I – Série statistique:

a) Définitions :

La **statistique** est le recueil et l'étude d'un ensemble de **données**, la plupart du temps en grand nombre et permettant de décrire l’état d’une population (personnes, objets, faits ...)

❖ Exemple 1 :

Voici le relevé du nombre de frères et sœurs pour les élèves d'une classe de seconde :
0 ; 3 ; 2 ; 4 ; 2 ; 0 ; 1 ; 1 ; 1 ; 0 ; 0 ; 2 ; 5 ; 1 ; 0 ; 1 ; 2 ; 1 ; 1 ; 1 ; 0 ; 0 ; 3 ; 1 ; 1 ; 0 ; 0 ; 2 ; 1 ; 2 ; 1 ; 1 ; 0 ; 3 ; 2

Cet ensemble de **données** est appelé une **série statistique**.

b) Vocabulaire :

L’ensemble sur lequel on travaille en statistique est appelé la **population**.
Un élément de cet ensemble est appelé un **individu.**
La propriété étudiée sur la population est appelée le **caractère**.
Ce caractère peut prendre différentes **valeurs**.

❖ Dans l’exemple 1 :

✓ La **population** étudiée est une classe.

✓ Les **individus** étudiés sont les élèves.

✓ Le **caractère** étudié est le nombre de frères et sœurs.

✓ Les **valeurs** prises par le caractère sont: 0 ; 1 ; 2 ; 3 ; 4 ; 5

Il y a 6 valeurs prises par le caractère.

c) Tableau des effectifs :

Le nombre d’individus (n_i) d’une même valeur est appelé **effectif**.
Le nombre total d’individus (N) de la population est appelé l’**effectif total**.

Méthode : Pour construire un tableau d'effectifs, on regroupe les résultats de la série dans un tableau donnant pour chaque valeur, son effectif :
- On place dans la première ligne les valeurs prises par le caractère par ordre croissant.
- On renseigne les effectifs correspondant aux valeurs dans la deuxième ligne.

Il y a 10 élèves qui ont 0 frère ou sœur d’où la première colonne :

Nombre de frères et sœurs x_i	0	1	2	3	4	5
Effectif n_i	10	13	7	3	1	1

✓ L'**effectif total N** est 35.

✓ L'**effectif** de la valeur 1 est 13.

Remarque : On vérifie l’effectif total N qui est la somme des effectifs de toutes les valeurs prises par le caractère. Sur l'exemple précédent on a bien N=10+13+7+3+1+1=35

❖ **Exercice 1 :** On donne les données suivantes donnant les âges des différents adhérents d'un club sportif :

14 ;14 ;14 ;12 ;13 ;15 ;14 ;12 ;13 ;12 ;13 ;15 ;14 ;12 ;14 ;15 ;13 ;13 ;14 ;12 ;12 ;15 ;12 ;14 ;15 ;12 ;13 ;13.

Compléter le tableau des effectifs :

Age des adhérents x_i	12	13	14	15
Effectif n_i	8	7	8	5

d) Effectifs cumulées :

L'effectif cumulé croissant de la valeur x_i, noté EEC, est la somme des effectifs de toutes les valeurs inférieures ou égales à x_i.

❖ **Exercice 2 :** Certaines séries statistiques sont regroupées en classe d'intervalles : Un biologiste étudie la taille de 100 pétales d'une fleur, il regroupe les résultats ainsi : Il y a 15 pétales qui mesurent entre 0 et 24 mm :

Taille d'un pétale en mm	[0;24[	[24;26[	[26;28[	[28;30[	[30;37[
Effectifs	15	27	20	18	20
ECC	15	42	62	80	100

II – Les paramètres d'une série statistique : Indicateurs de tendance centrale

a) La moyenne pondérée :

La moyenne pondérée de la série statistique (x_i) ci-contre est le nombre réel, noté $\boldsymbol{m}$ (ou $\bar{x}$), tel que :

$m = \frac{n_1x_1+n_2x_2+\cdots+n_px_p}{N}$ ou $m = f_1x_1 + f_2x_2 + \cdots f_px_p$

Valeur	x_1	x_2	...	x_p
Effectif	n_1	n_2	...	n_p
Fréquence	f_1	f_2	...	f_p

Effectif total : $N = n_1 + n_2 + \ldots + n_p$

❖ **Exemple 1 :**

Nombre de frères et sœurs x_i	0	1	2	3	4	5
Effectif n_i	10	13	7	3	1	1

La **moyenne** de la série est $m = \frac{10\times0+13\times1+7\times2+3\times3+4\times1+5\times1}{35}$ = 1,29 à 10^{-2} près.

En moyenne, un élève de la classe a 1,29 frère et sœur.

❖ **Exemple 2 : Calcul de la moyenne d'une série regroupée en classe d'intervalle :**

Il va falloir chercher le centre de chaque classe, en faisant la moyenne des deux nombres extrémités de chaque intervalle : Pour la classe : [24;26[on calcule (24+26)/2=25

Taille d'un pétale en mm	[0;24[	[24;26[	[26;28[	[28;30[	[30;37[
Centre de la classe	12	25	27	29	33,5
Effectifs	15	27	20	18	20

$$\mathbf{m} = \frac{12 \times 15 + 25 \times 27 + 27 \times 20 + 29 \times 18 + 33{,}5 \times 20}{100} = \frac{2587}{100} = 25{,}87$$

En moyenne un pétale mesure 25,9 mm

Remarque : La moyenne n'est pas forcément égale à une des valeurs prises par le caractère.

b) Linéarité de la moyenne :

a et b désignent des nombres réels.
Si la série (x_i) a pour moyenne m, alors la série $(ax_i + b)$ a pour moyenne $\boldsymbol{M = am + b}$.

x_i	x_1	x_2	...	x_p
$ax_i + b$	$ax_1 + b$	$ax_2 + b$	...	$ax_p + b$
Effectif	n_1	n_2	...	n_p

Démonstration

$$M = \frac{n_1(ax_1+b)+n_2(ax_2+b)+\ldots+n_p(ax_p+b)}{n_1+n_2+\ldots+n_p} = \frac{a(n_1x_1+n_2x_2+\ldots+n_px_p)+b(n_1+n_2+\ldots+n_p)}{n_1+n_2+\ldots+n_p}$$

$$M = a\times\frac{n_1x_1+n_2x_2+\ldots+n_px_p}{n_1+n_2+\ldots+n_p} + b\times\frac{n_1+n_2+\ldots+n_p}{n_1+n_2+\ldots+n_p} = am+b$$

c) La médiane :

On considère la série statistique dont les valeurs sont ordonnées par ordre **croissant.**

La médiane est le nombre **Me** qui partage une série ordonnée en deux groupes de même effectif.

- Si l'effectif est impair, $N = 2k + 1$ (k entier), la médiane est la valeur de la donnée au rang $k + 1$.
- Si l'effectif est pair, $N = 2k$ (k entier), la médiane est un nombre entre la valeur de la donnée au rang k et la valeur de la donnée au rang $k + 1$.

❖ **Exemple 1 :** On distingue deux cas, selon que l'effectif total de la série est pair ou impair.

Note de Léa (effectif pair) : 12; 9; 11; 9; 14; 9; 7; 11
On ordonne la série: 7; 9; 9; 9 ; 11; 11; 12; 14
L'effectif total est $N = 8$; $8 : 2 = 4$ donc la médiane est la demi-somme des 4ème et 5ème données

$$\underbrace{7;\ 9;\ 9;\ 9}_{\text{4 valeurs}}\ ;\ \underbrace{11;\ 11;\ 12;\ 14}_{\text{4 valeurs}}$$

Mediane= $\frac{9+11}{2} = 10$

Tous les nombres supérieurs à 9 et inférieurs à 11 partagent la série en deux séries d'effectif égal. On prend en général la moyenne des deux valeurs (ici 9 et 11), soit 10. La médiane des notes de Léa est 10, on dit aussi que c'est la note médiane.

Note de Tom (effectif impair) : 12; 9; 13; 11; 5; 7; 15
On ordonne la série: 5; 7; 9; 11 ; 12; 13; 15
L'effectif total est $N = 7$; $7 : 2 = 3{,}5$ donc la médiane est la 4ème donnée .

$$\underbrace{5;\ \ 7;\ \ 9;}_{\text{3 valeurs}}\ \ \mathbf{11}\ ;\ \underbrace{12;\ \ 13;\ \ 15}_{\text{3 valeurs}}$$

Médiane =11

Il y a 3 données inférieures ou égales à la médiane et 3 données supérieures ou égales à la médiane.

❖ **Exemple 2 : Série sous la forme d'un tableau d'effectif**

Lorsque la série est donnée par un tableau, on rajoute une ligne au tableau : la ligne des ECC

Consommation quotidienne de lait (en L)	**0,25**	**0,50**	**0,75**	**1**
effectif	**5**	**11**	**3**	**6**
ECC	**5**	**16**	**19**	**25**

L'effectif total est $N = 25$; $25 : 2 = 12{,}5$ donc la médiane correspond à la 13ème donnée .
La 13ème donnée se trouve dans la colonne 16 des effectifs cumulés croissants : elle correspond à la valeur 0,5. La médiane de la série « lait » est donc 0,5 L.

On peut dire que dans cette classe, il y a autant d'élèves qui boivent plus de 0,5 L de lait par jour que d'élèves qui en boivent moins.

d) Les quartiles :

Le **premier quartile Q_1** d'une série est la plus petite valeur prise par le caractère telle **qu'au moins 25 % des données** aient une valeur inférieure ou égale.

Le **troisième quartile Q_3** d'une série est la plus petite valeur prise par le caractère telle **qu'au moins 75 % des données** aient une valeur inférieure ou égale.

L'**écart interquartile est** la différence **Q_3 - Q_1**.

❖ **Exemple 1 :**

Notes de Léa : 7, 9, 9, 9, 11, 11, 12, 14

L'effectif total est $N = 8$.

$8 : 4 = 2$ Q_1 est donc la donc la donnée de rang 2,

soit $Q_1 = 9$

$\frac{3}{4} \times 8 = 6$ Q_3 est donc la donc la donnée de rang 6,

soit $Q_3 = 11$

L'écart interquartile est : $Q_3 - Q_1 = 11 - 9 = 2$

Notes de Tom : 5, 7, 9, 11, 12, 13, 15

L'effectif total est $N = 7$.

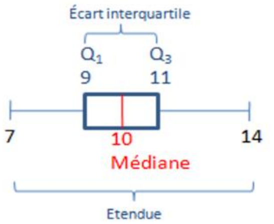
$7 : 4 = 1{,}75$ Q_1 est donc la donc la donnée de rang 2, soit $Q_1 = 7$

$\frac{3}{4} \times 7 = 5{,}25$ Q_3 est donc la donc la donnée de rang 6, soit $Q_3 = 13$

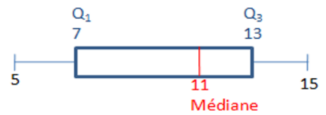
L'écart interquartile est : $Q_3 - Q_1 = 13 - 7 = 6$

Diagrammes en boîte résumant les indicateurs des deux séries :

Léa :



Tom :



❖ **Exemple 2 : Détermination des quartiles à l'aide du tableau des fréquences cumulées.**

✓ On calcule les fréquences cumulées (FCC)
✓ On lit la plus petite valeur telle que 25% (respectivement 75%) des données lui soit inférieures. C'est le 1er quartile (respectivement 3ème quartile)

Nombre de frères et sœurs x_i	0	1	2	3	4	5
Effectif n_i	10	13	7	3	1	1
ECC	10	23	30	33	34	35
FCC	28,6%	65,7%	85,7%	94,3%	97,2%	100%

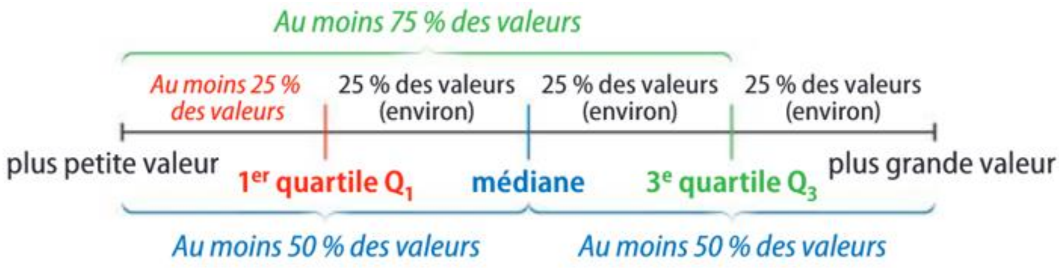
L'effectif total de cette série est $N = 35$.

On peut directement lire dans ce tableau la fréquence des données inférieures ou égales à une valeur donnée.
La médiane est : $M_e = 1$
En effet, au moins 50% des données sont inférieures ou égales à 1 et au moins 50% des données sont supérieures ou égales à 1.

Le premier quartile : $Q_1 = 0$
En effet, la plus petite valeur telle que 25% des données lui soit inférieures est 0.

Le troisième quartile : $Q_3 = 2$.
En effet, la plus petite valeur telle que 75% des données lui soit inférieures est 2.

Récapitulatif :



Remarques : Les valeurs extrêmes d'une série n'ont pas d'influence sur la médiane d'une série statistique alors qu'elles en ont sur la moyenne. Le mode d'une série est la valeur qui a le plus grand effectif.

III – Les paramètres d'une série statistique : Indicateurs de dispersion :

a) L'étendue :

L'étendue d'une série statistique est la différence entre la plus grande et la plus petite valeur .

❖ **Exemple :** On a relevé la température à différents moments d'une journée. Voici les températures en degrés Celsius : 5 ; 3 ; 6 ; 12 ; 15 ;9 ;-7.

Etendue = 15 – (-7) = 22°C

L'étendue de cette série est de 22°C.

b) La variance et écart-type :

Valeur	x_1	x_2	$\dots$	x_p
Effectif	n_1	n_2	$\dots$	n_p

La variance d'une série statistique est le réel noté V, défini par :

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

L'écart-type d'une série statistique est le réel noté s ou σ, défini par $s = \sqrt{V}$.

L'écart-type mesure la dispersion des valeurs autour de la moyenne : plus il est grand, plus les valeurs sont dispersées.

❖ **Exemple :** Soit la série statistique donnée par le tableau ci-dessous.

Valeur x_i	1	2	4	5
Effectif n_i	2	5	9	4

L'effectif total est $N = 20$

La moyenne est : $\bar{x} = 3{,}4$.

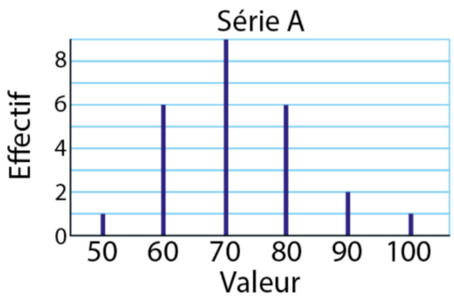
$$V = \frac{2\times(1-3{,}4)^2 + 5\times(2-3{,}4)^2 + 9\times(4-3{,}4)^2 + 4\times(5-3{,}4)^2}{20} = \frac{34{,}8}{20} = 1{,}74$$

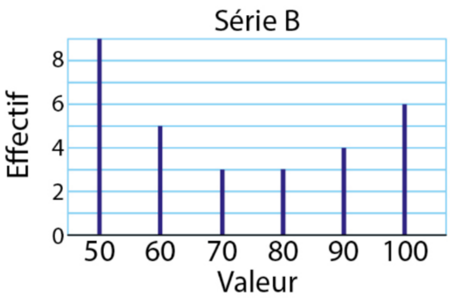
$s = \sqrt{1{,}74}$ soit $s = 1{,}32$ à 0,01 près.

c) Décrire les différences entre deux séries :

Pour étudier une série statistique, on utilise souvent un couple associant un indicateur de tendance centrale à un indicateur de dispersion ; généralement on choisit (médiane ; écart inter quartile) ou (moyenne ; écart-type).

- Exemple : On souhaite comparer les séries A et B suivantes





Compléter le tableau suivant et interpréter les résultats :

	Effectif total	Moyenne	Écart-type
Série A	25	$m_A = 72$	$s_A \approx 11{,}3$
Série B	30	$m_B = 72$	$s_B \approx 19{,}4$

La moyenne des deux séries est 72.
Les valeurs des écarts-types sont de 11,3 pour la série A et 19,4 pour la série B , ce qui confirme la plus grande dispersion des valeurs de la série B observée au niveau du graphique.

III – Etude d'une série statistique avec une calculatrice :

On cherche à calculer les différents paramètres de la série statistique ci-contre.

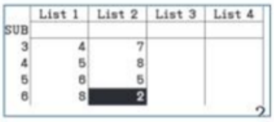
Valeur	1	3	4	5	6	8
Effectif	1	5	7	8	5	2

Avec une calculatrice Casio	Avec une calculatrice Texas

Saisie des données

Avec une calculatrice Casio

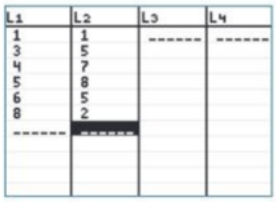
Accès au mode **statistique** : touche MENU puis sélectionner l'icône puis EXE.
Saisir les valeurs dans une liste, par exemple **List 1**, et les effectifs dans une autre liste, par exemple **List 2**.



Utiliser la touche EXE pour passer d'une valeur à l'autre et les flèches pour se déplacer dans les listes.

Avec une calculatrice Texas

Accès au mode **statistique** : touche stats.
Choisir le menu **EDIT**, puis sélectionner **1: Modifier** suivi de entrer.
Saisir les valeurs dans une liste, par exemple **L1**, et les effectifs dans une autre liste, par exemple **L2**.
Pour passer d'une valeur à l'autre ou pour changer de liste, se servir des flèches du curseur.

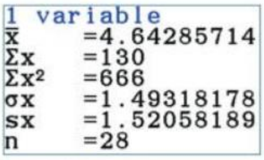


Calcul de paramètres statistiques – Séries avec effectifs

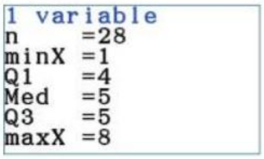
Avec une calculatrice Casio

Activer le menu **CALC** (touche F2), puis le menu **SET** (touche F6).
Choisir **List1** pour **1VarX List** et **List2** pour **1Var Freq** : **LIST** s'obtient à l'aide de la touche F2, qu'on complète ensuite avec le numéro de la liste. Valider en appuyant deux fois sur la touche EXE.

Sélectionner le menu **1-Var** (touche F1).
Les paramètres s'affichent.
$\bar{x}$ est la moyenne, Σx est la somme des valeurs, σx est l'écart-type et **n** est l'effectif total.

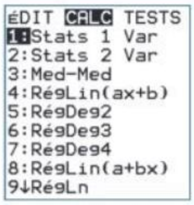


On utilise les flèches du curseur pour avoir d'autres résultats.
On peut alors lire la médiane (**Med**) les valeurs minimum (**minX**) et maximum (**maxX**) ainsi que les 1[er] et 3[e] quartiles (**Q_1** et **Q_3**).

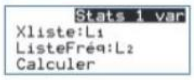


Avec une calculatrice Texas

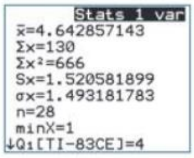
À l'aide de la touche stats, choisir le menu **CALC**, puis sélectionner **1: Stats 1-Var**, suivi de entrer :



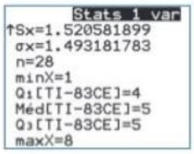
Stats 1-Vars apparaît à l'écran.
Compléter **L1** pour **Xliste** et **L2** pour **ListeFréq** (**L1** et **L2** s'obtiennent à l'aide des touches 2nde 1 et 2nde 2). Puis sélectionner **Calculer**, puis entrer.



Les paramètres s'affichent.
$\bar{x}$ est la moyenne, Σx est la somme des valeurs, σx est l'écart-type et **n** est l'effectif total.



On utilise les flèches du curseur pour avoir d'autres résultats.
On peut alors lire la médiane (**Med**) les valeurs minimum (**minX**) et maximum (**maxX**) ainsi que les 1[er] et 3[e] quartiles (**Q_1** et **Q_3**).



Calcul de paramètres statistiques – Série brute

Avec une calculatrice Casio

Quand la série est une liste dont les effectifs sont tous 1, il suffit de saisir les valeurs dans une liste, par exemple **List1** et de choisir « **1** » pour **1Var Freq**.

Avec une calculatrice Texas

Quand la série est une liste dont les effectifs sont tous 1, il suffit de saisir les valeurs dans une liste, par exemple **L1**, et de laisser vide la rubrique **ListFreq**.

Effacement des données

Avec une calculatrice Casio

Instruction **QUIT** (touches SHIFT EXIT).
Placer le curseur sur une valeur de la liste à effacer, taper F6 puis sélectionner **DEL-A** (touche F4).

Avec une calculatrice Texas

Placer le curseur sur le nom de la liste à effacer, puis utiliser la touche annul suivi de entrer.

⚠ Les valeurs des quartiles ne sont pas nécessairement les mêmes que celles obtenues avec la méthode utilisée dans le cours.