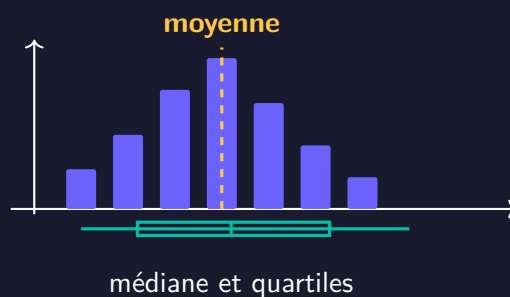


Statistiques à une variable

Décrire, résumer et comparer des données

Moyenne ■ Quartiles ■ Écart type



Seconde générale ■ Mathématiques ■ Programme officiel

 Cours complet

 15 exercices

 Problème défi

 Corrigés

Table des matières

1	 Pourquoi étudier les statistiques ?	3
2	 Vocabulaire et organisation des données	3
2.1	Population, individus et caractère	3
2.2	Effectifs et fréquences	4
3	 Représenter une série statistique	4
3.1	Diagramme en bâtons	4
3.2	Effectifs cumulés et fréquences cumulées	5
4	 La moyenne	5
4.1	Moyenne simple et moyenne pondérée	5
4.2	Linéarité de la moyenne	6
4.3	Ajouter, retirer ou réunir des données	7
5	 Médiane et quartiles	7
5.1	La médiane partage la série en deux	7
5.2	Les quartiles	8
5.3	Boîte à moustaches	8
6	 Mesurer la dispersion	9
6.1	Étendue et écart interquartile	9
6.2	Variance et écart type	9
7	 Données regroupées en classes	10
7.1	Pourquoi regrouper ?	10
7.2	Histogramme	10
7.3	Moyenne estimée avec les centres de classes	11
7.4	Classe médiane et interpolation	11
8	 Comparer deux séries	12
9	 Calculer avec Python	13
10	 Pour aller plus loin - Accessible à tous	14
10.1	Une formule plus rapide pour la variance	14
10.2	Transformer l'écart type	14
10.3	Pourquoi la moyenne est-elle un centre naturel ?	15
10.4	Score centré-réduit	15
11	 Boîte à outils et erreurs à éviter	16
12	 Exercices progressifs	17
13	 Problème défi : deux stations de mesure	20
14	 Corrigés détaillés	22
15	 Corrigé du problème défi	27
16	 Fiche-mémoire	30

1 ? Pourquoi étudier les statistiques ?

Une liste de centaines de valeurs est souvent illisible. Les statistiques cherchent à la résumer sans mentir : où se situe son centre, comment les valeurs se dispersent-elles, existe-t-il des valeurs inhabituelles, et comment comparer deux groupes ?

Intuition | Un résumé n'est jamais la série entière

La moyenne, la médiane ou l'écart type éclairent chacun un aspect des données. Aucun indicateur ne raconte tout. Une bonne analyse associe au moins un indicateur de position, un indicateur de dispersion et, si possible, un graphique.

Exemple | Même moyenne, histoires différentes

Les séries

$$A : 9, 10, 10, 10, 11 \quad \text{et} \quad B : 0, 5, 10, 15, 20$$

ont toutes deux pour moyenne 10. Pourtant, les valeurs de A sont très regroupées tandis que celles de B sont très dispersées. La moyenne seule ne suffit donc pas.

Attention | Les données ne parlent pas toutes seules

Avant de calculer, vérifie la population étudiée, l'unité, la façon dont les données ont été recueillies et les éventuelles valeurs manquantes. Un calcul exact sur des données biaisées peut conduire à une conclusion trompeuse.

2 Vocabulaire et organisation des données

2.1 Population, individus et caractère

Définition | Le langage statistique

- La **population** est l'ensemble étudié ; un **individu** est un de ses éléments.
- Le **caractère** ou la **variable** est l'information observée sur chaque individu.
- Une variable quantitative est **discrète** si elle prend des valeurs isolées, et **continue** si elle peut prendre toute valeur d'un intervalle.
- Une **modalité** est une valeur possible du caractère.

Exemple | Reconnaître la nature de la variable

Le nombre de livres lus en un mois est quantitatif discret. La taille d'un élève est quantitative continue. Le moyen de transport est qualitatif : il sera étudié plus précisément dans la fiche suivante.

2.2 Effectifs et fréquences

Définition | Effectif et fréquence

Une valeur x_i apparaît n_i fois. Son **effectif** est n_i . Si l'effectif total est

$$N = n_1 + n_2 + \cdots + n_p,$$

sa **fréquence** est

$$f_i = \frac{n_i}{N}.$$

★ Théorème | Somme des fréquences

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

En pourcentage, leur somme vaut 100 %.

 *Démonstration | Pourquoi la somme vaut-elle 1 ?*

$$f_1 + \cdots + f_p = \frac{n_1}{N} + \cdots + \frac{n_p}{N} = \frac{n_1 + \cdots + n_p}{N} = \frac{N}{N} = 1.$$

Exemple | Nombre de livres lus par 20 élèves

Nombre x_i	0	1	2	3	4	Total
Effectif n_i	2	5	7	4	2	20
Fréquence f_i	0,10	0,25	0,35	0,20	0,10	1

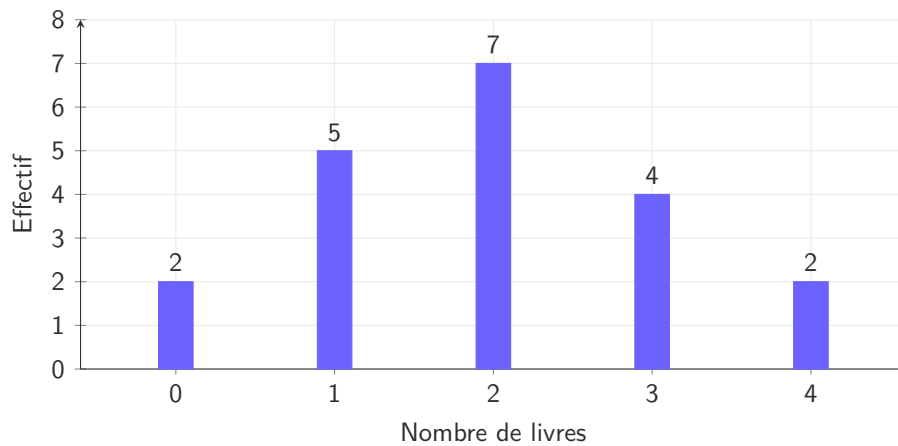
Méthode | Construire un tableau fiable

Trie les valeurs, compte chaque modalité, vérifie que la somme des effectifs vaut N , puis calcule les fréquences et vérifie qu'elles somment à 1. Ces deux contrôles détectent beaucoup d'erreurs.

3 Représenter une série statistique

3.1 Diagramme en bâtons

Pour une variable quantitative discrète, chaque valeur est représentée par un bâton dont la hauteur est proportionnelle à son effectif ou à sa fréquence.



⚠ Attention | Bâtons ou histogramme ?

Dans un diagramme en bâtons, les barres sont séparées car les valeurs sont isolées. Dans un histogramme, les rectangles se touchent car ils représentent des intervalles de valeurs continues.

3.2 Effectifs cumulés et fréquences cumulées

📖 Définition | Cumuler jusqu'à une valeur

L'effectif cumulé croissant jusqu'à x_i est le nombre de données inférieures ou égales à x_i :

$$N_i = n_1 + n_2 + \dots + n_i.$$

La fréquence cumulée correspondante est $F_i = N_i / N$.

Pour la série des livres, les effectifs cumulés sont 2, 7, 14, 18, 20. Ils permettent de localiser immédiatement la médiane et les quartiles.

4 ⚖ La moyenne

4.1 Moyenne simple et moyenne pondérée

📖 Définition | Moyenne

Pour N valeurs $y_1, \dots, y_N$, la moyenne est

$$\bar{x} = \frac{y_1 + y_2 + \dots + y_N}{N}.$$

Si la valeur x_i apparaît n_i fois, alors

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

On parle de moyenne pondérée par les effectifs.

 Démonstration | Pourquoi les deux formules sont-elles équivalentes ?

Dans la somme de toutes les données, la valeur x_1 est écrite n_1 fois : elle contribue donc pour $n_1 x_1$. De même, x_i contribue pour $n_i x_i$. Regrouper les termes égaux transforme la première formule en la seconde.

 Exemple | Moyenne du nombre de livres

$$\bar{x} = \frac{2 \times 0 + 5 \times 1 + 7 \times 2 + 4 \times 3 + 2 \times 4}{20} = \frac{39}{20} = \boxed{1,95}.$$

En moyenne, les élèves ont lu 1,95 livre. Cette valeur n'a pas besoin d'être une modalité observée.

 Intuition | La moyenne comme point d'équilibre

Imagine chaque donnée comme une masse posée sur un axe. La moyenne est le point où l'ensemble s'équilibre. Les écarts positifs et négatifs à la moyenne se compensent exactement.

 Théorème | Somme des écarts à la moyenne

$$\sum_{i=1}^p n_i (x_i - \bar{x}) = 0.$$

 Démonstration | Les écarts se compensent

$$\begin{aligned} \sum_{i=1}^p n_i (x_i - \bar{x}) &= \sum_{i=1}^p n_i x_i - \bar{x} \sum_{i=1}^p n_i \\ &= N\bar{x} - \bar{x}N = 0. \end{aligned}$$

4.2 Linéarité de la moyenne

 Théorème | Transformer toutes les données

Si chaque donnée x est remplacée par $ax + b$, alors la moyenne devient

$$\bar{x}_{\text{nouvelle}} = a\bar{x} + b.$$

 Démonstration | Distribuer la somme

Les nouvelles valeurs sont $ay_1 + b, \dots, ay_N + b$. Leur moyenne vaut

$$\frac{(ay_1 + b) + \dots + (ay_N + b)}{N} = a \frac{y_1 + \dots + y_N}{N} + \frac{Nb}{N} = a\bar{x} + b.$$

💡 Exemple | Convertir des températures

La conversion de degrés Celsius en degrés Fahrenheit est $F = 1,8C + 32$. Si la température moyenne est 15°C , la moyenne convertie vaut

$$1,8 \times 15 + 32 = 59^{\circ}\text{F}.$$

Il n'est pas nécessaire de convertir chaque mesure séparément.

4.3 Ajouter, retirer ou réunir des données

✓ Propriété | Ajouter une valeur

Une série de N données a pour moyenne $\bar{x}$. Après l'ajout d'une valeur a , la nouvelle moyenne est

$$\bar{x}' = \frac{N\bar{x} + a}{N + 1}.$$

✎ Démonstration | Reconstruire la somme

L'ancienne somme vaut $N\bar{x}$. Ajouter a donne la somme $N\bar{x} + a$, répartie entre $N + 1$ valeurs.

✓ Propriété | Réunir deux groupes

Deux groupes d'effectifs N_1, N_2 et de moyennes m_1, m_2 ont pour moyenne commune

$$m = \frac{N_1 m_1 + N_2 m_2}{N_1 + N_2}.$$

⚠ Attention | La moyenne des moyennes

On ne calcule $(m_1 + m_2)/2$ que si les deux groupes ont le même effectif. Sinon, chaque moyenne doit être pondérée par la taille de son groupe.

5 Médiante et quartiles

5.1 La médiane partage la série en deux

📖 Définition | Médiane

On range les N valeurs dans l'ordre croissant.

- Si N est impair, la médiane est la valeur de rang $(N + 1)/2$.
- Si N est pair, on prend usuellement la moyenne des valeurs de rangs $N/2$ et $N/2 + 1$.

Au moins la moitié des données est inférieure ou égale à la médiane, et au moins la moitié lui est supérieure ou égale.

💡 Exemple | Deux tailles de séries

Pour 3, 5, 8, 9, 12, la médiane est la troisième valeur : 8. Pour 3, 5, 8, 9, 12, 20, elle vaut

$$\frac{8+9}{2} = 8,5.$$

⚠ Attention | Toujours trier avant de chercher les rangs

Les formules de rang n'ont de sens que dans une liste ordonnée. Dans un tableau d'effectifs, utilise les effectifs cumulés pour repérer la valeur qui contient le rang recherché.

5.2 Les quartiles

📖 Définition | Premier et troisième quartiles

Dans la convention usuelle du lycée français :

- Q_1 est la plus petite valeur telle qu'au moins 25 % des données lui soient inférieures ou égales ; son rang est $\lceil N/4 \rceil$.
- Q_3 est la plus petite valeur telle qu'au moins 75 % des données lui soient inférieures ou égales ; son rang est $\lceil 3N/4 \rceil$.

Le symbole $\lceil u \rceil$ signifie : le plus petit entier supérieur ou égal à u .

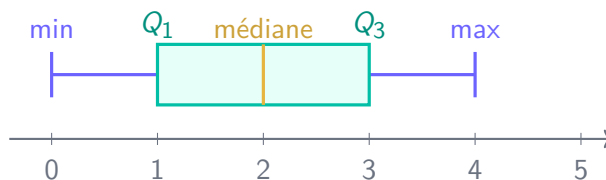
💡 Exemple | Quartiles de la série des livres

$N = 20$. Le rang de Q_1 est $20/4 = 5$ et celui de Q_3 est $3 \times 20/4 = 15$. Les effectifs cumulés 2, 7, 14, 18, 20 donnent

$$Q_1 = 1, \quad Q_3 = 3.$$

5.3 Boîte à moustaches

Une boîte à moustaches représente le minimum, Q_1 , la médiane, Q_3 et le maximum.



🧠 Intuition | Lire la boîte

La moitié centrale des données se trouve dans la boîte $[Q_1; Q_3]$. Une boîte courte indique une forte concentration centrale ; des moustaches longues signalent des valeurs éloignées.

6 ↔ Mesurer la dispersion

6.1 Étendue et écart interquartile

Définition | Deux indicateurs simples

$$\text{étendue} = \text{maximum} - \text{minimum},$$

$$\text{écart interquartile} = Q_3 - Q_1.$$

L'étendue utilise les deux valeurs extrêmes. L'écart interquartile mesure la largeur de la moitié centrale de la série.

Exemple | Série des livres

Le minimum est 0, le maximum 4, $Q_1 = 1$ et $Q_3 = 3$. Donc

$$\text{étendue} = 4, \quad Q_3 - Q_1 = 2.$$

Propriété | Résistance aux valeurs extrêmes

La médiane et l'écart interquartile sont peu sensibles à une valeur extrêmement grande ou petite. La moyenne et l'étendue y sont davantage sensibles. Le choix des indicateurs dépend donc de la forme des données.

6.2 Variance et écart type

Définition | Variance et écart type

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

$$V = \frac{1}{N} \sum_{i=1}^p n_i (x_i - \bar{x})^2.$$

L'écart type est

$$\sigma = \sqrt{V}.$$

Intuition | Pourquoi mettre les écarts au carré ?

La somme des écarts simples vaut toujours zéro. Les carrés empêchent les écarts positifs et négatifs de s'annuler et donnent plus de poids aux valeurs très éloignées. La racine carrée ramène ensuite l'écart type dans l'unité des données.

✂ Méthode | Calculer un écart type à la main

1. Calcule la moyenne $\bar{x}$.
2. Calcule chaque écart $x_i - \bar{x}$, puis son carré.
3. Pondère par n_i , additionne et divise par N : tu obtiens V .
4. Prends la racine carrée : $\sigma = \sqrt{V}$.

💡 Exemple | Écart type de la série des livres

La moyenne vaut 1,95. Un calcul direct donne

$$V = \frac{2(0 - 1,95)^2 + 5(1 - 1,95)^2 + 7(2 - 1,95)^2 + 4(3 - 1,95)^2 + 2(4 - 1,95)^2}{20} = 1,2475.$$

Ainsi

$$\sigma = \sqrt{1,2475} \approx 1,12 \text{ livre}.$$

⚠ Attention | Variance et écart type n'ont pas la même unité

Si les données sont en mètres, la variance est en mètres carrés, tandis que l'écart type est en mètres. Pour interpréter la dispersion avec l'unité d'origine, utilise l'écart type.

7 📊 Données regroupées en classes

7.1 Pourquoi regrouper ?

Pour une variable continue, les valeurs exactes peuvent être très nombreuses. On les regroupe en intervalles appelés **classes**. La classe $[10; 20[$ contient 10 mais pas 20.

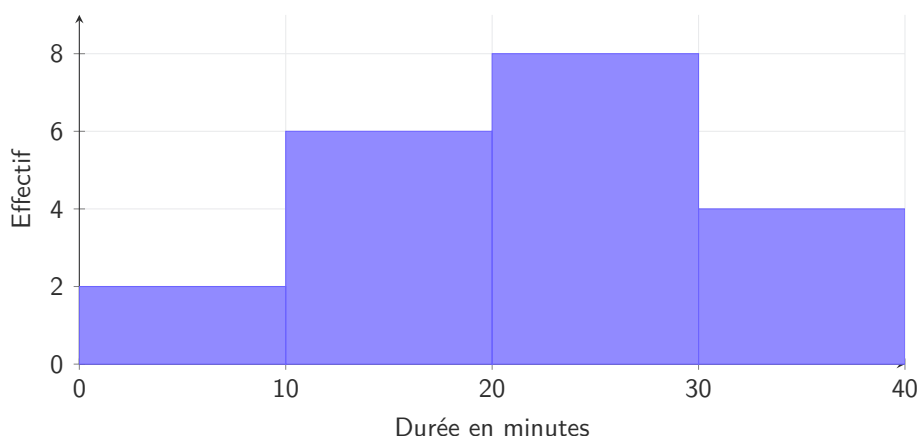
💡 Exemple | Durées de trajet de 20 élèves

Durée en minutes	$[0; 10[$	$[10; 20[$	$[20; 30[$	$[30; 40[$	Total
Effectif	2	6	8	4	20
Fréquence	0,10	0,30	0,40	0,20	1

7.2 Histogramme

📄 Définition | Histogramme pour des classes de même amplitude

Lorsque toutes les classes ont la même largeur, chaque classe est représentée par un rectangle dont la hauteur est proportionnelle à son effectif ou à sa fréquence. Les rectangles se touchent.



⚠ Attention | Si les amplitudes sont différentes

La hauteur ne peut plus être simplement l'effectif : c'est l'**aire** du rectangle qui doit être proportionnelle à l'effectif. Dans cette fiche, les calculs exigibles portent sur des classes de même amplitude.

7.3 Moyenne estimée avec les centres de classes

📖 Définition | Centre d'une classe

Le centre de $[a; b[$ est

$$c = \frac{a + b}{2}.$$

Si l'on suppose les valeurs uniformément réparties dans chaque classe, on les remplace par le centre pour estimer la moyenne.

🔧 Méthode | Moyenne d'une série groupée

Avec des centres c_i et des effectifs n_i ,

$$\bar{x} \approx \frac{n_1 c_1 + \dots + n_p c_p}{N}.$$

Le symbole $\approx$ rappelle qu'on a perdu l'information sur les valeurs exactes à l'intérieur des classes.

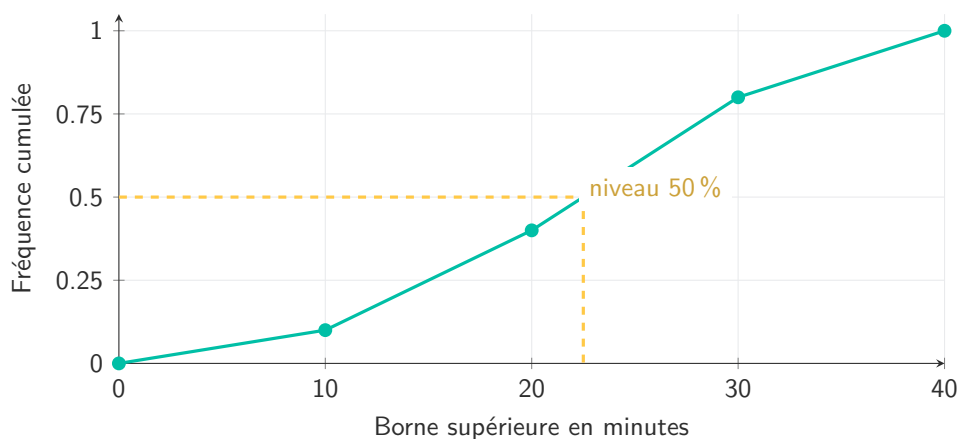
💡 Exemple | Temps de trajet moyen estimé

Les centres sont 5, 15, 25, 35. Ainsi

$$\bar{x} \approx \frac{2 \times 5 + 6 \times 15 + 8 \times 25 + 4 \times 35}{20} = \frac{440}{20} = \boxed{22 \text{ minutes}}.$$

7.4 Classe médiane et interpolation

Les fréquences cumulées de l'exemple sont 0,10, 0,40, 0,80, 1. La fréquence 0,50 est atteinte dans $[20; 30[$: c'est la **classe médiane**.



✓ Propriété | Estimation dans la classe médiane

Sous l'hypothèse d'une répartition uniforme dans la classe $[L; L + h[$ d'effectif n_c , si C valeurs la précèdent, on estime la médiane par

$$M \approx L + h \times \frac{N/2 - C}{n_c}.$$

Démonstration | Comprendre l'interpolation

Il faut avancer de $N/2 - C$ rangs à l'intérieur d'une classe contenant n_c valeurs. Cela représente la fraction $\frac{N/2 - C}{n_c}$ de la largeur h . On ajoute donc cette fraction de largeur à la borne gauche L .

Exemple | Estimer la médiane des trajets

Ici $L = 20$, $h = 10$, $C = 8$ et $n_c = 8$. Donc

$$M \approx 20 + 10 \times \frac{10 - 8}{8} = \boxed{22,5 \text{ minutes}}.$$

Ce nombre est une estimation fondée sur l'hypothèse d'uniformité dans la classe.

8 Comparer deux séries

Méthode | Deux couples d'indicateurs

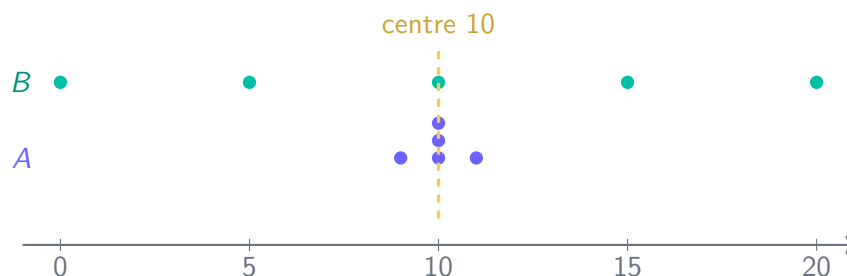
- **Moyenne et écart type** : ils utilisent toutes les valeurs et sont sensibles aux valeurs extrêmes.
- **Médiane et écart interquartile** : ils décrivent le centre et la moitié centrale, avec davantage de résistance aux valeurs extrêmes.

💡 Exemple | Même centre, dispersion différente

Les séries $A : 9, 10, 10, 10, 11$ et $B : 0, 5, 10, 15, 20$ ont moyenne et médiane égales à 10. Mais

$$\sigma_A = \sqrt{0,4} \approx 0,63, \quad \sigma_B = \sqrt{50} \approx 7,07.$$

La série A est beaucoup plus homogène.



⚠ Attention | Comparer ne signifie pas seulement classer

Une série peut avoir une moyenne plus haute mais aussi une dispersion beaucoup plus grande. Formule une phrase complète : *plus élevée en moyenne, plus homogène, moitié centrale plus resserrée*, etc.

9 </> Calculer avec Python

```

1 from math import sqrt
2
3 def indicateurs(valeurs):
4     n = len(valeurs)
5     moyenne = sum(valeurs) / n
6     variance = sum((x - moyenne)**2 for x in valeurs) / n
7     return moyenne, sqrt(variance)
8
9 serie = [0, 0, 1, 1, 1, 1, 1, 2, 2, 2,
10          2, 2, 2, 2, 3, 3, 3, 3, 4, 4]
11 print(indicateurs(serie)) # (1.95, environ 1.117)

```

⚠ Attention | Vérifier la convention de la calculatrice

Certaines machines proposent deux écarts types : celui de la population, qui divise par N , et celui d'un échantillon, qui divise par $N - 1$. Dans cette fiche et au programme de seconde, on utilise la formule divisée par N .

10 Pour aller plus loin - Accessible à tous

Intuition | Statut de cette partie

Le socle officiel est entièrement présenté dans les sections précédentes. Les résultats suivants vont plus loin, mais toutes les preuves utilisent seulement le calcul littéral étudié en seconde.

10.1 Une formule plus rapide pour la variance

★ Théorème | Moyenne des carrés moins carré de la moyenne

$$V = \frac{1}{N} \sum_{i=1}^p n_i x_i^2 - \bar{x}^2.$$

Démonstration | Développer les carrés

$$\begin{aligned} V &= \frac{1}{N} \sum_i n_i (x_i - \bar{x})^2 \\ &= \frac{1}{N} \sum_i n_i (x_i^2 - 2x_i \bar{x} + \bar{x}^2) \\ &= \frac{1}{N} \sum_i n_i x_i^2 - 2\bar{x} \underbrace{\frac{1}{N} \sum_i n_i x_i}_{\bar{x}} + \bar{x}^2 \underbrace{\frac{1}{N} \sum_i n_i}_1 \\ &= \frac{1}{N} \sum_i n_i x_i^2 - \bar{x}^2. \end{aligned}$$

10.2 Transformer l'écart type

★ Théorème | Effet de $x \mapsto ax + b$

Si toutes les données sont remplacées par $ax + b$, alors

$$\sigma_{\text{nouveau}} = |a|\sigma.$$

Ajouter b ne change pas la dispersion ; multiplier par a multiplie toutes les distances par $|a|$.

Démonstration | Transformer les écarts à la moyenne

La nouvelle moyenne vaut $a\bar{x} + b$. L'écart correspondant à x_i devient

$$(ax_i + b) - (a\bar{x} + b) = a(x_i - \bar{x}).$$

Son carré est multiplié par a^2 , donc la variance par a^2 . En prenant la racine carrée, l'écart type est multiplié par $\sqrt{a^2} = |a|$.

10.3 Pourquoi la moyenne est-elle un centre naturel ?

★ Théorème | La somme des carrés est minimale à la moyenne

Pour tout réel a ,

$$\sum_{j=1}^N (y_j - a)^2 = \sum_{j=1}^N (y_j - \bar{x})^2 + N(a - \bar{x})^2.$$

La somme est donc minimale lorsque $a = \bar{x}$.

Démonstration | Faire apparaître un carré positif

Écrivons $y_j - a = (y_j - \bar{x}) + (\bar{x} - a)$ et développons. En sommant, le terme croisé contient $\sum_j (y_j - \bar{x}) = 0$. Il reste

$$\sum_j (y_j - a)^2 = \sum_j (y_j - \bar{x})^2 + N(\bar{x} - a)^2.$$

Le dernier terme est positif ou nul, et il vaut zéro exactement pour $a = \bar{x}$.

10.4 Score centré-réduit

Définition | Comparer des positions dans des unités différentes

Si $\sigma > 0$, le score

$$z = \frac{x - \bar{x}}{\sigma}$$

indique à combien d'écarts types la valeur x se trouve de la moyenne. Un score positif est au-dessus de la moyenne ; un score négatif est en dessous.

Exemple | Comparer deux résultats

Un score de 14 dans une série de moyenne 10 et d'écart type 2 donne $z = 2$. Un score de 72 dans une série de moyenne 60 et d'écart type 8 donne $z = 1,5$. Relativement à son groupe, le premier résultat est plus éloigné au-dessus de la moyenne.

11 Boîte à outils et erreurs à éviter

Méthode | Fréquence

$$f_i = n_i / N \text{ et } \sum f_i = 1.$$

Méthode | Écart type

Moyenne, écarts, carrés, moyenne, racine.

Méthode | Moyenne

$$\bar{x} = \frac{\sum n_i x_i}{N}.$$

Méthode | Classes

Centre $(a + b)/2$; la moyenne est estimée.

Méthode | Rangs

Trie, puis utilise les effectifs cumulés.

Méthode | Comparer

Associe position et dispersion.

Méthode | Quartiles

Rangs $\lceil N/4 \rceil$ et $\lceil 3N/4 \rceil$.

Attention | Réflexes

Unité, tri, total des effectifs, somme des fréquences et arrondi annoncé.

Intuition | Le résumé le plus honnête

Donne un indicateur de centre, un indicateur de dispersion et un graphique. Puis explique en une phrase ce que chacun apporte.

12 Exercices progressifs

Intuition | Mode d'emploi

Les exercices sont classés par difficulté. Cherche d'abord sans regarder le corrigé, écris les étapes et contrôle toujours l'ordre de grandeur du résultat.

Exercice 1 : Vocabulaire statistique

On étudie le temps, en minutes, mis par chacun des 50 salariés d'une entreprise pour venir travailler.

Identifier la population, un individu, le caractère étudié, sa nature et une unité adaptée. Expliquer pourquoi le temps n'est pas, en théorie, un caractère discret.

Exercice 2 : Effectifs et fréquences

Le nombre de livres lus pendant un mois par 20 élèves est résumé ci-dessous.

Nombre de livres x_i	0	1	2	3	4
Effectif n_i	2	5	7	4	2

1. Vérifier l'effectif total.
2. Calculer les fréquences, puis les effectifs cumulés croissants.
3. Quelle proportion d'élèves a lu au moins deux livres ?

Exercice 3 : Calculer une moyenne

Avec les données de l'exercice 2, calculer le nombre moyen de livres lus. Interpréter le résultat sans prétendre qu'un élève a nécessairement lu ce nombre de livres.

Exercice 4 : Réunir deux groupes

Dans la classe A, 24 élèves ont obtenu une moyenne de 12,5. Dans la classe B, 36 élèves ont obtenu une moyenne de 14,5.

Calculer la moyenne des 60 élèves. Pourquoi la moyenne globale n'est-elle pas la moyenne simple de 12,5 et 14,5 ?

Exercice 5 : Retrouver une valeur

Neuf nombres ont pour moyenne 12. Huit d'entre eux sont

8, 9, 10, 11, 12, 13, 14, 15.

Déterminer le neuvième nombre et vérifier le résultat.

Exercice 6 : Médiane et quartiles

Reprendre la série de l'exercice 2. Déterminer la médiane, le premier quartile Q_1 , le troisième quartile Q_3 , l'étendue et l'écart interquartile. Justifier chaque réponse avec un rang ou un effectif cumulé.

Exercice 7 ★★☆☆ : Boîte à moustaches

On considère la série ordonnée

2, 3, 3, 4, 5, 6, 6, 6, 7, 8, 10, 12.

1. Déterminer le minimum, Q_1 , la médiane, Q_3 et le maximum.
2. Calculer l'étendue et l'écart interquartile.
3. Tracer la boîte à moustaches sur un axe gradué de 0 à 12.

Exercice 8 ★★☆☆ : Même moyenne, dispersions différentes

On compare les deux séries

$A : 9, 10, 10, 10, 11$ et $B : 0, 5, 10, 15, 20$.

Calculer, pour chacune, la moyenne, la variance et l'écart type. Interpréter la comparaison.

Exercice 9 ★★☆☆ : Changement d'unité

Une température X , exprimée en degrés Celsius, a pour moyenne 18 et pour écart type 3,5. On la convertit en degrés Fahrenheit par

$$Y = 1,8X + 32.$$

Déterminer la moyenne et l'écart type de Y . Expliquer pourquoi ajouter 32 ne modifie pas l'écart type.

Exercice 10 ★★☆☆ : Ajouter ou retirer une donnée

Une série de 30 valeurs a pour moyenne 14.

1. On ajoute la valeur 20. Calculer la nouvelle moyenne.
2. En repartant de la série initiale, on retire une valeur égale à 8. Calculer la nouvelle moyenne.
3. Dans chaque cas, expliquer le sens de la variation avant même de calculer.

Exercice 11 ★★☆☆ : Données regroupées en classes

Les durées, en minutes, de 20 trajets sont regroupées ainsi.

Durée	$[0; 10[$	$[10; 20[$	$[20; 30[$	$[30; 40[$
Effectif	2	6	8	4

1. Calculer les fréquences et les fréquences cumulées croissantes.
2. Tracer l'histogramme. Pourquoi peut-on prendre les effectifs comme hauteurs ici ?
3. Estimer la moyenne à l'aide des centres des classes.
4. Déterminer la classe médiane, puis estimer la médiane en supposant une répartition uniforme dans cette classe.

Exercice 12 ★★☆☆ : Comparer sans surinterpréter

Deux séries C et D ont les résumés suivants.

	Moyenne	Écart type	Médiane	Écart interquartile
C	50	4	49	5
D	53	12	50	6

Comparer le niveau central et la dispersion des deux séries. Donner ensuite une affirmation que ces seuls indicateurs ne permettent pas de justifier.

Exercice 13 ★★★ : Une moyenne pondérée trompeuse

Une usine possède deux lignes de production. La première fabrique 80 pièces de masse moyenne 102 g ; la seconde en fabrique 120 de masse moyenne 98 g.

Calculer la masse moyenne des 200 pièces. Un élève affirme : « La moyenne vaut 100 g, car 100 est au milieu de 98 et 102. » Expliquer précisément son erreur.

Exercice 14 ★★★ : Formule raccourcie de la variance

La série suivante est donnée par effectifs.

x_i	1	2	3
n_i	2	4	2

1. Calculer $\bar{x}$ et la moyenne des carrés $\frac{1}{N} \sum n_i x_i^2$.
2. Utiliser $V = \frac{1}{N} \sum n_i x_i^2 - \bar{x}^2$ pour calculer la variance et l'écart type.
3. Vérifier la variance avec la définition fondée sur les écarts à la moyenne.

Exercice 15 ★★★ : La meilleure valeur centrale

Pour les données 2, 5 et 8, on pose

$$S(a) = (2 - a)^2 + (5 - a)^2 + (8 - a)^2.$$

1. Calculer la moyenne de la série.
2. Montrer que $S(a) = 3(a - 5)^2 + 18$.
3. En déduire la valeur de a qui minimise $S(a)$ et interpréter ce résultat.

13 🦉 Problème défi : deux stations de mesure

🧠 Intuition | Objectif

Ce problème rassemble presque tout le chapitre. Il est long, mais chaque partie est guidée. Les résultats précédents peuvent être réutilisés : avance une question après l'autre.

Deux stations mesurent chaque jour une même grandeur. Durant une première période de 20 jours, on obtient les tableaux suivants.

Station A : valeur	10	12	14	16	18
Effectif	2	4	8	4	2

Station B : valeur	8	11	14	17	20
Effectif	2	4	8	4	2

Partie A — Étudier la station A

1. Vérifier que l'effectif total est 20.
2. Calculer les fréquences et les effectifs cumulés croissants.
3. Calculer la moyenne.
4. Déterminer la médiane.
5. Déterminer Q_1 , Q_3 et l'écart interquartile.
6. Calculer la variance et l'écart type, arrondi au centième.

Partie B — Comparer avec la station B

7. Calculer la moyenne de B.
8. Déterminer sa médiane, ses quartiles et son écart interquartile.
9. Calculer sa variance et son écart type, arrondi au centième.
10. Comparer les deux stations avec une phrase complète portant sur la position et la dispersion.

Partie C — Une seconde semaine, données regroupées

La station A fournit 20 nouvelles durées, regroupées en classes de même amplitude.

Durée	[0; 10[	[10; 20[	[20; 30[	[30; 40[
Effectif	2	6	8	4

11. Calculer les fréquences et les fréquences cumulées.
12. Tracer l'histogramme.
13. Estimer la moyenne à l'aide des centres des classes.
14. Identifier la classe médiane.
15. En supposant les valeurs uniformément réparties dans cette classe, estimer la médiane par interpolation linéaire.
16. Expliquer pourquoi les réponses aux questions 13 et 15 sont des estimations.

Partie D — Calibrer la station A

Après étalonnage, chaque valeur X de la première période est remplacée par

$$Y = 1,2X - 2.$$

17. Calculer la moyenne de Y .
18. Déterminer la médiane, Q_1 et Q_3 de Y .
19. Déterminer l'écart type de Y .
20. Déterminer sa variance.

Partie E — Regrouper les deux stations

On réunit les 20 mesures de A et les 20 mesures de B de la première période.

21. Calculer la moyenne des 40 mesures.
22. Les deux séries ayant la même moyenne, montrer que la variance de la série réunie est la moyenne de leurs variances. En déduire l'écart type global.
23. Classer A, B et la série réunie de la plus homogène à la plus dispersée. Justifier.

Attention | Avant de tourner la page

Vérifie que tes fréquences totalisent 1, que tes quartiles correspondent à des rangs, et que chaque comparaison mentionne à la fois le centre et la dispersion.

14 Corrigés détaillés

Corrigé 1

 *Démonstration / Traduire la situation*

- La **population** est l'ensemble des 50 salariés étudiés.
- Un **individu** est l'un de ces salariés.
- Le **caractère** est son temps de trajet domicile-travail.
- Il est **quantitatif continu** : en théorie, un temps peut prendre toutes les valeurs d'un intervalle, par exemple 18,4 minutes.
- Une unité adaptée est la minute, éventuellement la seconde si la mesure est plus précise.

Si les temps sont arrondis à la minute, le tableau n'affiche que des entiers ; cela ne change pas la nature théorique du caractère.

Corrigé 2

 *Démonstration / Du comptage aux fréquences*

L'effectif total est

$$N = 2 + 5 + 7 + 4 + 2 = 20.$$

On divise chaque effectif par 20.

x_i	0	1	2	3	4
n_i	2	5	7	4	2
f_i	0,10	0,25	0,35	0,20	0,10
Effectif cumulé	2	7	14	18	20

Contrôle : $0,10 + 0,25 + 0,35 + 0,20 + 0,10 = 1$.

Au moins deux livres signifie 2, 3 ou 4 livres. Il y a $7 + 4 + 2 = 13$ élèves, soit

$$\frac{13}{20} = 0,65 = \boxed{65\%}.$$

Corrigé 3

 *Démonstration / Une somme pondérée*

Il faut tenir compte du nombre d'élèves associé à chaque valeur :

$$\bar{x} = \frac{2 \times 0 + 5 \times 1 + 7 \times 2 + 4 \times 3 + 2 \times 4}{20} = \frac{39}{20} = \boxed{1,95}.$$

Les élèves ont donc lu en moyenne 1,95 livre durant le mois. Cette moyenne résume le groupe ; elle n'est pas obligatoirement une valeur réellement observée.

Corrigé 4

 *Démonstration / Pondérer par les effectifs*

La somme des notes de la classe A vaut $24 \times 12,5 = 300$. Celle de la classe B vaut $36 \times 14,5 = 522$. Ainsi,

$$\bar{x} = \frac{300 + 522}{24 + 36} = \frac{822}{60} = \boxed{13,7}.$$

La classe B compte davantage d'élèves : sa moyenne doit donc peser davantage. La moyenne simple $(12,5 + 14,5)/2 = 13,5$ ne conviendrait que si les deux classes avaient le même effectif.

Corrigé 5

 *Démonstration / Revenir à la somme*

Une moyenne de 12 pour 9 nombres correspond à une somme totale de

$$9 \times 12 = 108.$$

La somme des huit nombres connus est

$$8 + 9 + 10 + 11 + 12 + 13 + 14 + 15 = 92.$$

Le nombre manquant vaut donc

$$108 - 92 = \boxed{16}.$$

Vérification : $(92 + 16)/9 = 108/9 = 12$.

Corrigé 6

 *Démonstration / Lire des rangs dans les effectifs cumulés*

Il y a $N = 20$ valeurs.

— La médiane est la moyenne des valeurs de rangs 10 et 11. L'effectif cumulé atteint 14 pour la valeur 2 : ces deux rangs valent 2. Donc $\text{Med} = \boxed{2}$.

— Le rang de Q_1 est $\lceil 20/4 \rceil = 5$. Le 5^e terme vaut 1, donc $Q_1 = \boxed{1}$.

— Le rang de Q_3 est $\lceil 3 \times 20/4 \rceil = 15$. Le 15^e terme vaut 3, donc $Q_3 = \boxed{3}$.

Enfin,

$$\text{étendue} = 4 - 0 = \boxed{4}, \quad \text{écart interquartile} = 3 - 1 = \boxed{2}.$$

Corrigé 7

 *Démonstration / Construire les cinq nombres-clés*

La série contient $N = 12$ valeurs.

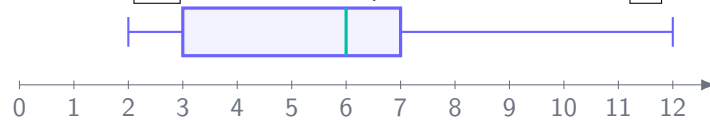
— minimum : 2 ; maximum : 12 ;

— Q_1 est la valeur de rang $\lceil 12/4 \rceil = 3$, donc $Q_1 = 3$;

— la médiane est la moyenne des 6^e et 7^e valeurs : $(6 + 6)/2 = 6$;

— Q_3 est la valeur de rang $\lceil 36/4 \rceil = 9$, donc $Q_3 = 7$.

Ainsi, l'étendue vaut $12 - 2 = 10$ et l'écart interquartile vaut $7 - 3 = 4$.



Corrigé 8

Démonstration / Le centre ne suffit pas

Les deux séries ont pour moyenne 10.

Pour A,

$$V_A = \frac{(9 - 10)^2 + 3(10 - 10)^2 + (11 - 10)^2}{5} = \frac{2}{5} = 0,4,$$

donc $\sigma_A = \sqrt{0,4} \approx 0,63$.

Pour B,

$$V_B = \frac{(-10)^2 + (-5)^2 + 0^2 + 5^2 + 10^2}{5} = \frac{250}{5} = 50,$$

donc $\sigma_B = \sqrt{50} \approx 7,07$.

Les centres sont identiques, mais A est beaucoup plus homogène : ses valeurs sont très proches de 10, contrairement à celles de B.

Corrigé 9

Démonstration / Appliquer les règles de transformation

Pour $Y = aX + b$,

$$\bar{Y} = a\bar{X} + b \quad \text{et} \quad \sigma_Y = |a|\sigma_X.$$

Ici,

$$\bar{Y} = 1,8 \times 18 + 32 = 64,4 \text{ } ^\circ\text{F},$$

et

$$\sigma_Y = 1,8 \times 3,5 = 6,3 \text{ } ^\circ\text{F}.$$

Ajouter 32 décale toutes les températures de la même quantité : les écarts entre les valeurs ne changent pas. Seule la multiplication par 1,8 agrandit les écarts.

Corrigé 10

Démonstration / Conserver la somme totale

La somme initiale est $30 \times 14 = 420$.

1. Après ajout de 20,

$$\bar{x}_{\text{nouvelle}} = \frac{420 + 20}{31} = \frac{440}{31} \approx 14,19.$$

Comme 20 est supérieur à l'ancienne moyenne, la moyenne augmente.

2. Après retrait de 8 à la série initiale,

$$\bar{x}_{\text{nouvelle}} = \frac{420 - 8}{29} = \frac{412}{29} \approx \boxed{14,21}.$$

Retirer une valeur inférieure à la moyenne fait aussi augmenter la moyenne.

Corrigé 11

 *Démonstration / Des classes aux estimations*

Les fréquences sont 0,10, 0,30, 0,40 et 0,20. Leurs valeurs cumulées sont

$$0,10, \quad 0,40, \quad 0,80, \quad 1.$$

Toutes les classes ont la même amplitude, 10 minutes : on peut donc prendre les effectifs comme hauteurs de l'histogramme. Sinon, il faudrait comparer des aires et utiliser des densités.

Les centres sont 5, 15, 25 et 35. La moyenne estimée vaut

$$\bar{x} \approx \frac{2 \times 5 + 6 \times 15 + 8 \times 25 + 4 \times 35}{20} = \frac{440}{20} = \boxed{22 \text{ min}}.$$

Le rang médian se situe autour de 10. Or 8 valeurs sont inférieures à 20 et 16 sont inférieures à 30 : la classe médiane est $\boxed{[20; 30[}$.

Il faut parcourir $10 - 8 = 2$ positions parmi les 8 valeurs supposées uniformément réparties dans cette classe, soit $2/8$ de sa largeur :

$$20 + \frac{2}{8} \times 10 = \boxed{22,5 \text{ min}}.$$

Ces deux résultats sont des estimations, car les valeurs exactes à l'intérieur des classes sont inconnues.

Corrigé 12

 *Démonstration / Séparer position et dispersion*

La série D a la moyenne la plus élevée : 53 contre 50. Cependant, les médianes sont proches : 50 contre 49. Cela suggère que le centre des données ordinaires est voisin, tandis que certaines valeurs de D peuvent tirer sa moyenne vers le haut.

La série C est nettement plus homogène selon l'écart type : 4 contre 12. Les écarts interquartiles, 5 et 6, sont beaucoup plus proches ; la moitié centrale des deux séries a donc une dispersion comparable. Avec ces seuls indicateurs, on ne peut pas affirmer, par exemple, que « toutes les valeurs de C sont inférieures à toutes celles de D », ni connaître la forme exacte des distributions.

Corrigé 13

 *Démonstration / Les deux lignes n'ont pas le même poids*

La masse totale produite est

$$80 \times 102 + 120 \times 98 = 8160 + 11760 = 19920 \text{ g.}$$

Donc

$$\bar{x} = \frac{19920}{200} = \boxed{99,6 \text{ g}}.$$

L'erreur consiste à accorder le même poids aux deux moyennes alors que la seconde ligne produit 120 pièces et la première seulement 80. La moyenne globale est donc plus proche de 98 g que de 102 g.

Corrigé 14

 *Démonstration / Deux chemins vers la même variance*

L'effectif total est $N = 8$. On trouve

$$\begin{aligned}\bar{x} &= \frac{2 \times 1 + 4 \times 2 + 2 \times 3}{8} = \frac{16}{8} = 2, \\ \frac{1}{N} \sum n_i x_i^2 &= \frac{2 \times 1^2 + 4 \times 2^2 + 2 \times 3^2}{8} = \frac{36}{8} = 4,5.\end{aligned}$$

La formule raccourcie donne

$$V = 4,5 - 2^2 = \boxed{0,5}, \quad \sigma = \sqrt{0,5} \approx \boxed{0,71}.$$

Avec la définition,

$$V = \frac{2(1-2)^2 + 4(2-2)^2 + 2(3-2)^2}{8} = \frac{4}{8} = 0,5.$$

Les deux méthodes concordent.

Corrigé 15

 *Démonstration / La moyenne minimise les carrés des écarts*

La moyenne vaut $\bar{x} = (2 + 5 + 8)/3 = 5$. Développons :

$$\begin{aligned}S(a) &= (a-2)^2 + (a-5)^2 + (a-8)^2 \\ &= 3a^2 - 30a + 93 \\ &= \boxed{3(a-5)^2 + 18}.\end{aligned}$$

Un carré étant toujours positif ou nul, $S(a)$ est minimale lorsque $(a-5)^2 = 0$, donc pour $\boxed{a=5}$. La valeur minimale est 18. Dans cette série, comme en général, la moyenne est le nombre qui minimise la somme des carrés des écarts aux données.

15 Corrigé du problème défi

Partie A — Station A

 *Démonstration* / Questions 1 à 6

1. L'effectif total est $2 + 4 + 8 + 4 + 2 = \boxed{20}$.

2. On obtient :

Valeur	10	12	14	16	18
Fréquence	0,10	0,20	0,40	0,20	0,10
Effectif cumulé	2	6	14	18	20

3. La distribution est symétrique autour de 14. Le calcul confirme :

$$\bar{x}_A = \frac{2 \times 10 + 4 \times 12 + 8 \times 14 + 4 \times 16 + 2 \times 18}{20} = \frac{280}{20} = \boxed{14}.$$

4. Les rangs 10 et 11 appartiennent au groupe de valeur 14, donc $\text{Med}_A = \boxed{14}$.

5. Le rang de Q_1 est 5 : $Q_{1,A} = 12$. Le rang de Q_3 est 15 : $Q_{3,A} = 16$. Ainsi,

$$IQR_A = 16 - 12 = \boxed{4}.$$

6. En exploitant les écarts à 14,

$$V_A = \frac{2(-4)^2 + 4(-2)^2 + 8(0)^2 + 4(2)^2 + 2(4)^2}{20} = \frac{96}{20} = \boxed{4,8}.$$

Donc

$$\sigma_A = \sqrt{4,8} \approx \boxed{2,19}.$$

Partie B — Station B

 *Démonstration* / Questions 7 à 10

7. La distribution est elle aussi symétrique autour de 14 :

$$\bar{x}_B = \frac{2 \times 8 + 4 \times 11 + 8 \times 14 + 4 \times 17 + 2 \times 20}{20} = \frac{280}{20} = \boxed{14}.$$

8. Les rangs 10 et 11 valent 14, donc $\text{Med}_B = 14$. Les rangs 5 et 15 valent respectivement 11 et 17 :

$$Q_{1,B} = 11, \quad Q_{3,B} = 17, \quad IQR_B = \boxed{6}.$$

9. Avec les écarts $-6, -3, 0, 3$ et 6 ,

$$V_B = \frac{2 \times 36 + 4 \times 9 + 8 \times 0 + 4 \times 9 + 2 \times 36}{20} = \frac{216}{20} = \boxed{10,8},$$

d'où

$$\sigma_B = \sqrt{10,8} \approx \boxed{3,29}.$$

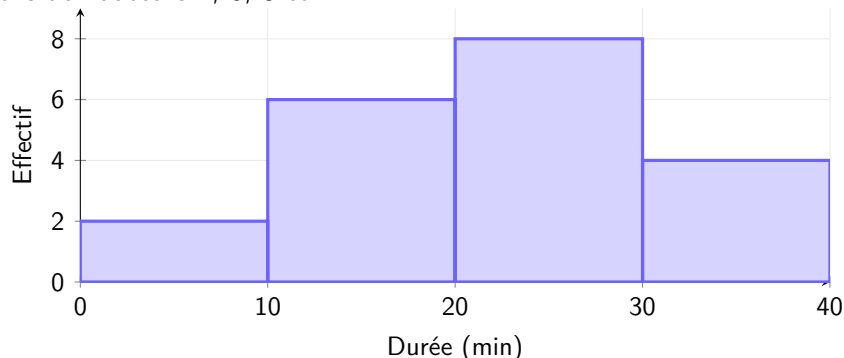
10. Les deux stations ont le même centre : moyenne et médiane égales à 14. En revanche, A est plus homogène, car son écart interquartile et son écart type sont plus petits : $4 < 6$ et $2,19 < 3,29$.

Partie C — Données regroupées

 Démonstration / Questions 11 à 16

11. Les fréquences sont 0,10, 0,30, 0,40 et 0,20 ; les fréquences cumulées sont 0,10, 0,40, 0,80 et 1.

12. Les classes ont toutes une largeur de 10 minutes. Un histogramme correct comporte donc quatre rectangles jointifs de hauteurs 2, 6, 8 et 4.



13. Les centres 5, 15, 25 et 35 donnent

$$\bar{x} \approx \frac{2 \times 5 + 6 \times 15 + 8 \times 25 + 4 \times 35}{20} = \boxed{22 \text{ min}}.$$

14. Le seuil médian est au rang 10. Il est atteint dans $\boxed{[20; 30[}$, car l'effectif cumulé passe de 8 à 16.

15. Il reste $10 - 8 = 2$ valeurs à parcourir parmi les 8 de cette classe. Sous l'hypothèse uniforme,

$$m \approx 20 + \frac{2}{8} \times (30 - 20) = \boxed{22,5 \text{ min}}.$$

16. Les positions exactes dans chaque classe sont inconnues. Remplacer chaque classe par son centre et supposer une répartition uniforme sont deux modèles utiles, mais ils ne reconstituent pas les données initiales.

Partie D — Calibrage affine

 Démonstration / Questions 17 à 20

La fonction $x \mapsto 1,2x - 2$ est croissante : elle conserve donc l'ordre des données.

17.

$$\bar{Y} = 1,2\bar{X} - 2 = 1,2 \times 14 - 2 = \boxed{14,8}.$$

18.

$$\text{Med}_Y = 1,2 \times 14 - 2 = \boxed{14,8},$$

$$Q_{1,Y} = 1,2 \times 12 - 2 = \boxed{12,4}, \quad Q_{3,Y} = 1,2 \times 16 - 2 = \boxed{17,2}.$$

19. Une translation ne change pas l'écart type et une multiplication par 1,2 le multiplie par 1,2 :

$$\sigma_Y = 1,2\sqrt{4,8} \approx \boxed{2,63}.$$

20.

$$V_Y = 1,2^2 V_X = 1,44 \times 4,8 = \boxed{6,912}.$$

On peut contrôler que $\sqrt{6,912} \approx 2,63$.

Partie E — Série réunie

 *Démonstration / Questions 21 à 23*

21. Les groupes ont chacun 20 valeurs et une moyenne égale à 14 :

$$\bar{x}_{A \cup B} = \frac{20 \times 14 + 20 \times 14}{40} = \boxed{14}.$$

22. Tous les écarts sont mesurés à la même moyenne 14. La somme des carrés des écarts de A vaut $20V_A$ et celle de B vaut $20V_B$. Ainsi,

$$V_{A \cup B} = \frac{20V_A + 20V_B}{40} = \frac{V_A + V_B}{2} = \frac{4,8 + 10,8}{2} = \boxed{7,8}.$$

Donc

$$\sigma_{A \cup B} = \sqrt{7,8} \approx \boxed{2,79}.$$

23. Les écarts types vérifient

$$2,19 < 2,79 < 3,29.$$

Le classement de la plus homogène à la plus dispersée est donc

$$\boxed{\text{station A} < \text{série réunie} < \text{station B}}.$$

La variance globale est bien intermédiaire, car elle est ici la moyenne des deux variances. Cette conclusion utilise le fait essentiel que les deux groupes ont la même moyenne.

16 Fiche-mémoire

Décrire une série = centre + dispersion + représentation

 Définition | Effectif et fréquence

$$N = \sum n_i, \quad f_i = \frac{n_i}{N}, \quad \sum f_i = 1.$$

 Définition | Moyenne

$$\bar{x} = \frac{\sum n_i x_i}{N} = \sum f_i x_i.$$

Elle utilise toutes les valeurs, mais elle est sensible aux valeurs extrêmes.

 Définition | Médiane et quartiles

Après avoir trié :

- médiane : milieu de la série ;
- Q_1 : rang $\lceil N/4 \rceil$;
- Q_3 : rang $\lceil 3N/4 \rceil$.

 Méthode | Données en classes

Centre de $[a; b]$: $(a + b)/2$.

$$\bar{x} \approx \frac{\sum n_i c_i}{N}.$$

Le résultat est une estimation.

 Définition | Dispersion

$$\text{étendue} = \max - \min, \quad IQR = Q_3 - Q_1.$$

$$V = \frac{1}{N} \sum n_i (x_i - \bar{x})^2, \quad \sigma = \sqrt{V}.$$

 Méthode | Formule raccourcie

$$V = \frac{1}{N} \sum n_i x_i^2 - \bar{x}^2.$$

Ne pas oublier le carré de la moyenne.

 Méthode | Transformation affine

$$\text{Si } Y = aX + b,$$

$$\bar{Y} = a\bar{X} + b, \quad \sigma_Y = |a|\sigma_X,$$

$$V_Y = a^2 V_X.$$

 Méthode | Comparer deux séries

1. comparer moyenne ou médiane ;
2. comparer écart type ou IQR ;
3. regarder le graphique ;
4. conclure avec une phrase nuancée.

 Attention | Les quatre pièges

Ne pas oublier de trier pour la médiane ; ne pas moyenner deux moyennes sans les effectifs ; ne pas confondre variance et écart type ; ne pas présenter comme exacte une moyenne calculée avec des centres de classes.

Idée à retenir : un indicateur résume, mais il ne raconte jamais toute l'histoire. Associe toujours un centre, une dispersion et une représentation.