## Statistiques

#### Statistique descriptive

On dispose de la série statistique suivante, le caractère étudié étant un caractère quantitatif discret :

Valeurs x <sub>i</sub>	$\chi_1$	$\chi_2$	х3	 $\chi_{\mathrm{p}}$
Effectifs n <sub>i</sub>	$n_1$	$\mathfrak{n}_2$	$n_3$	 $n_p$

L'effectif total de cette série est  $n=n_1+n_2+n_3+\ldots+n_p=\sum_{i=1}^p n_i.$ 

Moyenne, variance et écart-type

Définition (moyenne). – La moyenne de cette série statistique est le nombre réel

$$\bar{x} = \frac{\displaystyle\sum_{i=1}^p n_i \times x_i}{n} = \frac{n_1 \times x_1 + n_2 \times x_2 + \ldots + n_p \times x_p}{n}.$$

Définition (variance et écart-type). -

1. La variance de cette série statistique est le nombre réel V défini par

$$V = \frac{1}{n} \sum_{i=1}^{p} n_i (x_i - \bar{x})^2.$$

C'est la moyenne des carrés des écarts à la moyenne.

2. L'écart-type de cette série est le nombre réel  $\sigma$  défini par

$$\sigma = \sqrt{V}$$
.

L'écart-type mesure la dispersion des valeurs autour de la moyenne.

Proposition. – La variance est aussi égale à la moyenne des carrés moins le carré de la moyenne :

$$V = \frac{1}{n} \sum_{i=1}^{p} (n_i \times x_i^2) - \bar{x}^2.$$

Exemple. – On s'intéresse au temps total de transport des employés d'une usine pendant une semaine. Voici les résultats obtenus :

Temps en heures	1	2	3	4	5	6	7	8
Effectifs	2	3	6	8	10	15	24	16

Calculer la moyenne et l'écart-type de cette série, en arrondissant les résultats à 0, 1. On utilisera les deux formule données pour calculer la variance.	æ

#### Médiane et écart interquartile

On considère une série statistique dont les n termes  $a_1, a_2, \ldots, a_n$  sont rangés par ordre croissant, chaque valeur figurant autant de fois que son effectif. n est donc l'effectif total.

## Définition (médiane). -

- 1. Si l'effectif total est impair, la médiane M<sub>e</sub> est la valeur du caractère située au milieu.
- 2. Si l'effectif total est pair, la médiane M<sub>e</sub> est la demi-somme des deux valeurs situées au milieu.

## Définition (quartiles). -

- 1. Le premier quartile Q<sub>1</sub> est la plus petite valeur du caractère telle qu'au moins 25% des termes de la série aient une valeur qui lui soit inférieure ou égale.
- 2. Le troisième quartile Q<sub>3</sub> est la plus petite valeur du caractère telle qu'au moins 75% des termes de la série aient une valeur qui lui soit inférieure ou égale.
- 3. L'intervalle interquartile est l'intervalle  $[Q_1;Q_3]$ , le nombre  $I=Q_3-Q_1$  s'appelle l'écart interquartile.

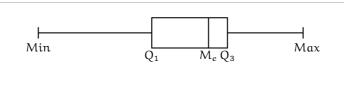
Remarque. - D'une manière analogue on peut définir les déciles et les centiles.

# Détermination pratique de Q<sub>1</sub> et Q<sub>3</sub>

- 1. Si  $\frac{n}{4}$  est un entier k,  $Q_1$  est le terme de la série de rang k et  $Q_3$  le terme de la série de rang 3k.
- 2. Si  $\frac{n}{4}$  n'est pas un entier,  $Q_1$  est le terme de la série de rang l'entier immédiatement supérieur à  $\frac{n}{4}$  et  $Q_3$  est le terme de rang l'entier immédiatement supérieur à  $\frac{3n}{4}$ .

#### Diagramme en boîte ou boîte à moustaches

C'est un outil graphique permettant d'étudier la répartition des valeurs d'une série. On y fait figurer les valeurs extrêmes Min et Max, la médiane Me, les premier et troisième quartiles Q1 et Q3, et parfois aussi les premier et neuvième déciles.



Exemple. – La température est relevée chaque heure pendant 4 jours dans une forêt. Les 97 résultats obtenus ont été triés et sont rassemblés dans le tableau suivant :

Température en °C	14,5	15	15,5	16	16,5	17	17,5	18	18,5	19	19,5
Nombre de fois où cette	Е	7	10	10	15	10	11	0	7	7	4
température a été relevée	5	1	10	12	15	10	11	9	'	1	4

Construire le diagramme en boîte de cette série statistique. On pourra utiliser les effectifs cumulés croissants.

## Influence d'une transformation affine

Soient  $(x_i; n_i)$  une série statistique,  $\bar{x}$  sa moyenne, V sa variance,  $\sigma$  son écart-type,  $Q_1$  son premier quartile et  $Q_3$  son troisième quartile.

Soit f une fonction affine définie sur  $\mathbb{R}$  par f(x) = ax + b. En posant pour tout i  $y_i = f(x_i)$ , on obtient une nouvelle série statistique  $(y_i; n_i)$ . Les valeurs ont été transformées mais les effectifs sont conservés.

#### Proposition. -

	Série (x <sub>i</sub> ; n <sub>i</sub> )	Série $(y_i; n_i)$ avec $y_i = ax_i + b$
Moyenne	$\bar{\chi}$	$\bar{y} = a\bar{x} + b$
Variance	V	$V'=\mathfrak{a}^2V$
Écart-type	σ	$\sigma' =  \mathfrak{a} \sigma$
Médiane	$M_e$	$M'_e = aM_e + b \text{ si } a > 0$
Premier quartile	$Q_1$	$Q_1' = \alpha Q_1 + b \text{ si } \alpha > 0$
Troisième quartile	Q <sub>3</sub>	$Q_3' = \alpha Q_3 + b \text{ si } \alpha > 0$
Écart interquartile	$Q_3 - Q_1$	$a(Q_3 - Q_1)$ si $a > 0$

Exemple. – Une série statistique $(x_i; n_i)$ a pour quartiles $Q_1 = 6$ et $Q_3 = 10$ , pour moyenne $\bar{x} = 7, 5$ et pour écart-type
$\sigma = 2, 8$ . On considère la fonction f définie sur $\mathbb{R}$ par $f(x) = 0, 5x + 3$ . En posant, pour tout i, $y_i = f(x_i)$ , on obtient
une nouvelle série statistique. Calculer les paramètres de cette série $(y_i, n_i)$ .