STATISTIQUES

— SECONDE —

1. VOCABULAIRE, EFFECTIFS ET FRÉQUENCES

- la **population** est l'ensemble étudié, ses éléments sont les **individus** ;
- le **caractère** étudié est **quantitatif** (nombre de frères, taille…) ou **qualitatif** (couleur, filière…) ;
- l'**effectif** n_i d'une valeur est le nombre d'individus qui la prennent ; $N = \sum n_i$ est l'effectif total.

$$f_i = \frac{n_i}{N} \qquad 0 \leq f_i \leq 1 \qquad \sum f_i = 1$$

Série d'étude : notes obtenues par 10 élèves.

valeur	4	5	5	7	8	9	12	15	20	22

La série est **rangée dans l'ordre croissant** : c'est le préalable à tout calcul de médiane ou de quartile.

2. INDICATEURS DE TENDANCE CENTRALE

Moyenne $\bar{x} = \dfrac{n_1x_1 + n_2x_2 + \cdots + n_px_p}{N} = \sum f_i x_i$

Médiane M_e : valeur qui partage la série ordonnée en **deux groupes de même effectif**.
N impair : la valeur de rang $\frac{N+1}{2}$.
N pair : la moyenne des valeurs de rangs $\frac{N}{2}$ et $\frac{N}{2}+1$.

Sur la série

$\bar{x} = \dfrac{107}{10} = 10{,}7 \qquad M_e = \dfrac{8+9}{2} = 8{,}5$

Linéarité : si $y_i = ax_i + b$ alors $\bar{y} = a\bar{x} + b$.

3. INDICATEURS DE DISPERSION

Étendue = valeur maximale − valeur minimale

Premier quartile Q_1 : plus petite valeur telle qu'au moins 25 % des données lui soient inférieures ou égales.
Troisième quartile Q_3 : de même avec 75 %.

Écart interquartile $= Q_3 - Q_1$: dispersion de la **moitié centrale** de la série.

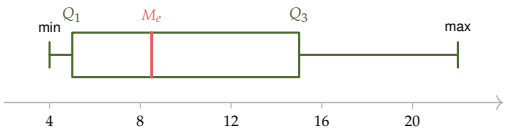
Écart-type σ : dispersion **autour de la moyenne** (à la calculatrice). Plus σ est petit, plus les valeurs sont resserrées autour de $\bar{x}$.

4. QUARTILES ET DIAGRAMME EN BOÎTE

Méthode de calcul

1. Ordonner la série et compter N.
2. Calculer $\dfrac{N}{4}$: si le résultat n'est pas entier, on **arrondit au rang supérieur** ; s'il est entier, on prend ce rang.
3. Q_1 est la valeur de ce rang ; on procède de même avec $\dfrac{3N}{4}$ pour Q_3.

Ici $N = 10$: $\frac{N}{4} = 2{,}5 \to$ rang 3, donc $Q_1 = 5$; $\frac{3N}{4} = 7{,}5 \to$ rang 8, donc $Q_3 = 15$.



La **boîte** contient 50 % des données, entre Q_1 et Q_3 ; les **moustaches** vont jusqu'aux valeurs extrêmes. Deux séries se comparent d'un coup d'œil.

5. DEUX VARIABLES QUALITATIVES

	Externes	D.-P.	Total
Filles	30	70	100
Garçons	50	50	100
Total	80	120	200

- **fréquence marginale** : par rapport au total général, $\frac{100}{200} = 50\,\%$ de filles ;
- **fréquence conditionnelle** : dans une ligne ou une colonne, $\frac{70}{100} = 70\,\%$ des filles sont demi-pensionnaires ;
- mais $\frac{70}{120} \approx 58\,\%$ des demi-pensionnaires sont des filles : les deux lectures diffèrent.

6. ERREURS À ÉVITER

La médiane n'est **pas** la moyenne : ici $M_e = 8{,}5$ alors que $\bar{x} = 10{,}7$.

Toujours **ordonner** la série avant de chercher médiane et quartiles.

Quartiles et médiane sont des **valeurs** du caractère, pas des rangs ni des effectifs.

Une moyenne de moyennes n'est pas la moyenne générale : il faut **pondérer** par les effectifs.

À RETENIR

Moyenne pondérée : $\bar{x} = \sum f_i x_i$; linéarité $\overline{ax+b} = a\bar{x} + b$.

Ordonner la série avant médiane et quartiles.

$Q_3 - Q_1$ mesure la dispersion des 50 % centraux.

Tableau croisé : distinguer fréquence marginale et conditionnelle.

pierre-carree.fr