

I – Série de données statistique :

1) Vocabulaire :

L'ensemble sur lequel porte l'étude statistique s'appelle la **population**. Un élément de cet ensemble est un **individu**. L'objet étudié s'appelle le **caractère**.

Si le caractère prend des **valeurs numériques**, on dit qu'il est **quantitatif**. Sinon, il est dit **qualitatif**.

Un caractère quantitatif peut être **discret** ou **continu** :

- discret s'il prend des **valeurs isolées** (0, 1, 2, ...).
- continu s'il peut prendre toutes les **valeurs dans un intervalle** appelé aussi **classe**.

2) Effectifs, Fréquences & Cumul :

L'**effectif** d'une valeur est le nombre d'individus ayant cette valeur.

L'**effectif total**, noté  $N$ , est le nombre d'individus de la population étudiée. Autrement dit, c'est la **somme de tous les effectifs**.

La **fréquence**, notée  $f$ , d'une valeur est le quotient de l'effectif de cette valeur par l'effectif total :

$$f = \frac{\text{effectif de la valeur}}{\text{effectif total}} = \frac{x}{N}$$

Pour un caractère quantitatif, on appelle **effectif cumulé croissant** (**fréquence cumulée croissante**) d'une valeur  $x$  le nombre (la fréquence) d'individus ayant une valeur inférieure ou égale à  $x$  dans la série.

3) Application :

Une machine réalise des pièces cylindriques. Lors d'un contrôle, on mesure le diamètre, en millimètres, d'un lot de pièces et on obtient la série statistique suivante :

Diamètre en mm	42,15	42,16	42,17	42,18	42,19	42,20	42,21	42,22
Effectif	7	20	42	88	52	24	12	5
Effectif cumulé croissant	7	27	69	157	209	233	245	250

La population étudiée est \_\_\_\_\_

Le caractère de la série statistique est \_\_\_\_\_

Le caractère est donc un caractère \_\_\_\_\_

L'effectif total est \_\_\_\_\_

La fréquence de la valeur 42,18 est \_\_\_\_\_

La fréquence cumulée croissante (F.C.C.) de la valeur 42,18 est la fréquence du nombre de pièces ayant un diamètre inférieur ou égal à 42,18 mm. Ce nombre vaut \_\_\_\_\_

## II – Représentations de données statistiques :

### 1 Nuage de points

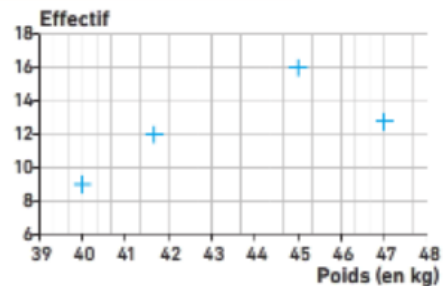
Pour représenter graphiquement une série statistique au caractère quantitatif discret, il existe différentes possibilités. L'une des plus utilisées est le **nuage de points**. Les valeurs de la série sont sur l'axe des abscisses. La fréquence ou l'effectif correspondant à chaque valeur est sur l'axe des ordonnées.

#### Exemple

On a relevé le poids (en g) à la naissance de 50 chatons. Les valeurs ont été regroupées dans le tableau ci-dessous.

Poids	40	42	45	47
Effectif	9	12	16	13

On représente cette situation par un **nuage de points**. Le poids est en abscisse et l'effectif en ordonnée.



### 2 Histogramme

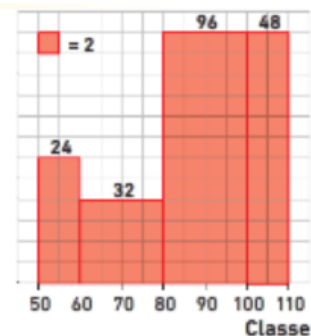
Pour représenter graphiquement une série statistique au caractère quantitatif continu, on utilise un **histogramme**. Un histogramme est constitué de rectangles. Pour chaque classe, l'aire de chaque rectangle est proportionnelle à l'effectif ou à la fréquence de la classe.

#### Exemple

Des tomates ont des calibres compris entre 50 mm et 110 mm. On a regroupé les valeurs par classe dans le tableau ci-dessous.

Calibre	[50 ; 60[	[60 ; 80[	[80 ; 100[	[100 ; 110[
Effectif	24	32	96	48

On construit l'histogramme correspondant avec l'unité d'aire donnée. Le rectangle correspondant à l'effectif 96 a bien une aire qui est le double de celle du rectangle correspondant à l'effectif 48.



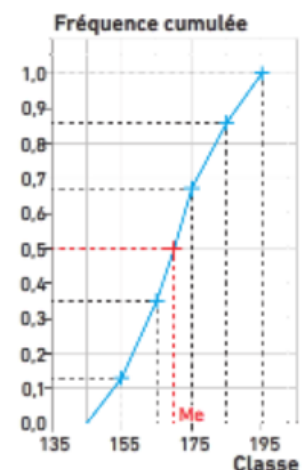
### 3 Courbe des fréquences cumulées croissantes (F.C.C.)

#### Exemple

50 personnes ont été mesurées. On a regroupé leur taille dans le tableau.

Taille en cm	[145 ; 155[	[155 ; 165[	[165 ; 175[	[175 ; 185[	[185 ; 195[
Fréquence	0,12	0,22	0,34	0,18	0,14
F.C.C.	0,12	0,34	0,68	0,86	1

La ligne des F.C.C. indique que 12% (34%) des personnes mesurent moins de 155 cm (165 cm) etc... On place alors les points de coordonnées (145 ; 0), (155 ; 0,12), (165 ; 0,34) etc ... Les segments reliant tous ces points constituent la courbe des F.C.C.



### III – Caractéristiques d'une série statistique :

#### 1) Paramètres de position d'une série statistique :

On considère la série statistique donnée par le tableau ci-contre et on note  $N = n_1 + n_2 + \dots + n_p$  l'effectif total.

Valeur	$x_1$	$x_2$	...	$x_p$
Effectif	$n_1$	$n_2$	...	$n_p$
Fréquence	$f_1$	$f_2$	...	$f_p$

##### a. La moyenne pondérée :

La **moyenne pondérée** de cette série, notée  $\bar{x}$ , est donnée par l'une des formules suivantes :

$$\bar{x} = \frac{n_1 \times x_1 + n_2 \times x_2 + \dots + n_p \times x_p}{N} \quad \text{ou} \quad \bar{x} = f_1 \times x_1 + f_2 \times x_2 + \dots + f_p \times x_p$$

Remarque : Si les données sont regroupées en classes, on prend alors les centres des classes comme valeurs.

##### b. La médiane :

On considère que les valeurs de la série sont **rangées dans l'ordre croissant**.

La **médiane**, notée **Me**, est une valeur telle qu'au moins la moitié des valeurs de la série lui sont inférieure ou égale et au moins la moitié des valeurs de la série lui soit supérieure ou égale.

$$Me = \frac{N}{2}^{\text{ème}} \text{ valeur de la série}$$

##### c. Les quartiles :

Le **premier quartile**, noté **Q<sub>1</sub>**, est la plus petite valeur de la série telle qu'au moins 25% ( $\frac{1}{4}$ ) des valeurs sont inférieures ou égale à Q<sub>1</sub>.

Le **troisième quartile**, noté **Q<sub>3</sub>**, est la plus petite valeur de la série telle qu'au moins 75% ( $\frac{3}{4}$ ) des valeurs sont inférieure ou égale à Q<sub>3</sub>.

#### 2) Paramètres de dispersion d'une série statistique :

##### a. L'étendue :

L'**étendue** est la différence entre la plus grande valeur (**valeur maximale**) et la plus petite valeur (**valeur minimale**).

##### b. L'écart interquartile :

L'**écart interquartile** d'une série statistique est la différence entre Q<sub>3</sub> et Q<sub>1</sub>, soit **Q<sub>3</sub> – Q<sub>1</sub>**.

Remarque : Cet écart interquartile donne une idée de la dispersion des valeurs autour de la médiane (il contient au moins 50% des valeurs de la série), en effet, plus l'écart est petit plus les valeurs se concentrent autour de la médiane.

### 3) Résumé d'une série statistique :

Pour comparer des séries, on utilise fréquemment un diagramme en boîte, qui résume les principaux paramètres de la série.

Le diagramme en boîte représente graphiquement une série statistique en indiquant : la valeur minimale, le 1<sup>er</sup> quartile, la médiane, le 3<sup>ème</sup> quartile, et la valeur maximale.

