

Boite à moustaches ou diagramme en boîte

I) Rappels de seconde

1) La médiane (paramètre de position)

a) Définition

La liste des N données est rangée par ordre croissant

- Si N est impair ($N = 2n + 1$) la médiane est la donnée de rang $n + 1$
- Si N est pair ($N = 2n$) la médiane est la demi somme des données de rang n et de rang $n + 1$.

Exemples :

Exemple 1 :

Un boulanger teste les masses (en grammes) de 30 baguettes qu'il vient de fabriquer, il obtient les résultats suivants :

235	235	237	238	238	239	239	239	240	241
241	243	245	247	247	249	250	205	250	250
250	251	251	253	253	255	255	255	257	260

Comme l'effectif total $N = 30$ est pair la médiane est la demi somme de la donnée de rang 15 et la donnée de rang 16 soit : $\frac{247 + 249}{2} = 248$

Exemple 2 :

Le tableau ci-dessous indique la durée (en minutes) de connexion internet par jour de 43 familles interrogées

Durée en minutes	40	60	80	120	180	200	240	300
Effectif	2	9	11	7	5	2	4	3

Comme l'effectif total $N = 43 = 2 \times 21 + 1$ est impair la médiane est la donnée de rang 22 soit 80 minutes

b) Propriétés

- Si on ajoute le même nombre k à toutes les valeurs de la série statistique, la médiane augmente de k .
- Si on multiplie toutes les valeurs de la série statistique par un même nombre k , la médiane est multipliée par k .

2) Les quartiles (paramètres de position)

La liste des N données est rangée par ordre croissant

Le **premier quartile** (Q_1) est la plus petite donnée de la liste telle qu'au moins un quart des données de la liste sont inférieures ou égales à Q_1 .

Le **troisième quartile** (Q_3) est la plus petite donnée de la liste telle qu'au moins les trois quarts des données de la liste sont inférieures ou égales à Q_3 .

Dans l'exemple 1 précédent portant sur les masses des baguettes le quart de l'effectif étant $\frac{30}{4} = 7,5$ Q_1 est la donnée de rang 8 soit $Q_1 = 239$ g et Q_3 est la donnée de rang 22 soit $Q_3 = 251$ g

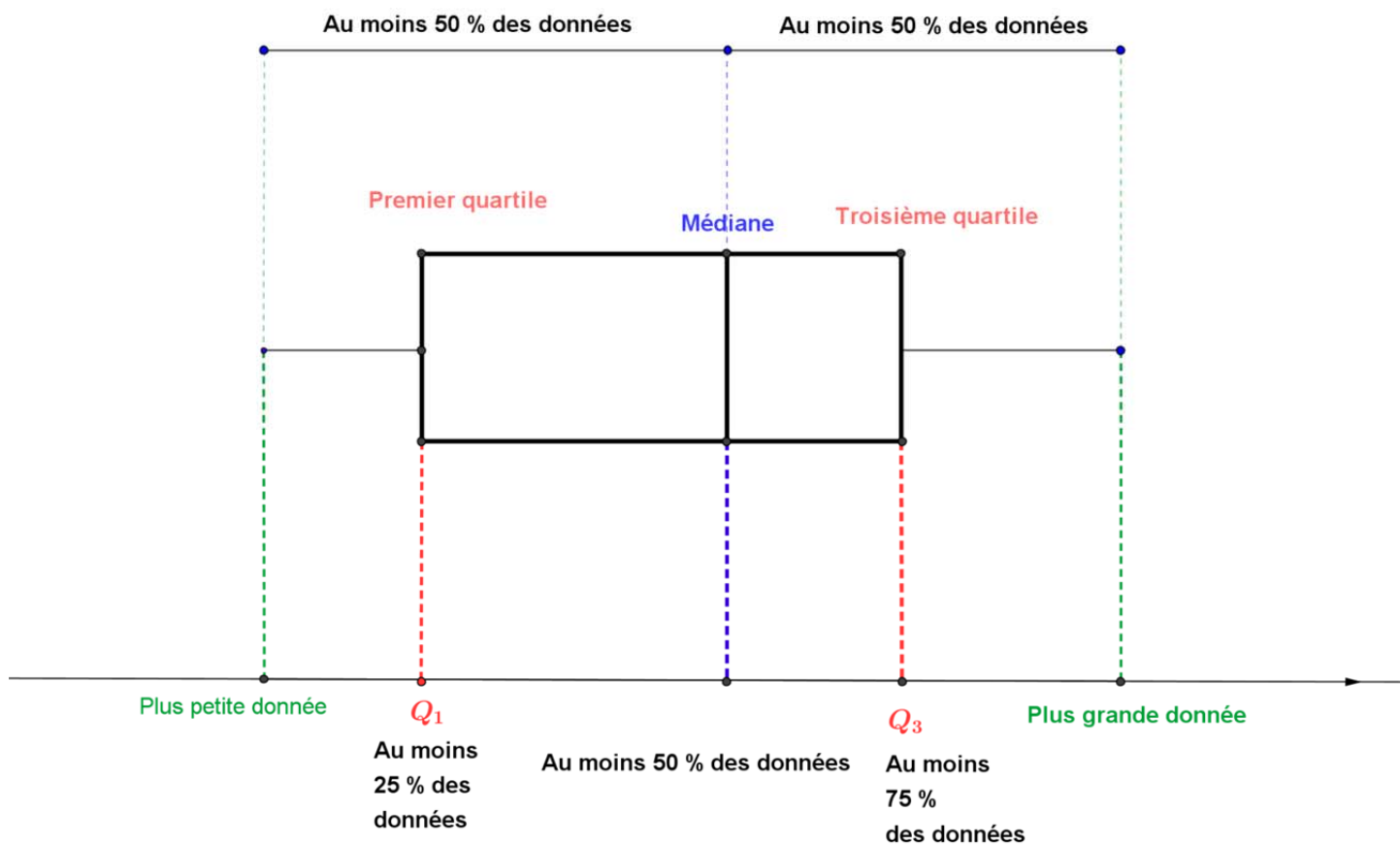
Dans l'exemple 2 précédent portant sur la durée de connexion internet le quart de l'effectif étant $\frac{43}{4} = 10,75$ Q_1 est la donnée de rang 11 soit $Q_1 = 60$ min et Q_3 est la donnée de rang 33 soit $Q_3 = 180$ min

II) Représentation graphique : Boite à moustaches

Une série statistique peut être représentée par un diagramme appelé « boite à moustache »

Définition

On appelle diagramme en boite ou boite à moustache d'une série , la représentation graphique ci-dessous. Elle est composée de deux rectangles et de deux segments dont les longueurs correspondent aux paramètres de la série, représentés sur un axe gradué



Remarques :

- Un tel diagramme est aussi appelé « diagramme en boîte », « boîte à pattes » ou encore « diagramme de Tukey » du nom de son concepteur.
- Lorsqu'on utilise une calculatrice ce diagramme porte le nom de « Box Plot ».
- Les boîtes à moustaches sont un moyen simple pour comparer un même caractère sur plusieurs séries statistiques.

Exemple 3

On a relevé les notes de 24 élèves d'une classe lors d'un examen noté sur 100 points

78	79	77	59	57	65	65	67
68	67	59	54	64	68	72	74
72	72	76	77	76	74	77	76

- 1) Déterminer la médiane et les quartiles de cette série
- 2) Dessiner la boîte à moustache de cette série
- 3) On peut comparer les résultats de cette classe avec les résultats d'une autre classe dont on sait que la note minimale est 47 , la note maximale est 85 , la médiane est 70, Q_1 est 67 et Q_3 est 76. Tracer sur le même graphique que dans la question 2 la boîte à moustache de cette nouvelle série.
- 4) Que peut-on dire sur les différences entre les deux classes ?

Solution :

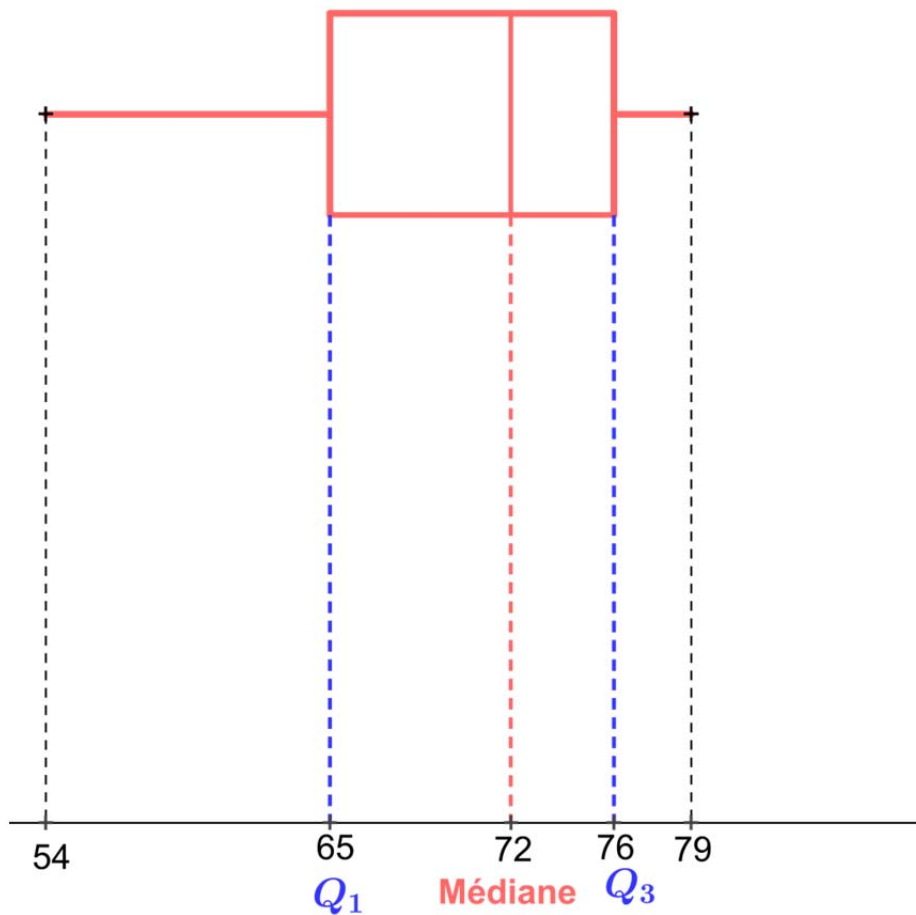
1) Trions les données de la série :

54	57	59	59	64	65	65	67
67	68	68	72	72	72	74	74
76	76	76	77	77	77	78	79

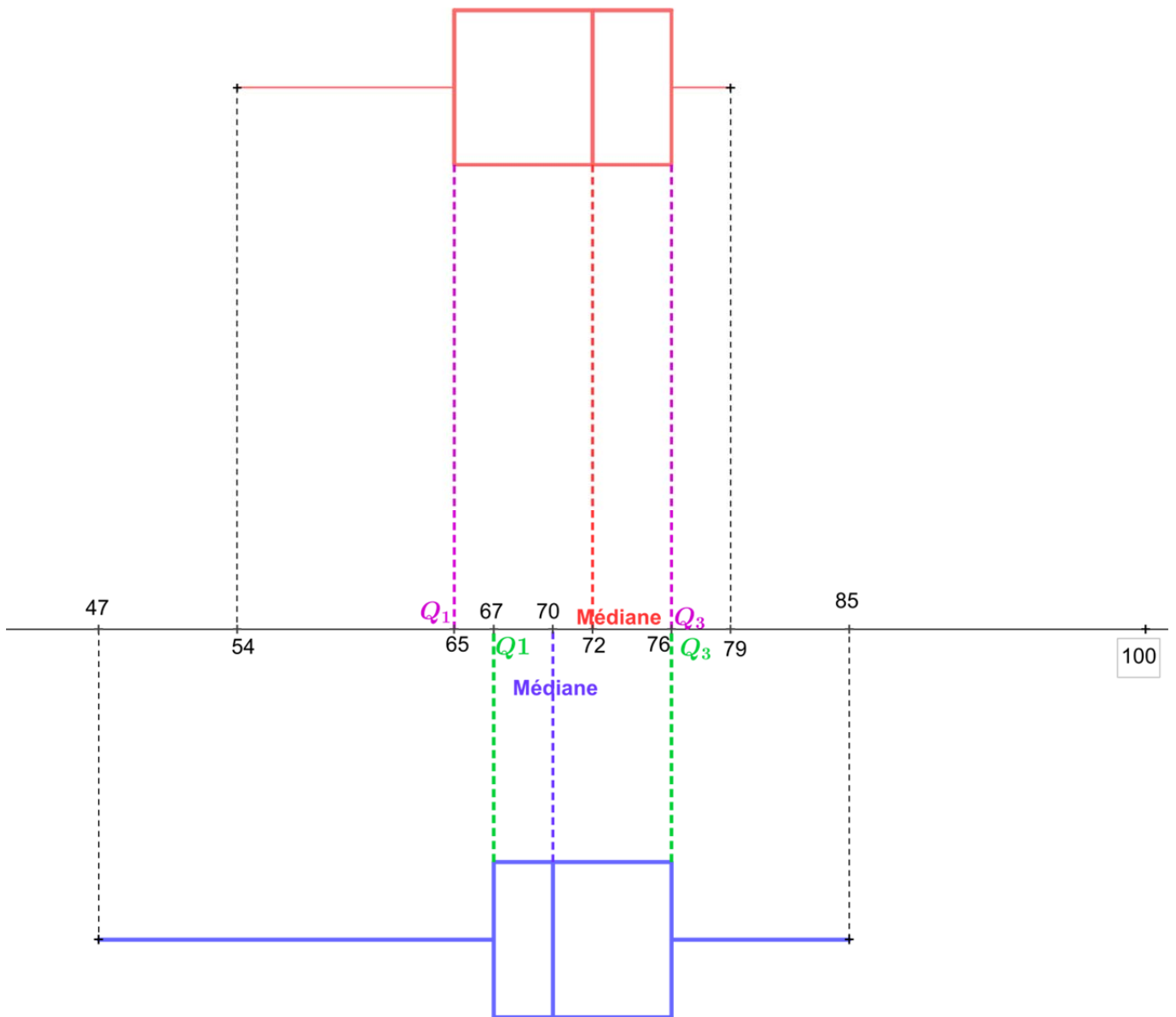
Comme il y a 24 valeurs la médiane est la moyenne entre la 12^{ème} et la 13^{ème} valeur

soit $M = \frac{72+72}{2} = 72$ le premier quartile est la 6^{ème} valeur soit $Q_1 = 65$ et le troisième quartile est la 18^{ème} valeur $Q_3 = 76$

2)



3)



4) Cette deuxième classe semble un peu plus hétérogène (un minimum inférieur et un maximum supérieur) mais pour 50 % des élèves (l'intérieur des boîtes) la deuxième classe est plus concentrée (boîte moins large).
 Pour les deux classes 75 % des élèves sont en dessous de 76 sur 100

III) Mesures de dispersion

a) L'étendue

L'étendue d'une série statistique est égale à la différence entre la plus grande et la plus petite des données de la série.

Dans l'**exemple 1** l'étendue $e = 260 - 235 = 25$

Dans l'**exemple 2** l'étendue $e = 300 - 40 = 260$

Dans l'**exemple 3** pour la première classe $e = 79 - 54 = 25$
pour la deuxième classe $e = 85 - 47 = 38$

b) l'écart interquartile

L'écart interquartile est égal à la différence $Q_3 - Q_1$

Dans l'**exemple 1** $Q_3 - Q_1 = 251 - 239 = 12$

Dans l'**exemple 2** $Q_3 - Q_1 = 180 - 60 = 120$

Dans l'**exemple 3** pour la première classe $Q_3 - Q_1 = 76 - 65 = 11$
pour la deuxième classe $Q_3 - Q_1 = 76 - 67 = 9$

