

# Chapitre 4 : Statistiques

## I Vocabulaire et représentations graphiques

### Définition

- La **population** est l'ensemble des personnes ou objets, appelés **individus**, qu'on étudie.
- Le **caractère** est la propriété étudiée sur chaque individu de la population.  
Le caractère peut-être :
  - **qualitatif** s'il ne prend pas que des valeurs numériques (couleur des yeux, marque de voiture).
  - **quantitatif discret** s'il ne peut prendre qu'un nombre fini de valeurs numériques (nombre d'enfants par foyer, note obtenue à un contrôle).
  - **quantitatif continu** s'il peut prendre une infinité de valeurs numériques (taille des personnes).  
On regroupe alors les valeurs par classes (intervalles).
- On appelle **modalités** les différentes valeurs prises par le caractère.  
Par exemple, si le caractère est la couleur des yeux, les modalités sont marron, bleu, vert, etc.  
Pour un caractère quantitatif, on parle des valeurs du caractère (ce sont des nombres).

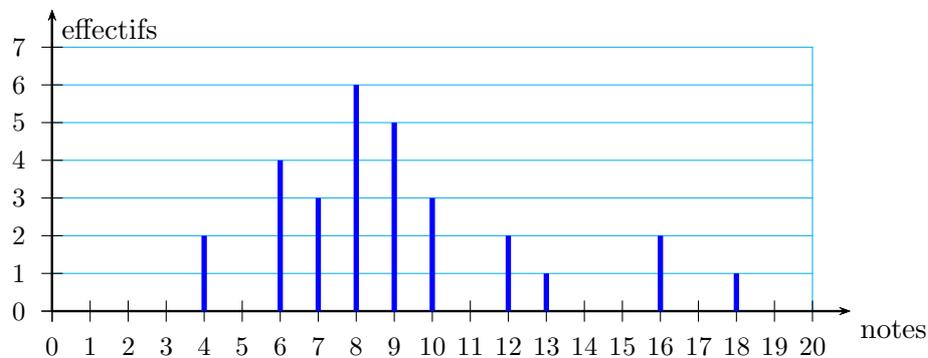
### I.1 Représentations graphiques

**Diagramme en bâtons :** utilisé plutôt pour des séries statistiques à caractère discret.

Les hauteurs des bâtons sont proportionnelles aux effectifs.

Exemple :

Notes $x_i$	4	6	7	8	9	10	12	13	16	18
Effectifs $n_i$	2	4	3	6	5	3	2	1	2	1



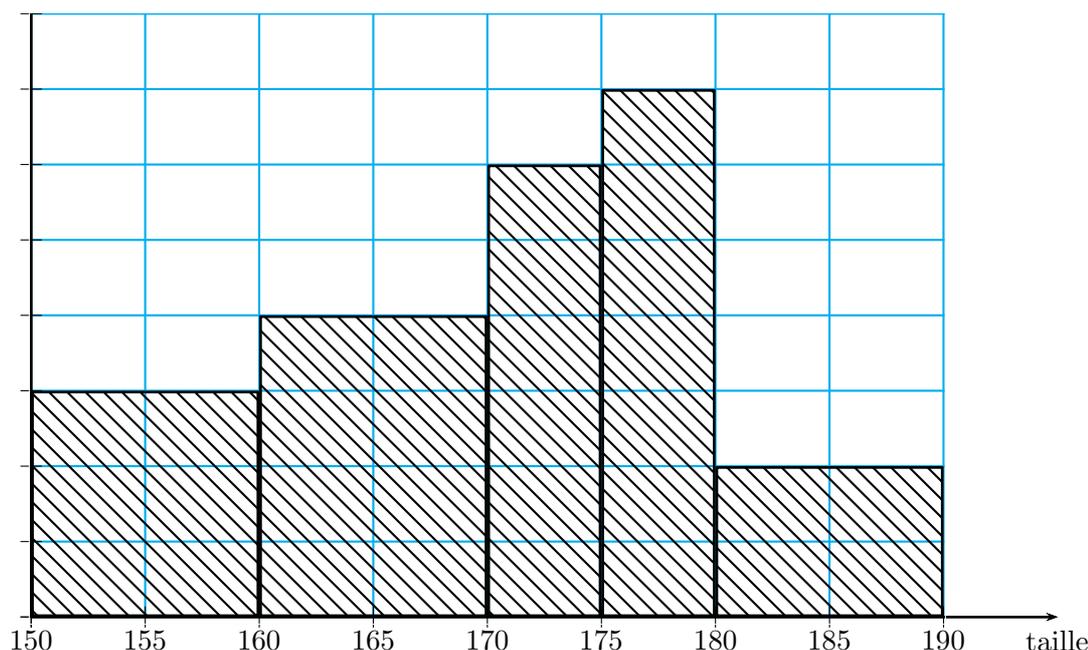
**Histogramme :** utilisé pour représenter les séries statistiques à caractère quantitatif continu.

La largeur du rectangle correspond à l'étendue de la classe.

Les aires des rectangles sont proportionnelles aux effectifs des classes.

Exemple :

Taille (en cm)	[150;160[	[160;170[	[170;175[	[175;180[	[180;190[
Nombre de personnes	6	8	6	7	4



Unité graphique : 1 individu.

**Remarque**

Attention si les classes n'ont pas toutes la même amplitude.

C'est l'aire des rectangles qui doit être proportionnelle à l'effectif.

Dans l'exemple précédent, la hauteur des rectangles n'est pas proportionnelle à l'effectif.

**Exercice 1**

1. Lire les effectifs sur un histogramme : [ressource 1111](#)
2. Représenter une série statistique dont les classes sont de même amplitude par un histogramme : [ressource 802](#)

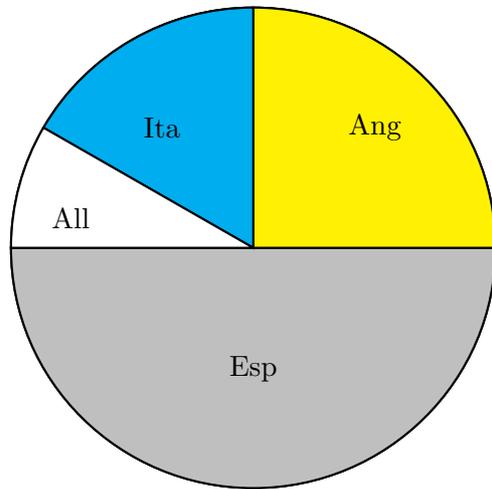
**Diagramme circulaire :** utilisé plutôt pour les séries à caractère quantitatif discret ou qualitatif.

L'angle d'ouverture de chaque secteur est proportionnel à l'effectif :

$$\text{angle du secteur (en } ^\circ) = \frac{\text{effectif de la modalité}}{\text{effectif total}} \times 360$$

Exemple :

Langue LV2	Espagnol	Anglais	Allemand	Italien	Total
Effectif	78	39	13	26	156
Angle en degrés	180	90	30	60	360



## II Indicateurs d'une série statistique

### II.1 effectif, fréquence

#### Définition

L'**effectif total** est le nombre d'individus de la population étudiée.

L'**effectif** d'une modalité est le nombre d'individus prenant cette modalité.

La **fréquence** d'une modalité est le quotient de son effectif par l'effectif total :

$$\text{Fréquence} = \frac{\text{effectif de la modalité}}{\text{effectif total}}$$

#### Remarque

Une fréquence est toujours comprise entre 0 et 1.

La somme de toutes les fréquences est égale à 1.

#### Exercice 2

1. Déterminer le ou les mode(s) d'une série statistique connaissant son tableau des effectifs : [ressource 803](#)
2. Déterminer le ou les mode(s) d'une série statistique connaissant sa représentation graphique par un diagramme en bâtons : [ressource 804](#)

#### Définition (Effectif cumulé croissant, fréquence cumulée croissante)

On considère une série statistique à caractère quantitatif (discret ou continu).

Soit  $p$  le nombre de modalités, on note  $x_1, x_2, \dots, x_p$  les différentes modalités.

L'**effectif cumulé croissant** d'une valeur  $x_i$  est égal à la somme des effectifs des valeurs inférieures ou égales à  $x_i$ .

La **fréquence cumulée croissante** de  $x_i$  est la somme des fréquences des valeurs inférieures ou égales à  $x_i$ .

#### Définition

Un **mode** est une modalité de plus grand effectif (il peut donc y avoir plusieurs modes).

On parle de classe modale dans le cas des séries à caractère continu.

#### Exercice 3

Voici les résultats portant sur le temps de traitement des dossiers d'une banque.

Durée en minutes	[0; 10[	[10; 20[	[20; 30[	[30; 40[	[40; 50[
Nombre	5	10	17	12	6
Effectifs cumulés croissants					

1. Calculer l'effectif de la série.
2. Quelle est la classe modale de la série ?
3. Compléter les effectifs cumulés croissants dans le tableau.
4. Calculer le pourcentage de dossiers dont la durée d'étude est strictement inférieure à 30 minutes.

### Remarque

Le dernier effectif cumulé croissant est l'effectif total  $N$ .

La dernière fréquence cumulée croissante est 1 (c'est la somme des fréquences).

### Exercice 4

1. Effectifs cumulés croissants : [ressource 3491](#)
2. Fréquences cumulées croissantes : [ressource 3493](#)

## II.2 Moyenne

### Définition (moyenne)

Soit une série quantitative discrète.

Notons  $x_1, x_2, \dots, x_p$  les valeurs et  $n_1, n_2, \dots, n_p$  les effectifs respectifs.

L'effectif total est alors  $N = n_1 + n_2 + \dots + n_p$ .

La moyenne (pondérée) est le nombre noté  $\bar{x}$  (prononcer "x barre") défini par :

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{N}$$

### Remarque (moyenne à partir des fréquences)

Si on note  $f_1, f_2, \dots, f_p$  les fréquences associées respectivement aux valeurs  $x_1, x_2, \dots, x_p$ , alors

$$\bar{x} = f_1x_1 + f_2x_2 + \dots + f_px_p$$

### Exercice 5

Calcul de moyenne : [819](#)

### Remarque

1. Dans le cas des séries à caractère continu, on prend les centres des classes pour jouer le rôle des valeurs. On obtient alors une estimation de la moyenne.
2. La moyenne est un indicateur de position. On la complète avec un indicateur de dispersion, comme l'étendue ( $Max - Min$ ).

### Propriété (linéarité de la moyenne)

Si l'on ajoute un même nombre  $a$  à chacune des valeurs de la série, alors la moyenne augmente de  $a$ .

Si l'on multiplie chaque valeur de la série par un même nombre  $b$ , alors la moyenne est multipliée par  $b$ .

## II.3 Médiane

### Définition (Médiane d'une série discrète)

On considère une série statistique discrète  $x_1, x_2, \dots, x_N$  de  $N$  valeurs .

On appelle médiane de la série et on note  $Me$  tout nombre tel que :

au moins 50% des valeurs de la série sont inférieures ou égales à  $Me$ ,

et au moins 50% des valeurs de la série sont supérieures ou égales à  $Me$ .

### Méthode de détermination

On range les  $N$  valeurs  $x_1, x_2, \dots, x_N$  dans l'ordre croissant.

Si l'effectif total  $N$  est pair, on choisit pour médiane la demi-somme des deux valeurs centrales.

Avec  $N = 2k$ ,  $Me = \frac{x_k + x_{k+1}}{2}$ .

Si l'effectif total  $N$  est impair, la médiane est la valeur centrale de la série.

Plus précisément, avec  $N = 2k + 1$ ,  $Me = x_{k+1}$ , c'est-à-dire la  $(k + 1)^e$  valeur.

### Remarque

Si  $N$  est pair et que les deux valeurs centrales sont distinctes, alors tout nombre compris strictement entre ces deux valeurs centrales peut être appelé médiane (il n'y a pas unicité).

Dans les autres cas, il n'y a qu'une médiane.

Exemple :

Valeurs $x_i$	35	36	37	38	39	40
Effectifs $n_i$	3	4	2	1	1	2
ECC						

$N = 13$ . L'effectif est impair, et  $13 = 2 \times 6 + 1$ .

$Me$  est la valeur centrale de la série, c'est-à-dire la 7e valeur.

$Me = 36$ .

Valeurs $x_i$	5	11	27	33	42
Effectifs $n_i$	4	3	1	5	3
ECC					

Ici,  $N = 16 = 2 \times 8$ . L'effectif est pair.

$Me$  est la demi-somme de la 8e valeur et de la 9e valeur.

Donc  $Me = \frac{x_8 + x_9}{2} = \frac{27 + 33}{2} = 30$ .

### Exercice 6

1. Médiane à partir des ECC : [ressource 811](#)
2. Médiane à partir des FCC : [ressource 812](#)
3. Médiane à partir du diagramme des FCC : [ressource814](#)

## II.4 Quartiles

### Définition

On considère une série statistique à caractère quantitatif discret.

**Le premier quartile**  $Q_1$  est la plus petite valeur de la série telle qu'au moins 25 % des valeurs soient inférieures ou égales à  $Q_1$ .

**Le troisième quartile**  $Q_3$  est la plus petite valeur de la série telle qu'au moins 75 % des valeurs soient inférieures ou égales à  $Q_3$ .

**L'intervalle interquartile** est  $[Q_1; Q_3]$ . Il contient environ 50 % des valeurs.

**L'écart interquartile** est le nombre  $Q_3 - Q_1$ .

### Méthode pour déterminer les quartiles :

Soit  $N$  l'effectif total de la série.

On range les  $N$  valeurs  $x_1, x_2, \dots, x_N$  dans l'ordre croissant.

Pour  $Q_1$ , on calcule  $\frac{N}{4}$ , ce qui donne, en arrondissant si besoin à l'entier supérieur, le rang de  $Q_1$ .

Pour  $Q_3$  on procède de même en remplaçant  $\frac{N}{4}$  par  $\frac{3N}{4}$ .

### Remarque

Les quartiles sont toujours des valeurs de la série.

Exemple :

Valeurs $x_i$	35	41	46	65	81
Effectifs $n_i$	5	3	2	3	2
Effectifs cumulés croissants					

L'effectif total est  $N = 5 + 3 + 2 + 3 + 2 = 15$ .

Déterminons le 1er quartile  $Q_1$ .

$\frac{N}{4} = \frac{15}{4} = 3.75$ . Donc  $Q_1$  est la ... valeur :  $Q_1 = \dots$

Déterminons le 3e quartile  $Q_3$ .

$\frac{3N}{4} = \frac{3 \times 15}{4} = 11.25$ . Donc  $Q_3$  est la ... valeur :  $Q_3 = \dots$

L'intervalle interquartile est donc ...

L'écart interquartile est ...

### Exercice 7

1. Médiane et quartiles à partir des effectifs : [ressource 3894](#)
2. Médiane et quartiles à partir des ECC : [ressource 3896](#)
3. Médiane et quartiles à partir des FCC : [ressource 3897](#)

### Remarque

Pour accompagner la médiane d'un indicateur de dispersion, on prend en général l'écart interquartile  $Q_3 - Q_1$ .

Plus l'écart interquartile est grand, plus les données sont dispersées.

## II.5 Utilisation de la calculatrice pour les statistiques

Note : Pour l'utilisation de la calculatrice en statistiques, voir aussi le rabat de la dernière page du manuel.

Texas

1. Pour entrer les données de la série :

**[Stats]**, puis EDIT, (edite...).

On entre les valeurs dans la liste  $L_1$ , et les effectifs dans la liste  $L_2$ .

Pour effacer les valeurs dans une liste, **[Stats]**, EDIT, Effliste, puis le nom de la liste à vider.

On peut aussi mettre directement la liste à vider en surbrillance et utiliser la touche **[annul]**.

2. Pour obtenir les indicateurs de la série :

**[Stats]** CALC Stats 1-Var

Xliste :  $L_1$

ListeFreq :  $L_2$

Ou bien, sur les anciens modèles,

**Stats** CALC Stats 1-Var  $L_1$  ,  $L_2$ .

Attention à bien utiliser la virgule pour séparer  $L_1$  et  $L_2$  (et pas le point).

Remarque : sur les anciens modèles TI82stats, les quartiles ne sont pas calculés suivant la même définition que dans le cours, et la calculatrice peut afficher des quartiles qui ne sont pas des valeurs de la série.

Casio

1. Pour entrer les données :

Aller dans le menu **List**.

Entrer les valeurs dans List 1, et les effectifs dans List 2.

Pour effacer :

— **DEL** pour effacer un élément dans une liste

— **DEL-A** pour effacer le contenu d'une liste.

2. Pour obtenir les indicateurs de la série :

Aller dans le menu **Stat**.

Si List 1 contient les valeurs  $x_1$ , et List 2 contient les effectifs  $n_i$ , alors, suivre

**SET** pour paramétrer les calculs de la façon suivante :

— 1Var X list : List 1

— 1Var Freq :List 2

Valider avec **EXE**.

Alors on obtient les indicateurs de la série par :

**CALC** **1 VAR**

### Exercice 8

Voici un tableau présentant les salaires dans une entreprise.

Salaires mensuels (euros)	1200	1650	2100	2400	6500
Effectifs	5	8	4	3	1
Effectifs cumulés croissants					

1. Compléter les effectifs cumulés.
2. Donner l'effectif total.
3. Quel est le mode ?
4. Quel est le salaire moyen ?
5. Quel est le salaire médian (la médiane de la série) ?
6. Que deviennent la médiane et la moyenne si l'on remplace la plus grande valeur par 500 000 ?

### Remarque

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

### Remarque

La médiane et les quartiles ne sont pas influencés par les valeurs extrêmes de la série.

La moyenne et l'étendue sont sensibles aux valeurs extrêmes.

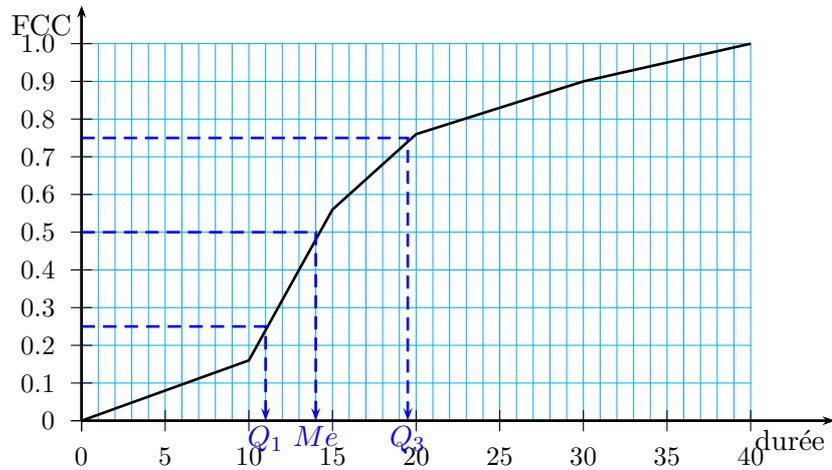
### Remarque (Le cas des séries à caractère continu)

1. Moyenne.

On prend les centres des classes pour jouer le rôle des valeurs. On a alors une estimation de la moyenne.

2. Médiane et quartiles.

On trace la courbe des fréquences cumulées croissantes.



La médiane  $Me$  est l'abscisse du point de la courbe qui a pour ordonnée  $\frac{1}{2}$ .  
 $Q_1$  est l'abscisse du point qui a pour ordonnée 0.25.  
 $Q_3$  est l'abscisse du point qui a pour ordonnée 0.75.

**II.6 Complément : diagramme en boîte**

Ce type de diagramme permet de résumer les principaux indicateurs de la série.

**Exercice 9**

Voici un relevé des températures durant le mois de février :

Température observée	-4	-3	-2	0	2	3	4	5	6	7	8	9
Nombre de jours	1	2	1	1	1	1	2	3	4	5	4	3
Effectifs cumulés croissants												

1. Compléter les effectifs cumulés croissants.

2. L'effectif total est  $N = \dots$

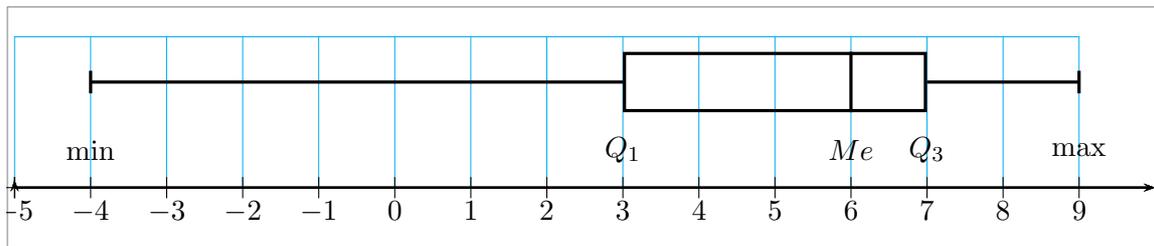
3.  $\frac{N}{4} = \dots$ , donc  $Q_1$  est la ...<sup>e</sup> valeur :  $Q_1 = \dots$

4.  $\frac{3N}{4} = \dots$ , donc  $Q_3$  est la ...<sup>e</sup> valeur :  $Q_3 = \dots$

5. Détermination de la médiane  $Me$

.....  
 .....  
 .....

On résume la série par le diagramme en boîte (ou boîte à moustache) suivant :



Cette série a pour valeur minimale  $-4$ , pour maximum  $9$ .  
Le premier quartile est  $3$ , la médiane est  $6$ , et le troisième quartile est  $7$ .

### Exercice 10

1. Lire un diagramme en boîte : [ressource 3517](#)
2. Construire un diagramme en boîte à partir des FCC : [ressource 3519](#)
3. Construire un diagramme en boîte à partir des ECC : [ressource 3520](#)