Statistiques descriptives

Exemple. On lance un dé cubique 40 fois, et on note le résultat dans ce tableau

Valeur	1	2	3	4	5	6
Effectif	5	10	5	7	3	10

Définition. **La moyenne** d'une série statistique de N valeurs $x_1; x_2; \dots; x_N$ est

$$m = \frac{x_1 + x_2 + \cdots + x_N}{N}$$

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

$$m = \frac{c_1 x_1 + c_2 x_2 + \cdots + c_N x_N}{c_1 + c_2 + \cdots + c_N}$$

Exemple. Dans l'exemple initial, $m = \frac{5\times1+10\times2+5\times3+7\times4+3\times5+10\times6}{40} = 3{,}575$

Propriété. **Linéarité.** Soit a et b deux nombres réels et $x_1, x_2, \dots, x_N$ une série statistique de moyenne m.
La moyenne de la série $ax_1; ax_2; \dots; ax_N$ est $a \times m$.
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 valeurs $x_1; x_2; \dots; x_N$ est

$$\sigma = \sqrt{\frac{(x_1 - m)^2 + (x_2 - m)^2 + \cdots + (x_N - m)^2}{N}}$$

Définition. L'écart-type pondéré d'une série de valeurs $x_1; x_2; \dots; x_N$ de poids respectifs $c_1; c_2; \dots; c_N$ est

$$\sigma = \sqrt{\frac{c_1(x_1 - m)^2 + c_2(x_2 - m)^2 + \cdots + c_N(x_N - m)^2}{c_1 + c_2 + \cdots + c_N}}$$

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

Exemple. $\sigma = \sqrt{\frac{5\times(1-3{,}575)^2+10\times(2-3{,}575)^2+5\times(3-3{,}575)^2+7\times(4-3{,}575)^2+3\times(5-3{,}575)^2+10\times(6-3{,}575)^2}{40}} \approx 1{,}46$

Définition. **La médiane M_e** d'une série est la plus petite valeur telle qu'au moins la moitié des valeurs sont $\leq$

Idée. Après tri, la médiane est la valeur qui sépare la série en deux parties de même taille (ou presque).
Par exemple, le salaire médian est le plus petit salaire tel qu'au moins la moitié des salaires sont $\leq$.
Si on choisit un salarié au hasard, il y a 50% de chance que le salaire soit $\leq$ au salaire médian.
Exemple. Dans l'exemple initial, la médiane est $M_e = 3$, car la moitié des valeurs $1; 2; 3$ sont ≤ 3.

Définition. **Le quartile Q_1** d'une série est la plus petite valeur telle qu'au moins 25 % des valeurs sont $\leq$
Définition. **Le quartile Q_2** d'une série est la plus petite valeur telle qu'au moins 50 % des valeurs sont $\leq$
Définition. **Le quartile Q_3** d'une série est la plus petite valeur telle qu'au moins 75 % des valeurs sont $\leq$
Définition. **L'écart interquartile** d'une série statistique est $Q_3 - Q_1$. Il s'agit d'un indicateur de dispersion.
Exemple. On considère la série ordonnée des 9 valeurs suivantes : $1; 3; 7; 8; 10; 11; 12; 12; 58$.
Alors $0{,}25 \times 9 = 2{,}25 < 3$. Donc Q_1 est la 3ème valeur. $0{,}75 \times 9 = 6{,}75 < 7$. Donc Q_3 est la 7ème valeur.

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

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

L'écart interquartile est donc $Q_3 - Q_1 = 12 - 7 = 5$.

Définition. **Le $\boldsymbol{k}$-ème décile** ($k \in \{1; 2; \dots; 9\}$) d'une série est la plus petite valeur telle qu'au moins $\frac{k}{10}$ des valeurs sont $\leq$.

Définition. **Le $\boldsymbol{k}$-ème percentile** ($k \in \{1; 2; \dots; 99\}$) d'une série statistique est la plus petite valeur telle qu'au moins k % des valeurs sont $\leq$.