

2- Indicateurs de tendance centrale et de dispersion des séries statistiques à une variable

thème : développement durable – vie sociale et loisirs

Date de début :

1 – Que sont les indicateurs de tendance centrale ?

On utilise une mesure de tendance centrale pour obtenir une idée de l'ordre de grandeur des valeurs de l'échantillon. Le nombre obtenu indique la position où semble se rassembler ces valeurs. Nous verrons ici deux mesures de tendance centrale : la moyenne et la médiane.

2 - La moyenne

c'est le plus célèbre de tous les indicateurs. Faire une phrase de définition n'est pas si simple que cela :

"Valeur qu'aurait chacune des parties d'une somme si, la somme restante égale, toutes les parties étaient égales entre elles".

On donne couramment cette définition :

$$\bar{x} = \frac{\sum n_i x_i}{N} = \frac{n_1 x_1 + n_2 x_2 + n_3 x_3 + \dots}{N}$$

avec N : effectif total

\bar{x} : la moyenne

Σ : est la lettre grecque *sigma* majuscule qui a donné S en français et qui signifie ici ; "*faire la Somme de*"

Exemple : un parc automobiles de 150 véhicules est classé en fonction de la quantité de CO₂ émise par km . Les résultats sont donnés dans le tableau ci-dessous.

i	Classes de valeurs en g de CO ₂ par km	Centres de classes en g de CO ₂ par km x_i	Effectif n_i	$n_i x_i$
1	[100 ; 120[110	12	1320
2	[120 ; 140[130	26	3380
3	[140 ; 160[150	73	10 950
4	[160 ; 180[170	32	5440
5	[180 ; 200[190	7	1330
			N = 150	$\Sigma n_i x_i = 22420$

i joue le rôle d'un compteur"

Les centres de classe sont les **milieux des classes de valeurs**. Pour la première ligne 110 est le **milieu** de la classe [100 ; 120[.

Chaque valeur de cette colonne est obtenu en multipliant les valeurs des deux colonnes précédentes. Donc pour la 1^{ère} ligne :
1320 = 12 × 110

Calcul de la quantité moyenne de CO₂ émise : x_i

$$\bar{x} = \frac{\sum n_i x_i}{N} = \frac{22420}{150} = 149,5 \text{ g de CO}_2 \text{ à } 10^{-1} \text{ près.}$$

*Remarque : cela signifie que si tous les véhicules émettaient 149,47 g au km, l'émission totale serait toujours de 22420 g au km car :
 $149,47 \times 150 = 22420$ g au km ce qui correspond bien à l'émission réelle des 150 véhicules.*

Avec l'application Sinequanon

- ① Ouvrir le logiciel puis *Définir – série statistiques simple*
- ② Choisir l'onglet *Valeurs regroupées en classe*
- ③ Compléter comme dans l'image ci-dessous :

Statistiques à une variable

Variable non numérique | Valeurs isolées | **Valeurs regroupées en classes** | Boîtes à moustaches multiples

n°	borne inférieure	borne supérieure	Effectifs (entiers)
1	100	120	12
2	120	140	26
3	140	160	73
4	160	180	32
5	180	200	7
6			
7			
8			
9			
10			
11			
12			
13			150.00

Calculs :

Moyenne	149.467	1er décile	122.308
Écart type	18.8957	1er quartile	139.615
Effectif total	150	Médiane	150.137
Minimum	100	3ème quartile	160.938
Maximum	200	9ème décile	175

Options: ☒ Visualiser les paramètres, ☐ Médiane seulement, ☒ Tous les paramètres

Titre du graphique : [] Police (Times New Roman) [Effacer tout] [OK]

Avec la calculatrice Casio Graph 25+

① COMMENT CHOISIR LE MODE STATISTIQUE ?



Allumer la calculatrice **AC/ON**
Choisir le menu **STAT** (en haut au milieu) puis **EXE**

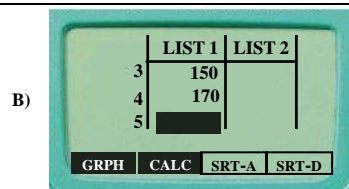
② COMMENT SAISIR LES DONNEES DU TABLEAU



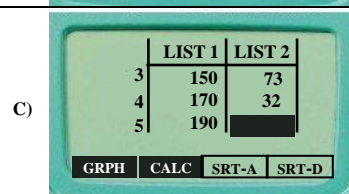
Si vous obtenez l'écran ci-contre, vous êtes prêt pour entrer les données

Sinon **▶** **F2** (DEL•A) **F4** (YES) pour chaque liste non vide.
Finir en appuyant sur **QUIT**

Remarque : il y a 6 listes visitables par déplacement avec) et (



Entrer les valeurs de la colonne x_i dans la **LIST 1** . Chaque valeur est validée par un appui sur **EXE**
Vous obtenez l'écran ci-contre.



Entrer les valeurs préalablement déterminées de la colonne n_i dans la **LIST 2** . Chaque valeur est validée par un appui sur **EXE**
Vous obtenez l'écran ci-contre.

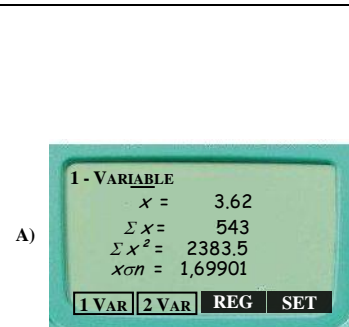
Valeurs : 12 - 26 - 73 - 32 - 7

③ COMMENT AFFECTER LES BONNES LISTES ?



Appuyer sur **F2** (CALC) puis sur **F4** (SET)
Choisir **LIST 1** **F1** pour 1 Var X, car les valeurs de x_i sont dans la liste 2
Choisir **LIST 2** **F2** pour 1 Var F, car les valeurs de n_i sont dans la liste 1
Puis valider par **EXE**

④ COMMENT OBTENIR LES RESULTATS ?



QUIT **F2** (CALC) **F1** (1 VAR) La machine affiche les résultats
1-Variable

$\bar{x} = 149.466$ est la moyenne


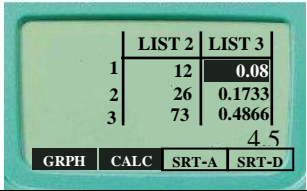
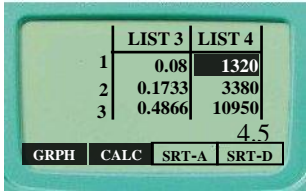
$\Sigma x = 22420$ est la somme des $n_i \times x_i$

puis en descendant vous obtenez :

$n = 150$ est l'effectif total N

$med = 150$ est la médiane

⑤ COMMENT CONSTRUIRE LES AUTRES COLONNES DU TABLEAU ?

<p>A)</p> 	<p>L'appui sur QUIT permet de retrouver l'écran 2A. Sélectionner LIST 3 à l'aide des touches de déplacement. Vous obtenez l'écran ci-contre ;</p>
Colonne des fréquences f_i	
<p>B)</p> 	<p>Appuyer successivement sur OPTN F1 (LIST) F1 (LIST) 2 ÷ 150 EXE (ici 150 est lu dans les résultats et c'est N) Vous obtenez l'écran ci-contre . Vous avez donc en liste 3 la colonne des fréquences. La machine a effectué $\frac{n}{N}$</p>
Colonne des $n_i \times x_i$	
<p>C)</p> 	<p>QUIT puis, sélectionner cette fois LIST 4 Appuyer successivement sur OPTN F1 (LIST) F1 (LIST) 1 × F1 (LIST) 2 EXE Vous obtenez l'écran ci-contre . Vous avez réalisé le produit des nombres de la liste 1 par ceux de la liste 2 et donc effectué le produit $n_i \times x_i$. <i>remarque : la somme des nombres de la liste 4 redonne la valeur de Σn trouvée sur l'écran 4A.</i></p>

Exercice : une enquête sur la consommation hebdomadaire de cigarettes au lycée donne les résultats suivants :

Nombres de cigarettes	[1 ; 5[[5 ; 9 [[9 ; 13 [[13 ; 17[[17; 21[
Nombres d'élèves	170	108	87	15	20

Remarque : les élèves consommant plus de 25 cigarettes ne sont pas pris en compte

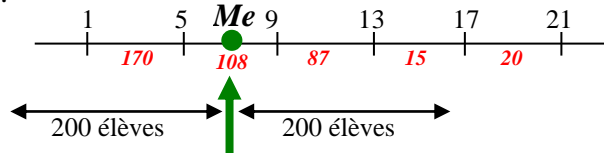
- Construire un tableau permettant de calculer le nombre moyen de cigarettes consommées par semaine . Il devra comporter une ligne (ou une colonne !) pour : les centres de classes x_i ,les effectifs n_i la fréquence f_i et les $n_i \times x_i$.
- Calculer la consommation moyenne hebdomadaire de cigarettes.

3 - La médiane Me

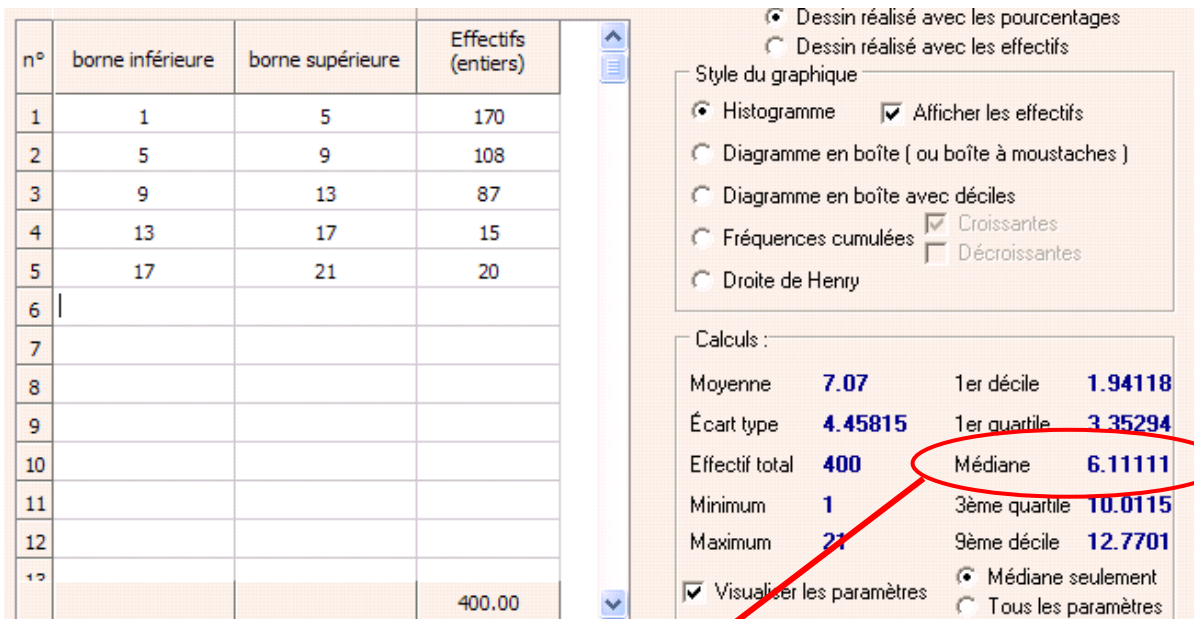
La **médiane** est un nombre qui permet de partager la population en deux groupes de même effectif. La moitié des valeurs est inférieure à Me et l'autre moitié lui est supérieure.

Exemple :

Dans l'exercice précédent la population de fumeurs est de 400 élèves. La médiane Me est le nombre de cigarettes fumées telle que 200 élèves fument **moins** de Me cigarettes et 200 fument plus de Me cigarettes.



Me est comprise dans la classe [5 ; 9] donc $Me = 7$ cigarettes



La détermination de la médiane avec l'application *Sine qua non* donne $Me = 6,11$ cigarettes.

Pourquoi une telle différence ?

4 - L'étendue e :

L'**étendue** est la différence entre les deux valeurs extrêmes prises par les valeurs du caractère étudiée.

Exemple :

○ Pour la consommation de cigarettes précédentes, les valeurs extrêmes sont 1 et 21 cigarettes, donc $e = 21 - 1 = 20$ cigarettes.

○ Pour l'émission de CO₂ du parc automobile précédent, les valeurs extrêmes sont 100 et 200 g de CO₂ au km, donc $e = 200 - 100 = 100$ g de CO₂ au km.

Exercice : tourisme mondial en 2007

Rang 2007	Pays visité	Recettes 2007 (milliards \$ US)
1	États-Unis	96,7
2	Espagne	57,8
3	France	54,2
4	Italie	42,7
5	Chine	41,9
6	Royaume-Uni	37,6
7	Allemagne	36,0
8	Australie	22,2
9	Autriche	18,9

① Calculer la recette moyenne du tourisme pour les neuf pays ci-dessous en 2007.

② déterminer le pays "médian", c'est-à-dire le pays pour lequel 4 autres pays ont des recettes supérieures aux siennes et 4 autres pays des recettes inférieures.

③ quelle est l'étendue des recettes de ces 9 pays ?

④ Question ouverte :
quel est le véritable intérêt des deux calculs précédents ?

Rang 2007	Pays visité	Arrivées 2007 (millions)
1	France	81,9
2	Espagne	59,2
3	États-Unis	56,0
4	Chine	54,7
5	Italie	43,7
6	Royaume-Uni	30,7
7	Allemagne	24,4
8	Ukraine	23,1
9	Turquie	22,2

⑤ Calculer le nombre moyen de touristes pour les neuf pays ci-dessous en 2007;

⑥ Déterminer le pays "médian", c'est-à-dire le pays pour lequel 4 autres pays ont un nombre de touristes supérieur au sien et 4 pays un nombre inférieur.

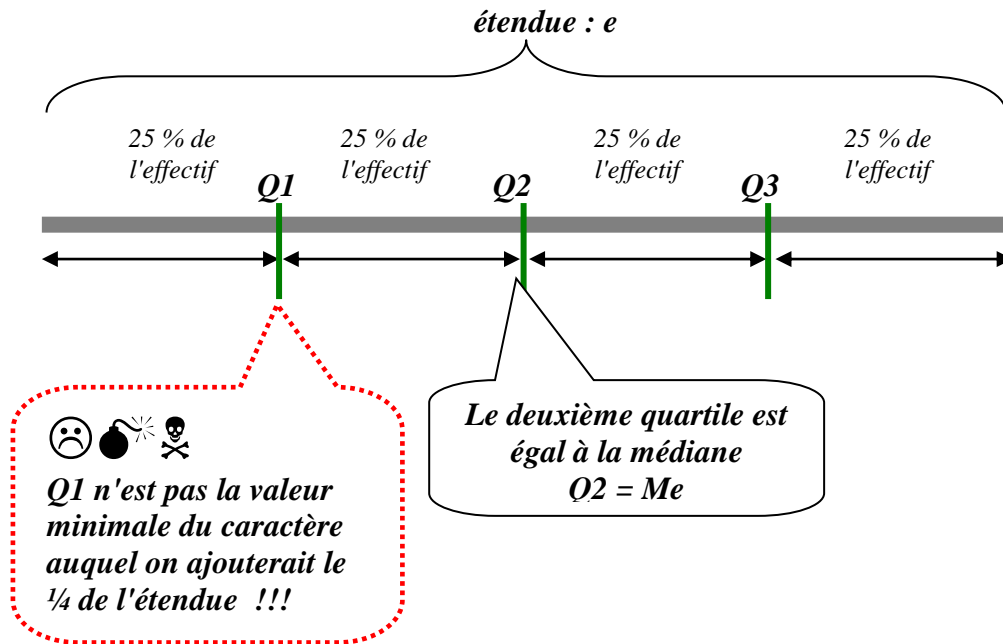
⑦ quelle est l'étendue du nombre de touristes ?

⑧ Question ouverte :
quel est le véritable intérêt des deux calculs précédents ?

⑨ Question ouverte : en comparant les résultats de la France et des États-Unis dans les deux tableaux ci-dessus, quelle conclusion chiffrée peut-on établir.

5 - Les quartiles $Q1$, $Q2$, $Q3$

les quartiles sont les valeurs du caractère qui partagent les valeurs en quatre parties d'effectifs égaux.



Exemple : si on reprend l'exercice de la consommation de cigarettes, calculons les différents quartiles.

Nombres de cigarettes	[1 ; 5[[5 ; 9 [[9 ; 13 [[13 ; 17[[17; 21[
Nombres d'élèves	170	108	87	15	20

Calcul de $Q1$: l'effectif $N = 400$, donc le quart de l'effectif est donc $\frac{N}{4} = 100$. La classe [1 ; 5[ayant un effectif de 170, c'est dans cette classe que se situe **$Q1$** . Par **simplification** on choisit le centre de cette classe comme valeur de $Q1$.

$$\mathbf{Q1 = 3 \text{ cigarettes}}$$

Calcul de $Q2 = Me$: la médiane a déjà été calculé précédemment et sa valeur simplifiée est

$$\mathbf{Me = Q2 = 7 \text{ cigarettes}}$$

Calcul de $Q3$: ce troisième quartile se situe au $\frac{3}{4}$ de l'effectif, donc pour un effectif de $\frac{3N}{4} = 300$.

Il y a $170 + 108 = 278$ élèves qui fument moins de 9 cigarettes.

Il y a $170 + 108 + 87 = 365$ élèves qui fument moins de 13 cigarettes.

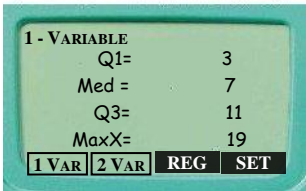
Donc c'est dans cette classe que se situe $Q3$, et par simplification la valeur du centre de classe sera adoptée.

$$\mathbf{Q3 = 11 \text{ cigarettes}}$$

Vérification Avec la Calculatrice Casio Graph 25+

En entrant la série statistique dans cette calculatrice comme précédemment permet de retrouver les résultats ci-dessus.

A)



QUIT **F2 (CALC)** **F1 (1 VAR)** La machine affiche les résultats

1-Variable

$Q1 = 3$ est le premier quartile

$Med = 7$ est le deuxième quartile et la médiane.

$Q3 = 11$ est le troisième quartile

$MaxX = 19$ est le centre de la dernière classe de valeurs.

Vérification Avec sine qua non

L'application *Sine qua non* calcule avec précision les valeurs de $Q1$, $Q2$ et $Q3$.

Calculs :

Moyenne	7.07	1er décile	1.94118
Écart type	4.45815	1er quartile	3.35294
Effectif total	400	Médiane	6.11111
Minimum	1	3ème quartile	10.0115
Maximum	21	9ème décile	12.7701

☒ Visualiser les paramètres
 ☐ Médiane seulement
☒ Tous les paramètres

$Q1 = 3,35$ cigarettes

$Me = Q2 = 6,11$ cigarettes

$Q3 = 10$ cigarettes

Que faire alors ?

- Les valeurs simplifiées et donc la méthode simplifiée ci-dessous est la seule exigée des élèves.
- Si vous donnez les valeurs exactes calculées avec un outil mathématique précis comme Sine qua non, très bien mais la méthode d'interpolation linéaire n'est pas exigible.

Exercice : Les pointures de 19 des élèves d'une classe de 2nd professionnelle Bac pro se répartissent de la manière suivante:

Pointure x_i	Nombre d'élèves n_i
38	1
39	2
40	3
41	1
42	3
43	3
44	5
45	1

- ① Calculer la pointure moyenne des élèves au dixième près.
- ② Déterminer la pointure médiane à l'unité près.
- ③ déterminer les valeurs des quartiles $Q1$, $Q2$ et $Q3$.

6 – Exercices

N°1 Les pourcentages de surfaces plantées en agriculture biologique en Europe en 2000 et 2007 pour quelques pays d'Europe sont fournis dans le tableau ci-dessous :

Année	2000	2007	Surface agricole totale en milliers d'hectares			
Allemagne	3.2	5.1	17 038			
Grèce	0.7	6.9	3 575			
Espagne	1.5	4	29 596			
France	1.3	2	27 856			
Italie	8	9	15 355			
Pays-Bas	1.6	2.5	933			
Autriche	8.1	11.7	3 375			
Suède	5.7	9.9	3 054			
Royaume-Uni	3.7	4.1	15 799			

① Calculer les pourcentages moyens de surface plantée en agriculture biologique pour 2000 et 2007 de ces pays.

② Déterminer le pays médian en 2007.

En vous aidant éventuellement des colonnes libres :

③ Déterminer le pays qui en pourcentage a le plus développé l'agriculture biologique.

④ Déterminer le pays ayant en la plus grande surface agricole biologique.

⑤ Déterminer la surface agricole biologique totale en milliers d'hectares de ces pays.

N°2 La taille des cabosses des cacaoyers d'une exploitation agricole vénézuélienne travaillant en commerce équitable et produisant du criollo, un cacao de luxe (3% de la production mondiale) est résumé dans le tableau suivant :



Tailles de cabosses en cm : x_i	Effectif : n_i	f_i (%)	ECC	
[10 ; 15[7			
[15 ; 20[50			
[20 ; 25[68			
[25 ; 30[90			
[30 ; 35[35			
	$N =$			

- ① Construire le tableau des fréquences et des effectifs cumulés croissants
- ③ Déterminer la taille moyenne et la taille médiane des cabosses, en vous aidant éventuellement de la colonne vide du tableau.
- ③ Quel est le pourcentage de cabosses dont la taille est inférieure à 25 cm ?
- ④ Imaginer une méthode simple pour évaluer le pourcentage de cabosse dont la taille est inférieure à 17,5 cm.
- ⑤ Calculer le premier et le troisième quartile, Q_1 et Q_3 . Que pensez vous de la validité de ces deux indicateurs dans cet exercice.

N°3

Les salaires de 2007 en France sont présentés, entre autres, de la façon ci-dessous :

Population salariée en 2007	Effectifs (en milliers)	Salaire (€) Premier quartile Q_1	Salaire (€) médian Me	Salaire (€) Troisième quartile Q_3
Ensemble	22 844	1 112	1 450	1 964
Cadres	3 604	2 000	2 567	3 315
Professions intermédiaires	5 782	1 386	1 700	2 117
Employés	7 624	850	1 192	1 451
Ouvriers	5 834	1 100	1 307	1 555

- ① Pour chaque case encadrée ci-dessus, faites une phrase expliquant la signification du chiffre concerné.
- ② **Plus dur !** Construire un histogramme présentant la ligne "employés" avec les effectifs en ordonnées (\uparrow) et quatre classes de valeurs inégales en abscisse (\leftrightarrow)

N°5

Soit le tableau statistique suivant :

I	Classes de valeurs	Valeur centrale x_i	Effectif : n_i	$n_i \times x_i$
1	[0 ; 4[2	7	14
2	[4 ; 8[6	21	126
3	[8 ; 12[10	13	130
4	[12 ; 16[14	36	504
5	[16 ; 20[18	3	54
			80	828

Relier les définitions de droite avec les calculs de gauche avec des flèches.

Moyenne \bar{x}	[4; 8[
Classe contenant la médiane	$20 - 0 = 20$
Classe du premier quartile Q_1	[8;12[
Effectif total N	$\frac{828}{80}$
Etendue e	6
Fréquence f_4	[12;16[
Valeur centrale x_2	$\frac{36}{80}$
Classe contenant le troisième quartile Q_3	$7+21+13+36+3$