Statistiques descriptives

Définition. La moyenne d'une série statistique de N valeurs $x_1; x_2; \dots; x_N$ est $m = \frac{x_1+x_2+\dots+x_N}{N}$

Définition. La moyenne pondérée d'une série statistique de N valeurs $x_1; x_2; \dots; x_N$ d'effectifs respectifs $n_1; n_2; \dots; n_N$ est

$$m = \frac{n_1x_1 + n_2x_2 + \dots + n_Nx_N}{n_1 + n_2 + \dots + n_N}$$

Définition. La moyenne pondérée d'une série statistique de N valeurs $x_1; x_2; \dots; x_N$ de poids respectifs $c_1; c_2; \dots; c_N$ est

$$m = \frac{c_1x_1 + c_2x_2 + \dots + c_Nx_N}{c_1 + c_2 + \dots + c_N}$$

Exemple. Ce trimestre, Emilia a eu quatre contrôles de mathématiques (notés sur 20). Elle a obtenu les notes suivantes :

Note	8	9	20	5
Coefficient	1	1,5	4	0,5

Sa moyenne en mathématique est $m = \frac{1\times 8+1,5\times 9+4\times 20+0,5\times 5}{1+1,5+4+0,5} \approx 14,9$

Propriété. Linéarité de la moyenne

Soit a et b deux nombres réels et $x_1, x_2, \dots, x_N$ une série statistique de moyenne m.

• Si on multiplie par a toutes les valeurs de la série, on obtient la moyenne de la nouvelle série en multipliant par a la moyenne de la série de départ. La moyenne de la série $ax_1; ax_2; \dots; ax_N$ est am.

• Si on ajoute b à toutes les valeurs de la série, on obtient la moyenne de la nouvelle série en ajoutant b à la moyenne de la série de départ. La moyenne de la série $x_1 + b; x_2 + b; \dots; x_N + b$ est $m + b$.

Définition. L'écart-type d'une série statistique de N valeurs $x_1; x_2; \dots; x_N$ d'effectifs (ou de poids) respectifs $n_1; n_2; \dots; n_N$ est $s = \sqrt{\frac{n_1(x_1-m)^2+n_2(x_2-m)^2+\dots+n_N(x_N-m)^2}{m}}$

Remarque. L'écart-type d'une série statistique est un indicateur de dispersion de cette série statistique autour de la moyenne. Plus l'écart-type s d'une série est petit, plus les valeurs de la série sont concentrées autour de la moyenne, donc plus la série est homogène.

Définition. La médiane d'une série statistique est la plus petite valeur de la série telle qu'au moins 50 % (la moitié) des valeurs de la série lui soient inférieures ou égales.
Pour une série d'effectif total n, la médiane est la valeur de rang le plus petit entier $\geq \frac{n}{2}$.
Donc c'est la valeur de rang $\frac{n}{2}$ si n est pair, et la valeur de rang $\frac{n+1}{2}$ si n est impair.

Définition. Le 1er quartile Q_1 (resp. **le 3e quartile Q_3**) d'une série statistique est la plus petite valeur de la série telle qu'au moins 25 % (resp. 75 %) des valeurs de la série lui soient $\leq$.

Remarque. Pour une série ordonnée d'effectif total n, Q_1 (resp. Q_3) est la k-ième valeur où k est le plus petit entier supérieur ou égal à $\frac{n}{4}$ (resp. $\frac{3n}{4}$).

Exemple. On considère la série ordonnée de 9 valeurs $1; 3; 7; 8; 10; 11; 12; 12; 58$. On a alors :

$\overbrace{1;3;7}^{\text{Plus de 25 \% des valeurs}}$;8;10;11;12;12;58 donc $Q_1 = 7$ — $\underbrace{1;3}_{\text{Moins de 25 \% des valeurs}}$

$\overbrace{1;3;7;8;10;11;12}^{\text{Plus de 75 \% des valeurs}}$;12;58 donc $Q_3 = 12$ — $\underbrace{1;3;7;8;10;11}_{\text{Moins de 75 \% des valeurs}}$

Définition. L'écart interquartile d'une série statistique est $Q_3 - Q_1$. Il s'agit d'un indicateur de dispersion.

Définition. Le k-ème décile ($k \in \{1; 2; \dots; 9\}$) d'une série statistique est la plus petite valeur de la série telle qu'au moins $10k$ % des valeurs de la série lui soient $\leq$.
Pour une série d'effectif total n, le k-ième décile est la valeur de rang le plus petit entier $\geq \frac{kn}{10}$.

Définition. Le k-ème percentile ($k \in \{1; 2; \dots; 99\}$) d'une série statistique est la plus petite valeur de la série telle qu'au moins k % des valeurs de la série lui soient $\leq$.
Pour une série d'effectif total n, le k-ième percentile est la valeur de rang le plus petit entier $\geq \frac{kn}{100}$.