# STATISTIQUES DESCRIPTIVES.

## 1 Définitions et vocabulaire des statistiques

**Definition 1.1** *On considère une série statistique à caractère quantitatif, dont les $p$ valeurs sont données par :*
*$x_1$, $x_2$, ..., $x_p$ d'effectifs associés $n_1$, $n_2$, ..., $n_p$ avec $n_1 + n_2 + ... + n_p = N$.*

- *A chaque valeur (ou classe) est associée une fréquence $f_i$ : c'est la proportion d'individus associés à cette valeur.*
- *$f_i = \dfrac{n_i}{N}$ est un nombre compris entre 0 et 1, que l'on peut écrire sous forme de pourcentage.*
- *L'ensemble des fréquences de toutes les valeurs du caractère s'appelle la caractère s'appelle la distribution des fréquences de la série statistique.*

**Exemple(s) 1.1** *Voici les notes obtenues à un contrôle dans une classe de 30 élèves : (**Série A :**)*
2−3−3−4−5−6−6−7−7−7−8−8−8−8−8−9−9−9−9−9−9−10−10−11−11−11−13−13−15−16
*On peut représenter cette série par un tableau d'effectifs, et le compléter par la distribution des fréquences :*

| *Notes* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Eff.* | 0 | 1 | 2 | 1 | 1 | 2 | 3 | 5 | 6 | 2 | 3 | 0 | 2 | 0 | 1 | 1 | 0 | 0 | 0 |
| *Fréq. en %* | 0 | 3 | 7 | 3 | 3 | 7 | 10 | 17 | 20 | 7 | 10 | 0 | 7 | 0 | 3 | 3 | 0 | 0 | 0 |

**Remarque(s) 1.1** *On peut vérifier que la somme des fréquences est égale à 1 (ou à 100 si on les exprime en pourcentages).*

**Exemple(s) 1.2** *On peut aussi faire un regroupement par classe, ce qui rend l'étude moins précise, mais qui permet d'avoir une vision plus globale. Toujours pour la **série A**, si on regroupe les données par classes d'amplitude 5 points, on obtient :*

| *Notes* | [ 0 ; 5 [ | [ 5 ; 10 [ | [ 10 ; 15 [ | [ 15 ; 20 [ | *total* |
|---|---|---|---|---|---|
| *Effectif* | 4 | 17 | 7 | 2 | 30 |
| *Fréquence* | 0,13 | 0,57 | 0,23 | 0,07 | 1 |

# 2 Caractéristiques de position

## 2.1 Moyenne

**Definition 2.1** ***valeurs discrêtes.***

*Soit une série statistique à caractère quantitatif, dont les $p$ valeurs sont données par $x_1$, $x_2$, ..., $x_p$ d'effectifs associés $n_1$, $n_2$, ..., $n_p$ avec $n_1 + n_2 + ... + n_p = N$.*
*La moyenne pondérée de cette série est le nombre noté $\overline{x}$ qui vaut*

$$\overline{x} = \frac{n_1x_1 + n_2x_2 + ... + n_px_p}{n_1 + n_2 + ... + n_p} = \frac{1}{N}\sum_{i=1}^{p} n_ix_i.$$

❒ ***Valeurs regroupés en classes.***

*Soit la série statistique suivante :*

| *Classes du caractère* | $[a_0; a_1[$ | $[a_1; a_2[$ | $[a_2; a_3[$ | ... | $[a_{p-1}; a_p]$ |
|---|---|---|---|---|---|
| *Effectifs* | $n_1$ | $n_2$ | $n_3$ | ... | $n_p$ |
| *Centre des classes* | $c_1$ | $c_2$ | $c_3$ | ... | $c_p$ |

*La moyenne de cette série statistique est le nombre noté $\overline{x}$ défini par :*

$$\overline{x} = \frac{n_1c_1 + n_2c_2 + n_3c_3 + ... + n_pc_p}{n_1 + n_2 + n_3 + ... + n_p}.$$

*Où chaque centre des classes $c_i$ est égal à* $\frac{a_{i-1} + a_i}{2}$.

**Remarque(s) 2.1** *Lorsque la série est regroupée en classes, on calcule la moyenne en prenant pour valeurs $x_i$ le* ***centre de chaque classe****; ce centre est obtenu en faisant la moyenne des deux extrémités de la classe.*

**Exemple(s) 2.1** ➔ *Dans la* ***série A****, la moyenne du contrôle est égale à* $\overline{m} = \frac{2 \times 1 + 3 \times 2 + \cdots + 16 \times 1}{30} = \frac{254}{30} \approx 8,47$,

➔ *si on regroupe par classe d'amplitude 5 points, une estimation de la moyenne est :* $\overline{m}\frac{2,5 \times 4 + 7.5 \times 17 + \cdots + 17,5 \times 2}{30} = \frac{260}{30} \approx 8,67$.

**Remarque(s) 2.2** *On peut aussi calculer une moyenne à partir de la distribution de fréquences :*

$$\overline{x} = f_1x_1 + f_2x_2 + \cdots + f_px_p = \sum_{i=1}^{p} f_ix_i.$$

**Proposition 2.1** ***Linéarité de la moyenne***

- ♦ *Si on ajoute (ou soustrait) un même nombre* $\boxed{k}$ *à toutes les valeurs d'une série, alors la moyenne de cette série se trouve augmentée (resp. diminuée) de* $\boxed{k}$.
- ♦ *Si on multiplie (ou divise) par un même nombre non nul* $\boxed{k}$ *toutes les valeurs d'une série, alors la moyenne de cette série se trouve multipliée (resp. divisée) par* $\boxed{k}$.

**Exercice(s) 2.1** *On considère la* ***série A*** *:*

- ➔ *Si on ajoute* $1,5$ *points à chaque note du contrôle, alors la moyenne de classe devient* $\overline{m} = 8,47 + 1,5 = 9,97$.
- ➔ *Si on augmente chaque note de* $10\%$*, cela revient à multiplier chaque note par* $1,1$*, ce qui donne* $\overline{m} = 8,47 \times 1,1 = 9,32$.

**Proposition 2.2** ***Moyenne par sous-groupes***
*Soit une série statistique, d'effectif total* $N$*, de moyenne* $\overline{x}$*.*
*Si on divise cette série en deux sous-groupes* ***disjoints*** *d'effectifs respectifs* $p$ *et* $q$ *(avec* $p+q=N$*) de moyennes respectives* $\overline{x}_1$ *et* $\overline{x}_2$*, alors on a :*

$$\boxed{\overline{x} = \frac{p}{N} \times \overline{x}_1 + \frac{q}{N} \times \overline{x}_2.}$$

**Exemple(s) 2.2** *On suppose par exemple que les* $12$ *garçons de la classe de la* ***série A*** *ont obtenu une moyenne globale de* $8$ *sur* $20$.

- ➔ *La moyenne du groupe formé par les filles de la classe vérifie :* $9,47 = \frac{12}{30} \times 8 + \frac{18}{30} \times \overline{m}_f$ .
- ➔ *Soit* $\overline{m}_f = \frac{30}{18}\left(9,47 - \frac{12}{30} \times 8\right) = 10,45$.

## 2.2 Médiane

**Definition 2.2** *Soit une série statistique ordonnée dont les $n$ valeurs sont $x_1 \leqslant x_2 \leqslant x_3 \leqslant \cdots \leqslant x_n$.*
*La médiane est un nombre $M$ qui permet de diviser cette série en deux sous-groupes de même effectif.*

- ➤ *Si $n$ est impair, $n$ est la valeur de cette série qui est située au milieu, à savoir la valeur dont le rang est $\dfrac{n+1}{2}$, notée $x_{\frac{n+1}{2}}$.*
- ➤ *Si $n$ est pair, $n$ est le centre l'intervalle médian, qui est l'intervalle formé par les deux nombres situés « au milieu » de la série, à savoir $x_{\frac{n}{2}}$ et $x_{\frac{n}{2}+1}$.*

**Exercice(s) 2.2**

- ➔ *La médiane de la série « $2-5-6-8-9-9-10$ » est* 8.
- ➔ *La médiane de la série « $2-5-6-8-9-9$ » est* 7.
- ➔ *La médiane de la série « $2-5-6-6-9-10$ » est* 6.

## 2.3 Quartiles

**Definition 2.3** *Le premier quartile et le troisiéme quartile de la série statistique $(x_1, x_2, x_3, ..., x_n)$ sont les nombres que l'on note souvent $Q_1$ et $Q_3$ définis de la façon suivante :*

1. *On range les valeurs de la série statistique par ordre croissant, chaque nombre figurant le nombre de fois égal à son effectif.*
2. *$Q_1$ est la **plus petite valeur de la série** telle qu'au moins 25% des valeurs lui soient inférieures ou égales.*
3. *$Q_3$ est la **plus petite valeur de la série** telle qu'au moins 75% des valeurs lui soient inférieures ou égales.*
4. *L'intervalle $]Q_1; Q_3[$ est appelé **intervalle interquartile**. Le nombre $Q_3 - Q_1$ est appelé **écart interquartile**.*

**Exemple(s) 2.3** *Voici les tailles en cm de 22 enfants de 8 ? 11 ans :*
*123, 116, 113, 133, 129, 142, 119, 137, 146, 139, 112, 121, 130, 150, 128, 127, 147, 126, 119, 150, 133, 117.*
*Déterminer la médiane et les premier et troisième quartiles.*

***Utilisation de la calculatrice.***

## 3 Caractéristiques de dispersion

**Definition 3.1** *On appelle étendue d'une série discrète $X$ le réel défini par* $\boxed{E(X) = \max(X) - \min(X)}$.

## 4 Variance, écart type d'une série statistique.

**Definition 4.1** *Soit la série statistique suivante :*

| *Valeur du caractére* | $x_1$ | $x_2$ | $x_3$ | ... | $x_p$ |
|---|---|---|---|---|---|
| *Effectifs* | $n_1$ | $n_2$ | $n_3$ | ... | $n_p$ |
| *Fréquences* | $f_1$ | $f_2$ | $f_3$ | ... | $f_p$ |

*On* $N = n_1 + n_2 + ... + n_p$ *et pour tout* $i \in \mathbb{N}$, $1 \leq i \leq p$, *on a* $\boxed{f_i = \frac{n_i}{N}}$.
*La* ***Variance*** *de cette série statistique est le nombre noté* $V$ *ou* $V(x)$ *est défini par :*

$$V = \frac{n_1(x_1 - \overline{x})^2 + n_2(x_2 - \overline{x})^2 + n_3(x_3 - \overline{x})^2 + ... + n_p(x_p - \overline{x})^2}{n_1 + n_2 + n_3 + ... + n_p} = \Sigma_{i=1}^{p} f_i(x_i - \bar{x})^2.$$

*Afin de mesurer la dispersion d'une série statistique avec la même unité que les valeurs de la série, on définit* ***l'écart type*** *de la série par* $\boxed{\sigma = \sqrt{V}}$.

**Interprétation graphique** :

Considérons les deux série statistiques suivantes :

| $x_i$ | 1 | 2 | 3 | 17 | 20 |
|---|---|---|---|---|---|

$\bar{x} =$ .....

| $y_i$ | 9 | 10 | 11 | 12 | 12,5 |
|---|---|---|---|---|---|

Calculons $V(x)$ et $V(y)$ :

$\bar{y} =$ .....

$V(x) =$ .....

$V(y) =$ .....