[$: il manque la classe $[140;150[$.

B. 12

Juste.
Juste : $3+9=12$.

C. 28

Faux.

C'est le nombre d'élèves de moins de 170 cm.

Correction

Les élèves de moins de 160 cm sont ceux des classes $[140;150[$ et $[150;160[$. On additionne leurs effectifs : $3+9=12$.

C'est l'effectif cumulé croissant de la classe $[150;160[$; on l'associe à sa borne droite, 160.

Définition : Polygone des effectifs cumulés croissants

On considère une série regroupée en classes $[a_0;a_1[$, $[a_1;a_2[$, ..., $[a_{p-1};a_p[$, d'effectif total N. Pour chaque i, on note N_i l'effectif cumulé croissant de la classe $[a_{i-1};a_i[$: c'est le nombre d'individus dont la valeur est strictement inférieure à a_i.

Le **polygone des effectifs cumulés croissants** est la ligne brisée qui joint, dans l'ordre, les points de coordonnées $(a_0,0)$, (a_1,N_1), (a_2,N_2), ..., (a_p,N_p), avec $N_p=N$.

En remplaçant les effectifs cumulés par les fréquences cumulées, on obtient le **polygone des fréquences cumulées croissantes**, qui se termine au point $(a_p,1)$.

Exemple : les points du polygone

On cumule les effectifs des tailles, puis on divise par $N=40$.

Borne droite	150	160	170	180	190
Effectif cumulé N_i	3	12	28	36	40
Fréquence cumulée	0,075	0,3	0,7	0,9	1

D'après la définition du polygone, celui des effectifs cumulés part de $(140,0)$. Il passe ensuite par $(150,3)$, $(160,12)$, $(170,28)$, $(180,36)$ et $(190,40)$.

Celui des fréquences cumulées se termine en $(190,1)$.

À toi : Une autre classe

Dans une classe de 25 élèves, les tailles sont regroupées ainsi : 6 élèves dans $[150;160[$, 11 dans $[160;170[$ et 8 dans $[170;180[$. Donne l'amplitude et le centre de chaque classe, puis les points du polygone des effectifs cumulés croissants. Écris ta réponse avant de dévoiler la correction.

Indice 1

L'amplitude est la différence des bornes ; le centre est leur moyenne.

Indice 2

Cumule les effectifs : 6, puis $6+11$, puis le total.

Indice 3

Point délicat : le polygone commence au point $(150, 0)$, et chaque cumul se place à la borne **droite** de sa classe.

Réponse

Les trois classes ont pour amplitude 10 cm et pour centres 155, 165 et 175.

Les effectifs cumulés croissants valent 6, $6+11=17$ et $17+8=25$; le dernier est bien $N=25$.

D'après la définition du polygone, il joint les points $(150, 0)$, $(160, 6)$, $(170, 17)$ et $(180, 25)$.

La moyenne

La moyenne d'un tableau

Devine : Un rappel du collège

Tu l'as vu au collège : la moyenne est la somme des valeurs divisée par leur nombre. Dans un groupe, 4 élèves ont eu 10 à un devoir et 1 élève a eu 15. Quelle est la moyenne du groupe ?

A. 11

Juste.

Juste : $\dfrac{4\times 10+15}{5}=11$.

B. $12{,}5$

Faux.

C'est la moyenne de 10 et 15 : elle oublie que 10 compte quatre fois.

C. 55

Faux.

C'est la somme des cinq notes : il reste à diviser par 5.

D. $27{,}5$

Faux.

Tu as divisé la somme 55 par 2, le nombre de notes différentes, au lieu de 5.

Correction

Le groupe compte $4+1=5$ élèves. La somme de leurs notes vaut $4\times 10+1\times 15=55$. La moyenne vaut donc :

$$\frac{55}{5}=11.$$

Chaque note compte autant de fois que d'élèves l'ont obtenue.

Définition : Moyenne arithmétique

La **moyenne** d'une série statistique d'un caractère quantitatif de valeurs x_1, x_2, ..., x_p, d'effectifs n_1, n_2, ..., n_p et d'effectif total N est le nombre

$$\overline{x} = \frac{n_1 x_1 + n_2 x_2 + \cdots + n_p x_p}{N}.$$

Lorsque la série est regroupée en classes, on calcule la moyenne en remplaçant chaque classe par son centre : on obtient une **valeur approchée** de la moyenne de la série.

Exemple : des notes, puis des tailles

Des notes. Voici les notes d'un devoir dans une classe de 20 élèves.

Note x_i	7	10	12	15	18
Effectif n_i	2	5	6	4	3
Produit $n_i x_i$	14	50	72	60	54

D'après la définition de la moyenne :

$$\begin{aligned}\overline{x} &= \frac{14 + 50 + 72 + 60 + 54}{20} \\ &= \frac{250}{20} \\ &= 12{,}5.\end{aligned}$$

Des tailles. Pour les 40 tailles en classes, on remplace chaque classe par son centre. La somme des produits vaut :

$$\begin{aligned}&3 \times 145 + 9 \times 155 \\ &+ 16 \times 165 + 8 \times 175 \\ &+ 4 \times 185 = 6610.\end{aligned}$$

Donc :

$$\overline{x} = \frac{6610}{40} = 165{,}25.$$

C'est une valeur approchée : arrondie au centimètre, la taille moyenne vaut environ 165 cm.

Attention : chaque note compte selon son effectif

✗ « La moyenne du devoir vaut $\dfrac{7 + 10 + 12 + 15 + 18}{5} = 12{,}4$. »

✓ Ce calcul donne le même poids aux cinq notes, quel que soit leur effectif. Or 6 élèves ont 12, et seulement 2 élèves ont 7.

Il faut multiplier chaque note par son effectif : la moyenne vaut $12{,}5$.

La moyenne avec les fréquences

Parfois, un relevé ne donne que les fréquences, sans l'effectif total. Peut-on encore calculer la moyenne ?

Pourquoi est-ce vrai ? Diviser une somme par N, c'est diviser chacun de ses termes par N. Or $\frac{n_i}{N}$ est justement la fréquence f_i.

Propriété : Moyenne et fréquences

Avec les notations de la définition de la moyenne, si f_1, f_2, ..., f_p sont les fréquences des valeurs x_1, x_2, ..., x_p, alors

$$\overline{x} = f_1 x_1 + f_2 x_2 + \cdots + f_p x_p.$$

Démonstration

Étape 1. Par définition de la moyenne, on divise la somme $n_1 x_1 + n_2 x_2 + \cdots + n_p x_p$ par N ; on peut diviser chaque terme par N :

$$\overline{x} = \frac{n_1 x_1}{N} + \frac{n_2 x_2}{N} + \cdots + \frac{n_p x_p}{N}.$$

Étape 2. Pour chaque i, $\frac{n_i x_i}{N} = \frac{n_i}{N} \times x_i = f_i x_i$, par définition de la fréquence f_i. Donc $\overline{x} = f_1 x_1 + f_2 x_2 + \cdots + f_p x_p$.

À retenir : Ce qu'il faut retenir de la preuve

Le geste de la preuve : **couper la fraction** en autant de fractions qu'il y a de termes. Puis **reconnaître** $\frac{n_i}{N} = f_i$ dans chacune.

La moyenne devient une somme de produits $f_i x_i$. Les fréquences jouent le rôle des effectifs, et l'on ne divise plus par N.

Vérifie : L'idée de la preuve

Dans la preuve, quelle définition permet de remplacer chaque quotient $\frac{n_i}{N}$ par f_i ?

A. La définition de l'effectif total.

Faux.
Elle dit que $N = n_1 + n_2 + \cdots + n_p$, mais ne fait pas apparaître f_i.

B. La définition de la moyenne.

Faux.
Elle sert au point de départ de la preuve, pas à cette étape.

C. La définition de la fréquence.

Juste.
Juste : par définition, $f_i = \dfrac{n_i}{N}$.

D. La définition du mode.

Faux.
Le mode ne joue aucun rôle dans le calcul d'une moyenne.

Correction
La preuve écrit chaque terme $\dfrac{n_i x_i}{N}$ sous la forme $\dfrac{n_i}{N} \times x_i$.

Par la définition de la fréquence, $\dfrac{n_i}{N} = f_i$. Chaque terme devient donc $f_i x_i$, et la moyenne est la somme de ces produits.

Exemple : la moyenne du devoir par les fréquences

Les fréquences des notes 7, 10, 12, 15 et 18 valent $0{,}1$, $0{,}25$, $0{,}3$, $0{,}2$ et $0{,}15$. D'après la propriété de la moyenne et des fréquences :

$$\begin{aligned}\overline{x} &= 0{,}1 \times 7 + 0{,}25 \times 10 \\ &\quad + 0{,}3 \times 12 + 0{,}2 \times 15 \\ &\quad + 0{,}15 \times 18 \\ &= 12{,}5.\end{aligned}$$

On retrouve la moyenne calculée avec les effectifs.

Les fréquences doivent bien avoir pour somme 1. Ici :

$$\begin{aligned}&0{,}1 + 0{,}25 + 0{,}3 \\ &+ 0{,}2 + 0{,}15 = 1.\end{aligned}$$

Changer d'unité

En physique, une température x en degrés Celsius devient $y = 1{,}8x + 32$ en degrés Fahrenheit. Faut-il convertir chaque jour, puis refaire la moyenne ?

Pourquoi est-ce vrai ? Chaque valeur subit le même calcul $ax_i + b$. En développant, les termes en a redonnent $\overline{x}$, et les termes en b redonnent b.

Propriété : Moyenne et changement affine des valeurs

Soit une série statistique d'un caractère quantitatif de valeurs x_1, x_2, ..., x_p, d'effectifs n_1, n_2, ..., n_p et d'effectif total N, de moyenne $\overline{x}$, et soit a et b deux réels. On considère la série obtenue en remplaçant chaque valeur x_i par $y_i = ax_i + b$, avec les mêmes effectifs. Sa moyenne est

$$\overline{y} = a\,\overline{x} + b.$$

Si $a = 0$, toutes les valeurs y_i sont égales à b : chacune compte encore avec l'effectif n_i, et $\overline{y} = b$.

Démonstration

Étape 1. Par définition de la moyenne, appliquée à la nouvelle série :

$$\overline{y} = \frac{n_1(ax_1 + b) + n_2(ax_2 + b) + \cdots + n_p(ax_p + b)}{N}.$$

Étape 2. On développe chaque produit et on regroupe les termes en a et les termes en b :

$$\overline{y} = a \times \frac{n_1 x_1 + n_2 x_2 + \cdots + n_p x_p}{N} + b \times \frac{n_1 + n_2 + \cdots + n_p}{N}.$$

Étape 3. Par définition de la moyenne, la première fraction vaut $\overline{x}$; par définition de l'effectif total, la seconde vaut $\dfrac{N}{N} = 1$. Donc $\overline{y} = a\,\overline{x} + b$.

À retenir : Ce qu'il faut retenir de la preuve

Le geste de la preuve : **développer** chaque produit $n_i(ax_i + b)$, puis **regrouper** les termes en a et les termes en b.

On reconnaît la moyenne $\overline{x}$, puis $\dfrac{N}{N} = 1$. La moyenne suit donc le même changement que les valeurs.

Vérifie : Le terme en b

Dans la preuve, pourquoi le dernier terme devient-il simplement b ?

A. Parce que b est petit devant les valeurs.

Faux.

La preuve ne fait aucune approximation : l'égalité est exacte pour tout b.

B. Parce que b vaut 0.

Faux.

Dans la conversion en degrés Fahrenheit, $b = 32$: b n'est pas nul en général.

C. Parce que la moyenne des x_i vaut 1.

Faux.

La fraction multipliée par b ne contient aucune valeur x_i.

D. Parce que $\dfrac{n_1 + n_2 + \cdots + n_p}{N} = 1$.

Juste.

Juste : par la définition de l'effectif total.

Correction

Après regroupement, b est multiplié par :

$$\frac{n_1 + n_2 + \cdots + n_p}{N}.$$

Par la définition de l'effectif total, ce numérateur vaut N. La fraction vaut 1, et le terme se réduit à b .

Exemple : Exemple de physique : des degrés Fahrenheit

Les températures de la ville A ont pour moyenne $\overline{x} = 24$ °C. En degrés Fahrenheit, chaque valeur devient $y_i = 1{,}8x_i + 32$.

D'après la propriété de la moyenne et du changement affine :

$$\overline{y} = 1{,}8 \times 24 + 32 = 75{,}2.$$

La moyenne vaut $75{,}2$ °F, sans convertir les sept températures.

Contrôle : converties une à une, elles ont pour somme $526{,}4$, et $\dfrac{526{,}4}{7} = 75{,}2$.

La méthode, puis à toi

Méthode : calculer une moyenne

1. **Ajoute** au tableau une ligne des produits $n_i x_i$; pour des classes, prends les centres.
2. **Additionne** ces produits, puis divise par N.
3. **Simplifie** si besoin par un changement $y_i = ax_i + b$: calcule $\overline{y}$, puis reviens à $\overline{x}$.
4. **Vérifie** que le résultat se situe entre la plus petite valeur et la plus grande.

À toi : Un deuxième devoir

Au devoir suivant, les 20 élèves ont obtenu 3 fois la note 8, 7 fois 11, 6 fois 13 et 4 fois 16. Calcule la moyenne de la classe. Le professeur ajoute ensuite 1 point à chaque copie : quelle est la nouvelle moyenne ? Écris ta réponse avant de dévoiler la correction.

Indice 1
Complète une ligne des produits $n_i x_i$, puis divise leur somme par 20.
Indice 2
Commence par le premier produit : $3 \times 8 = 24$; continue pour chaque note.
Indice 3
Point délicat : inutile de tout recalculer pour la nouvelle moyenne. C'est un changement $y = x + 1$, avec $a = 1$ et $b = 1$.

Réponse
D'après la définition de la moyenne, on additionne d'abord les produits $n_i x_i$:

$$\begin{aligned}&3\times 8+7\times 11\\&+6\times 13+4\times 16\\&=243.\end{aligned}$$

Donc :

$$\overline{x}=\frac{243}{20}=12{,}15.$$

Contrôle : $12{,}15$ est entre 8 et 16.

Chaque note devient $y_i=x_i+1$. D'après la propriété de la moyenne et du changement affine, $\overline{y}=\overline{x}+1=13{,}15$.

Médiane et mode

La médiane

Vérifie : Un rappel du collège

Tu l'as vu au collège : la médiane partage une série rangée en deux moitiés. Voici les températures de la ville A, jour par jour : 22, 24, 25, 23, 25, 24, 25. Quelle est leur médiane ?

A. 23

Faux.
C'est la valeur du milieu de la liste, qui n'est pas rangée.

B. $23{,}5$

Faux.
C'est le milieu entre la plus petite et la plus grande valeur : ce n'est pas la médiane.

C. 25

Faux.
C'est la valeur relevée le plus grand nombre de jours.

D. 24

Juste.
Juste : c'est la quatrième valeur de la liste rangée.

Correction
On range d'abord les sept valeurs : 22, 23, 24, 24, 25, 25, 25. La valeur du milieu est la quatrième : 24.

Il y a trois valeurs avant elle et trois après. Donc la médiane vaut 24 °C.

Définition : Médiane

On range les N valeurs d'une série statistique, chacune répétée autant de fois que son effectif, dans l'ordre croissant.

- Si N est impair, la **médiane** M_e est la valeur de rang $\dfrac{N+1}{2}$.
- Si N est pair, la **médiane** M_e est la moyenne des valeurs de rangs $\dfrac{N}{2}$ et $\dfrac{N}{2}+1$.

La médiane partage ainsi la série ordonnée en deux groupes de même effectif. Lorsque la série est regroupée en classes, on prend pour **valeur approchée** de la médiane l'abscisse du point du polygone des effectifs cumulés croissants dont l'ordonnée est $\dfrac{N}{2}$.

Exemple : un effectif impair, puis pair

Effectif impair. Les sept températures de la ville B, rangées, sont $18, 21, 22, 24, 26, 27, 30$. Ici $N = 7$ et $\dfrac{N+1}{2} = 4$: d'après la définition de la médiane, $M_e = 24$ °C.

Effectif pair. On ajoute le lundi suivant, à 29 °C. La série rangée devient $18, 21, 22, 24, 26, 27, 29, 30$. Ici $N = 8$: les rangs du milieu sont 4 et 5. Donc $M_e = \dfrac{24+26}{2} = 25$ °C.

Cette médiane n'est pas une température relevée : c'est permis.

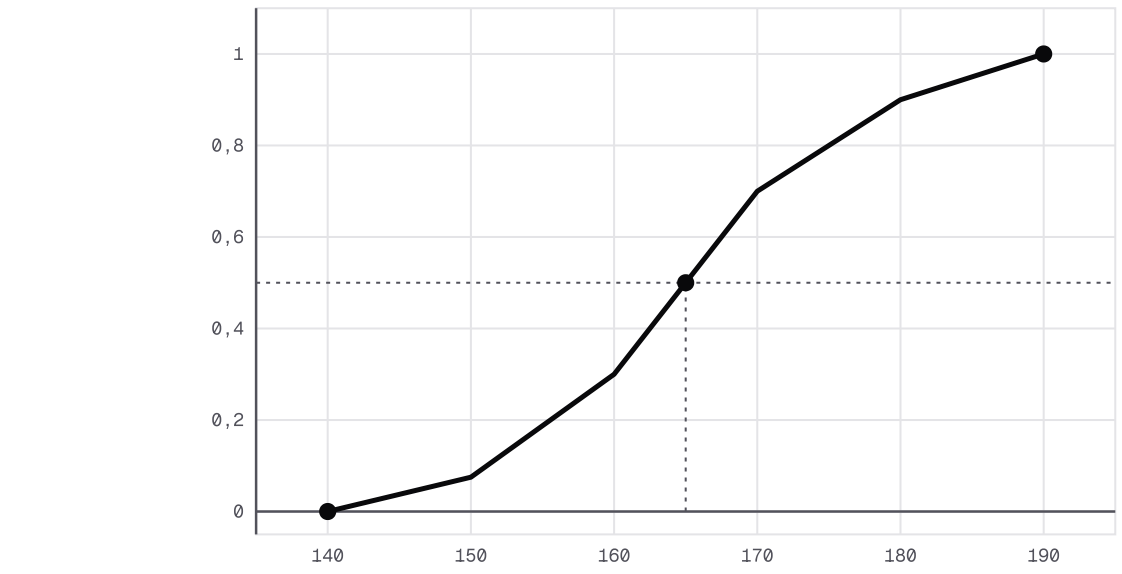
Attention : la médiane d'un tableau

✗ « Pour les familles du quartier, les valeurs vont de 0 à 5 enfants : la médiane est $2{,}5$. »

✓ La médiane se cherche dans la série **rangée**, chaque valeur comptée selon son effectif. Ici $N = 50$: les rangs du milieu sont 25 et 26.

Les effectifs cumulés croissants valent 18 pour 1 enfant et 33 pour 2 enfants. Les deux rangs correspondent donc à 2 enfants : $M_e = 2$.

La médiane d'une série en classes



Polygone des fréquences cumulées croissantes des tailles des 40 élèves, et droite horizontale d'ordonnée 0,5 pour lire la médiane.

Méthode : déterminer une médiane

1. **Calcule** N et les effectifs cumulés croissants.
2. **Série de valeurs** : repère le rang du milieu, ou les deux rangs du milieu si N est pair. Pour chaque rang, lis la première valeur pour laquelle le cumul croissant atteint ou dépasse ce rang. Si N est pair, M_e est la moyenne des deux valeurs obtenues.
3. **Série en classes** : note N_a et N_b les cumuls en a et en b, et repère la classe $[a;b[$ telle que $N_a < \dfrac{N}{2} \leq N_b$. Calcule $M_e = a + (b-a) \times \dfrac{\frac{N}{2} - N_a}{N_b - N_a}$: la fraction est la part de la classe à parcourir pour atteindre $\dfrac{N}{2}$.
4. **Vérifie**, pour une série en classes, que $a < M_e \leq b$.

Exemple : la médiane des tailles

Pour les 40 tailles, $\dfrac{N}{2} = 20$. Les effectifs cumulés valent 12 en 160 et 28 en 170. Comme $12 < 20 \leq 28$, la médiane est dans la classe $[160;170[$.

Sur ce segment du polygone, on interpole :

$$\begin{aligned} M_e &= 160 + 10 \times \frac{20 - 12}{28 - 12} \\ &= 160 + 10 \times \frac{8}{16} \\ &= 165. \end{aligned}$$

La médiane vaut environ 165 cm : c'est une valeur approchée, comme la lecture sur le polygone.

Avec les fréquences cumulées, le quotient est le même : $\dfrac{0{,}5-0{,}3}{0{,}7-0{,}3}=\dfrac{8}{16}$.

Le mode

Devine : Un caractère qualitatif

Pour les groupes sanguins de la classe, on ne peut calculer ni moyenne ni médiane. Quel renseignement garde un sens ?

A. Le groupe qui a le plus grand effectif.
Juste.
Juste : comparer des effectifs reste possible.

B. La moyenne des quatre groupes.
Faux.
On ne fait pas la moyenne de O, A, B et AB : ce ne sont pas des nombres.

C. Le groupe placé au milieu du tableau.
Faux.
L'ordre des colonnes O, A, B, AB est arbitraire : il n'a pas de milieu naturel.

D. L'écart entre le plus grand et le plus petit groupe.
Faux.
Un écart se calcule entre des nombres, pas entre des groupes sanguins.

Correction
Le caractère est qualitatif : ses modalités ne sont pas des nombres. On ne peut ni les additionner, ni les ranger de façon naturelle.

En revanche, on peut chercher la modalité la plus représentée, en comparant les effectifs.

Définition : Mode et classe modale

Le **mode** d'une série statistique est une valeur (ou une modalité) du caractère dont l'effectif est le plus grand. Une série peut avoir plusieurs modes.

Lorsque la série est regroupée en classes, une **classe modale** est une classe dont le rectangle est le plus haut dans l'histogramme. Si toutes les classes ont la même amplitude, c'est une classe de plus grand effectif.

Exemple : un mode, une classe modale

- Dans la classe de 40 élèves, le groupe O a le plus grand effectif, 18. D'après la définition du mode, le mode de la série est le groupe O.
- Pour les tailles, les classes ont toutes l'amplitude 10 cm. La classe modale est donc celle de plus grand effectif : $[160;170[$, avec 16 élèves. Sur l'histogramme, c'est le rectangle le plus haut.
- Les températures de la ville A, 22, 23, 24, 24, 25, 25, 25, ont pour mode 25 °C.

Vérifie : Le mode des familles

Quel est le mode de la série du nombre d'enfants des 50 familles du quartier ?

A. 15

Faux.
C'est l'effectif de la valeur la plus représentée, pas la valeur elle-même.

B. 5

Faux.
C'est la plus grande valeur, prise par seulement 3 familles.

C. 2

Juste.
Juste : 2 enfants est la valeur de plus grand effectif.

D. $2{,}06$

Faux.
C'est la moyenne de la série, pas son mode.

Correction
Le plus grand effectif du tableau est 15 ; il correspond à la valeur 2. D'après la définition du mode, le mode est donc 2 enfants.

Le mode est une valeur du caractère, jamais un effectif.

À toi, puis le coin devoir

À toi : Médiane et mode du deuxième devoir

Reprends le deuxième devoir : 3 fois la note 8, 7 fois 11, 6 fois 13 et 4 fois 16, pour 20 élèves. Détermine la médiane et le mode de la série. Écris ta réponse avant de dévoiler la correction.

Indice 1
Calcule les effectifs cumulés croissants : 3, 10, puis la suite.
Indice 2
Avec $N = 20$, les rangs du milieu sont 10 et 11.
Indice 3

Point délicat : le rang 10 et le rang 11 ne tombent pas sur la même note.

Réponse
Les effectifs cumulés croissants valent 3, 10, 16 et 20.

Ici $N = 20$ est pair : d'après la définition de la médiane, on prend les rangs 10 et 11. La valeur de rang 10 est 11, car le cumul atteint 10 à la note 11. La valeur de rang 11 est 13, car le cumul passe de 10 à 16 à la note 13. Donc :

$$M_e = \frac{11 + 13}{2} = 12.$$

Le plus grand effectif est 7, pour la note 11 : le mode est 11.

Remarque : Coin devoir : rédiger une interpolation

En devoir, une médiane par interpolation se rédige en trois temps : la classe, le calcul, l'arrondi.

Reprends la classe de 25 élèves, avec 6, 11 et 8 élèves dans $[150; 160[$, $[160; 170[$ et $[170; 180[$. Écris : « $\frac{N}{2} = 12{,}5$. Les effectifs cumulés valent 6 en 160 et 17 en 170 ; comme $6 < 12{,}5 \leq 17$, M_e est dans $[160; 170[$. Par interpolation linéaire, $M_e = 160 + 10 \times \frac{12{,}5 - 6}{17 - 6} \approx 165{,}9$ cm, arrondi au dixième. »

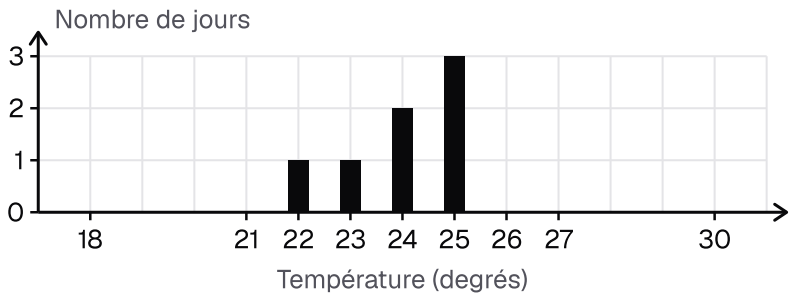
Vérifie que le résultat est bien dans la classe repérée.

Dispersion : étendue, écart moyen, variance, écart type

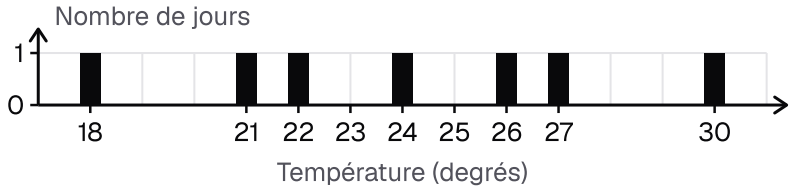
Mesurer l'étalement

En clair. Reviens aux deux villes du début. Elles ont la même moyenne, 24 °C, et la même médiane, 24 °C. Les paramètres de position ne les distinguent pas.

Pourtant, les températures de la ville B s'éloignent beaucoup plus de 24 °C. Il faut un nombre qui mesure cet éloignement : un **paramètre de dispersion**.



Ville A, semaine du début du chapitre : les sept relevés se tassent autour de 24 degrés, entre 22 et 25.



Ville B sur le même axe et le même quadrillage : ses sept relevés s'étalent de 18 à 30 degrés.

Devine : Des écarts qui s'annulent

Pour la ville A, les écarts à la moyenne $x_i - 24$ valent, jour par jour : $-2, 0, 1, -1, 1, 0, 1$. Que vaut leur somme ?

A. 6

Faux.
C'est la somme des distances, sans les signes : les écarts négatifs ont été comptés positifs.

B. 3

Faux.
C'est la somme des écarts positifs seulement : il faut aussi ajouter -2 et -1.

C. 0

Juste.
Juste : les écarts positifs et négatifs se compensent.

D. 24

Faux.
24 est la moyenne elle-même, pas la somme des écarts.

Correction
La somme vaut :

$$-2 + 0 + 1 - 1 + 1 + 0 + 1 = 0.$$

Les écarts au-dessus et au-dessous de la moyenne se compensent. Leur somme ne mesure donc pas la dispersion.

On prend plutôt les distances $|x_i - \overline{x}|$, ou les carrés $(x_i - \overline{x})^2$.

Définition : Étendue, écart moyen, variance, écart type

On considère une série statistique d'un caractère quantitatif de valeurs x_1, x_2, ..., x_p, d'effectifs n_1, n_2, ..., n_p et d'effectif total N, de moyenne $\overline{x}$.

- L'**étendue** est la différence entre la plus grande et la plus petite valeur de la série.

- L'**écart moyen** est la moyenne des distances des valeurs à la moyenne :

$$e = \frac{n_1|x_1 - \overline{x}| + n_2|x_2 - \overline{x}| + \cdots + n_p|x_p - \overline{x}|}{N}.$$

- La **variance** est la moyenne des carrés des écarts à la moyenne :

$$V = \frac{n_1(x_1 - \overline{x})^2 + n_2(x_2 - \overline{x})^2 + \cdots + n_p(x_p - \overline{x})^2}{N}.$$

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

Exemple : la dispersion de la ville A

Pour la ville A, $\overline{x} = 24$. On complète le tableau des valeurs rangées.

Température x_i	22	23	24	25
Effectif n_i	1	1	2	3
Distance $\lvert x_i - 24\rvert$	2	1	0	1
Carré $(x_i - 24)^2$	4	1	0	1

L'étendue vaut $25 - 22 = 3$ °C.

D'après la définition de l'écart moyen, on additionne les produits $n_i|x_i - 24|$:

$$1 \times 2 + 1 \times 1 + 2 \times 0 + 3 \times 1 = 6.$$

Donc $e = \dfrac{6}{7} \approx 0{,}86$ °C.

D'après la définition de la variance, on additionne de même les produits $n_i(x_i - 24)^2$:

$$1 \times 4 + 1 \times 1 + 2 \times 0 + 3 \times 1 = 8.$$

Donc $V = \dfrac{8}{7}$ et $\sigma = \sqrt{\dfrac{8}{7}} \approx 1{,}07$ °C. Les deux valeurs sont arrondies au centième.

Une formule plus rapide

Quand la moyenne n'est pas un nombre entier, calculer chaque $(x_i - \overline{x})^2$ devient long. Une autre formule évite ces soustractions.

Pourquoi est-ce vrai ? On développe chaque carré avec une identité remarquable du collège. On reconnaît ensuite la moyenne et l'effectif total.

Propriété : Variance : moyenne des carrés moins carré de la moyenne

Avec les notations de la définition de la variance,

$$V = \frac{n_1 x_1^2 + n_2 x_2^2 + \cdots + n_p x_p^2}{N} - \overline{x}^{\,2}.$$

Autrement dit, la variance est la moyenne des carrés des valeurs, diminuée du carré de la moyenne.

Démonstration

Étape 1. Pour chaque i, l'identité remarquable $(u-v)^2 = u^2 - 2uv + v^2$ donne $(x_i - \overline{x})^2 = x_i^2 - 2\,\overline{x}\,x_i + \overline{x}^{\,2}$.

Étape 2. Par définition de la variance, V s'obtient en multipliant chaque $(x_i - \overline{x})^2$ par n_i, en additionnant et en divisant par N. On fait de même avec les membres de droite et on regroupe les termes :

$$V = \frac{n_1 x_1^2 + \cdots + n_p x_p^2}{N} - 2\,\overline{x} \times \frac{n_1 x_1 + \cdots + n_p x_p}{N} + \overline{x}^{\,2} \times \frac{n_1 + \cdots + n_p}{N}.$$

Étape 3. Par définition de la moyenne, la deuxième fraction vaut $\overline{x}$; par définition de l'effectif total, la troisième vaut 1. Donc $V = \dfrac{n_1 x_1^2 + \cdots + n_p x_p^2}{N} - 2\,\overline{x}^{\,2} + \overline{x}^{\,2}$, et comme $-2\,\overline{x}^{\,2} + \overline{x}^{\,2} = -\overline{x}^{\,2}$, on obtient la formule annoncée.

À retenir : Ce qu'il faut retenir de la preuve

Le geste de la preuve : **développer** $(x_i - \overline{x})^2$, puis **séparer** la somme en trois morceaux. Le deuxième fait apparaître $\overline{x}$, le troisième $\dfrac{N}{N} = 1$.

Il reste la moyenne des carrés, moins $2\,\overline{x}^{\,2}$, plus $\overline{x}^{\,2}$.

Vérifie : L'identité employée

Quelle identité remarquable la preuve emploie-t-elle pour développer $(x_i - \overline{x})^2$?

A. $(u-v)^2 = u^2 - 2uv + v^2$

Juste.

Juste : avec $u = x_i$ et $v = \overline{x}$.

B. $(u-v)^2 = u^2 - v^2$

Faux.

Il manque le double produit : par exemple, $(3-1)^2 = 4$ et non 8.

C. $(u-v)^2 = u^2 + 2uv + v^2$

Faux.

C'est le développement de $(u+v)^2$: le signe du double produit est faux.

D. $(u-v)(u+v) = u^2 - v^2$

Faux.

Cette identité ne développe pas un carré.

Correction

La preuve utilise $(u-v)^2 = u^2 - 2uv + v^2$, avec $u = x_i$ et $v = \overline{x}$. Elle obtient :

$$(x_i - \overline{x})^2 = x_i^2 - 2\,\overline{x}\,x_i + \overline{x}^{\,2}.$$

Les autres écritures sont fausses : avec $u = 3$ et $v = 1$, on a $(u-v)^2 = 4$, alors que $u^2 - v^2 = 8$.

Exemple : la ville B par la formule

Pour la ville B, $\overline{x} = 24$. On ajoute une ligne des carrés des valeurs.

Température x_i	**18**	**21**	**22**	**24**	**26**	**27**	**30**
Carré x_i^2	324	441	484	576	676	729	900

La somme des carrés vaut 4130. D'après la formule de la variance :

$$\begin{aligned} V &= \frac{4130}{7} - 24^2 \\ &= 590 - 576 \\ &= 14. \end{aligned}$$

Donc $\sigma = \sqrt{14} \approx 3{,}74$ °C, arrondi au centième.

La méthode, un piège, puis à toi

Méthode : calculer une variance et un écart type

1. **Calcule** d'abord la moyenne $\overline{x}$.
2. **Ajoute** une ligne des produits $n_i x_i^2$, puis fais leur somme.
3. **Applique** la formule de la variance ; si $\overline{x}$ est un nombre simple, la définition convient aussi.
4. **Prends** la racine carrée pour obtenir σ, dans l'unité des valeurs, et annonce l'arrondi.

Attention : l'ordre de la soustraction

✗ « Pour la ville B, $V = 24^2 - \dfrac{4130}{7} = -14$, donc $\sigma = 14$ °C. »

✓ Deux erreurs. On retranche le **carré de la moyenne** à la **moyenne des carrés** :
$V = 590 - 576 = 14$. Une variance n'est jamais négative.

Ensuite, l'écart type est la racine carrée de la variance : $\sigma = \sqrt{14} \approx 3{,}74$ °C, arrondi au centième.

La variance s'exprime en degrés Celsius au carré ; seul l'écart type est dans l'unité des valeurs.

À toi : La dispersion des notes

Reprends le premier devoir : 2 fois la note 7, 5 fois 10, 6 fois 12, 4 fois 15 et 3 fois 18, de moyenne $12{,}5$. Calcule son étendue, sa variance et son écart type, arrondi au centième. Écris ta réponse avant de dévoiler la correction.

Indice 1
Suis la méthode : la moyenne est donnée, ajoute la ligne des produits $n_i x_i^2$.
Indice 2
Commence par le premier produit : $2 \times 7^2 = 98$; continue pour chaque note.
Indice 3
Point délicat : on retranche $12{,}5^2 = 156{,}25$, et non $12{,}5$.

Réponse
L'étendue vaut $18 - 7 = 11$ points.

Les produits $n_i x_i^2$ valent 98, 500, 864, 900 et 972 ; leur somme est 3334. D'après la formule de la variance :

$$\begin{aligned} V &= \frac{3334}{20} - 12{,}5^2 \\ &= 166{,}7 - 156{,}25 \\ &= 10{,}45. \end{aligned}$$

Donc $\sigma = \sqrt{10{,}45} \approx 3{,}23$ points, arrondi au centième.

Comparer deux séries

Méthode : comparer deux séries

1. **Compare les positions** : moyennes et médianes, dans la même unité.
2. **Compare les dispersions** : étendues, puis écarts types.
3. **Conclus** en une phrase sur la situation : où les valeurs sont-elles les plus hautes, où sont-elles les plus régulières ?

Exemple : les deux villes

Ville	Moyenne	Médiane	Étendue	Écart type
A	24 °C	24 °C	3 °C	1,07 °C environ
B	24 °C	24 °C	12 °C	3,74 °C environ

Les positions sont identiques : en moyenne, les deux villes ont eu aussi chaud.

L'écart type de la ville B vaut environ trois fois et demie celui de la ville A.

Conclusion. Cette semaine, la ville A a eu le climat le plus régulier. C'est bien la réponse prévue au début du chapitre.

Remarque : changer toutes les valeurs

Si tu ajoutes un même nombre à toutes les valeurs, la moyenne augmente de ce nombre et chaque écart à la moyenne reste le même : l'écart type ne change pas. Avec un point de plus sur chaque copie du premier devoir, il vaut toujours environ $3{,}23$ points.

Si tu multiplies toutes les valeurs par un nombre a, l'écart type est multiplié par $|a|$. Tu le vérifieras dans un exercice d'entraînement, pour un exemple.

Par exemple, en degrés Fahrenheit, l'écart type de la ville A vaut $1{,}8 \times \sqrt{\dfrac{8}{7}} \approx 1{,}92$ °F, arrondi au centième.

Remarque : Coin devoir : interpréter un écart type

En devoir, un écart type s'interprète par une phrase sur la situation, pas seulement par un nombre.

Une autre classe a eu, au même devoir, une moyenne de $12{,}5$ et un écart type de $1{,}5$ point. Écris :

« Les deux classes ont la même moyenne, $12{,}5$. L'écart type de la première classe, environ $3{,}23$ points, est plus du double de celui de la seconde, $1{,}5$ point. Les notes de la première classe sont donc plus dispersées autour de la moyenne : la seconde classe est plus homogène. »

Ce qu'il faut savoir faire

À retenir : quel paramètre pour quelle question

- **Organiser** : un tableau d'effectifs, puis les fréquences $f_i = \dfrac{n_i}{N}$, de somme 1.
- **Compter « au plus » ou « au moins »** : les effectifs cumulés croissants ou décroissants.
- **Représenter une série en classes** : l'histogramme, où l'aire de chaque rectangle représente l'effectif.
- **Situer les valeurs** : la moyenne $\overline{x}$, la médiane M_e, le mode.
- **Mesurer la dispersion** : l'étendue, l'écart moyen e, la variance V et l'écart type $\sigma = \sqrt{V}$.
- **Comparer deux séries** : d'abord la position, puis la dispersion, puis une phrase sur la situation.

Le mot du prof. Une série statistique ne se résume jamais à un seul nombre. La moyenne dit où se placent les valeurs ; l'écart type dit si l'on peut s'y fier.

Les deux villes du début le montrent : même moyenne, climats très différents. Avec ces outils, tu sais lire un relevé de physique ou de SVT, et en tirer une conclusion.

SOURCES

Référentiel BURHAN du chapitre « Statistique » (tc · sciences · mathematiques)