

Chapitre 5

STATISTIQUES DESCRIPTIVES

I. Vocabulaire et premiers outils de comparaison

CONTEXTE : Lors d'une épreuve orale, 2 professeurs A et B ont interrogé respectivement 15 et 16 élèves.

Voici les notes des élèves :

Série du professeur A : 7-7-7-8-8-9-10-11-11-11-12-12-12-12-13

Série du professeur B : 2-3-4-4-5-6-7-10-12-12-13-13-16-17-18-18

Le but de ce chapitre est de COMPARER les résultats de ces deux groupes d'élèves.

1. Vocabulaire et notations

Population : ensemble sur lequel porte l'étude (attention la population n'est pas forcément humaine !)
Dans notre ex : 2 groupes d'élèves

Caractère : aspect étudié ou objet de l'étude. *Dans notre ex : note à un oral*

Nature du caractère : quantitatif (qui se mesure) ou qualitatif (qui ne se mesure pas)
Dans notre ex : la variable est quantitative

Modalités : différentes valeurs prises par le caractère. Notation x_1, x_2, \dots, x_p (p est le nombre de modalités différentes)

Effectif : - L'effectif total est l'effectif de la population, noté N . C'est le nombre total d'individus.

- L'effectif d'une valeur est le nombre de fois que cette valeur apparaît dans l'étude. On le note n_i .

Fréquence : quotient de l'effectif de cette valeur par l'effectif total : $f_i = \frac{n_i}{N}$.

Propriété sur les fréquences :

1. Une fréquence est un nombre toujours compris entre 0 et 1
2. La somme des fréquences d'une série statistique est égale à 1

Résumé pour la série A

Modalités	$x_1=7$	$x_2=8$	$x_3=9$	$x_4=10$	$x_5=11$	$x_6=12$	$x_7=13$
Effectifs	$n_1=3$	$n_2=2$	$n_3=1$	$n_4=1$	$n_5=3$	$n_6=4$	$n_7=1$

Effectif total $N=3+2+1+1+3+4+1=15$

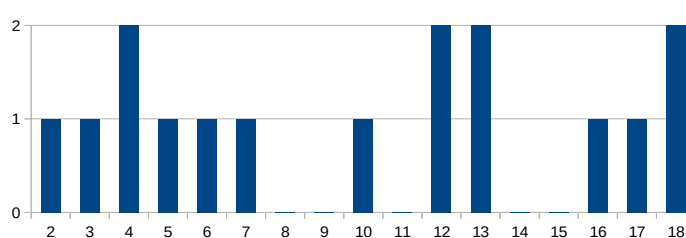
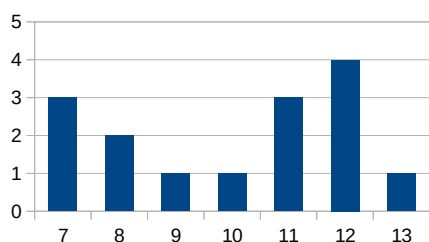
Résumé pour la série B : à compléter

Modalités	2	3	4	5	6	7	10	12	13	16	17	18
Effectifs	1	1	2	1	1	1	1	2	2	1	1	2

Effectif total $N=16$

Représenter ci-dessous ces deux séries par deux diagrammes en bâtons.

(en abscisses : les notes ; en ordonnées : les effectifs)



2. Des outils de comparaison : indicateurs de position

a. Moyenne

Définition : Soit une série statistique dont les valeurs du caractère sont x_1, x_2, \dots, x_k et n_1, n_2, \dots, n_k les effectifs associés.

La **moyenne pondérée** de la série statistique, notée \bar{x} , a pour valeur : $\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_k x_k}{N}$

x_1	x_2	...	x_k
n_1	n_2		n_k

Dans notre exemple (*faire les calculs !*) :

la moyenne de la série A est 10

$$\frac{3 \times 7 + 2 \times 8 + 1 \times 9 + 1 \times 10 + 3 \times 11 + 4 \times 12 + 1 \times 13}{15} = \frac{150}{15} = 10$$

la moyenne de la série B est 10

$$\frac{1 \times 2 + 1 \times 3 + 2 \times 4 + \dots + 2 \times 18}{16} = \frac{160}{16} = 10$$



Coup de pouce : Si tu n'arrives pas à calculer la moyenne, consulte la vidéo suivante :

📺 Vidéo https://youtu.be/88_16UbkdZM

b. Médiane

Définition : La **médiane** (M_e) d'une série statistique est le plus petit nombre tel que 50% au moins des individus aient une valeur du caractère inférieure ou égale à ce nombre.

Détermination de la médiane : On range la série par ordre croissant de valeurs.

- si l'effectif total N est impair alors la médiane est la valeur centrale, c'est-à-dire le terme de rang $\frac{N+1}{2}$
- si l'effectif total N est pair alors la médiane est la demi somme des termes de rang $\frac{N}{2}$ et $\frac{N}{2} + 1$

SÉRIE A (N=15) : compléter la ligne des Effectifs Cumulés Croissants (ECC) pour vous aider.

Modalités	$x_1=7$	$x_2=8$	$x_3=9$	$x_4=10$	$x_5=11$	$x_6=12$	$x_7=13$
Effectifs	$n_1=3$	$n_2=2$	$n_3=1$	$n_4=1$	$n_5=3$	$n_6=4$	$n_7=1$
ECC	3	5	6	7	10	14	15

Dans notre exemple (*faire les calculs !*) :

la médiane de la série A est 11

l'effectif total $N=15$, N est impair donc la médiane

est le terme de rang $\frac{15+1}{2} = \frac{16}{2} = 8$

le terme de rang 8 est la valeur 11.

(repérer le premier ECC au moins 8, c'est pour la modalité x_5)

la médiane de la série B est 11

l'effectif total $N=16$, N est pair . $\frac{N}{2} = 8$, $\frac{N}{2} + 1 = 9$

La médiane est la moyenne des termes de rang 8 et 9

$$\frac{10+12}{2} = 11$$



Coup de pouce : Si tu n'arrives pas à calculer la médiane, consulte la vidéo suivante :

📺 Vidéo <https://youtu.be/kr90dXv0NFY>

CONCLUSION : ces deux séries ont la même moyenne et la même médiane pourtant elles semblent différentes quant à leur dispersion. Nous verrons par la suite d'autres indicateurs.

c. Quartiles

Définitions :

- Le **1^{er} quartile Q1** est la plus petite valeur de la série telle qu'au moins 25% (soit $\frac{1}{4}$) des valeurs soient inférieures ou égales à Q1.
- Le **3^{ème} quartile Q3** est la plus petite valeur de la série telle qu'au moins 75% (soit $\frac{3}{4}$) des valeurs soient inférieures ou égales à Q3.

Dans notre exemple : série A rangée dans l'ordre croissant (faire les calculs pour la série B)

$$N=15; \frac{15}{4}=3.75; \text{ Q1 est la 4^{ème} valeur soit Q1=8}$$

$$3 \times \frac{15}{4}=11.25; \text{ Q3 est la 12^{ème} valeur soit Q3=12. (On arrondit toujours au-dessus)}$$

Série B :

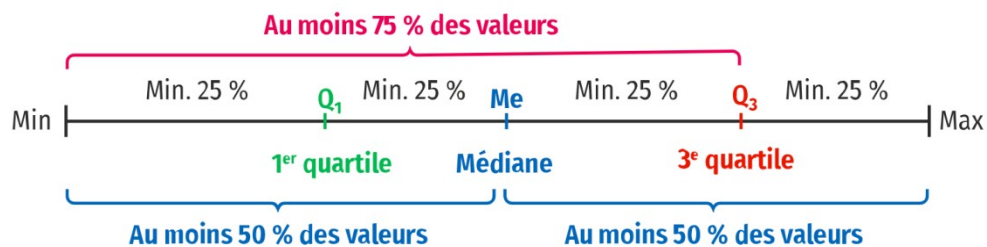
$$N=16; \frac{16}{4}=4; \text{ Q1 est la 4^{ème} valeur soit Q1=4}$$

$$3 \times \frac{16}{4}=12; \text{ Q3 est la 12^{ème} valeur soit Q3=13}$$



Coup de pouce : Si tu n'arrives pas à calculer les quartiles, consulte la vidéo suivante :

Vidéo <https://youtu.be/2jbpNjXMdSA>



II. Indicateurs de dispersion

1. L'écart interquartile

Définition :

- L'intervalle [Q1 ; Q3] s'appelle **l'intervalle interquartile**. Il contient au moins 50% des valeurs de la série.
- Le nombre Q3-Q1 s'appelle **l'écart interquartile**.

Dans notre exemple (faire les calculs !) :

l'écart interquartile de la série la série A est :4

$$Q3-Q1=12-8=4$$

l'écart interquartile de la série la série B est :9

$$Q3-Q1=13-4=9$$



Coup de pouce : Si tu n'arrives pas à calculer l'écart interquartile, consulte la vidéo suivante :

Vidéo <https://youtu.be/ljsDK0ODwIw>

Remarques :

- L'écart interquartile mesure la **dispersion** des valeurs autour de la médiane ; **plus l'écart est petit, plus les valeurs de la série appartenant à l'intervalle interquartile sont concentrées autour de la médiane.**
- L'écart interquartile élimine les valeurs extrêmes qui peuvent être douteuses, cependant il ne tient compte que de 50% de l'effectif.

2. L'écart-type

Au lieu de comparer et de classer les entre-elles, on s'intéresse maintenant à leurs écarts à la moyenne.

Définitions :

- La **variance V** d'une série statistique de moyenne \bar{x} dont les valeurs du caractère sont $x_1, x_2, x_3, \dots, x_k$ et les effectifs correspondants sont $n_1, n_2, n_3, \dots, n_k$ est égale à :
$$V = \frac{n_1 \times (x_1 - \bar{x})^2 + n_2 \times (x_2 - \bar{x})^2 + \dots + n_k \times (x_k - \bar{x})^2}{n_1 + n_2 + \dots + n_k}$$

- **L'écart-type σ** d'une série statistique de variance V est égal à : $\sigma = \sqrt{V}$.

Dans notre exemple (*faire les calculs !*) :

la variance et l'écart type de la série la série A est :

Ici , $n_1 + n_2 + \dots + n_k = N = 15$ (effectif total)

et la moyenne est $\bar{x} = 10$

La variance est :

$$V = \frac{3 \times (7 - 10)^2 + 2 \times (8 - 10)^2 + \dots + 1 \times (13 - 10)^2}{15}$$

$$V = \frac{3 \times (-3)^2 + 2 \times (-2)^2 + \dots + 1 \times (3)^2}{15}$$

$$V = \frac{27 + 8 + 1 + 0 + 3 + 16 + 9}{15} = \frac{64}{15} \approx 4.26$$

L'écart Type est :

$$\sigma = \sqrt{\frac{64}{15}} \approx 2.07$$

la variance et l'écart type de la série la série B est :

Ici $N = 16$ et et la moyenne est aussi $\bar{x} = 10$

La variance est :

$$V = \frac{1 \times (2 - 10)^2 + 1 \times (3 - 10)^2 + \dots + 2 \times (18 - 10)^2}{16}$$


$$V = \frac{1 \times (-8)^2 + 1 \times (-7)^2 + \dots + 2 \times (8)^2}{16}$$

$$V = \frac{64 + 49 + 72 + 25 + 16 + 9 + 0 + 8 + 18 + 36 + 49 + 128}{16}$$

$$V = \frac{474}{16} = 29.625$$

L'écart Type est :

$$\sigma = \sqrt{29.625} \approx 5.44$$

 **Coup de pouce :** Si tu n'arrives pas à calculer la variance et l'écart type, consulte la vidéo suivante :

 **Vidéo** <https://youtu.be/CiFoBkipJQk>

Remarques :

• L'écart interquartile mesure la **dispersion** des valeurs autour de la moyenne ; **plus l'écart type est petit, plus les valeurs de la série sont concentrées autour de la moyenne.**

CONCLUSION :

La série A est donc bien **moins** dispersée autour de sa moyenne que la série B, le groupe d'élèves **B** est plus hétérogène.

Remarque :

Votre calculatrice donne la moyenne pondérée, la médiane, les quartiles et l'écart-type d'une série de données quantitatives. Nous verrons son utilisation en classe.

Pour les impatients, consulter la vidéo suivante :

 **Vidéo** <https://www.youtube.com/watch?v=uwL6J3Sp0Qw>

Erratum