Chapitre V : Les statistiques

I. Etude d'un caractère discret

a) moyenne

Exemple : Voici les notes à un devoir commun des 23 élèves de Seconde A.

0 – 12 – 9 – 10,5 – 2,5 – 8 – 3 – 8 – 3 – 14 – 6 – 2,5 – 6 – 16,5 – 14 – 6 – 9 – 3 – 6 – 14 – 12 – 3 – 9

On se propose de ranger ces valeurs dans un tableau :

note	0	2,5	3	6	8	9	10,5	12	14	16,5
effectif	1	2	4	4	2	3	1	2	3	1

La moyenne des notes à ce devoir est :

$$\overline{x} = \frac{0 \times 1 + 2,5 \times 2 + 3 \times 4 + 6 \times 4 + 8 \times 2 + 9 \times 3 + 10,5 \times 1 + 12 \times 2 + 14 \times 3 + 16,5 \times 1}{24} \approx 7,7$$

Définition : On se donne une série statistique :

Valeur	x_1	x_2	…	x_p
effectif	n_1	n_2	…	n_p

La moyenne de cette série est le nombre réel, noté $\overline{x}$, tel que :

$$\overline{x} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_p x_p}{N}$$ où N est l'effectif total ; $N = n_1 + n_2 + \ldots + n_p$.

On note souvent $N = \sum_{i=1}^{p} n_i$ (somme des n_i de $i = 1$ à p) ; $\overline{x} = \frac{1}{N} \sum_{i=1}^{p} n_i x_i$.

• Comme la moyenne au devoir de la classe a été relativement faible, le professeur propose trois possibilités pour augmenter cette moyenne :

→ Il décide d'ajouter 1,5 points à chaque devoir :
Que devient la moyenne ? Elle augmente de 1,5 points.

Propriété : k désigne un réel quelconque. Si $\overline{a}$ est la moyenne des nombres $a_1, a_2, \ldots, a_n$, alors la moyenne des nombres $a_1 + k, a_2 + k, \ldots, a_n + k$ est égale à $\overline{a} + k$.

→ Il décide de multiplier chaque note par 1,2.
Que devient la moyenne ? Elle est multipliée par 1,2.

Propriété : λ désigne un réel quelconque.
Si $\overline{a}$ est la moyenne des nombres $a_1, a_2, \ldots, a_n$, alors la moyenne des nombres $\lambda a_1, \lambda a_2, \ldots, \lambda a_n$ est égale à $\lambda \overline{a}$.

→ Il décide de faire une interrogation notée sur 10 dont voici les résultats des élèves classés dans le même ordre que pour le premier devoir : 6 – 10 – 8 – 8,5 – 4 – 7,5 – 6 – 7,5 – 8 – 9,5 – 6,5 – 7,5 – 6,5 – 10 – 9 – 9 – 7,5 – 9 – 9 – 10 – 10 – 6 – 9,5.

- Quelle est la moyenne sur 10 de cette interrogation ?

$\overline{y} \approx 8$.

- Ajouter chaque note de l'interrogation à la note du devoir commun pour obtenir une troisième série de notes, mais sur un total de 30. Déterminer la moyenne de cette nouvelle série.

La moyenne sur 30 de ces notes est environ 15,7, c'est à dire $\overline{x} + \overline{y}$.

Propriété : Si $\overline{a}$ désigne la moyenne des nombres $a_1, a_2, \ldots, a_n$ et $\overline{b}$ celle des nombres $b_1, b_2, \ldots, b_n$, alors la moyenne des nombres $a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n$ est égale à $\overline{a} + \overline{b}$.

→ Dans la première série de notes, 0 est une valeur exceptionnelle. On peut calculer la moyenne de la série privée de cette valeur. On dit qu'il s'agit d'une **moyenne élaguée**.
Cette moyenne est alors égale à environ 8.

b) autres calculs de moyenne

• A partir des moyennes de sous-groupes

pour le même devoir commun, la classe de seconde B contenant 27 élèves a eu une moyenne de 9,2. Quelle est la moyenne à ce devoir des élèves des deux classes ?

$$m = \frac{23 \times 7{,}7 + 27 \times 9{,}2}{23 + 27} = 8{,}51.$$

Propriété : On répartit N nombres en k **sous-groupes disjoints**. Le premier groupe contient N_1 éléments, de moyenne m_1, le deuxième N_2 éléments, de moyenne m_2, …, le dernier N_k éléments, de moyenne m_k. Alors la moyenne m des N nombres peut-être calculée par :

$$m = \frac{m_1 N_1 + m_2 N_2 + \ldots + m_k N_k}{N}$$

• A partir de la distribution des fréquences

notes	0	2,5	3	6	8	9	10,5	12	14	16,5
effectif	1	2	4	4	2	3	1	2	3	1
fréquence	0,043	0,087	0,174	0,174	0,087	0,131	0,043	0,087	0,131	0,043

La moyenne est alors $\overline{x} \approx 0 \times 0{,}043 + 2{,}5 \times 0{,}087 + 3 \times 0{,}174 + \ldots \approx 7{,}7$.

Propriété : On se donne la série statistique :

Valeurs	x_1	x_2	…	x_p
effectif	n_1	n_2	…	n_p
fréquence	f_1	f_2	…	f_p

La moyenne de cette série peut être calculée par :

$$\overline{x} = x_1 f_1 + x_2 f_2 + \ldots + x_p f_p = \sum_{i=1}^{p} x_i f_i$$

Remarque : $f_1 + f_2 + \ldots + f_p = 1$.

c) médiane dans le cas d'un caractère discret

Exemple 1 : Considérons les 5 nombres rangés dans l'ordre croissant : 2 ; 3 ; 5 ; 10 ; 12.
Il y a autant de nombres supérieurs à 5 que de nombres inférieurs à 5.
On dit que la médiane de cette suite est 5.
Exemple 2 : Considérons les 6 nombres rangés dans l'ordre croissant : 2 ; 3 ; 5 ; 10 ; 12 ; 14.
Pour tous les nombres de l'intervalle]5 ; 10[, trois nombres de la liste lui sont inférieurs et trois nombres lui sont supérieurs. Pour la médiane, on choisit par convention le milieu de l'intervalle [5 ; 10].
On dit que 7,5 est la médiane de cette suite de nombres.

Définition :
La médiane d'une série statistique partage cette série en deux parties de telle sorte que :
- au moins la moitié des valeurs sont inférieures ou égales à la médiane ;
- au moins la moitié des valeurs sont supérieures ou égales à la médiane.

méthode : Si la série contient n valeurs rangées dans l'ordre croissant :
- si n est impair, on prend la $\frac{n+1}{2}$ ème valeur pour médiane.
- si n est pair, on prend pour médiane la moyenne entre la $\frac{n}{2}$ ème et la $\frac{n}{2} + 1$ ème valeur.

Dans la série de notes de la classe de seconde A, on a 23 valeurs.
On prend alors la $\frac{23+1}{2}$ ème, c'est à dire la $12^{\text{ème}}$ valeur.
La médiane de cette série est alors 8.

d) mode et étendue

Définition : On appelle **mode** d'une série statistique une valeur du caractère dont l'effectif associé est le plus grand.

La série de notes de la seconde A admet deux modes : 3 et 6.

Définition : On appelle **étendue** d'une série statistique la différence entre la plus grande valeur du caractère et la plus petite .

L'étendue de la série de notes de la seconde A est : 16,5 – 0 = 16,5.

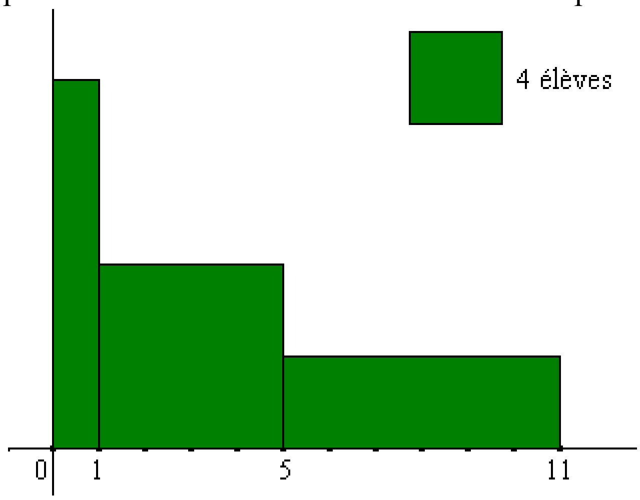
II. Cas d'un regroupement par classes de valeurs

Le tableau suivant donne la distance entre le domicile et le lycée pour les élèves d'une classe.

distance (en km)	[0 ; 1[	[1 ; 5[	[5 ; 11[
nombre d'élèves	8	16	12

a) histogramme

Dans le cas d'un histogramme, les effectifs (ou les fréquences) et les aires des rectangles sont proportionnels.
Lorsque les classes n'ont pas la même amplitude, on ne peut pas proposer d'unité sur l'axe des ordonnées.
Construction de l'histogramme représentant l'exemple :
On choisit 0,5 cm pour 1 km sur l'axe des abscisses et 1 cm² représente 4 élèves.



b) moyenne

Pour calculer la moyenne, on se ramène à un caractère discret en remplaçant **chaque classe par son centre**.

Dans notre exemple, on obtient : $\overline{x} = \dfrac{0,5 \times 8 + 3 \times 16 + 8 \times 12}{36} \approx 4,1$.

La distance moyenne est environ 4,1 km.

c) médiane

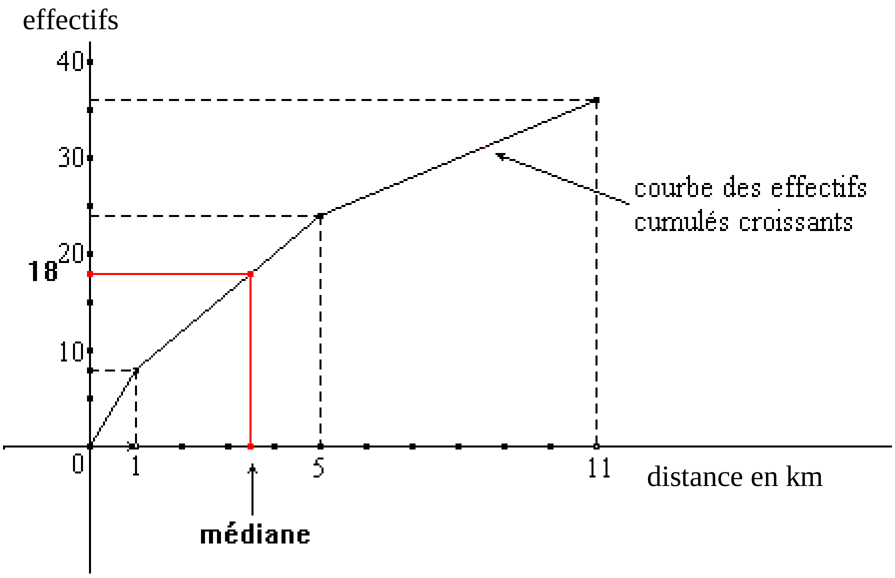
On construit tout d'abord le tableau des effectifs cumulés croissants (ou celui des fréquences cumulées croissantes)

distance (en km)	[0 ; 1[	[1 ; 5[	[5 ; 11[
effectifs cumulés	8	24	36
fréquences cumulées (en %)	22,2	66,7	100

On place dans un repère orthogonal les points (0 ; 0), puis (1 ; 8), (5 ; 24) et (11 ; 36).
On admet que la répartition dans chaque classe est uniforme, ainsi on joint ces points par des segments.

$\frac{36}{2} = 18.$

La médiane est l'abscisse du point de la courbe d'ordonnée 18.
Ici, la distance médiane est environ 3,5 km.



Remarque : on obtient la même courbe et la même médiane en utilisant les fréquences cumulées croissantes. Seules les ordonnées sont modifiées.

d) classe modale

On appelle **classe modale** d'une série statistique, une classe dont l'effectif associé est le plus grand.

Dans notre exemple, la classe modale est la classe [1 ; 5[.