Série statistique :

On veut étudier un ***caractère*** qui concerne une certaine population statistique (lorsque ***l'effectif*** de cette population est trop important, on en étudie des ***échantillons***). La population est composée ***d'individus*** .

Le ***caractère étudié*** peut être ***qualitatif*** ou ***quantitatif*** :

Exemple :

Soit l'ensemble des élèves d'une classe de Seconde (*population*) .

Le 1er caractère étudié est la couleur de leurs yeux (*caractère qualitatif*) {bleu, noir, vert, marron} .

Le 2ème caractère étudié est l'âge des élèves (*caractère quantitatif*) , valeurs numériques, de 15 à 18 ans.

Une série statistique est la suite des valeurs du caractère étudié, pour chaque individu de la population.
L'effectif total de la série statistique est le nombre d'individus dans la population étudiée.

Exemple précédent : Effectif total $N = 10$ élèves.

1ère série statistique { Bleu, Noir, N, Vert, B, Marron, V, V, M, M},

2ème série statistique {17, 15, 17, 16, 16, 18, 16, 17, 16, 16}

On regroupe généralement une série statistique par valeurs identiques x_i au sein d'un tableau, en précisant l'effectif n_i de chaque valeur, ou la fréquence de ces valeurs $f_i = \frac{n_i}{N}$.

1ère série :

Couleur	Bleu	Noir	Vert	Marron	*Totaux*
Effectif	2	2	3	3	$N = 10$
Fréquence	$\frac{2}{10} = 0{,}20$	$\frac{2}{10} = 0{,}20$	$\frac{3}{10} = 0{,}30$	$\frac{3}{10} = 0{,}30$	$\frac{10}{10} = 1$ (100%)

2ème série :

Age x_i	$x_1 = 15$	$x_2 = 16$	$x_3 = 17$	$x_4 = 18$	*Totaux*
Effectif n_i	$n_1 = 1$	$n_2 = 5$	$n_3 = 3$	$n_4 = 1$	$N = 10$
Fréquence $f_i = \frac{n_i}{N}$	$\frac{1}{10} = 0{,}10$	$\frac{5}{10} = 0{,}50$	$\frac{3}{10} = 0{,}30$	$\frac{1}{10} = 0{,}10$	$\frac{10}{10} = 1$ (100%)

Série statistique regroupée en classes :

Une série statistique quantitative (numérique) est regroupée en classes, si ses valeurs sont regroupées par intervalles, et non séparées (discrètes)

Exemple : Soit l'ensemble des filles d'une classe de Seconde (*population d'effectif total* $N = 10$).
Le caractère étudié est la taille de ces élèves (*caractère quantitatif*).

Taille en cm	[155 ; 160[	[160 ; 165[	[165 ; 170[	[170 ; 175[	*Total*
Effectif n_i	$n_1 = 1$	$n_2 = 3$	$n_3 = 4$	$n_4 = 2$	$N = 10$

<u>*Paramètres d'une série statistique*</u> :

Il existe deux types de paramètres (indicateurs) :

- Indicateurs de <u>***valeur centrale***</u> : Moyenne, Médiane.
- Indicateurs de <u>***dispersion***</u> : Etendue, Intervalle interquartile, Ecart-type.

Il faut comprendre que dire que deux classes de Seconde ont obtenu une <u>même moyenne</u> 11/20 au dernier contrôle de Maths, indicateur de valeur centrale, est une information insuffisante.

Il est souhaitable de savoir en plus comment se sont réparties les notes dans les deux classes, entre 02 et 18 en Seconde 1 , et entre 05 et 16 en Seconde 2 (<u>étendue</u> : indicateur de dispersion).

Indicateurs de Valeur Centrale

<u>*Moyenne arithmétique*</u> *:* Notée $\overline{X}$ ou $E(X)$ (espérance mathématique).

La moyenne est la somme des valeurs de la série statistique, divisée par l'effectif de la série.

{4 ; 7 ; 12 ; 15} admet pour moyenne $\overline{X} = \frac{4 + 7 + 12 + 15}{4} = \frac{38}{4} = 9{,}75$.

On prendra note que <u>**la moyenne *se calcule à partir des valeurs***</u> *du caractère, alors que l'on verra que* <u>**la médiane *se calcule à partir des effectifs***</u> *de la série, sans tenir compte des valeurs.*

{4 ; 7 ; 18} admet pour *moyenne* $X = \frac{4 + 7 + 16}{3} = \frac{27}{3} = 9$.

{4 ; 7 ; 18} admet pour *médiane* $Me = 7$ (position médiane).

<u>*Moyenne arithmétique sous forme de valeurs groupées*</u> *:*

$$\overline{X} = \frac{S}{N} = \frac{\sum_{i=1}^{p} n_i x_i}{N} = \frac{1}{N} \sum_{i=1}^{p} n_i x_i = \frac{n_1x_1 + n_2x_2 + \ldots + n_px_p}{n_1 + n_2 + \ldots + n_p} .$$

x_i *(années)*	15	16	17	18	
n_i	1	5	3	1	$N = 10$
n_ix_i	15	80	51	18	$S = 164$

Moyenne : $\overline{X} = \frac{S}{N} = \frac{164}{10} = 16{,}4$ ans .

<u>*Moyenne arithmétique sous forme de classes*</u> *:*

On prend pour valeurs x_i les <u>centres de classe</u> (milieu des intervalles : $\frac{x_i + x_{i+1}}{2}$, comme $\frac{155 + 160}{2} = 157{,}5$).

x_i *(taille cm)*	157,5	162,5	167,5	172,5	
n_i	1	3	4	2	$N = 10$
$n_i\, x_i$	157,5	487,5	670	345	$S = 1660$

Moyenne : $\overline{X} = \frac{S}{N} = \frac{1660}{10} = 166$ cm .

Autre présentation : En fréquences f_i

x_i *(taille cm)*	157,5	162,5	167,5	172,5	
n_i	1	3	4	2	$N = 10$
f_i	1/10	3/10	4/10	3/10	$s = 1$ (100%)
$f_i\, x_i$	15,75	48,75	67	34,5	$S = 166$

Moyenne : $\overline{X} = \sum_{i=1}^{p} f_i\, x_i = f_1\, x_1 + f_2\, x_2 + \ldots + f_4\, x_4 = 166$ cm .

Médiane :

On doit tout d'abord ranger les valeurs du caractère dans l'ordre croissant.

La médiane *Me* divise la série en deux sous-populations de même effectif.

Si la série comporte un nombre *impair* de valeurs, la *Médiane* et la valeur centrale.

{2 ; 3 ; **7** ; 8 ; 12} admet pour médiane $Me = 7$.

Si la série comporte un nombre *pair* de valeurs, la *Médiane* est la moyenne des deux valeurs centrales.

{2 ; 3 ; **7 ; 8** ; 12, 18} admet pour médiane $Me = \dfrac{7+8}{2} = 7{,}5$.

A l'inverse de la moyenne, qui tient compte des valeurs du caractère dans son calcul, la médiane ne tient pas compte des valeurs, mais de leur classement et de leur nombre.

Exemple : Dans une entreprise, les salaires mensuels nets des 5 employés sont :

{1.200 € ; 1.400 € ; **1.800 €** ; 2.000 € ; 4.300 €}

Pour évaluer le niveau de salaire du personnel :

- Le représentant du personnel parlera du salaire *médian* : 1.800 € ☹ .
- Le président parlera du salaire *moyen* : $\dfrac{1.200 + 1.400 + 1.800 + 2.000 + 4.300}{5} = 2.140$ € ☺ .

On a compris que l'un des salaires étant beaucoup plus important que les autres, en utilisant le salaire moyen (basé sur les valeurs) on remonte la moyenne.

Par contre, la médiane, qui n'est basée que sur le nombre de salaires, n'est pas influencée par ce haut salaire, donc donne un résultat plus faible (chacun utilise ce qui l'arrange).

Indicateurs intermédiaires – Les Quartiles Q_1 et Q_3 :

La médiane *Me* indique la valeur du caractère telle que **50%** (moitié) des valeurs lui soient inférieures, et 50% lui soient supérieures.

Dans l'exemple précédent, la moitié des employés (50%) gagnent 1.800 € mensuels ou moins, et l'autre moitié gagne 1.800 € ou plus.

Le 1er Quartile Q_1 (Quartile = Quart) indique la valeur du caractère telle que **25%** (un quart) des valeurs lui soient inférieures, et 75% lui soient supérieures.

Le 3ème Quartile Q_3 indique la valeur du caractère telle que **75%** (trois quarts) des valeurs lui soient inférieures, et 25% lui soient supérieures.

On pourrait dire que la médiane est le 2ème Quartile $Me = Q_2$ (50% = 2 quarts).

Exemple :
Soit la série classée dans l'ordre croissant {2 ; 4 ; **6 ; 7** ; 9 ; 9 ; **11** ; 12 ; 13 ; **15 ; 16** ; 16 ; 18} .
Elle comporte $N = 13$ valeurs.

- 25% de 13 = 0,25 × 13 = 3,25 . On prendra comme 1er quartile Q_1 le milieu des 3ème et 4ème valeurs.

$Q_1 = \dfrac{6+7}{2} = 6{,}5$ (25% des valeurs sont inférieures à 6,5 et 75% lui sont supérieures).

- De façon très symétrique, on prendra comme 3ème quartile Q_3 le milieu des 11ème et 12ème valeurs.

$Q_3 = \dfrac{15+16}{2} = 15{,}5$ (75% des valeurs sont inférieures à 15,5 et 25% lui sont supérieures).

- La médiane est $Me = 11$ (50% des valeurs sont inférieures à 11 et 50% lui sont supérieures).

Indicateurs de Dispersion

Etendue de la série : Notée e .

L'étendue est celle sur laquelle s'étale la série, entre sa valeur la plus basse (mini) et sa valeur la plus haute (maxi).

$\boldsymbol{e = x_{Max} - x_{min}}$.

Exemple : Soit la série classée dans l'ordre croissant {2 ; 4 ; **6 ; 7** ; 9 ; 9 ; **11** ; 12 ; 13 ; **15 ; 16** ; 16 ; 18} .

Son ***étendue*** est $e = 18 - 2 = 16$.

On prend en compte l'ensemble des valeurs, même les valeurs extrêmes, moins significatives.

Intervalle Interquartile : Noté i .

L'intervalle interquartile indique sur quelle étendue se trouvent les 50% de valeurs centrales de la série.

$\boldsymbol{i = Q_3 - Q_1}$.

Exemple : Soit la série classée dans l'ordre croissant {2 ; 4 ; **6 ; 7** ; 9 ; 9 ; **11** ; 12 ; 13 ; **15 ; 16** ; 16 ; 18} .

Son ***intervalle interquartile*** est $i = 15{,}5 - 6{,}5 = 9$.

On élimine les 25% les plus basses, et les 25% les plus hautes, pour ne garder que les 50% centrales, plus significatives.

Ecart-Type : Noté s ou σ .

L'écart-type est la racine carrée de la *Variance V* .

$$\sigma = \sqrt{V} = \sqrt{\frac{1}{N}\sum_{i=1}^{p} n_i\,(x_i - \overline{X})^2} = \sqrt{\sum_{i=1}^{p} f_i\,(x_i - \overline{X})^2} \;.$$

Autre notation : $\sigma = \sqrt{\overline{X^2} - (\overline{X})^2}$.

Représentations graphiques d'une série statistique :

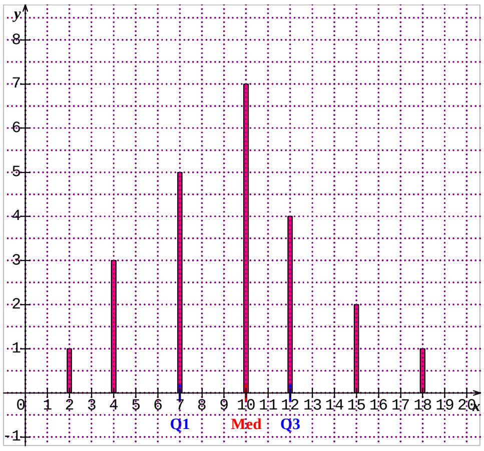
Diagramme en bâtons :

Ne s'utilise que pour des valeurs distinctes de la série (**valeurs discrètes**) , jamais si on travaille en classes.

A chaque valeur de la série, on associe un bâton, dont la hauteur est proportionnelle à l'effectif de la valeur.

Exemple : Soit la série suivante :

x_i	2	4	7	10	12	15	18
n_i	1	3	7	5	4	2	1



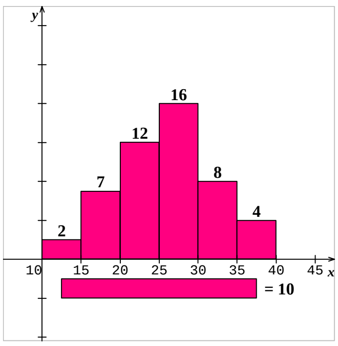
Histogramme :

Ne s'utilise que pour des valeurs en classes de la série.

A chaque classe (*de même largeur*) de la série, on associe un rectangle, dont la surface (*aire*) est proportionnelle à l'effectif de la classe.

Exemple : Soit la série suivante :

Classe	[10 ; 15[	[15 ; 20[	[20 ; 25[	[25 ; 30[	[30 ; 35[	[35 ; 40[
Effectif	2	7	12	16	8	4



Camembert (secteurs angulaires) :

Que ce soit pour des valeurs discrètes ou en classes, on peut donner une présentation en secteurs angulaires (camembert).

On effectue une règle de trois pour traduire la fréquence d'une valeur en pourcentage angulaire de 360°, tour complet du cercle, qui correspond à 100%.

Ainsi une fréquence $f = 0{,}30$ se traduira en un secteur de $0{,}30 \times 360° = 108°$.

Les secteurs angulaires sont généralement classés dans l'ordre décroissant.

Exemple : Exercice corrigé

Un groupe d'élèves a analysé un lot de pelotes de réjection d'une chouette effraie.

Leurs observations sont résumées dans le tableau suivant :

	Campagnols	**Mulots**	**Musaraignes**	**Divers**	**TOTAL**
Effectifs	244	114	106	14	**478**
Fréquence en %	**51**	**24**	**22**	**3**	100
Angles en Degrés	**184**	**86**	**80**	**10**	**360**

On veut représenter ces fréquences par un diagramme circulaire.

1/ Recopier et compléter ce tableau (fréquences et angles seront arrondis à l'unité).

$N = 244 + 114 + 106 + 14 = 478$.

$\frac{244}{478} = 0{,}51$ par défaut $\Rightarrow$ angle : $360° \times \frac{244}{478} = 184°$ par excès.

$\frac{114}{478} = 0{,}24$ par excès $\Rightarrow$ angle : $360° \times \frac{114}{478} = 86°$ par excès.

$\frac{106}{478} = 0{,}22$ par excès $\Rightarrow$ angle : $360° \times \frac{106}{478} = 80°$ par excès.

$\frac{14}{478} = 0{,}03$ par excès $\Rightarrow$ angle : $360° \times \frac{14}{478} = 10°$ par excès.

2/ Dessiner le diagramme sur la copie en prenant un diamètre de 10 cm.