

Chapitre 10 : Statistiques descriptives.

Eliot ROY

2022-2023

Points du programme.

Contenus.

- Indicateurs de tendance centrale d'une série statistique : moyenne pondérée.
- Linéarité de la moyenne.
- Indicateur de dispersion : écart-type.

Capacités attendues.

- Décrire verbalement les différences entre deux séries statistiques, en s'appuyant sur des indicateurs ou sur des représentations graphiques données.
- Pour des données réelles ou issues d'une simulation, lire et comprendre une fonction écrite en Python renvoyant la moyenne m , l'écart-type s , et la proportion d'éléments appartenant à $[m - 2s; m + 2s]$.

Activité 1 : diapo de rappels sur le calculs d'une moyenne.

Dans ce cours, on considérera la série statistique donnée par le tableau ci-contre et on notera $N = n_1 + n_2 + \dots + n_p$ l'effectif total.

Valeur	x_1	x_2	...	x_p
Effectif	n_1	n_2	...	n_p

I Moyenne pondérée.

I.1 Définition.

Définition : La moyenne pondérée de la série statistique est le nombre noté $\bar{x}$ défini par :

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{N}$$

Exemple : Voici les résultats d'une interrogation de classe de seconde, notée sur 5 points :

Note	1	2	4	5
Effectif	2	5	9	4

L'effectif total de cette série statistique est $N = 2 + 5 + 9 + 4 = 20$ et la moyenne pondérée est :

$$\bar{x} = \frac{2 \times 1 + 5 \times 2 + 9 \times 4 + 4 \times 5}{20} = \frac{68}{20} = 3,4.$$

La moyenne de cette interrogation est donc de $3,4/5$.

Remarque : On dit que la moyenne est pondérée car chaque valeur est multipliée par un poids (son effectif ici).

Pour une série regroupée par classes, on obtient une valeur approchée de la moyenne en prenant pour x_i le centre des classes.

Exemple : Le tableau ci-dessous donne la durée de vie (en années) de 20 télévisions :

Durée de vie	$[0; 1[$	$[1; 2[$	$[2; 3[$	$[3; 4[$	$[4; 5[$
Effectif	3	2	7	3	5

Pour calculer $\bar{x}$, on va utiliser le centre de chaque classe :

$$\bar{x} = \frac{3 \times 0,5 + 2 \times 1,5 + 7 \times 2,5 + 3 \times 3,5 + 5 \times 4,5}{20} = \frac{55}{20} = 2,75.$$

I.2 Propriétés.

Définition : La fréquence f_i d'une valeur x_i de la série statistique est $f_i = \frac{n_i}{N}$.

Exemple : Reprenons l'exemple précédent, on obtient pour chacune des valeurs les fréquences suivantes :

Note	1	2	4	5
Effectif	2	5	9	4
Fréquence	$\frac{2}{20} = 0,1$	$\frac{5}{20} = 0,25$	$\frac{9}{20} = 0,45$	$\frac{4}{20} = 0,2$

Par exemple, 25% des élèves de cette classe ont eu une note égale à 2/5.

On remarque que la somme des fréquences est égale à 1 :

$$\frac{2}{20} + \frac{5}{20} + \frac{9}{20} + \frac{4}{20} = \frac{20}{20} = 1$$

Propriété : Soit f_i la fréquence de chacune des valeurs x_i de la série statistique, alors :

$$\bar{x} = f_1x_1 + f_2x_2 + \dots + f_px_p$$

Exemple : Reprenons l'exemple précédent,

$$0,1 \times 1 + 0,25 \times 2 + 0,45 \times 4 + 0,2 \times 5 = 3,4 = \bar{x}.$$

Propriété - linéarité de la moyenne : Soient a et b deux nombres réels et m la moyenne de la série statistique (x_i) .

La moyenne de la série $(ax_i + b)$ est $M = am + b$.

Exemple : A un examen, la moyenne des notes des candidats est 7,6.

Si le jury augmente de 25% toutes les notes et ajoute 2,4 points à chacun des candidats, alors la nouvelle moyenne vaut $7,6 \times 1,25 + 2,4 = 9,5 + 2,4 = 11,9$.

Objectif : Calculer la moyenne d'une série statistique.

Communs : Exercices 18 page 296, 25 page 297.

Pour les plus rapides : Exercices 19, 20, 21, 23, 24 page 296.

Activité 2 : Quel élève a autant d'amis plus petit que lui que d'amis plus grand que lui ?

II Médiane et Quartile.

II.1 Médiane.

Définition : La médiane de la série statistique **rangée par ordre croissant** est le nombre Me tel qu'au moins 50% des valeurs de la série lui sont inférieures ou égales, et au moins 50% lui sont supérieures ou égales.

Méthode : ✎ Déterminer la médiane.

- Si l'effectif N de la série est impair, alors la médiane est la valeur centrale de la série.
- Si N est pair, la médiane est la demi-somme des valeurs centrales de la série.

Exemple : On considère la série statistique suivante ordonnée par ordre croissant :

7-7-8-9-10-10-11-11-11-12-14-15

L'effectif 12 est pair et $12 = 2 \times 6$ donc la médiane est la demi-somme de la 6^{eme} et 7^{eme} valeur de la série : $Me = \frac{10+11}{2} = 10,5$.

Remarque : La médiane ne dépend pas des valeurs extrêmes de la série, contrairement à la moyenne. Elle est la valeur qui partage une série en deux groupes de même effectif.

Définition : Soit une série statistique dont les valeurs sont ordonnées par ordre croissant.

Le premier quartile, noté Q_1 , est la plus petite valeur de la série telle qu'au moins 25% des valeurs lui sont inférieures ou égales.

Le troisième quartile, noté Q_3 , est la plus petite valeur de la série telle qu'au moins 75% des valeurs lui sont inférieures ou égales.

Méthode : ✎ Déterminer le premier et le troisième quartile.

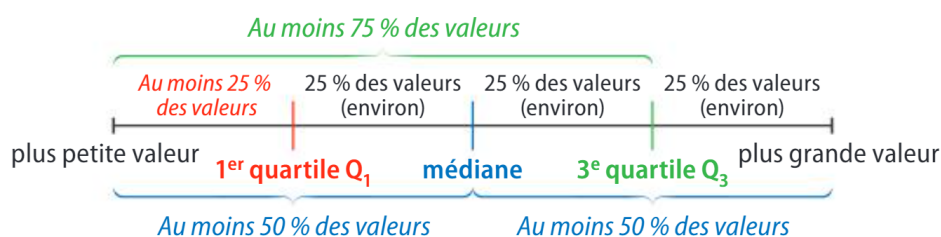
- Le rang du premier quartile d'une série d'effectif total N est le plus petit entier supérieur ou égal à $\frac{N}{4}$.
- Le rang du troisième quartile est le plus petit entier supérieur ou égal à $\frac{3}{4}N$.

Exemple : On reprend les données de l'exemple précédent.

L'effectif total est 12 et $\frac{12}{4} = 3$ donc le premier quartile est la 3^{ème} valeur, soit $Q_1 = 8$.

De même, comme $\frac{3}{4} \times 12 = 9$, le troisième quartile est la 9^{ème} valeur, soit $Q_3 = 11$.

Remarque : La médiane et les quartiles donnent des indications sur la répartition des valeurs d'une série statistique.



Définition : L'écart inter-quartile d'une série statistique est $Q_3 - Q_1$.

Exemple : Dans l'exemple précédent, l'écart-interquartile vaut $Q_3 - Q_1 = 11 - 8 = 3$.

Objectif : Calculer des quartiles d'une série statistique.

Communs : Exercices 33, 35

Pour les plus rapides : Exercices 34, 36 page 297.

Activité 3.

III Variance et écart-type.

Définition : La variance d'une série statistique est le réel noté V défini par :

$$V = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_p(x_p - \bar{x})^2}{N}$$

L'écart-type d'une série statistique est le réel $\sigma = \sqrt{V}$.

Exemple : Reprenons les résultats de l'interrogation notée sur 5 points :

Note	1	2	4	5
Effectif	2	5	9	4

La variance de cette série statistique est :

$$V = \frac{2 \times (1-3,4)^2 + 5 \times (2-3,4)^2 + 9 \times (4-3,4)^2 + 4 \times (5-3,4)^2}{20} = \frac{34,8}{20} = 1,74.$$

Donc $\sigma = \sqrt{1,74} \approx 1,32$.

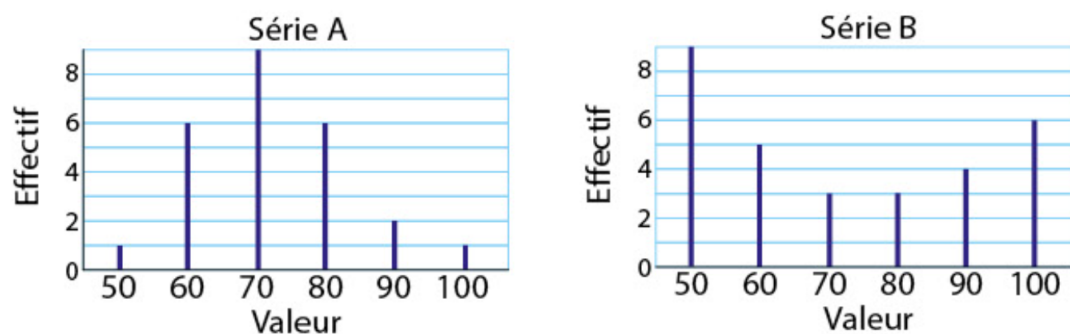
Remarques :

- L'écart-type a la même unité que les x_i et mesure la dispersion des valeurs autour de la moyenne : plus il est grand, plus les valeurs sont dispersées et éloignées de la moyenne.
- La variance et l'écart-type sont des nombres positifs.

IV Décrire les différences entre deux séries.

On résume souvent une série statistique par un couple associant un indicateur de tendance centrale à un indicateur de dispersion : généralement, on choisit (médiane ; écart-interquartile) ou (moyenne ; écart-type). Cela permet de comparer deux séries statistiques.

Exemple : Considérons les séries statistiques A et B étudiant le même caractère dans deux populations différentes :



Dans le tableau ci-dessous, on a regroupé la moyenne et l'écart-type des deux séries :

	Effectif total	Moyenne	Ecart-type
Série A	25	$m_A = 72$	$\sigma_A = 11,3$
Série B	30	$m_B = 72$	$\sigma_B = 19,4$

La moyenne des deux séries est identique mais on remarque sur les diagrammes en bâtons que les valeurs de la série B semblent plus dispersées que celles de la série A. Cela est confirmé par l'écart-type de la série B qui est plus élevé que celui de la série A.

Objectifs : Calculer un écart-type, décrire verbalement les différences entre deux séries statistiques.

Communs : Exercice 31 page 297, 52, 54 page 299.

Pour les plus rapides : Exercices 29, 30 page 297, 53, 51 page 299.