

# Chapitre 2: Statistiques

Dans tous le chapitre, on considère une série statistique à caractère quantitatif d'effectif total  $N$ .

## 1 Diagramme en boîte

### A) Médiane et quartiles

Intuitivement, les quartiles sont les nombres qui partagent la série statistique en quatre parties qui ont toutes sensiblement le même nombre de termes, c'est à dire 25% de l'effectif total.

#### Définition:

La médiane d'une série statistique est le nombre, noté  $M_e$ , tel que :

- 50% au moins des individus ont une valeur du caractère inférieure ou égale à  $M_e$  ;
- 50% au moins des individus ont une valeur du caractère supérieure ou égale à  $M_e$  ;

#### Remarque:

Le liste des  $N$  valeurs étant rangée par ordre croissant, chacune figurant un nombre de fois égal à son effectif :

→ Si  $N$  est impair, la médiane est la valeur de rang  $\frac{N+1}{2}$  ;

→ Si  $N$  est pair, la médiane est la demi-somme des valeurs de rang  $\frac{N}{2}$  et  $\frac{N}{2} + 1$ .

#### Définition:

La liste des  $N$  valeurs est rangée dans l'ordre croissant :

- Le premier quartile est la plus petite valeur  $Q_1$  de la liste telle qu'au moins un quart des valeurs de la liste sont inférieurs ou égale à  $Q_1$ .
- Le troisième quartile est la plus petite valeur  $Q_3$  de la liste telle qu'au moins trois quarts des valeurs de la liste sont inférieurs ou égale à  $Q_3$ .

#### Remarque:

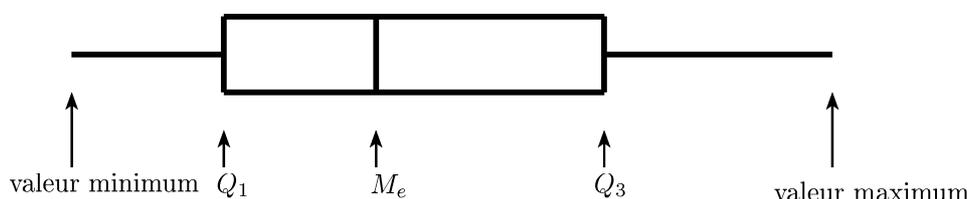
Lorsqu'on a ordonné dans l'ordre croissant la liste des valeurs, chacune d'elles répétée autant de fois que son effectif, alors  $N$  est l'effectif total et :

→ Si  $N$  est multiple de 4, le premier quartile  $Q_1$  est le terme de rang  $\frac{N}{4}$  de la liste et le troisième quartile  $Q_3$  est le terme de rang  $\frac{3N}{4}$  de la liste ;

→ Si  $N$  n'est pas multiple de 4, le premier quartile  $Q_1$  est le terme de rang immédiatement supérieur à  $\frac{N}{4}$  de la liste et le troisième quartile  $Q_3$  est le terme de rang immédiatement supérieur à  $\frac{3N}{4}$  de la liste ;

### B) Représentation par un diagramme en boîte

La répartition des données peut-être représentée par un diagramme dit en boîte, qui résume me caractère quantitatif étudié par les valeurs extrêmes, la médiane et les quartiles.



## 2 Paramètres de dispersion

### A) Étendue et écart interquartile

Pour rappel, deux paramètres de dispersion vus en seconde :

**Définition:**

- *L'étendue est la différence entre les valeurs extrêmes de la série.*

$$e = x_{max} - x_{min}$$

- *L'écart interquartile est la différence entre les quartiles  $Q_3$  et  $Q_1$ .*

$$E_i = Q_3 - Q_1$$

### B) Variance et écart-type

Les valeurs  $x_i$  de la série statistique à caractère quantitatif d'effectif total  $N$  ont pour effectifs respectifs  $n_i$ . On a ainsi :

$$N = \sum_{i=1}^p n_i$$

où  $p$  est le nombre de valeurs différentes de la série.

**Définition:**

On appelle **variance** d'une série statistique  $(x_i; n_i)$  avec  $1 \leq i \leq p$ , la moyenne des carrés des écarts  $x_i - \bar{x}$  :

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

On appelle **écart-type** d'une série statistique la racine de la variance :

$$s = \sqrt{V}$$

**Remarque:**

L'écart-type est un nouveau paramètre qui permet de mesurer la dispersion des valeurs  $x_i$  autour de la moyenne  $\bar{x}$ .

**Propriété:**

Si  $(x_i; n_i)$  avec  $1 \leq i \leq p$  est une série statistique de variance  $V$  alors :

$$V = \frac{1}{N} \sum_{i=1}^p n_i x_i^2 - \bar{x}^2$$

## 3 Résumé d'une série statistique

Pour indiquer la répartition des données d'une série statistique, on utilise l'un des deux couples suivants, composé chacun d'un paramètre de tendance centrale et d'un paramètre de dispersion :

- $(E_i; M_e)$  Ce couple est peu sensible aux valeurs extrêmes.
- $(s; \bar{x})$  Ce couple est sensible aux valeurs extrêmes.