

Statistique descriptive

Pour reprendre contact n°1 – 3 – 4 p 179

I. Autour de la médiane

A. Rappels

Définition

La liste des N données est rangée **dans l'ordre croissant**.

- **la médiane Me** est
 - si N est impair, la valeur « du milieu », de rang $\frac{N+1}{2}$
 - si N est pair, la demi-somme des valeurs « du milieu » de rang $\frac{N}{2}$ et de rang $\frac{N}{2} + 1$
- **Le premier quartile** est la plus petite donnée Q_1 de la liste telle qu'**au moins un quart** des données de la liste sont **inférieures ou égales à Q_1** .
- **Le troisième quartile** est la plus petite donnée Q_3 de la liste telle qu'**au moins trois quarts** des données de la liste sont **inférieures ou égales à Q_3** .
- **L'écart interquartile** est la différence entre le troisième quartile et le premier quartile : $Q_3 - Q_1$

Remarque

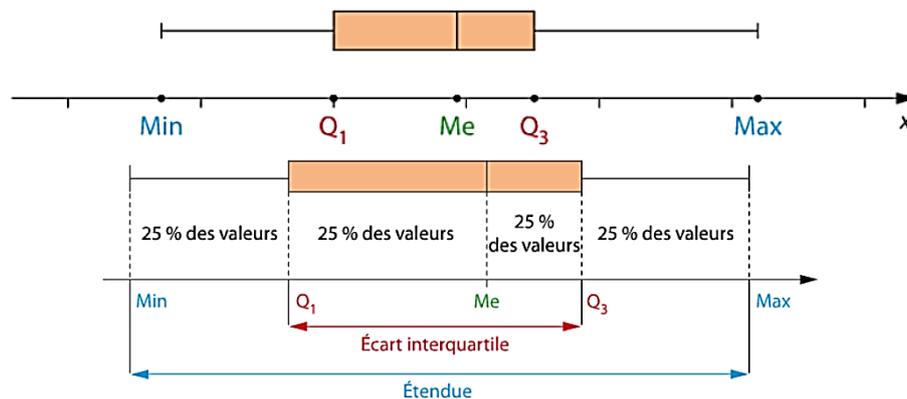
La médiane est une « valeur centrale » d'une série. L'écart interquartile est un indicateur de dispersion.

Exercices n°11 – 12 p 191

B. Diagramme en boîte

Activité 1 p 180

Les nombres Q_1, Me, Q_3, Min et Max valeurs extrêmes de la série) constituent un bon résumé d'une série statistique et on les représente sur **un diagramme en boîte** (ou **boîte à moustaches**)



Exercice résolu n°1 p 183

Exercices n°13 à 16 p 191 – 192

II. Autour de la moyenne

A. La moyenne

On utilise la série statistique suivante :

Valeur	x_1	x_2	...	x_p	TOTAL
Effectif	n_1	n_2	...	n_p	$N = n_1 + n_2 + \dots + n_p$
Fréquence	f_1	f_2	...	f_p	

Définition

La **moyenne** de cette série statistique est le nombre réel, noté \bar{x} , tel que

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_p x_p}{N} = f_1 x_1 + f_2 x_2 + \dots + f_p x_p$$

Propriété

Une série statistique partagée en p sous-séries disjointes de moyennes et d'effectifs respectifs (\bar{x}_1, n_1) , (\bar{x}_2, n_2) , ..., (\bar{x}_p, n_p) a pour moyenne $\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \dots + n_p\bar{x}_p}{n_1 + n_2 + \dots + n_p}$

B. Variance et écart-type**Définition**

- La **variance** de cette série statistique est le nombre réel, noté V , tel que :

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

- L'écart-type** d'une série statistique est le nombre réel, noté σ , tel que $\sigma = \sqrt{V}$

Remarque

La variance et l'écart-type permettent de mesurer la dispersion des valeurs par rapport à la valeur moyenne. L'avantage de l'écart-type est de s'exprimer dans la même unité que les x_i .

Exercice résolu n°2 p 185

Exercices n°17 à 24 p 191 – 192

III. Choisir un résumé d'une série statistique

Pour résumer une série statistique, il faut choisir un indicateur de position : médiane ou moyenne, auquel on associe un indicateur de dispersion (écart interquartile ou écart-type) pour rendre compte de la répartition des valeurs autour de cette valeur « centrale »

A. Le couple (médiane, écart interquartile)

Plus l'écart interquartile (qui contient la moitié centrale des valeurs) est petit, plus les valeurs centrales de la série se concentrent autour de la médiane.

B. Le couple (moyenne, écart-type)

Plus l'écart-type est petit, plus les valeurs se concentrent autour de la moyenne.

C. Choix du couple➤ **Le couple (médiane, écart interquartile)**

Il est assez facile à interpréter. Il est peu sensible aux valeurs extrêmes.

Par contre, il ne se calcule pas par paquets : si on connaît les médianes de sous-séries disjointes constituant une série, on ne peut pas en déduire la médiane de cette série.

➤ **Le couple (moyenne, écart-type)**

Il est souvent utilisé dans les statistiques théoriques comme les sondages.

On peut calculer par paquets la moyenne et l'écart-type (voir ex 54 p198).

Cependant, l'écart-type est très sensible aux valeurs extrêmes, le choix de ce couple n'est donc pertinent que quand la série est assez symétrique (avec une courbe en cloche)

Exercices n°27 – 30 - 31 p 194 – 195

Exercices n°58 – 60 p 199

Prendre des initiatives n°61 à 63 p 200