STATISTIQUES

Le chapitre s'appuie sur la série du tableau ci-dessous qui présente le nombre de buts par match durant la Coupe du monde de football de 2010 :

Nombre de buts x_i	0	1	2	3	4	5	6	7
Nombre de matchs n_i	7	17	13	14	8	6	0	1

Les valeurs x_i du caractère étudié sont les "nombres de buts".

Les effectifs n_i correspondants sont les "nombres de matchs".

1. Paramètre de position

1.1 Moyenne

Exemple
La moyenne de buts par match est égale à :

$$\bar{x} = \frac{7\times 0+17\times 1+13\times 2+14\times 3+8\times 4+6\times 5+0\times 6+1\times 7}{7+17+13+14+8+6+1} = \frac{154}{66} \approx 2,3$$

<u>Définition :</u> La moyenne $\bar{x}$ d'une série statistique dont les valeurs du caractère sont x_1, x_2,….,x_k et les effectifs correspondants sont n_1,n_2,…,n_k est égale à :

$$\bar{x} = \frac{x_1\times n_1 + x_2\times n_2 + \ldots + x_k\times n_k}{n_1+n_2+\ldots+n_k} = \frac{\sum_{i=1}^{k} x_i\times n_i}{\sum_{i=1}^{k} n_i}$$

1.2 Médiane

Pour obtenir la médiane d'une série, on range les valeurs de la série dans l'ordre croissant. La médiane est la valeur qui partage la série en deux populations d'effectif égal.

Exemple :

Pour la série étudiée dans le chapitre, l'effectif total est égal à 66. La médiane se trouve donc entre la 33^e et 34^e valeur de la série.

On écrit les valeurs de la série dans l'ordre croissant :

0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 **2** **2** 2 2 2 3 3 3 3 3 …

↑

La 33^e et la 34^e valeur sont égales à 2. La médiane est donc également égale à 2.

On en déduit que durant la Coupe du monde 2010, il y a eu autant de matchs dont le nombre de buts était supérieur à 2 que de matchs dont le nombre de buts était inférieur à 2.

1.3 Quartiles

Le premier quartile Q_1 est la plus petite valeur de la série telle qu'au moins 25% des valeurs sont inférieures ou égales à Q_1.

Le troisième quartile Q_3 est la plus petite valeur de la série telle qu'au moins 75% des valeurs sont inférieures ou égales à Q_3.

Exemple

Pour la série étudiée dans le chapitre, l'effectif total est égal à 66.

Le premier quartile Q_1 est valeur 17e valeur. En effet, $\frac{1}{4} \times 66 = 16{,}5 \rightarrow 17$. Donc Q_1 = 1.

Le troisième quartile Q_3 est valeur 50e valeur. En effet, $\frac{3}{4} \times 66 = 49{,}5 \rightarrow 50$. Donc Q_3 = 3.

2. Paramètre de dispersion

2.1 Écart interquartile

2.1.1 Définition :

L'écart interquartile d'une série statistique de premier quartile Q_1 et de troisième quartile Q_3 est égal à la différence Q_3 - Q_1.

2.1.2 Exemple

Pour la série étudiée dans le chapitre, l'écart interquartile est égal :
Q_3 - Q_1 = 3 – 1 = 2.

2.1.3 Remarque

L'écart interquartile d'une série mesure la dispersion autour de la médiane. Il contient au moins 50% des valeurs de la série.

L'écart interquartile n'est pas influencé par les valeurs extrêmes de la série.

2.2 Variance

2.2.1 Définition

- La variance V d'une série statistique de moyenne $\overline{x}$ dont les valeurs du caractère sont x_1, x_2, x_3, …, x_k et les effectifs correspondants sont n_1, n_2, n_3, …, n_k est égale à :

$$V = \frac{n_1 \times \left(x_1 - \overline{x}\right)^2 + n_2 \times \left(x_2 - \overline{x}\right)^2 + \ldots + n_k \times \left(x_k - \overline{x}\right)^2}{n_1 + n_2 + \ldots + n_k} = \frac{\sum_{i=1}^{k} n_i \times \left(x_i - \overline{x}\right)^2}{\sum_{i=1}^{k} n_i}$$

- L'écart-type σ d'une série statistique de variance V est égal à : $\sigma = \sqrt{V}$

2.2.2 Exemple

Pour la série étudiée dans le chapitre, la variance est égale à :

$$V = \frac{7\times\left(0-\frac{7}{3}\right)^2+17\times\left(1-\frac{7}{3}\right)^2+13\times\left(2-\frac{7}{3}\right)^2+14\times\left(3-\frac{7}{3}\right)^2+8\times\left(4-\frac{7}{3}\right)^2+6\times\left(5-\frac{7}{3}\right)^2+0\times\left(6-\frac{7}{3}\right)^2+1\times\left(7-\frac{7}{3}\right)^2}{66}$$

$$\approx 2,4646$$

$$\sigma \approx \sqrt{2,4646} \approx 1,57$$

L'écart-type possède la même unité que les valeurs de la série.
Ainsi pour la série étudiée, l'écart-type est environ égal à 1,57 buts.

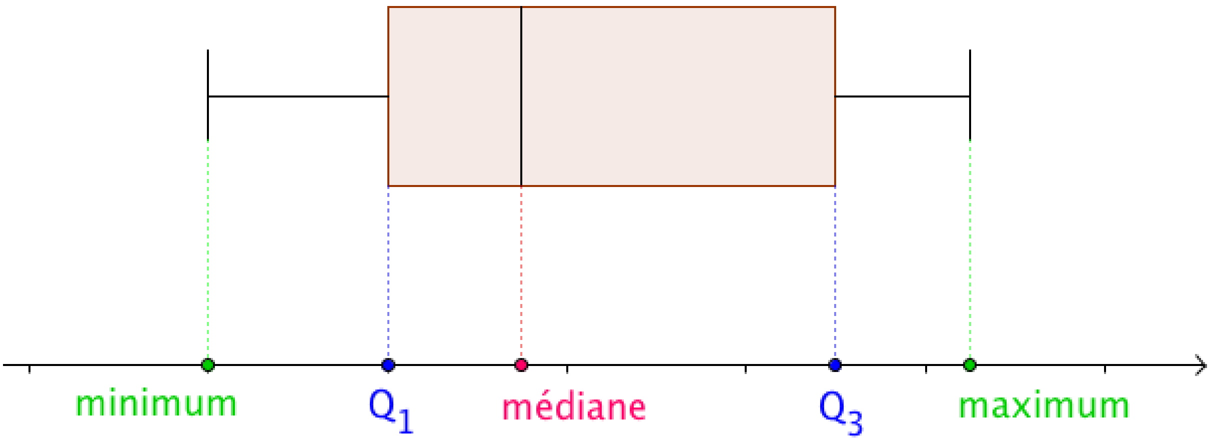
2.2.3 Remarque

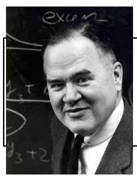
L'écart-type exprime la dispersion des valeurs d'une série statistique autour de sa moyenne. Les valeurs extrêmes influencent l'écart-type.

2.2.4 Propriété

$$V = \frac{n_1\times x_1^{\,2}+n_2\times x_2^{\,2}+\ldots+n_k\times x_k^{\,2}}{n_1+n_2+\ldots+n_k}-\left(\overline{x}\right)^2 = \frac{\sum_{i=1}^{k} n_i\times x_i^{\,2}}{\sum_{i=1}^{k} n_i}-\left(\overline{x}\right)^2$$

3. Diagramme en boîte





Ce type de diagramme porte également le nom de *boîte à moustaches* ou *diagramme de Tukey*.
John Wilder Tukey (1915 – 2000) était un statisticien américain.

Exemple

Pour la série étudiée dans le chapitre :