

IV. Médiane ; quartiles ; déciles

On ne s'intéresse à présent qu'à des séries statistiques à caractère quantitatif.

1. Définition : médiane

Dans une série statistique **ordonnée**, la médiane partage l'ensemble des valeurs prises en deux groupes de même effectif.

Point méthode : Si les valeurs sont peu nombreuses, il est simple d'ordonner la série et de séparer les valeurs prises par le caractère en deux groupes de même effectif.

Si l'effectif total est impair : une valeur restera entre les deux demi-groupes. Cette valeur est la médiane.

Si l'effectif total est pair : aucune valeur de la série ne sépare la série en deux groupes de même effectif. Par convention, on considère que la médiane est la moyenne de la dernière valeur du premier groupe et de la première valeur du deuxième groupe.

2. Définition : quartiles

Le **premier quartile** d'une série statistique, noté Q_1 est la plus petite valeur prise par le caractère telle **qu'au moins un quart soit au moins 25%** des valeurs lui sont **inférieures ou égales**.

Le **troisième** d'une série statistique, noté Q_3 est la plus petite valeur prise par le caractère telle **qu'au moins trois quarts soit au moins 75%** des valeurs lui sont **inférieures ou égales**.

Remarque : la différence $Q_3 - Q_1$ est appelée **écart interquartiles**.

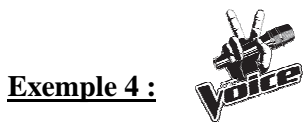
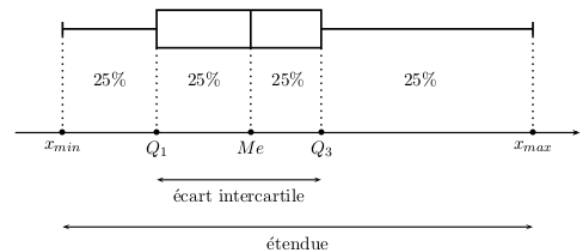
L'intervalle $[Q_3 - Q_1]$ est appelé **intervalle interquartiles** : plus l'écart interquartile est grand, plus les valeurs de la série sont dispersées autour de la médiane.

3. Diagramme en boîte :

On peut résumer une série statistique par un diagramme gradué faisant apparaître les indicateurs : $Min, Max, Q_1 ; Q_3$ et Me .

Un tel diagramme est appelé **diagramme en boîte** ou **boîte à moustache**.

Remarque : on peut facilement comparer plusieurs séries statistiques en superposant leurs diagrammes en boîte.



Exemple 4 :

Les élèves de 2^{nde} 12 du Lycée P. Picasso ont reçu une invitation pour assister aux enregistrements de l'émission de télévision The Voice. Les spectateurs sont placés sur deux rangs suivant leur taille. Aide les organisateurs en déterminant la taille médiane, puis donne le diagramme en boîte.

172 ; 162 ; 190 ; 190 ; 169 ; 164 ; 177 ; 181 ;
189 ; 161 ; 164 ; 182 ; 185 ; 188 ; 169 ; 180 ;
193 ; 189 ; 179 ; 180 ; 173 ; 193 ; 166 ; 164 ;
163 ; 164 ; 190 ; 176 ; 176 ; 192 ; 173 ; 194 ;
165 ; 172

Correction :

Voici la liste ordonnée des tailles des élèves de 2^{nde} 12 :

161 ; 162 ; 163 ; 164 ; 164 ; 164 ; 164 ; 165 ; 166 ; 169 ; 169 ; 172 ; 172 ; 173 ;
173 ; 176 ; 176 ; 177 ; 179 ; 180 ; 180 ; 181 ; 182 ; 185 ; 188 ; 189 ; 189 ; 190 ;
190 ; 190 ; 192 ; 193 ; 193 ; 194.

L'effectif total de cette série est 34. Une fois la série ordonnée, les 17^{ème} et 18^{ème} valeurs partagent cette série en deux groupes de 17 élèves.

Leur moyenne est $Me = \frac{x_{17} + x_{18}}{2} = \frac{176 + 177}{2} = 176,5$

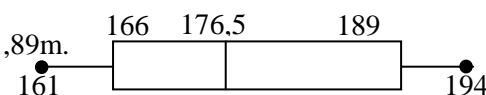
Une valeur possible de la médiane est 176,5cm.

$\frac{N}{4} = 8,5$ donc $Q_1 = x_9 = 166$: au moins 25% des élèves de 2^{nde} 12 ont

une taille inférieure ou égale à 1,66m.

$\frac{3N}{4} = 25,5$ donc $Q_3 = x_{26} = 189$: au moins 75% des élèves de 2^{nde} 12

ont une taille inférieure ou égale à 1,89m.



Point méthode : déterminer les quartiles avec les fréquences

Exemple 5 :

Le tableau ci-dessous indique la répartition du nombre d'enfants de moins de 25 ans dans les familles du Val de Marne en 2009. Donner la distribution des fréquences. En déduire la médiane et les quartiles. Interpréter.

Nombre de familles avec	en 2009
Aucun enfant	244 918
1 enfant	131 271
2 enfants	109 776
3 enfants	35 907
4 enfants ou plus	13 311
Ensemble	535 183

Correction On utilise la colonne des cumulés.

- Les 25% de la population sont atteints pour les familles sans enfant. Donc le 1^{er} quartile est 0 enfant.
- Les 50% de la population sont atteints pour les familles avec un enfant ou moins. Donc la médiane est 1 enfant.
- Les 75% de la population sont atteints pour les familles avec deux enfants. Donc le 3^e quartile est 2 enfants.

Fréquences en % des familles avec	en 2009	cumulés
Aucun enfant	45,76	45,76
1 enfant	24,53	70,29
2 enfants	20,51	90,80
3 enfants	6,71	97,51
4 enfants ou plus	2,49	100
Ensemble	100	

Remarque : pour une série dont les données sont regroupées en classe, les valeurs brutes prises par le caractère ne sont pas accessibles. Il est possible d'obtenir une approximation de la médiane et des quartiles par lecture graphique sur le polygone des effectifs (ou fréquences) cumulés.

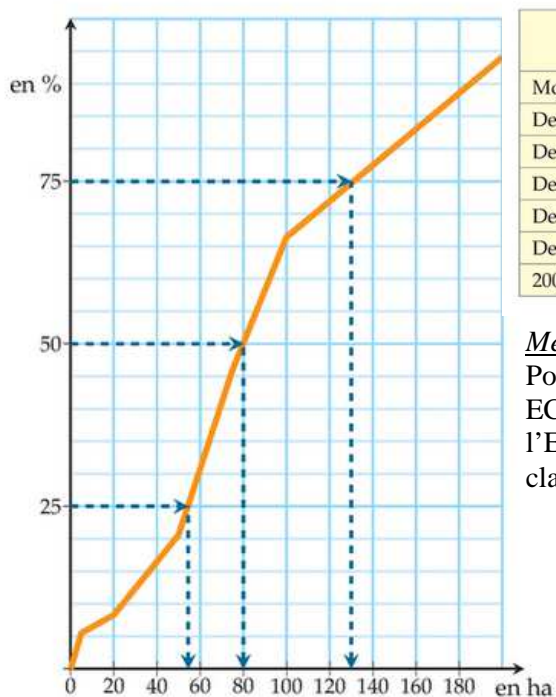
Exemple 6 :

Déterminer par lecture graphique la médiane et les quartiles de la série constituée par les tailles des exploitations agricoles professionnelles en Franche-Comté.

	Effectifs
Moins de 5 ha	370
De 5 à moins de 20 ha	190
De 20 à moins de 50 ha	840
De 50 à moins de 75 ha	1720
De 75 à moins de 100 ha	1380
De 100 à moins de 200 ha	1880
200 ha et plus	400
Ensemble	6780

Correction :

Polygone des fréquences cumulées croissantes des tailles des exploitations agricoles de la région Franche-Comté en 2006.



	F.C.C. en %
Moins de 5 ha	5,5
De 5 à moins de 20 ha	8,3
De 20 à moins de 50 ha	20,6
De 50 à moins de 75 ha	46
De 75 à moins de 100 ha	66,4
De 100 à moins de 200 ha	94,1
200 ha et plus	100

Méthode :

Pour tracer le polygone des ECC (ou FCC), on associe l'ECC (ou la FCC) de la classe $[a ; b[$ à la borne b .

V. Moyenne

1. Définition :

La moyenne d'une série statistique quantitative est souvent notée \bar{x} .

Avec les notations précédemment introduites, on a :

$$\bar{x} = \frac{\text{somme totale des valeurs prises par le caractère}}{\text{nombre de valeurs}} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_p x_p}{N} = f_1 x_1 + f_2 x_2 + \dots + f_p x_p$$

avec $N = n_1 + n_2 + \dots + n_p$

2. Calculs de moyennes :

Méthode : *exercices n°24-25-26* de la feuille n°1. Voir aussi l'utilisation de la calculatrice en mode STAT