2021/2022

STATISTIQUES DESCRIPTIVES

I Rappels : Étude d'une série statistique calcul des caractéristiques de position et de dispersion

I.1 La moyenne

Exemple

Soit la série statistique suivante : $1; 7; 15; 18; 98; 7; 10$

La moyenne s'obtient en additionnant l'ensemble des valeurs précédentes et en divisant le résultat par **l'effectif total** de la série.

$M = \frac{1+7+15+18+98+7+10}{7} = \frac{156}{7} \approx 22{,}3$

La moyenne est une caractéristique de position.

I.2 La médiane

Définition

La médiane m est une valeur telle que la moitié au moins de l'effectif ait des valeurs inférieures ou égales à m, l'autre moitié des valeurs supérieures ou égales à m.
La médiane est une caractéristique de position.

Exemples

Soient les deux séries de notes suivantes :
Paul : 12 ; 15 ; 20 ; 8 ; 10
Mélanie : 12 ; 14 ; 19 ; 5 ; 17 ; 13
Il faut commencer par ranger la série dans l'ordre croissant
Paul : 8 ; 10 ; 12 ; 15 ; 20 La troisième valeur coupe l'effectif en deux, donc $m_{Paul} = 12$

Mélanie : 5 ; 12 ; 13 ; 14 ; 17 ; 19

L'effectif est paire, il n'y a donc pas de valeur appartenant à la série qui coupe celle-ci en deux.
La médiane s'obtient donc en faisant la moyenne des deux valeurs "du milieu"
$m_{Melanie} = \frac{13+14}{2} = 13{,}5$

I.3 L'étendue

Définition

L'étendue d'une série statistique est la différence entre la plus grande valeur et la plus petite valeur de la série
L'étendue est une caractéristique de dispersion.

Exemples

L'étendue des notes de Paul est $E_{Paul} = 20 - 8 = 12$
L'étendue des notes de Mélanie est $E_{Melanie} = 19 - 5 = 14$

I.4 Les quartiles et l'écart interquartile

Définition

Premier quartile
Le premier quartile est la plus petite valeur de la série telle qu'au moins 25% des autres valeurs de la série sont inférieures ou égales à cette valeur.

Troisième quartile
Le troisième quartile est la plus petite valeur de la série telle qu'au moins 75 % des autres valeurs de la série sont inférieures ou égales à cette valeur.

Écart interquartile
L'écart interquartile d'une série statistique de premier quartile $Q1$ et de troisième quartile $Q3$ est égal à la différence $Q3 - Q1$.

Remarque

L'écart interquartile d'une série mesure la dispersion autour de la médiane. Il contient au moins 50% des valeurs de la série.
L'écart interquartile n'est pas influencé par les valeurs extrêmes de la série.
Les quartiles sont des caractéristiques de position.
L'écart interquartile est une caractéristique de dispersion.

Exemples

Pour déterminer les quartiles, il faut **ordonner** les séries.
Le premier quartile est la valeur de la série se trouvant **au quart** de l'effectif.
Le troisième quartile est la valeur de la série se trouvant au **trois-quarts** de l'effectif.

La série de notes de Paul contient 5 valeurs :
$\frac{1}{4} \times 5 = 1{,}25$ Pour avoir au moins un quart des valeurs inférieures à $Q1$ il faut prendre la deuxième valeur.
$\frac{3}{4} \times 5 = 3{,}75$ Pour avoir au moins les trois quart des valeurs inférieures à $Q3$ il faut prendre la quatrième valeur.
Paul : 8 ; 10 ; 12 ; 15 ; 20 $\quad Q1 = 10$ et $Q3 = 15$

Pour Mélanie : $\frac{1}{4} \times 6 = 1{,}5$ $\frac{3}{4} \times 6 = 4{,}5$
Mélanie : 5 ; 12 ; 13 ; 14 ; 17 ; 19 $\quad Q1 = 12$ et $Q3 = 17$

I.5 Interprétations des données

Exemples

On peut commencer par calculer les moyennes de Mélanie et Paul.
$M_{Paul} = \frac{8+10+12+15+20}{5} = 13$
$M_{Melanie} = \frac{5+12+13+14+17+19}{6} \approx 13{,}3$

Les moyennes sont environ égales et pourtant les notes ne se répartissent pas de la même manière autour de cette caractéristique de position. Les étendues sont très différentes (12 et 14).

Dire que Paul à une médiane égale à 12 signifie que Paul a obtenu autant de notes au-dessus de 12 que de notes en-dessous de 12.

Dire que le premier quartile de Mélanie est égal à 10 signifie qu'au moins un quart des notes de Mélanie sont inférieures à 10. Dire que le troisième quartile de Paul est égal à 17 signifie qu'au moins trois quarts des notes de Paul sont inférieurs à 17.

L'écart interquartile de Mélanie est égal à $15 - 10 = 5$ signifie qu'au moins 50% des notes de Mélanie sont comprises entre 10 et 15 (les quartiles).

II Quand les données sont pondérées

Exemple

On a sondé plusieurs cyclistes afin de savoir la distance qu'il parcourait sur une semaine. Les distances sont données en km :

`150 ; 250 ; 90 ; 120 ; 400 ; 350 ; 150 ; 150 ; 200 ; 120 ; 200 ; 350 ; 150 ; 200 ; 175 ; 150 ; 175 ; 50 ; 100 ; 200 ; 100.`

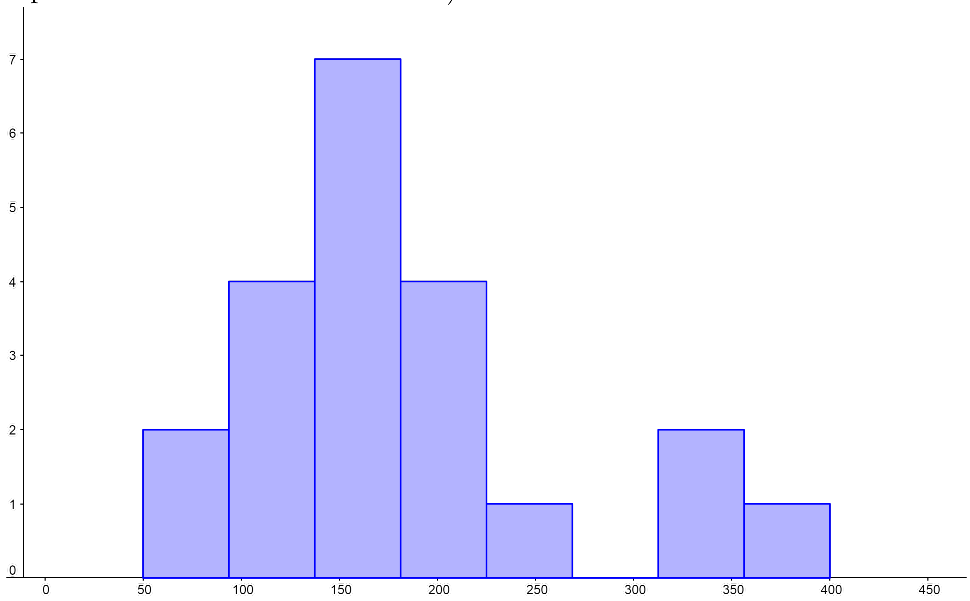
Pour interpréter ces résultats, il faut commencer par les ranger dans un tableau.

Distance (km)	50	90	100	120	150	175	200	250	350	400
Effectifs	1	1	2	2	5	2	4	1	2	1
Fréquences	$\frac{1}{21}$	$\frac{1}{21}$	$\frac{2}{21}$	$\frac{2}{21}$	$\frac{5}{21}$	$\frac{2}{21}$	$\frac{4}{21}$	$\frac{1}{21}$	$\frac{2}{21}$	$\frac{1}{21}$

L'effectif total est ici de 21.
A chaque distance on a attribué l'effectif que l'on a comptabilisé. Cette effectif est donc un **poids** pour la valeur (distance) qui lui correspond. Cette série statistique est donc maintenant **pondérée**.

Remarque

Il est intéressant de tracer l'histogramme représentatif de ces données statistiques. (Ici nous les représentons à l'aide de **`GeoGebra`**)



Dans un histogramme, l'aire des rectangles est proportionnelle à l'effectif (ou à la fréquence)

Définition

La moyenne d'une série statistique dont les valeurs sont $x_1, x_2, \ldots, x_k$ et les effectifs correspondants $n_1, n_2, \ldots, n_k$ est notée $\overline{x}$ et est égale à :

$$\overline{x} = \frac{n_1 x_1 + \ldots + n_k x_k}{n_1 + \ldots + n_k}$$

.

Exemple

Pour la série précédente :

$$\overline{x} = \frac{50 \times 1 + 90 \times 1 + 100 \times 2 + 120 \times 2 + 150 \times 5 + 175 \times 2 + 200 \times 4 + 250 \times 1 + 350 \times 2 + 400 \times 1+}{1+1+2+2+5+2+4+1+2+1}$$

$\approx 182{,}4$ km

Propriété

Si une série de valeurs x_i a pour moyenne $\overline{x}$, alors la série de valeurs $ax_i + b$ avec a et b réels, a pour moyenne $a\overline{x} + b$

On dit que la moyenne est une caractéristique **linéaire**.

Définition

La variance

La variance V d'une série statistique de moyenne $\overline{x}$ dont les valeurs du caractère sont $x_1, x_2, x_3, \ldots, x_k$ et les effectifs correspondants sont $n_1, n_2, n_3, \ldots, n_k$ est égale à :

$$V = \frac{n_1 \times (x_1 - \overline{x})^2 + n_2 \times (x_2 - \overline{x})^2 + \ldots + n_k \times (x_k - \overline{x})^2}{n_1 + n_2 + \ldots + n_k}$$

L'écart-type

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

Exemple

Pour la série précédente :

$V = \frac{1\times(50-182{,}4)^2+...+1\times(400-182{,}4)^2}{21} \approx 7720{,}3$
$\sigma = \sqrt{7720{,}3} \approx 87{,}9$ km

Remarque

L'écart-type possède la même unité que les valeurs de la série.
L'écart-type exprime la dispersion des valeurs d'une série statistique autour de sa moyenne. Les valeurs extrêmes influencent l'écart-type.