Statistiques

Christophe ROSSIGNOL*

Année scolaire 2020/2021

Table des matières

1 Moyenne et écart-type

1.1 Moyenne pondérée

Définition : On considère la série statistique suivante :

valeur du caractère	x_1	x_2	x_3	$\cdots$	x_p
effectif	n_1	n_2	n_3	$\cdots$	n_p

L'effectif total est : $N = n_1 + n_2 + n_3 + \ldots + n_p$.
La moyenne de la série est :

$$m = \frac{n_1x_1 + n_2x_2 + \ldots + n_px_p}{N}$$

Remarques :

1. Pour une série statistique simple (non regroupée suivant les effectifs) x_1, x_2,…,x_n la formule de la moyenne est plus simplement : $m = \frac{x_1+x_2+\ldots+x_n}{n}$.
2. Pour une série dont les valeurs sont regroupées en classes, on utilise le centre de chaque classe comme valeur de x_i dans le calcul de la moyenne.
3. La moyenne est très sensible aux valeurs extrêmes.

On considère la série statistique suivante :

Exemple 1 :

valeur	30	45	50	60	61
effectif	2	3	2	2	2

La moyenne est :

$$m_1 = \frac{2\times 50 + 3\times 45 + 2\times 30 + 2\times 60 + 2\times 61}{11} \simeq 48,8$$

Exemple 2 : On considère la série statistique suivante :

valeur	2	5	6	8	9
effectif	3	3	1	3	2

La moyenne est :

$$m_2 = \frac{3\times 2 + 3\times 5 + 1\times 6 + 3\times 8 + 2\times 9}{12} = 5,75$$

Exercices : 1 page 294 ; 17, 18, 21 page 296 [1] – 43, 44 page 298 [2] – 48 page 298 [3] [Magnard]

1.2 Linéarité de la moyenne

Activité : Activité 2 page 288 [4] [Magnard]

Propriété : On considère une série statistique de moyenne m.

— Si on multiplie (ou on divise) toutes les valeurs de la série par le même nombre a, la moyenne de la nouvelle série est elle-aussi multipliée (ou divisée) par a.
La moyenne de la nouvelle série est donc $a \times m$ (ou $\frac{m}{a}$ dans le cas d'une division) ;

— Si on ajoute (ou on soustrait) à toutes les valeurs de la série le même nombre b, la moyenne de la nouvelle série est elle-aussi obtenue en ajoutant (ou en soustrayant) le nombre b.
La moyenne de la nouvelle série est donc $m + b$ (ou $m - b$ dans le cas d'une soustraction).

Exemple 1 : On reprend la série statistique de l'exemple 1 du 1.1, et on multiplie toutes les valeurs de la série par 2. On obtient le tableau suivant :

1. Calcul de moyenne.
2. Avec des données inconnues
3. Cas d'une série regroupée en classes.
4. Découvrir la linéarité de la moyenne.

valeur	60	90	100	120	122
effectif	2	3	2	2	2

La moyenne sera alors $m'_1 = 2 \times m_1 \simeq 97,6$.

Exemple 2 : On reprend la série statistique de l'exemple 2 du 1.1, et on ajoute à toutes les valeurs de la série le nombre 6. On obtient le tableau suivant :

valeur	8	11	12	14	15
effectif	3	3	1	3	2

La moyenne sera alors $m'_2 = m_2 + 6 \simeq 11,75$.

Exercices : 23, 24 page 296 ; 42 page 298[5] – 25, 26 page 297[6] [Magnard]

1.3 Écart-type

Activité : Activité 3 page 289[7] [Magnard]

Définition : L'écart-type d'une série statistique est un nombre positif, noté s, qui est un indicateur de la dispersion de cette série autour de la moyenne :

- plus l'écart-type s d'une série statistique est petit, plus les valeurs de la série sont concentrées autour de la moyenne, donc plus la série est homogène ;
- plus l'écart-type s d'une série statistique est grand, plus les valeurs de la série sont éloignées de la moyenne, donc plus la série est hétérogène.

Remarques :

1. On utilisera la calculatrice pour déterminer l'écart-type (et même parfois la moyenne) d'une série statistique. Voir pour cela le TP 1 page 304 [Magnard].
2. Il existe une formule pour calculer l'écart-type, mais celle-ci n'est pas exigible en classe de Seconde :

   On considère la série statistique suivante :

valeur du caractère	x_1	x_2	x_3	$\cdots$	x_p
effectif	n_1	n_2	n_3	$\cdots$	n_p

   - L'effectif total est : $N = n_1 + n_2 + n_3 + \ldots + n_p = \sum_{i=1}^{p} n_i$.
   - La moyenne est :
   $$m = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_p x_p}{N}$$
   - L'écart-type s est :
   $$s = \sqrt{\frac{n_1 (x_1 - m)^2 + n_2 (x_2 - m)^2 + \ldots + n_p (x_p - m)^2}{N}}$$
3. Le couple (moyenne , écart-type) permet de comparer plusieurs séries statistiques.

On reprend la série statistique de l'exemple 1 du 1.1.

En utilisant la calculatrice, on trouve que l'écart-type de cette série est $s_1 \simeq 10,9$.

Exemple 1Exemple 2 : On reprend la série statistique de l'exemple 2 du 1.1.

En utilisant la calculatrice, on trouve que l'écart-type de cette série est $s_2 \simeq 2,6$.

Exercices : 3 page 294 ; 29, 30, 32 page 297[8] – 27, 28 page 297 et 52, 53 page 299[9] [Magnard]

Module : TP2 page 305[10] [Magnard]

5. Linéarité de la moyenne.
6. Avec des évolutions en pourcentage.
7. Découvrir l'écart-type.
8. Calculer l'écart-type.
9. Utiliser l'écart-type.
10. « Forme » d'une série.

2 Médiane, quartiles, écart-interquartile

2.1 Médiane

Définition : On considère une série statistique dont les valeurs du caractère étudié ont été rangés dans l'ordre croissant :

$$x_1 \leq x_2 \leq x_3 \leq \cdots \leq x_n$$

On appelle médiane la valeur centrale de cette série, c'est-à-dire celle qui la sépare en deux parties de même effectif.
On la note : **Me**.

Remarques : 1. Si l'effectif total est impair, la médiane correspond à la valeur centrale.
Si l'effectif total est pair, la médiane correspond à la demi-somme des deux valeurs centrales.

2. Au moins 50 % des valeurs de la série sont inférieures (ou égales) à la médiane et au moins 50 % des valeurs de la série lui sont supérieures (ou égales).
3. La médiane est beaucoup moins sensible aux valeurs extrêmes que la moyenne.

Exemple 1 : On reprend la série statistique de l'exemple 1 du 1.1.
L'effectif total est $N = 11$, il est impair. La médiane est donc la $6^{\text{ème}}$ valeur, classées dans l'ordre croissant.
La médiane est $\mathbf{Me} = 50$.

Exemple 2 : On reprend la série statistique de l'exemple 2 du 1.1.
L'effectif total est $N = 12$, il est pair. La médiane est donc située entre la $6^{\text{ème}}$ et la $7^{\text{ème}}$ valeur, classées dans l'ordre croissant.
La médiane est $\mathbf{Me} = \frac{5+6}{2} = 5,5$.

2.2 Quartiles, écart interquartile

Définition : On considère une série statistique dont les valeurs du caractère étudié ont été rangés dans l'ordre croissant :

$$x_1 \leq x_2 \leq x_3 \leq \cdots \leq x_n$$

1. Le premier quartile est la plus petite valeur Q_1 de la liste telle qu'au moins un quart des valeurs de la liste sont inférieures ou égales à Q_1.
2. Le troisième quartile est la plus petite valeur Q_3 de la liste telle qu'au moins les trois quart des valeurs de la liste sont inférieures ou égales à Q_3.
3. On appelle écart interquartile la quantité $Q_3 - Q_1$.

Exemple 1 :

On reprend la série statistique de l'exemple 1 du 1.1.

— L'effectif total est $N = 11$.
— $\frac{N}{4} = \frac{11}{4} = 2,75$. Comme au moins un quart des valeurs doit être inférieure à Q_1, Q_1 est donc la $3^{\text{ème}}$ valeur (classée dans l'ordre croissant...). On a donc $Q_1 = 45$.
— $\frac{3N}{4} = \frac{33}{4} = 8,25$. Comme au moins les trois quart des valeurs doit être inférieure à Q_3, Q_3 est donc la $9^{\text{ème}}$ valeur (classée dans l'ordre croissant...). On a donc $Q_3 = 60$.
— L'écart interquartile est donc $Q_3 - Q_1 = 60 - 45 = 15$.

Exemple 1 :

On reprend la série statistique de l'exemple 2 du 1.1.

— L'effectif total est $N = 12$.
— $\frac{N}{4} = \frac{12}{4} = 3$. Comme au moins un quart des valeurs doit être inférieure à Q_1, Q_1 est donc la $3^{\text{ème}}$ valeur (classée dans l'ordre croissant...). On a donc $Q_1 = 2$.
— $\frac{3N}{4} = \frac{36}{4} = 9$. Comme au moins les trois quart des valeurs doit être inférieure à Q_3, Q_3 est donc la $9^{\text{ème}}$ valeur (classée dans l'ordre croissant...). On a donc $Q_3 = 8$.
— L'écart interquartile est donc $Q_3 - Q_1 = 8 - 2 = 6$.

Remarques :

1. On peut utiliser la calculatrice pour déterminer la médiane et les quartiles d'une série statistique. Voir pour cela le TP 1 page 304 [Magnard].
2. On a donc partagé la série en quatre parties de même effectif, comme indiqué sur le schéma suivant :

$$x_{min}\underbrace{\cdots\cdots\cdots\cdots\cdots}_{25\ \text{\%de l'effectif}}\underbrace{\mathbf{Q_1}\underbrace{\cdots\cdots\cdots\cdots\cdots}_{25\ \text{\%de l'effectif}}\mathbf{Me}\underbrace{\cdots\cdots\cdots\cdots\cdots}_{25\ \text{\%de l'effectif}}\mathbf{Q_3}}_{50\ \text{\%de l'effectif}}\underbrace{\cdots\cdots\cdots\cdots\cdots}_{25\ \text{\%de l'effectif}}x_{max}$$

   — 25 % de l'effectif a une valeur du caractère comprise entre x_{min} et $\mathbf{Q_1}$;
   — 50 % de l'effectif a une valeur du caractère comprise entre $\mathbf{Q_1}$ et $\mathbf{Q_3}$;
   — 25 % de l'effectif a une valeur du caractère comprise entre $\mathbf{Q_3}$ et x_{max}.
3. Entre Q_1 et Q_3, on trouve donc les 50 % de l'effectif dont les valeurs sont « les plus proches » de la médiane.
   L'écart interquartile, qui est une mesure de la longueur de cet intervalle, est donc une mesure de la dispersion des données autour de la médiane :
   — plus il est grand, plus les données sont dispersées autour de la médiane ;
   — plus il est petit, plus les données sont proches de la médiane.

Exercices : 5, 6 page 195 ; 35, 36 page 297 [11] [Magnard]

Exercices de synthèse : 58, 59 page 300 [12] – 61, 63 page 301 [13] – 66 page 302 [14] – 68 page 302 [15] [Magnard]

Références

[Magnard] Maths 2$^{\text{de}}$, MAGNARD, 2019

2, 3, 5

11. Calculer des quartiles et l'écart-interquartile.
12. Choisir le bon indicateur.
13. Étude de série statistiques.
14. Augmentation de salaire.
15. Une question de débit.