Objectifs

Être capable :

1. de calculer une moyenne, un écart type ;
2. de calculer une médiane,
3. de calculer une étendue, des quartiles, un interquartile ;
4. de calculer des déciles, un interdécile ;
5. d'exploiter un diagramme en boite à moustaches.

I. Vocabulaire et représentations graphiques

1) Vocabulaire

Définitions

Une **population** est un ensemble de personnes ou d'objets, appelés **individus**, définis par une propriété commune. Pour une population choisie, on peut étudier un caractère de ses individus, il est :

- **quantitatif** quand il est mesurable :
  - → **discret** si les valeurs sont des nombres isolés ;
  - → **continu** si les valeurs ne sont pas isolées. Les valeurs sont regroupées en **classes** ou **intervalles** $[a\,;b[$
- **qualitatif** quand il n'est pas mesurable.

L'**effectif** n_i est le nombre d'individus correspondant à une valeur du caractère. L'**effectif total** N est le nombre total d'individus de la population étudiée. Pour chaque valeur du caractère la **fréquence** f_i est calculée en divisant l'effectif correspondant à la valeur par l'effectif total ($\frac{n_i}{N}$).

2) Représentation graphique

Á retenir

- Le **diagramme en secteurs (ou circulaire)** est une représentation adaptée une série à **caractère qualitatif**.
- Le **diagramme en bâtons (ou en barres)** est une représentation adaptée pour une série à **caractère quantitatif discret**.
- L'**histogramme** est utilisé pour représenter les séries à **caractère quantitatif continu**.

II. Indicateurs de tendance centrale

1) Moyenne

Activité 1 page 76

1° Calcul de la distance moyenne à la piscine pour cet ensemble de neuf lycées :

$$\begin{aligned}\bar{x} &= \frac{1{,}8+1{,}0+20{,}2+0+0{,}6+0+0{,}8+2{,}6+0}{9}\\ \bar{x} &= \frac{27}{9}\\ \bar{x} &= 3\end{aligned}$$

La distance moyenne à la piscine pour ces neuf lycées est de 3 km, il faut donc les classer dans la troisième catégorie, distance supérieure à 2,5 km.

2° Calcul de la distance moyenne à la piscine pour cet ensemble de neuf lycées en prenant en compte le nombre d'élèves :

$$\begin{aligned}\bar{x} &= \frac{930\times 1{,}8+1130\times 1{,}0+\ldots+1250\times 0}{930+1130+\ldots+530+1250}\\ \bar{x} &= \frac{15072}{10770}\\ \bar{x} &\approx 1{,}4\end{aligned}$$

En tenant compte du nombre d'élèves de chaque lycée, on obtient une distance moyenne à la piscine d'environ 1,4 km.

3° Pour estimer les frais supplémentaires créés par les déplacement entre les lycées et les piscines il faut tenir compte du nombre d'élèves donc la deuxième distance moyenne est la plus appropriée.

4° Calcul de la distance moyenne à la piscine en prenant en compte les nouveaux effectifs :

$$
\begin{aligned}
\bar{x} &= \frac{1450 \times 1{,}8 + 1130 \times 1{,}0 + ... + 530 \times 0}{930 + 1130 + ... + 530 + 1250} \\
\bar{x} &= \frac{48864}{10770} \\
\bar{x} &\approx 5{,}54
\end{aligned}
$$

Cette modification de la répartition des élèves dans les lycées multiplie la distance moyenne à la piscine par 3,2.

Á retenir

Pour calculer une moyenne d'une série quantitative de N éléments, notée $\bar{x}$, on distingue trois cas :

- 1^{er} **cas :** la population est donnée par la liste de ses N éléments : x_1, x_2, ..., x_N.

$$
\begin{aligned}
\bar{x} &= \frac{Somme\ des\ élements}{Nombre\ d'éléments} \\
\bar{x} &= \frac{x_1 + x_2 + ... + x_N}{N}
\end{aligned}
$$

- 2^e **cas :** la population est donnée par le tableau des effectifs n_i de chacune des p valeurs x_i : x_1, x_2, ..., x_N.

$$
\begin{aligned}
\bar{x} &= \frac{Somme\ des\ produits}{Effectif\ total} \\
\bar{x} &= \frac{x_1 \times n_1 + x_2 \times n_2 + ... + x_p \times n_p}{N}
\end{aligned}
$$

- 3^e **cas :** la population est donnée par le tableau des effectifs n_i de chacune des p classes $[a_i; b_i[$ de centre $c_i = \frac{a_i+b_i}{2}$: On se ramène au cas discret en remplaçant chaque classe par son centre.

$$
\begin{aligned}
\bar{x} &= \frac{Somme\ des\ produits}{Effectif\ total} \\
\bar{x} &= \frac{c_1 \times n_1 + x_2 \times n_2 + ... + c_p \times n_p}{N}
\end{aligned}
$$

Exemple

On a relevé la taille en cm de 20 personnes :
Dans ce cas, il faut déterminer le centre de la classe.

Classe	$[145;155[$	$[155;165[$	$[165;175[$	$[175;185[$	$[185;195[$
Centre de classe	150	160	170	180	190
Effectif	2	5	8	4	1

On remarque que l'effectif total est 20, la moyenne des tailles est :

$$\begin{aligned}\bar{x} &= \frac{150\times 2+160\times 5+170\times 8+180\times 4+190\times 1}{20}\\ \bar{x} &= 168{,}5\end{aligned}$$

2) Médiane

Définition

Soit une série quantitative ordonnée. La médiane, notée Me est un nombre qui sépare la population en **deux sous-ensembles de même effectif**.

Méthode

Calculer la médiane d'une série

1. Classer les valeurs par ordre croissant ;
2. Déterminer l'effectif total de la série N ;
3. - Si N est impair, alors la médiane est la valeur de rang $\frac{N+1}{2}$
   - Si N est pair, alors la médiane se trouve entre les valeurs de rang $\frac{N}{2}$ et $\frac{N}{2}+1$.

Exemples

1. On considère la série des notes suivantes :
   10 ; 12 ; 15 ; 17 ; 12,5 ; 9 ; 13 ; 18,5 ; 16,5
   - Je range, ces notes par ordre croissant : 9 ; 10 ; 12 ; 12,5 ; 13 ; 15 ; 16,5 ; 17 ; 18,5 ;
   - Il y a neuf notes, donc $N = 9$, c'est un nombre impair ;
   - $\frac{9+1}{2} = 5$, donc la médiane est la 5^{eme} note ;
   - $Me = 13$.

III. Indicateurs de dispersion

1) Étendue

Définition

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

2) Écart type

Définition

L'**écart type** σ (sigma), fourni par la calculatrice ou le tableur, mesure la dispersion de la série autour de la moyenne $\bar{x}$.
Plus l'écart type σ est grand, plus les valeurs sont «**dispersées**» autour de la moyenne.
Inversement, plus l'écart type σ est grand, plus les valeurs sont «**resserrées**» autour de la moyenne.

3) Quartiles

Définition

- Le **premier quartile** Q_1, est la plus petite valeur à laquelle un quart (ou 25 %) des valeurs sont inférieures ou égales.
- Le **troisième quartile** Q_3, est la plus petite valeur à laquelle trois quarts (ou 75 %) des valeurs sont inférieures ou égales.
- L'**écart interquartile** $Q_3 - Q_1$ est la différence entre les 3^e et 1^{er} quartiles : $Q_3 - Q_1$. Il regroupe au moins 50 % des effectifs de la série avec un nombre égal de valeurs réparties de part et d'autre de la médiane Me.

Méthode

Calculer le premier et le troisième quartile d'une série

1. Classer les valeurs par ordre croissant ;
2. Déterminer l'effectif total de la série N ;
3. - Le premier quartile Q_1 est la valeur de rang $0{,}25 \times N$ (arrondi à l'entier supérieur au besoin).
   - Le troisième quartile Q_3 est la valeur de rang $0{,}75 \times N$ (arrondi à l'entier supérieur au besoin).

Exemple

On considère la série des notes suivantes :
10 ; 12 ; 15 ; 17 ; 12,5 ; 9 ; 13 ; 18,5 ; 16,5

- Je range, ces notes par ordre croissant : 9 ; 10 ; 12 ; 12,5 ; 13 ; 15 ; 16,5 ; 17 ; 18,5 ;
- Il y a neuf notes, donc $N = 9$;
- $0{,}25 \times 9 = 2{,}25$ et $0{,}75 \times 9 = 6{,}75$, donc le premier quartile est la 3^{eme} note et le troisième quartile est la 7^{eme} note ;
- $Q_1 = 15$ et $Q_3 = 16{,}5$.

4) Diagramme en boite

Définition

Le **diagramme en boîte à moustaches** est un dessin à l'échelle, où la «**boîte**» est un rectangle limité par Q_1 et Q_3, et regroupe donc 50 % des valeurs.
La médiane Me est repérée par un segment dans le rectangle.
Le minimum x_{min} et le maximum x_{max} correspondent aux extrémités des «**moustaches**».