Statistiques descriptives

I - Description d'une série par les quantiles

Définition *Caractéristiques de position, ou de tendance centrale*

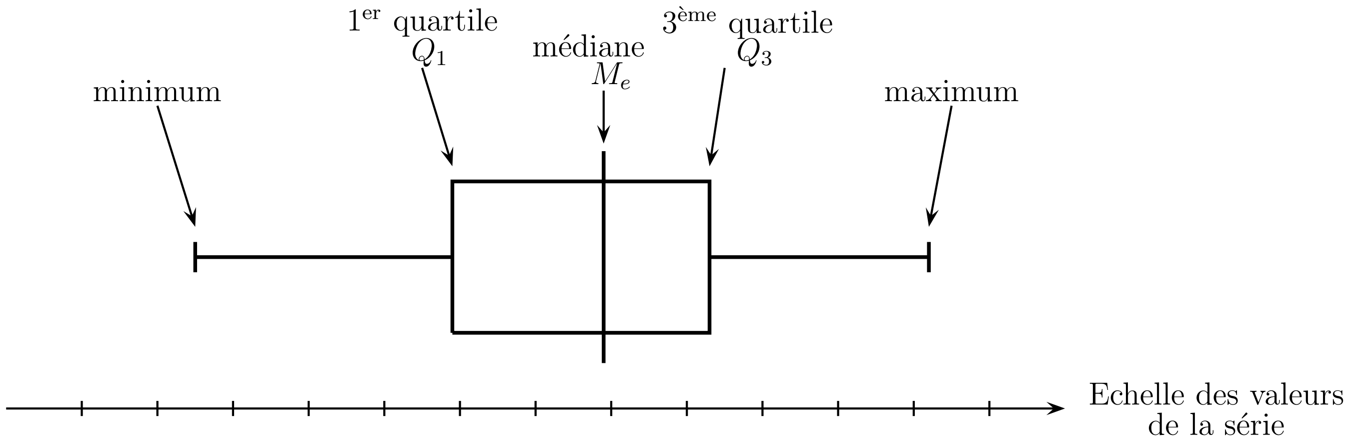
On considère une série statistique d'effectif total N.

- *La* **médiane** M_e *est une valeur qui partage la série ordonnée en deux séries de même effectif.*
  Si N est impair, $N = 2n + 1$, alors la médiane est la $n^{\text{ème}}$ valeur de la série ordonnée.
  Si N est pair, $N = 2n$, alors la médiane est la moyenne de la $n^{\text{ème}}$ et de la $(n+1)^{\text{ème}}$ valeur.
- *Le* **premier quartile** Q_1 *est la plus petite valeur de la série telle qu'au moins un quart des données de la série sont inférieures ou égales à Q_1.*
  Le **trosième quartile** Q_3 *est la plus petite valeur de la série telle qu'au moins les trois quarts des données de la série sont inférieures ou égales à Q_3.*

Définition *Caractéristiques de dispersion*

- **L'étendue** d'une série est la différence entre sa plus grande et sa plus petite valeur.
- L'**écart inter-quartile** est la différence entre le troisième et le premier quartile : $Q_3 - Q_1$.

Diagrammes en boîte (boites à moustaches) On peut alors représenter les données de la série statistique par un diagramme en boîte, aussi connu sous le nom de "boîte à moustaches" :



Exercice 1 Le tableau suivant donne les notes des élèves d'une classe.

Elèves	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
Notes	15	10	12	8	10	18	12	8	8	15	10	8	6	18	12	8	12

On ordonne la série :

Notes x_i						
Effectifs n_i						
Effectifs cumulés croissants						

L'effectif total de la série : $N = \ldots$

La médiane de la série : $M_e = \ldots$

Les 1^{er} et $3^{\text{ème}}$ quartiles sont : $Q_1 = \ldots$, $Q_3 = \ldots$

L'étendue de la série est : $\ldots$

L'écart inter-quartile est : $\ldots$

Exercice 2 On compare les températures moyennes (en ° C) de chaque mois de l'année pour deux communes de Haute-Savoie situées à 1000 m d'altitude : Chamonix et La Clusaz.

Mois	1	2	3	4	5	6	7	8	9	10	11	12
Chamonix	1,5	4	7,5	12	15,5	20	23	22	19	14	6,5	2
La Clusaz	2,5	3,5	6	9,5	14	17	20,5	20,0	17	13	7	3,5

Déterminer pour ces deux communes la médiane et les quartiles des températures.
Tracer ensuite les diagrammes en boîte de ces deux séries en utilisant la même échelle, de manière à pouvoir les comparer.

II - Description par la moyenne et son écart type

On considère une série statistique générale :

Valeur	x_1	x_2	x_3	...	x_p
Effectifs	n_1	n_2	n_3	...	n_p

d'effectif total : $N = n_1 + n_2 + \cdots + n_p$.

Définition *La moyenne de la série est :* $\overline{x} = \dfrac{n_1x_1 + n_2x_2 + \cdots + n_px_p}{N}$

Exercice 3 Déterminer la moyenne des séries suivantes :

<u>S_1 :</u> 1 ; 8 ; 10 ; 10 ; 12 ; 19 $\quad\overline{x} = \ \ldots.$

<u>S_2 :</u> 9 ; 9,5 ; 10 ; 10,5 ; 11 $\quad\overline{x} = \ \ldots.$

<u>S_3 :</u> 10 ; 10 ; 10 ; 10 ; 10 ; $\quad\overline{x} = \ \ldots.$

<u>S_4 :</u> 10 $\quad\overline{x} = \ \ldots.$

La moyenne suffit-elle, à elle-seule, à décrire une sréie de données ?

Définition • *La variance de la série est la moyenne des carrés des écarts à la moyenne :*

$$V = \frac{n_1(x_1 - \overline{x})^2 + n_2(x_2 - \overline{x})^2 + \cdots + n_p(x_p - \overline{x})^2}{N}$$

• *L'écart type d'une série est la racine carrée de la variance :* $\sigma = \sqrt{V}$.

<u>Exemple :</u> Soit la série statistique :

Valeurs x_i	5	9	12	18
Effectifs n_i	2	4	3	1

L'effectif total de cette série est : $N = 2 + 4 + 3 + 1 = 10$.

La moyenne de cette série est : $\overline{x} = \dfrac{2\times5 + 4\times9 + 3\times12 + 1\times18}{10} = 10$

Sa variance est alors : $V = \dfrac{2\times(5-10)^2 + 4\times(9-10)^2 + 3\times(12-10)^2 + 1\times(18-10)^2}{10} = 13$,
et son écart type : $\sigma = \sqrt{13} \simeq 3,6$.

Exercice 4 Calculer l'écart type de chacune des séries S_1, S_2, S_3 et S_4 de l'exercice précédent.

Exercice 5 Le tableau suivant donne les tailles de 30 élèves d'une classe.

taille(cm)	145	146	151	152	155	160	165	170	172	176	180	186	188	190	193
effectif	1	1	1	2	3	3	5	2	6	3	3	2	1	1	1

Calculer la moyenne et l'écart type de cette série.