



THÈME : Statistiques et Probabilités



Séquence 7 : Statistiques à une variable

A la fin de cette séquence je sais :



- Lire, interpréter et représenter des données sous forme de diagrammes circulaires, diagrammes en bâtons et histogrammes.
- Calculer et interpréter la médiane d'une série de données de petit effectif total.

Je connais :



- La formule de la moyenne et de la moyenne pondérée
- Les noms des différents types de représentations graphiques.
- La méthode pour déterminer la médiane d'une série de données de petit effectif total.

A. Vocabulaire

Les élèves d'une classe 4ème réalisent trois enquêtes dont les informations sont données dans les tableaux suivants.

► Tableau 1 :

Notes obtenues par ³⁰31 élèves de la classe de 1ère A lors de l'évaluation de français :

Note x_i	3	6	8	9	10	11	12	13	14	15	16	
Effectif n_i	1	2	3	7	5	3	3	1	2	2	1	30

1. Quelle est la population étudiée ? ... Les élèves de la classe de 1ère A
2. Quel est l'effectif de la population ? ... L'effectif de la population est de 30
3. Quel est le caractère étudié (variable) ? ... La note obtenue à l'évaluation de français
4. Le caractère étudié peut-il être mesurable ? ... Oui, on parle de caractère quantitatif
5. Si oui, prend-t-il des valeurs isolées (une seule valeur à la fois) ? ... Oui, c'est le cas

► Tableau 2 :

Temps consacré chaque semaine par les élèves du lycée à regarder la télévision :

Duree h_i	[0 ; 4[[4 ; 8[[8 ; 12[[12 ; 20[[20 ; 28[Total
Effectif n_i	40	80	160	200	140	620

1. Quelle est la population étudiée ? ... Les élèves du lycée
2. Quel est l'effectif de la population ? ... 620
3. Quel est le caractère étudié (variable) ? ... Le temps consacré chaque semaine à regarder la télévision
4. Le caractère étudié peut-il être mesurable ? ... Oui
5. Si oui, prend-t-il des valeurs isolées ? ... Non, il s'agit d'intervalles ou de classes

► Tableau 3 :

Types de musique préférés des élèves du lycée :

Type de musique	Rock	Rap	Techno	Variété française	Variété étrangère	Autre	Total
Effectif ni	180	120	80	120	80	40	620

1. Quelle est la population étudiée ? *Les élèves du lycée*
2. Quel est l'effectif de la population ? *620*
3. Quel est le caractère étudié (variable) ? *Type de musique préféré*
4. Le caractère étudié peut-il être mesurable ? *Non, ici c'est un caractère qualitatif*

→ Remarques :

- l'effectif de la population est souvent noté N.
- Pour **identifier le caractère étudié (la variable statistique)** on peut se poser la question suivante :
« Pour réaliser ce tableau ou cette enquête, quelle question a été posée aux personnes interrogées »

La réponse à cette question.... c'est la variable statistique !

B. Situation concrète

Deux classes de 4^{ème} ont effectué un cycle d'endurance en EPS. Leurs professeurs ont consigné les valeurs de vitesse maximale aérobie (VMA) des élèves dans les tableaux suivants :

Messieurs D et K souhaitent analyser ces résultats pour déterminer laquelle des deux classes a mieux réussi le test, en comparant non seulement le niveau global de performance suite à ce cycle d'endurance, mais aussi la répartition des résultats au sein d'une classe.

4ème Bleue - VMA en km/h										
19,8	15	18,7	9,5	19,5	18,5	21	17,8	10,5	20,5	13,8
13,5	9,5	13,5	11,5	17,5	11,7	11,5	17	10	18,3	10,8

4ème Rouge - VMA en km/h										
19	16	18	8	19	17,8	19,8	16,7	10,5	19,5	17
16	15,3	9,4	15	13,5	16	14	13,2	16,5	10	11,2

Comment comparer de manière pertinente les résultats obtenus par les deux classes ?

1) Appropriation de la situation

a) Quelle est la population étudiée dans cette étude statistique ?

Les élèves de deux classes de 4^{ème}

b) Quel est le caractère étudié (la variable statistique) ?

La VMA (en km/h)

c) De quel type de caractère s'agit-il ?

C'est un caractère quantitatif

2) Hypothèses

▪ **Proposer**, par écrit, vos hypothèses afin de répondre à la question de départ. Attention, il ne s'agit pas de mettre en œuvre une méthode mais de proposer des pistes de méthodes !

1. Niveau global de performance : faire une moyenne

2. Répartition des résultats : faire une médiane* et faire des intervalles (exemple : regrouper les chiffres 5 par 5 : $[0; 5[$; $[5; 10[$; $[10; 15[$ etc.) et trouver la différence entre le plus rapide et le moins rapide

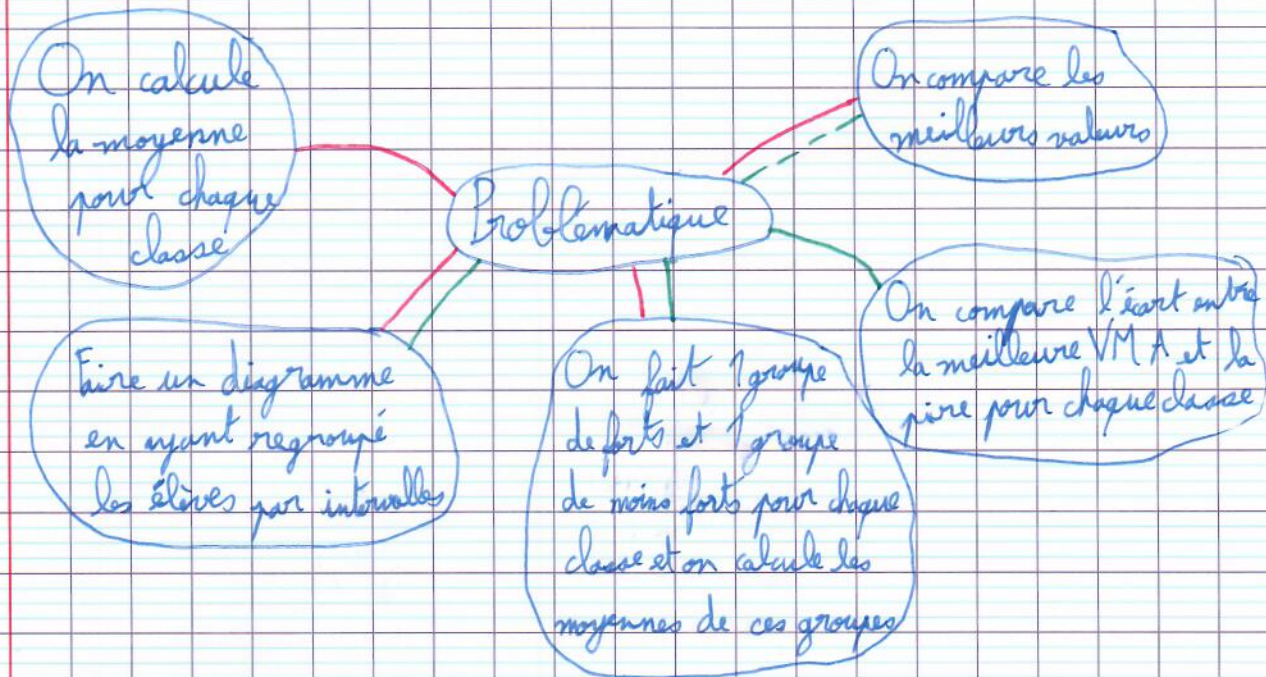
* et la comparer à la moyenne

▪ Mise en commun en binôme. Comparer vos propositions et tenter de retenir les pistes qui vous paraissent les plus pertinentes.

Mise en commun avec la classe

Performance

Répartition
des résultats



4) Exploitation des résultats et bilan

Les représentations graphiques nous permettent d'avoir une idée de la répartition des performances. On observe beaucoup d'élèves forts et beaucoup d'élèves lents dans la classe bleue (forte polarisation). Dans la classe rouge, les résultats sont beaucoup plus homogènes à l'exception de l'intervalle 16 à 18 km/h qui regroupe une dizaine d'élèves.

La formule de la moyenne sera à revoir.

D'après nos calculs, les VMA moyennes sont quasiment identiques pour les deux classes. En revanche, les plus rapides de la 4^{ème} bleue sont meilleurs que ceux de la 4^{ème} rouge et les plus lents de la 4^{ème} bleue sont plus lents que ceux de la 4^{ème} rouge. Cela confirme la tendance observée sur les graphiques.

C. Fréquence d'une valeur

► **Définition :** La fréquence d'une valeur est égale au quotient :

$$\text{fréquence} = \frac{\text{effectif d'une valeur}}{\text{effectif total}}$$

Une fréquence est un nombre inférieur à 1. On peut aussi l'exprimer en pourcentage.

- **Exemple 1 :** Dans le tableau 1 page 1 : la fréquence des élèves ayant obtenu 12/20 est $\frac{3}{30} = \frac{1}{10} = \frac{10}{100} = 10\%$
- **Exemple 2 :**

Lors d'un sondage, on a demandé aux élèves d'un collège combien de fois par semaine ils utilisent le cahier de textes sur Pronote. Le tableau indique les réponses.

Nombre d'utilisations	0	1	2	3	4	5	6
Effectifs	20	42	60	64	26	16	12

- Quelle est la fréquence des élèves qui utilisent 4 fois Pronote par semaine?

$$\frac{26}{240} \approx 0,1083 \approx \frac{10,83}{100} \approx 10,83\%$$

- Si on transforme cette fréquence en pourcentage : 10,83%

D. Indicateurs de tendance centrale

- Les indicateurs de tendance centrale permettent de déterminer des valeurs particulières de la variable statistique autour de laquelle des données ont tendance à se rassembler.
- Nous en utiliserons deux : la médiane et la moyenne.

1) La médiane

a) Définition

► **Définition :** La médiane d'une série de valeurs rangées par ordre croissant est la valeur de la variable qui partage l'effectif total en deux groupes de même effectif. Elle est notée Me .

→ **Signification :** pour une série de valeurs rangées par ordre croissant, 50% des valeurs de la série sont inférieures ou égales à la médiane et 50% des valeurs sont supérieures ou égales à la médiane.

b) Méthode de détermination

On souhaite déterminer la médiane d'une série de données d'effectif total n .

Étape 1 : Classer les données de la série par ordre croissant.

Étape 2 : Repérer ou calculer la valeur médiane.

- **Si n est impair :**

On identifie facilement la valeur centrale qui correspond au rang $\frac{(n+1)}{2}$ dans la liste triée.

▪ **Exemple 3 :** 5 - 5 - 6 - 7 - 8 - 9 - 11 $M_e = 7$

- **Si n est pair :**

Trouver les deux valeurs centrales de la série. **la médiane est la moyenne des deux valeurs centrales de la série.**

▪ **Exemple 4 :** 8 - 9 - 11 - 12 - 14 - 17 - 18 - 18 $M_e = \frac{12+14}{2} = 13$

c) Applications directes

Chaque tableau regroupe les notes obtenues par différentes classes lors de la même évaluation.

▪ Déterminer la médiane des séries statistiques suivantes autrement dit la note médiane obtenue par chaque classe.

Note	2	4	5	7	9	Total élèves
Effectif	1	4	3	4	1	13

2 - 4 - 4 - 4 - 4 - 5 - 5 - 5 - 7 - 7 - 7 - 7 - 9

$M_e = 5$

Note	2	4	5	7	9	Total élèves
Effectif	2	3	3	5	3	16

2 - 2 - 4 - 4 - 4 - 5 - 5 - 5 - 7 - 7 - 7 - 7 - 7 - 9 - 9 - 9

E. La moyenne

1) La moyenne

\bar{x}

$$M_e = \frac{5+7}{2} = 6$$

► **Définition :** La moyenne, notée \bar{M} , d'une série statistique est un paramètre qui donne une tendance centrale de la série.

► **Calcul :** On additionne toutes les valeurs du caractère de la série statistique et on divise cette somme par l'effectif total de la série de données (nombre de valeurs de la série de données).

► Pour N valeurs $(x_1; x_2; \dots; x_n)$, la moyenne \bar{M} est donnée par $\bar{M} = \frac{x_1 + x_2 + \dots + x_n}{N}$

▪ Lorsque la série contient beaucoup de données, cet indicateur de tendance centrale sera calculé grâce aux fonctions statistiques d'une calculatrice ou à l'aide d'un tableur. Dans des cas plus simples nous pouvons la calculer à la main.

▪ **Exemple 5 :** On donne les tailles de cinq joueurs d'une équipe de Basket. 1,95m ; 2,10m ; 1,85m ; 1,92m ; 1,81 m

- Déterminer la taille moyenne des joueurs de l'équipe. $M = 1,926 \text{ m}$
- Déterminer la médiane de la série statistique. $M_e = 1,92 \text{ m}$

2) La moyenne pondérée

a) Méthode de calcul

On additionne les produits de chaque valeur de la variable par son effectif.
On divise la somme obtenue par l'effectif total de la série statistique.

Si n_1, n_2, \dots, n_p sont les effectifs associés aux valeurs x_1, x_2, \dots, x_p du caractère et N l'effectif total de la série statistique alors la moyenne pondérée se calcule comme suit :

$$M = \frac{n_1 x_1 + n_2 x_2 + \dots + n_p x_p}{N}$$

▪ **Exemple 6 :** On retrouve les notes obtenues par les élèves d'une classe lors d'une évaluation. Déterminer la moyenne de cette série statistique, autrement dit la moyenne obtenue par la classe à cette évaluation.

Note (x_i)	2	4	5	7	9	Total élèves
Effectif (n_i) (nbre d'élèves)	1	4	3	4	1	13

$$M = \frac{1 \times 2 + 4 \times 4 + 5 \times 3 + 7 \times 4 + 9 \times 1}{13} \approx 5,38$$

→ Remarques :

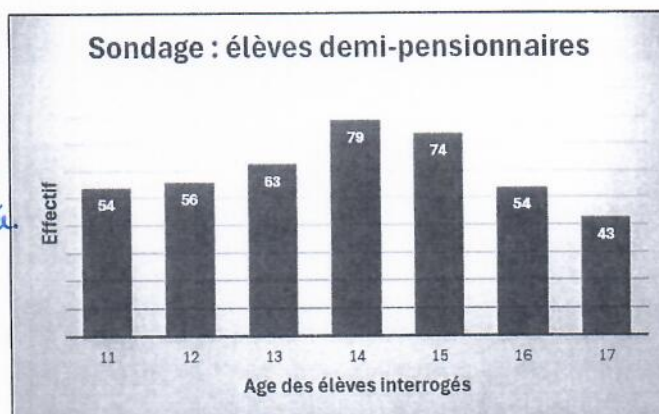
- La moyenne prend en compte toutes les valeurs et peut-être très **influencée par des valeurs extrêmes voire aberrantes du caractère**. Par exemple, si un élève obtient une note très élevée ou très basse, cela peut faire varier la moyenne de manière significative.
- La médiane **n'est pas sensible aux valeurs extrêmes**. C'est un indicateur plus adapté lorsque les valeurs de la variable ne sont **pas homogènes** ou lorsque les données **présentent des valeurs aberrantes**. Par exemple, dans un relevé quotidien de températures pendant un mois, une température très basse par rapport aux autres relevés n'influencera pas la médiane.

F. Étendue

► **Définition :** L'étendue (ou amplitude) d'une série statistique est la différence entre la plus grande et la plus petite valeur de la variable statistique étudiée.
On la note : e

▪ **Exemple 7 :** le diagramme ci-contre été obtenu en effectuant un sondage auprès des élèves d'un Collège-Lycée.

Le caractère étudié (la variable statistique) est l'âge des élèves et son étendue est $e = 6$ ans
 $e = 17 - 11 = 6$ ans



G. Choisir la bonne représentation graphique

1) Différentes représentations des données

- On choisira une représentation graphique en fonction du type de caractère étudié et de l'objectif de la série statistique.

Diagramme à lignes brisées	Diagrammes circulaires ou en secteurs
<ul style="list-style-type: none"> Pour mettre en avant une évolution chronologique. L'unité de temps est en abscisse. 	<ul style="list-style-type: none"> Convient pour tous les caractères Les mesures des angles sont proportionnelles aux effectifs ou aux fréquences.
<p>Température à midi</p>	<p>Ventes par représentant</p>
Diagramme en barres (ou bâtons)	Histogramme
<ul style="list-style-type: none"> Peut être utilisé pour un caractère qualitatif ou quantitatif dont les valeurs sont isolées. 	<ul style="list-style-type: none"> Peut être utilisé pour un caractère quantitatif continu présenté sous la forme de classes (intervalles).
<ul style="list-style-type: none"> Les hauteurs des colonnes (ou bâtons) sont proportionnelles aux effectifs ou aux fréquences. 	

2) Le cas du diagramme circulaire

- Dans un diagramme circulaire (familièrement appelé « camembert ») les angles de chaque secteur sont proportionnels aux effectifs ou aux fréquences (pourcentages) de la série statistique étudiée.

Pour calculer l'angle de chaque secteur, on utilise donc un tableau de proportionnalité dans lequel l'angle de 360° correspond à la fréquence 100% (ou à l'effectif total de la population étudiée).

Fréquence en %	100	20	40	35	5
Angle sur le diagramme en degrés	360	72	144	126	18

Proportionnalité fréquence/angles

