

Chapitre 13: statistiques

I – Médiane – quartiles :

1.Définitions – 1^{ère} propriété :

Définitions :

Soit une série rangée par ordre croissant

Premier quartile Q_1 : c'est la plus petite donnée de la série qui soit supérieure ou égale à au moins 25% des données.

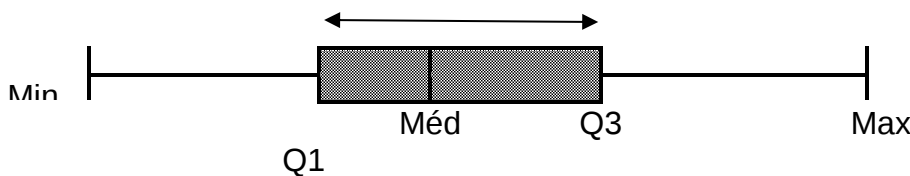
Médiane Me : c'est une valeur qui partage la série en deux ensembles de données de **même effectif**

Troisième quartile Q_3 : c'est la plus petite donnée de la série qui soit supérieure ou égale à au moins 75% des données.

L'écart interquartile est $Q_3 - Q_1$

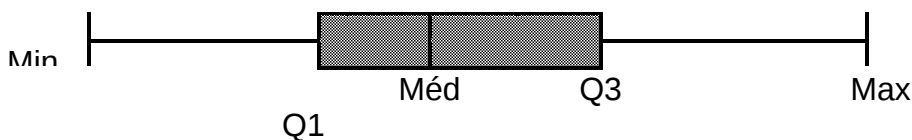
L'intervalle interquartile est $[Q_1 ; Q_3]$

On représente ces informations sous la forme d'un **diagramme en boîte** :



Remarques :

- Il faut représenter le diagramme en boîte à l'échelle.
- La médiane n'est pas nécessairement une valeur de la série contrairement aux quartiles.
- Le diagramme en boîte se nomme aussi « boîte à moustache », « diagramme de Tukey » ou « box-plot »
- Les quartiles et la médiane divisent la série en 4



Propriété :

L'intervalle interquartile contient 50 % des données.

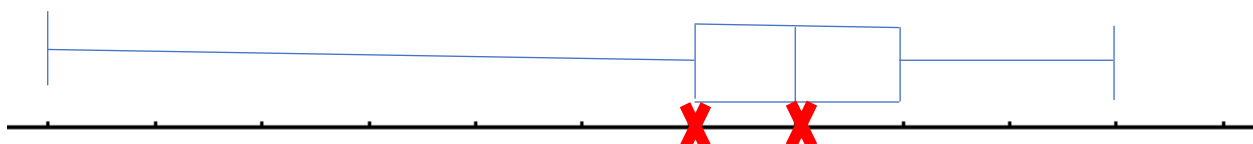
Intérêt des diagrammes en boîte :

Un diagramme en boîte est un outil qui permet de comparer efficacement une valeur donnée à toutes les valeurs d'une série statistique.

Exemple : on s'intéresse aux notes en spécialité maths en 2023.

On obtient : $Q_1=12$; Méd=14 ; $Q_3=16$

Yves a eu 12 et Tristan 14. Donner une interprétation.



75% des candidats ont eu une note supérieure ou égale à celle de Yves. 50% des candidats ont eu une note inférieure ou égale à celle de Tristan.

2.Des exemples

- On considère un échantillon statistique dont les valeurs rangées par ordre croissant sont les suivantes. Déterminer le diagramme en boîte.

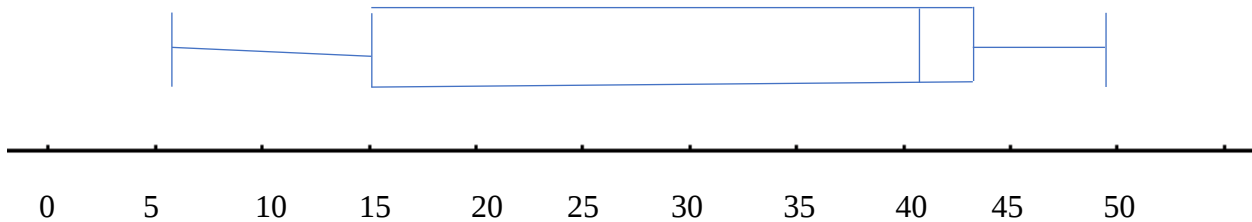
6 ; 7 ; 15 ; 36 ; 39 ; 41 ; 41 ; 43 ; 43 ; 47 ; 49 (effectif :11)

Min=6 ; Max=49

$Q_1=15$ $0,25 \times 11 = 2,75$

Méd=41

$Q_3=43$ $0,75 \times 11 = 8,25$



- On considère un échantillon statistique dont les valeurs rangées par ordre croissant sont les suivantes. Déterminer le diagramme en boîte.

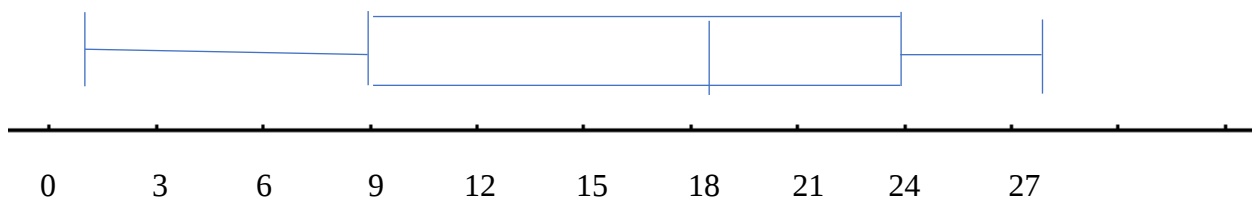
1 ; 3 ; 5 ; 7 ; 9 ; 11 ; 13 ; 15 ; 17 ; 18 ; 19 ; 20 ; 21 ; 22 ; 24 ; 24 ; 26 ; 26 ; 27 ; 28 (effectif :20)

Min=1 ; Max=28

$Q_1=9$ $0,25 \times 20 = 5$

$\text{Méd} = \frac{18+19}{2} = 18,5$

$Q_3=24$ $0,75 \times 20 = 15$



Propriété : Soit une série d'effectif N.

Si $N = 2n$ (N pair), la **médiane** est la moyenne des données de rang n et $n + 1$.

Si $N = 2n + 1$ (N impair), la **médiane** est la donnée de rang n .

3.Diagramme en boîte à partir du tableau des effectifs cumulés:

Point méthode : on dispose du tableau des effectifs

On rajoute une ligne au tableau, ligne dans laquelle on calcule les **effectifs cumulés croissants**.

La médiane s'obtient à l'aide de la propriété précédente.

On cherche à quoi correspondent 25% et 75% des données.

Q_1 s'obtient en prenant la plus petite valeur supérieure à 25% des données.

Q_3 s'obtient en prenant la plus petite valeur supérieure à 75% des données.

Exercice : construire le diagramme en boîte des 3 séries ci-dessous :

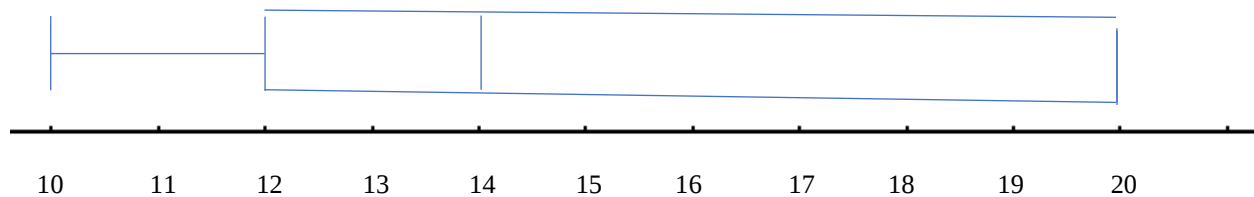
Construire le diagramme en boîte dans chacun des cas suivants (vérifier à l'aide de la calculatrice)

1.

Valeur du caractère	10	14	12	20	15
Effectif	2	3	6	5	3

Valeur du caractère	10	12	14	15	20
Effectif	2	6	3	3	5
Effectifs cumulés croissants	2	8	11	14	19

- Calcul de la médiane Me : $N=19=2\times 9+1=9+1+9$ Me est la 10^{ème} valeur de la série soit 14.
- Calcul de Q_1 : $0,25\times 19=4,75$ Q_1 est la 5^{ème} valeur de la série soit 12.
- Calcul de Q_3 : $0,75\times 19=14,25$ Q_3 est la 15^{ème} valeur de la série soit 20.
- Intervalle interquartile : [12 ; 20]
- Ecart interquartile : $20-12=8$

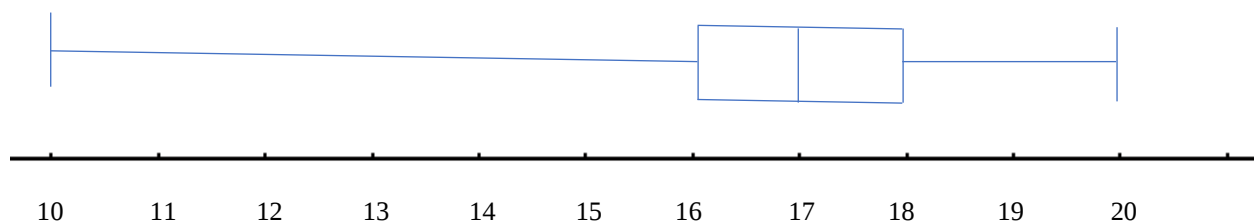


2.

Valeur du caractère	10	15	12	14	20
Effectif	2	5	7	3	3

Valeur du caractère	10	12	14	15	20
Effectif	2	7	3	5	3
Effectifs cumulés croissants	2	9	12	17	20

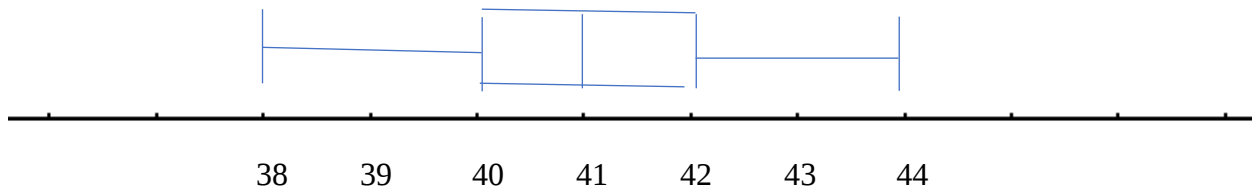
- Calcul de la médiane Me : $N=20=2\times 10$ Me est la moyenne entre la 10^{ème} valeur et la 11^{ème} valeur de la série soit $\frac{14+14}{2} = 14$.
- Calcul de Q_1 : $0,25\times 20=5$ Q_1 est la 5^{ème} valeur de la série soit 12.
- Calcul de Q_3 : $0,75\times 20=15$ Q_3 est la 15^{ème} valeur de la série soit 15.
- Intervalle interquartile : [12 ; 15]
- Ecart interquartile : $15-12=3$



3. on a relevé la pointure des chaussures des élèves garçons de 4 classes de seconde d'un lycée.

Valeurs	38	39	40	41	42	43	44
Effectifs	5	10	14	23	12	8	3
Effectifs cumulés croissants	5	15	29	52	64	72	75

- Calcul de la médiane Me : $N=75=2 \times 37+1=37+1+37$ Me est la 38^{ème} valeur de la série soit 41.
- Calcul de Q_1 : $0,25 \times 75=18,75$ Q_1 est la 19^{ème} valeur de la série soit 40.
- Calcul de Q_3 : $0,75 \times 75=56,25$ Q_3 est la 57^{ème} valeur de la série soit 42.
- Intervalle interquartile : [40 ; 42]
- Ecart interquartile : $42-40=2$



Vérification à l'aide de la calculatrice :

	TI83 premium	
Entrée des données	Aller dans stats Menu EDIT puis 1-Modifier Entrer les valeurs x_i dans L_1 Entrer les effectifs n_i dans L_2	
Obtention des paramètres	Aller dans stats Menu CALC puis 1-Var stats Taper L_1 , L_2 puis entrer (L_2 : faire 2 ^{nde} 2)	
Obtention du diagramme en boîte	Définir les paramètres de la fenêtre graphique avec fenetre Aller dans graph stats puis faire enter puis graphe	

II- Moyenne - variance-écart-type :

1.Définition de la moyenne :

Définition :

On considère la série statistique suivante :

Valeur	x_1	x_2	...	x_p	total
Effectif	n_1	n_2	...	n_p	N

La **moyenne** est le réel, noté $\bar{x}$, donné par la formule :

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{N}$$

Défi :

Anatole a eu les notes suivantes : 12 ,13 , 15 et 14.en devoirs maisons . Les devoirs maisons ont le coefficient 1.

Il a eu 7 en devoir surveillé . Les devoirs surveillés sont coefficientés 4.

Quelle note doit il avoir au prochain ds pour avoir 10 de moyenne.

2. Variance – écart-type :

Définition :

On considère la série statistique suivante :

Valeur	x_1	x_2	...	x_p	total
Effectif	n_1	n_2	...	n_p	N

La **variance** est le réel, noté V, donné par la formule :

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

L'**écart-type** est le réel, noté σ , donné par la formule :

$$\sigma = \sqrt{V}$$

Remarques :

- Autre formule : $V = \frac{n_1x_1^2 + n_2x_2^2 + \dots + n_px_p^2}{N} - \bar{x}^2$
- La **variance** et l'**écart-type** sont des réels positifs.
- L'**écart-type** est un indice de **dispersion**. Il mesure l'écart des valeurs d'une série par rapport à la moyenne. Ainsi plus l'écart type d'une série est grand et plus les données sont dispersées.
- La **variance** et l'écart type se calculent en général à l'aide de la calculatrice (effectif grand).

Exemple 1 :Déterminer l'écart-type de la série suivante :

Temps quotidiennement consacré au travail à la maison quotidien dans une classe de seconde.

Travail à la maison (en h)	0	1	2	3
Effectifs	2	9	16	5

$$\bar{x} = \frac{2 \times 0 + 9 \times 1 + 16 \times 2 + 5 \times 3}{32} = \frac{56}{32} = 1,75$$

$$V = \frac{2 \times (0-1,75)^2 + 9 \times (1-1,75)^2 + 16 \times (2-1,75)^2 + 5 \times (3-1,75)^2}{32} = \frac{20}{32} = 0,625$$

$$\sigma = \sqrt{0,625} \approx 0,791$$

Exemple 2 :Donner à l'aide la calculatrice l'écart-type de la série suivante :

Temps consacré trajet domicile/lycée dans cette classe

Trajet domicile/lycée (en min)	[0 ; 5[	[5 ; 10[	[10 ; 15[	[15 ; 30[	[30 ; 60[	[60 ; 90[
Effectifs	8	3	5	10	2	2

L1	L2		
2.5	8	Stats 1 var	
7.5	3	Xliste:L1	
12.5	5	ListeFréq:L2	
22.5	10	Calculer	
45	2		
75	2		

$$\bar{x} = 19$$

$$\sigma \approx 18,737$$

III-Résumé numérique d'une série statistique :

Une **série statistique** est une série de nombres qui sont souvent difficiles à interpréter.

On peut résumer une série statistique à l'aide d'un couple de deux paramètres

- Un paramètre qui mesure la **tendance centrale** ou la **position** (mode, médiane, moyenne)
- Un paramètre qui mesure la **dispersion** (écart interquartile , écart type)

Indicateurs de position :	Indicateurs de dispersion :
Médiane (classe médiane)	écart interquartile
Moyenne	écart type