Cours

Statistiques. Indicateurs d'une série statistique

Moyenne

Soit $x_1, x_2 \dots x_N$ une série statistique quantitative de N valeurs. La moyenne de la série notée $\bar{x}$ est le quotient de la somme des valeurs par l'effectif total N :	$\bar{x} = \dfrac{x_1 + x_2 + \dots x_N}{N}$
Soit $x_1, x_2 \dots x_p$ une série statistique quantitative de p valeurs affectée des coefficients $n_1, n_2 \dots n_p$ La moyenne pondérée de la série notée $\bar{x}$ est le quotient :	$\bar{x} = \dfrac{n_1 * x_1 + n_2 * x_2 + \dots n_p * x_p}{n_1 + n_2 + \cdots n_p}$
Exemple : Audrey prend souvent le train venant de Rouen en direction de Paris. En rentrant dans le wagon, elle compte le nombre de places disponibles. Après 20 trajets, elle obtient les résultats ci-contre. Le nombre moyen de places assises disponibles sur ces 20 trajets est égal à : $\dfrac{5*0+1*1+3*2+1*5+5*6+4*7+1*10}{5+1+3+1+5+4+1} = 4$	(voir tableau ci-dessous)

Valeur	0	1	2	5	6	7	10
Effectif	5	1	3	1	5	4	1

Propriété : Linéarité de la moyenne

Soient a et b deux nombres réels et $x_1, x_2 \dots x_N$ une série statistique de moyenne m

- Si on multiplie par a toutes les valeurs de la série, on obtient la nouvelle moyenne en multipliant par a l'ancienne moyenne.
- Si on ajoute b à toutes les valeurs de la série, on obtient la nouvelle moyenne en ajoutant b à l'ancienne moyenne.

La moyenne de la série $ax_1 + b, ax_2 + b, \dots ax_N + b$ est am+b

Démonstration

$$\frac{(ax_1+b)+(ax_2+b)+\dots(ax_N+b)}{N} = \frac{(ax_1+ax_2+\dots ax_N)+(b+b+\cdots b)}{N} = \frac{a(x_1+x_2+\dots x_N)+(b+b+\cdots b)}{N}$$

$$= \frac{a(x_1+x_2+\dots x_N)}{N} + \frac{b+b+\cdots b}{N} = am + \frac{Nb}{N} = am + b$$

Ecart type

Soit $x_1, x_2 \dots x_N$ une série statistique quantitative de N valeurs et de moyenne m

Définition

Pour la série $x_1, x_2 \dots x_N$ de moyenne m, l'écart type σ est égal à la racine carrée de la variance V

$$V = \frac{(x_1-m)^2+(x_2-m)^2+\dots(x_N-m)^2}{N} \qquad \sigma = \sqrt{V} = \sqrt{\frac{(x_1-m)^2+(x_2-m)^2+\dots(x_N-m)^2}{N}}$$

Pour la série donnée ci-bas de moyenne m et d'effectif total $n_1 + n_2 + \cdots + n_p = N$

Valeur	x_1	x_2	…	x_p
Effectif	n_1	n_2	…	n_p

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

$$\sigma = \sqrt{V}$$

Remarque

L'écart type d'une série statistique est un indicateur de dispersion de cette série autour de la moyenne. Concrètement il donne une certaine mesure de l'écart entre les valeurs de la série et la moyenne de celle-ci

- Plus l'écart type s d'une série 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 est grand, plus les valeurs de la série sont dispersées autour de la moyenne, donc plus la série est hétérogène

Exemple

On considère deux entreprises de 10 employés :

- L'entreprise 1 dans laquelle 5 employés gagnent 2500€ et 5 employés gagnent 3000€ par mois
- L'entreprise 2 dans laquelle 9 employés gagnent 1200€ et 1 employés gagne 19200€ par mois

Après calcul (ou utilisation de la calculatrice) il s'avère que le salaire moyen dans les deux entreprises est 3000€ par mois. Pourtant $\sigma_1 = 500€$ alors que $\sigma_2 = 5400€$. Cela confirme que la série 2 est beaucoup plus hétérogène que la série 1.

Médiane	
Soit $x_1, x_2 \dots x_N$ une série statistique quantitative de N valeurs	
Définition	On appelle Médiane de la série notée Me, une valeur réelle telle que : • 50% au moins des valeurs de la série soient inférieures ou égales à Me • 50% au moins des valeurs de la série soient supérieures ou égales à Me
Propriété	Détermination de la médiane : On range la série de N valeurs par ordre croissant des valeurs. • Si l'effectif total est impair, la médiane est la valeur de la série de rang $\frac{N+1}{2}$ • Si l'effectif total est pair, la médiane est la moyenne des valeurs de rang $\frac{N}{2}$ et $\frac{N+1}{2}$
Exemple	On considère la série ordonnée de 9 valeurs : 1 ;3 ;7 ;8 ;10 ;11 ;12 ;12 ;58. Effectif = 20. La médiane est la moyenne des 10e et 11e valeurs, c'est-à-dire : $\frac{5+6}{2} = \frac{11}{2} = 5{,}5$

Quartiles		
Soit $x_1, x_2 \dots x_N$ une série statistique quantitative de N valeurs		
Définitions	On appelle premier quartile de la série notée Q1, la plus petite valeur de la série telle que : • 25% au moins des valeurs de la série soient inférieures ou égales à Q1 • 75% au moins des valeurs de la série soient supérieures ou égales à Q1	
	On appelle troisième quartile de la série notée Q3, la plus petite valeur de la série telle que : • 75% au moins des valeurs de la série soient inférieures ou égales à Q3 • 25% au moins des valeurs de la série soient supérieures ou égales à Q3	
Propriétés	Pour une série ordonnée d'effectif n, Q1 (resp. Q3) est la kième valeur où k est le plus petit entier supérieur ou égal à $\frac{n}{4}$ (resp. $\frac{3n}{4}$)	
Exemple	On reprend la série du nombre de places disponibles dans les trains (ci-contre) Effectif = 20 $\frac{n}{4} = 5$ donc Q1 est la 5ième valeur. Q1=0. $\frac{3n}{4} = 15$ donc Q3 est la 15ième valeur. Q1=6.	(voir tableau ci-dessous)

Valeur	0	1	2	5	6	7	10
Effectif	5	1	3	1	5	4	1

Caractéristiques de dispersion
• L'étendue (différence entre la valeur maximale et minimale de la série) est une caractéristique de dispersion. • L'écart inter quartile (Q3-Q1) est aussi une caractéristique de dispersion. • Moyenne et étendue sont sensibles aux valeurs extrêmes, alors que médiane et écart interquartiles ne le sont pas.