

Statistiques en Seconde

DURÉE ESTIMÉE

30 minutes

OBJECTIFS DU CHAPITRE

Calculer une moyenne pondérée

Utiliser la linéarité de la moyenne

Calculer la moyenne d'une série regroupée en classes

Déterminer la médiane d'une série statistique

Calculer les quartiles et l'écart interquartile

Calculer la variance et l'écart type à la calculatrice

1 - Organisation des données

Définition

- Les statistiques permettent d'étudier un **caractère** d'une **population**.
- Le nombre d'éléments de la population s'appelle l'**effectif total** (noté N).
- Pour une valeur de caractère donnée, l'**effectif** n_i est le nombre d'éléments correspondant à cette valeur.
- La **fréquence** d'une valeur est le quotient de son effectif par l'effectif total : $f_i = \frac{n_i}{N}$. La somme des fréquences est égale à 1.
- Une **série statistique** est un tableau donnant les effectifs (ou les fréquences) pour chacune des valeurs du caractère.

Exemple

On étudie l'âge des élèves d'un lycée. Le **caractère** est l'âge, la **population** est l'ensemble des élèves.

Âge (en années)	14	15	16	17	18	19	20
Effectif	3	22	65	82	59	35	2

L'effectif total est $N = 3 + 22 + 65 + 82 + 59 + 35 + 2 = 268$.

La fréquence de l'âge 17 ans est $f = \frac{82}{268} \approx 0,306$, soit environ 30,6 %.

Effectifs cumulés croissants

L'**effectif cumulé croissant** (ECC) d'une valeur est la somme des effectifs de toutes les valeurs inférieures ou égales à cette valeur.

Exemple

En reprenant l'exemple précédent :

Âge	14	15	16	17	18	19	20
Effectif	3	22	65	82	59	35	2
ECC	3	25	90	172	231	266	268

Par exemple, 90 élèves ont 16 ans ou moins.

2 - Indicateurs de tendance centrale

Moyenne pondérée

La **moyenne pondérée** d'une série statistique dont les valeurs sont $x_1, x_2, \dots, x_p$ avec les effectifs $n_1, n_2, \dots, n_p$ est :

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

où $N = n_1 + n_2 + \dots + n_p$ est l'effectif total.

Remarque

Pour une série regroupée en classes, on obtient une valeur approchée de la moyenne en remplaçant chaque classe par son centre.

Exemple

Les notes à un contrôle dans une classe de 21 élèves sont :

5 ; 14 ; 13 ; 16 ; 9 ; 8 ; 18 ; 2 ; 13 ; 12 ; 15 ; 12 ; 8 ; 6 ; 5 ; 17 ; 3 ; 19 ; 9 ; 13 ; 14

On les ordonne et on les regroupe dans un tableau :

Note	2	3	5	6	8	9	12	13	14	15	16	17	18	19
Effectif	1	1	2	1	2	2	2	3	2	1	1	1	1	1

La moyenne est :

$$\bar{x} = \frac{1 \times 2 + 1 \times 3 + 2 \times 5 + \dots + 1 \times 18 + 1 \times 19}{21} = \frac{231}{21} = 11$$

La moyenne des notes est 11.

Linéarité de la moyenne

Si une série (x_i) a pour moyenne $\bar{x}$, alors la série $(ax_i + b)$ (où a et b sont des nombres réels) a pour moyenne $a\bar{x} + b$.

Démonstration

Calculons la moyenne de la série $(ax_i + b)$:

$$\begin{aligned} & \frac{n_1(ax_1 + b) + \cdots + n_p(ax_p + b)}{N} \\ &= \frac{a(n_1x_1 + \cdots + n_px_p) + bN}{N} \\ &= a \times \frac{n_1x_1 + \cdots + n_px_p}{N} + b = a\bar{x} + b \end{aligned}$$

Exemple

Si les 21 notes précédentes sont augmentées de 2 points, la nouvelle moyenne est :

$$\bar{x}' = \bar{x} + 2 = 11 + 2 = 13$$

Il est inutile de recalculer toute la moyenne.

Médiane

La **médiane** d'une série statistique est la valeur du caractère qui partage la population ordonnée en deux groupes de même effectif.

Déterminer la médiane

On ordonne les valeurs du caractère dans l'ordre croissant.

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

Lorsque l'effectif est élevé, on utilise les effectifs cumulés croissants pour localiser la médiane.

Exemple

Reprenons les 21 notes ordonnées :

2 ; 3 ; 5 ; 5 ; 6 ; 8 ; 8 ; 9 ; 9 ; 12 ; **12** ; 13 ; 13 ; 13 ; 14 ; 14 ; 15 ; 16 ; 17 ; 18 ; 19

$N = 21$ (impair), la médiane est la valeur de rang $\frac{21 + 1}{2} = 11$.

La 11e note est 12 : il y a 10 notes en dessous et 10 notes au-dessus.

La médiane est 12.

Supposons qu'il n'y ait que 20 élèves (on retire l'élève qui a eu 2) :

3 ; 5 ; 5 ; 6 ; 8 ; 8 ; 9 ; 9 ; 12 ; **12** ; **13** ; 13 ; 13 ; 14 ; 14 ; 15 ; 16 ; 17 ; 18 ; 19

$N = 20$ (pair), la médiane est la moyenne des valeurs de rang $\frac{20}{2} = 10$ et $\frac{20}{2} + 1 = 11$.

La 10e note est 12 et la 11e est 13.

La médiane est $\frac{12 + 13}{2} = 12,5$.

La médiane est 12,5.

3 - Indicateurs de dispersion

Étendue

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

Quartiles

Les valeurs de la série étant rangées par ordre croissant :

- Le **premier quartile** Q_1 est la plus petite valeur de la série telle qu'au moins 25 % des valeurs lui soient inférieures ou égales.
- Le **troisième quartile** Q_3 est la plus petite valeur de la série telle qu'au moins 75 % des valeurs lui soient inférieures ou égales.

Exemple

Reprenons les 21 notes ordonnées.

Pour Q_1 : il faut qu'au moins $\frac{1}{4} \times 21 = 5,25$ notes soient inférieures ou égales. Le premier quartile est la 6e note.

2 ; 3 ; 5 ; 5 ; 6 ; 8 ; 8 ; 9 ; 9 ; 12 ; 12 ; 13 ; 13 ; 13 ; 14 ; 14 ; 15 ; 16 ; 17 ; 18 ; 19

Le premier quartile est $Q_1 = 8$.

Pour Q_3 : il faut qu'au moins $\frac{3}{4} \times 21 = 15,75$ notes soient inférieures ou égales. Le troisième quartile est la 16e note.

2 ; 3 ; 5 ; 5 ; 6 ; 8 ; 8 ; 9 ; 9 ; 12 ; 12 ; 13 ; 13 ; 13 ; 14 ; 14 ; 15 ; 16 ; 17 ; 18 ; 19

Le troisième quartile est $Q_3 = 14$.

Écart interquartile

L'**écart interquartile** est la différence $Q_3 - Q_1$.

L'**intervalle interquartile** est l'intervalle $[Q_1 ; Q_3]$.

Au moins 50 % des valeurs de la série appartiennent à l'intervalle interquartile.

Exemple

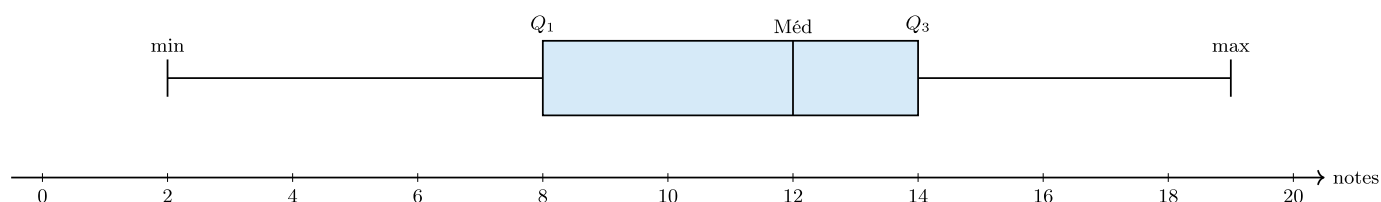
Dans l'exemple précédent, l'écart interquartile vaut $Q_3 - Q_1 = 14 - 8 = 6$.

L'intervalle interquartile est $[8 ; 14]$: au moins la moitié des notes se situent entre 8 et 14.

Diagramme en boîte

Le **diagramme en boîte** (ou **boîte à moustaches**) résume une série à partir de cinq valeurs : le minimum, le premier quartile Q_1 , la médiane, le troisième quartile Q_3 et le maximum.

La **boîte** s'étend de Q_1 à Q_3 et est partagée par un trait à la médiane : elle contient au moins la moitié des valeurs. Les **moustaches** relient la boîte au minimum et au maximum de la série.



Exemple

Pour les 21 notes précédentes : minimum = 2, $Q_1 = 8$, médiane = 12, $Q_3 = 14$, maximum = 19. Le diagramme en boîte ci-dessus représente cette répartition.

Variance et écart type

La **variance** d'une série statistique de moyenne $\bar{x}$ est le nombre positif :

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

L'**écart type** est le nombre positif $s = \sqrt{V}$.

Remarque

- L'écart type mesure la dispersion des valeurs autour de la moyenne : plus il est grand, plus les valeurs sont éloignées de la moyenne.
- L'écart type est exprimé dans la **même unité** que le caractère étudié.
- En pratique, le calcul de l'écart type se fait à la calculatrice.

Exemple

Un professeur a relevé le nombre de fautes par dictée pour 20 élèves :

Nombre de fautes	0	1	2	3	4
Effectif	3	6	5	4	2

La moyenne est :

$$\bar{x} = \frac{3 \times 0 + 6 \times 1 + 5 \times 2 + 4 \times 3 + 2 \times 4}{20} = \frac{36}{20} = 1,8$$

La variance est :

$$V = \frac{3(0 - 1,8)^2 + 6(1 - 1,8)^2 + \dots + 2(4 - 1,8)^2}{20}$$

$$V = \frac{3 \times 3,24 + 6 \times 0,64 + 5 \times 0,04 + 4 \times 1,44 + 2 \times 4,84}{20} = \frac{29,2}{20} = 1,46$$

L'écart type est $s = \sqrt{1,46} \approx 1,21$.

En moyenne, le nombre de fautes s'écarte d'environ 1,21 faute de la moyenne.

Intervalle à deux écarts types

Pour de nombreuses séries statistiques, la proportion de valeurs appartenant à l'intervalle $[\bar{x} - 2s; \bar{x} + 2s]$ est d'environ 95 %.

1. Comment calculer une moyenne pondérée ?

On multiplie chaque valeur par son effectif (ou son coefficient), on additionne les produits obtenus, puis on divise par l'effectif total.

Voir la fiche méthode : [Calculer une moyenne pondérée](#)

2. Comment trouver la nouvelle moyenne après une transformation affine ?

La linéarité de la moyenne permet de l'obtenir directement : si la transformation est $x \mapsto ax + b$, la nouvelle moyenne vaut $a\bar{x} + b$, sans recalculer la série entière.

Voir la fiche méthode : [Utiliser la linéarité de la moyenne](#)

3. Comment calculer la moyenne d'une série regroupée en classes ?

On remplace chaque classe par son centre (moyenne des deux bornes), puis on calcule la moyenne pondérée avec les effectifs de chaque classe. On obtient une valeur approchée de la moyenne réelle.

Voir la fiche méthode : [Calculer la moyenne d'une série regroupée en classes](#)

4. Comment déterminer la médiane d'une série statistique ?

Les valeurs étant rangées dans l'ordre croissant, la médiane est la valeur de rang $\frac{N+1}{2}$ si N est impair, ou la moyenne des valeurs de rang $\frac{N}{2}$ et $\frac{N}{2} + 1$ si N est pair. Pour un grand effectif, on utilise les ECC.

Voir la fiche méthode : [Déterminer la médiane d'une série statistique](#)

5. Comment calculer les quartiles et l'écart interquartile ?

Le rang de Q_1 est le plus petit entier supérieur ou égal à $\frac{N}{4}$, celui de Q_3 est le plus petit entier supérieur ou égal à $\frac{3N}{4}$. L'écart interquartile est $Q_3 - Q_1$.

Voir la fiche méthode : [Calculer les quartiles et l'écart interquartile](#)

6. Comment calculer la variance et l'écart type à la calculatrice ?

On entre les valeurs et les effectifs dans deux listes, puis on utilise le mode statistique à une variable (1-VAR). La calculatrice affiche $\bar{x}$ et σx ; la variance est $V = \sigma x^2$.

Voir la fiche méthode : [Calculer la variance et l'écart type à la calculatrice](#)

↓ Télécharger en PDF