

## Seconde

# Chapitre 4 : Statistiques descriptives

## 1- Série statistique :

### a) Définitions :

- On étudie un caractère qualitatif ou quantitatif  $x$  sur une certaine population.
- On obtient donc une série de  $p$  valeurs prises par le caractère, notées  $x_1, x_2, x_3, \dots, x_p$ .
- Pour chaque valeur  $x_i$  prise par ce caractère, on obtient un effectif  $n_i$  de population.
- L'effectif total est le nombre d'individus de la population étudiée : il est égal à la somme  $n_1 + n_2 + n_3 + \dots + n_p = N$ .
- La fréquence d'une valeur  $x_i$  est le rapport entre l'effectif de cette valeur par l'effectif total : on note  $f_i = \frac{n_i}{N}$ . On a toujours  $0 \leq f_i \leq 1$ .

Les fréquences sont en fait l'écriture décimale du pourcentage de la population pour la valeur donnée.

On construit souvent le tableau suivant :

$x_i$		
$n_i$		
$f_i$		

- L'effectif cumulé croissant en  $x_i$  est la somme des effectifs des valeurs inférieures ou égales à  $x_i$ .
- La fréquence cumulée croissante en  $x_i$  est la somme des fréquences des valeurs inférieures ou égales à  $x_i$ .



*Rq: les mêmes définitions existent avec « décroissant »*

- Lorsque les données sont en très grand nombre, on les regroupe en classes dans des intervalles du type  $[a; b[$ . La différence  $b - a$  s'appelle l'amplitude de la classe, et le centre de la classe est la valeur  $\frac{a+b}{2}$

### b) Représentation graphique :

Divers types de graphiques permettent de représenter une série statistique :

- **diagramme en bâtons** (caractère qualitatif ou quantitatif discret) dans lequel l'abscisse est  $x_i$  et dont la hauteur est proportionnelle à  $n_i$ .
- **histogramme** (caractère quantitatif continu : valeurs regroupées en classes) dans lequel on construit des rectangles dont l'aire est proportionnelle à  $n_i$  (la hauteur de chaque rectangle est proportionnelle à  $n_i$  dans le cas le plus fréquent où les classes sont de même amplitude).
- **diagramme circulaire ou semi-circulaire** dans lequel l'angle au centre de chaque portion est proportionnel à  $n_i$ .

- **diagramme cumulatif**, lorsque la série est regroupée en classes : on place en abscisse les  $x_i$ , en ordonnées les  $f_i$  cumulées et on relie les points obtenus par des segments (on suppose donc que la répartition est uniforme dans chaque classe).

## 11- Caractéristiques d'une série :

Afin de pouvoir interpréter des données souvent en très grand nombre, on utilise un certain nombre d'outils numériques permettant de synthétiser les données recueillies, de les résumer. Ces outils permettent également de comparer des séries dont les données pourraient parfois sembler proches, ou au contraire très différentes.

### a) Caractéristiques de position :

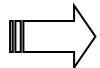
- La **moyenne pondérée** de la série, notée  $\bar{x}$  :

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{N} = \frac{\sum_{i=1}^p n_i x_i}{N}$$

Pté : On peut calculer la moyenne à partir de la distribution des fréquences :

$$\bar{x} = f_1x_1 + f_2x_2 + \dots + f_px_p = \sum_{i=1}^p f_i x_i$$

- La **médiane**  $m$  de la série statistique : d'un caractère quantitatif, dont les **valeurs sont ordonnées**, **tout nombre** qui partage cette série en deux groupes de même effectif.



En pratique :

- lorsque la série a un effectif total impair, la médiane est la valeur centrale
- lorsque l'effectif total est pair, la médiane est la demi somme des deux valeurs centrales.

Remarque : la valeur de la médiane n'est pas influencée par les valeurs extrêmes du caractère alors que celle de la moyenne l'est fortement.

Exemple : La série 3, 5, 12 a pour moyenne : .....et pour médiane :.....

En modifiant les valeurs extrêmes : la série 1, 5, 6 a pour moyenne : ..... et pour médiane .....

- Le **premier quartile** d'une série statistique est la plus petite valeur de la série, notée  $Q_1$ , telle qu'au moins 25% des valeurs de la série sont inférieures ou égales à  $Q_1$ .
- Le **troisième quartile** d'une série statistique est la plus petite valeur de la série, notée  $Q_3$ , telle qu'au moins 75% des valeurs de la série sont inférieures ou égales à  $Q_3$ .

On parle aussi parfois du **mode** ou de la classe modale d'une série : c'est la valeur de la série dont l'effectif est le plus grand. Le mode n'a d'intérêt que si l'effectif est réellement beaucoup plus grand que les autres. Il peut y avoir plusieurs modes pour une même série...

### b) Caractéristique de dispersion :

L'**étendue** d'une série statistique est la différence entre la plus grande et la plus petite des valeurs prises par cette série.

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