Fiche n° 27. Statistiques

I Les paramètres de position (ou de tendance centrale)

A La moyenne

Définition 1

On suppose que la série statistique x prend n valeurs **pas nécessairement deux à deux distinctes**. La **moyenne** de la série statistique x, notée $\overline{x}$ (ou aussi m ou aussi μ (lettre de l'alphabet grec qui se lit « mu »)) est

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

Définition 2

On suppose que la série statistique x prend p valeurs **deux à deux distinctes**, d'effectifs respectifs $n_1, \ldots, n_p$ ou de fréquences respectives $f_1, \ldots, f_p$. La **moyenne** de la série statistique x, notée $\overline{x}$ (ou aussi m ou aussi μ) est

$$\overline{x} = \frac{n_1 \times x_1 + n_2 \times x_2 + \ldots + n_p \times x_p}{n_1 + n_2 + \ldots + n_p} = f_1 \times x_1 + f_2 \times x_2 + \ldots + f_p \times x_p.$$

Le théorème suivant porte le nom « linéarité de la moyenne ».

Théoreme 1

(Linéarité de la moyenne) Soient a et b deux réels. La moyenne de la série $ax + b$ est $a\overline{x} + b$.

B La médiane

Définition 3

On suppose que la série statistique x prend n valeurs **pas nécessairement deux à deux distinctes** $x_1, x_2, \ldots, x_n$, **classées dans l'ordre croissant** : $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$.

Si n est impair, on pose $n = 2p + 1$ où $p \in \mathbb{N}$. La **médiane** est x_{p+1} (ou encore $x_{\frac{n+1}{2}}$).
Si n est pair, on pose $n = 2p$ où $p \in \mathbb{N}^*$. La **médiane** est la moyenne des deux nombres x_p et x_{p+1} c'est-à-dire $\frac{x_p + x_{p+1}}{2}$ (ou encore $\frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n+1}{2}}\right)$). Ce n'est pas forcément une valeur de la série statistique.
Dans tous les cas, on notera $\mathrm{M}e$ la médiane.

Théoreme 2

Au moins 50% des valeurs de la série x sont inférieures ou égales à $\mathrm{M}e$ et au moins 50% des valeurs de la série x sont supérieures ou égales à $\mathrm{M}e$.

C Les quartiles

Définition 4

On suppose que la série statistique x prend n valeurs **pas nécessairement deux à deux distinctes** $x_1, x_2, \ldots, x_n$, **classées dans l'ordre croissant** : $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$.

Le **premier quartile**, noté Q_1, est la première valeur (en partant de x_1) de la série telle que au moins 25% des valeurs de la série lui soient inférieures ou égales.
Le **troisième quartile**, noté Q_3, est la première valeur (en partant de x_1) de la série telle que au moins 75% des valeurs de la série lui soient inférieures ou égales.

II Les paramètres de dispersion

A Variance et écart-type

Définition 5

On suppose que la série statistique x prend n valeurs **pas nécessairement deux à deux distinctes**. La **variance** de la série statistique, notée V, est la moyenne des carrés des écarts à la moyenne c'est-à-dire

$$\mathrm{V} = \frac{1}{n}\left(\left(x_1 - \overline{x}\right)^2 + \ldots + \left(x_n - \overline{x}\right)^2\right)$$

L'**écart-type** de la série, noté s ou σ (lettre s de l'alphabet grec qui se lit « sigma ») est la racine carrée de la variance :

$$s = \sqrt{\mathrm{V}} = \sqrt{\frac{1}{n}\left(\left(x_1 - \overline{x}\right)^2 + \ldots + \left(x_n - \overline{x}\right)^2\right)}.$$

Définition 6

On suppose que la série statistique x prend p valeurs **deux à deux distinctes**, d'effectifs respectifs $n_1, \ldots, n_p$ ou de fréquences respectives $f_1, \ldots, f_p$. La **variance** de la série statistique, notée V, est la moyenne des carrés des écarts à la moyenne c'est-à-dire

$$\mathrm{V} = \frac{1}{n}\left(n_1 \times \left(x_1 - \overline{x}\right)^2 + \ldots + n_p\left(x_p - \overline{x}\right)^2\right) = f_1 \times \left(x_1 - \overline{x}\right)^2 + \ldots + f_p\left(x_p - \overline{x}\right)^2.$$

L'**écart-type** de la série, noté s ou σ est la racine carrée de la variance :

$$s = \sqrt{\mathrm{V}} = \sqrt{\frac{1}{n}\left(n_1 \times \left(x_1 - \overline{x}\right)^2 + \ldots + n_p\left(x_p - \overline{x}\right)^2\right)} = \sqrt{f_1 \times \left(x_1 - \overline{x}\right)^2 + \ldots + f_p\left(x_p - \overline{x}\right)^2}.$$

B L'intervalle interquartile

Définition 7

L'**écart interquartile** de la série est la distance entre le premier et le troisième quartile c'est-à-dire $\mathrm{Q}_3 - \mathrm{Q}_1$.