

STATISTIQUES DESCRIPTIVES

I. Caractéristiques d'une série statistique

1) Série statistique

Voici les séries de notes obtenues par 3 élèves :

Jérôme : 4 ; 6 ; 18 ; 7 ; 17 ; 12 ; 12 ; 18

Bertrand : 13 ; 13 ; 12 ; 10 ; 12 ; 3 ; 14 ; 12 ; 14 ; 15

Julie : 15 ; 9 ; 14 ; 13 ; 10 ; 12 ; 12 ; 11 ; 10

2) Moyenne

Méthode 1 : Calculer une moyenne

Calculer la moyenne pour chaque série de notes de Jérôme, de Bertrand et de Julie.

La moyenne est une caractéristique de position.

3) 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.

Méthode 2 : Calculer une médiane

Calculer la médiane pour chaque série de notes de Jérôme, de Bertrand et de Julie.

Résumé : Pour déterminer les notes médianes, il faut ordonner les séries. La médiane partage l'effectif en deux.

4) Étendue

Définition : L'**étendue** d'une série statistique est la différence entre la plus grande valeur et la plus petite valeur de la série.

Méthode 3 : Calculer une étendue

Calculer l'étendue pour chaque série de notes de Jérôme, de Bertrand et de Julie.

(cas particulier de la série de Bertrand : élagage possible)

L'étendue est une caractéristique de dispersion.

5) Quartiles, écart interquartile

Définitions :

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.

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.

Définition : L'**écart interquartile** d'une série statistique de premier quartile Q_1 et de troisième quartile Q_3 est égal à la différence $Q_3 - Q_1$.

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.

Méthode 4 : Calculer les quartiles

Calculer les quartiles pour chaque série de notes de Jérôme, de Bertrand et de Julie.

Résumé :

Pour déterminer les quartiles, il faut ordonner les séries.

Le premier quartile est la donnée de la série se trouvant au quart de l'effectif.

Le troisième quartile est la donnée de la série se trouvant au trois-quarts de l'effectif.

Les quartiles sont des caractéristiques de position.

L'écart interquartile est une caractéristique de dispersion.

6) Interprétations

	Moyenne	Médiane	Etendue	Q_1	Q_3	Écart interquartiles
Jerôme						
Bertrand						
Julie						

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.

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

Dire que le premier quartile de Bertrand est égal à 12 signifie qu'au moins un quart des notes de Bertrand sont inférieures à 12.

Dire que le troisième quartile de Julie est égal à 13 signifie qu'au moins trois quarts des notes de Julie sont inférieurs à 13.

L'écart interquartile de Jérôme est égal à 11 signifie qu'au moins 50% des notes de Jérôme sont comprises entre 6 et 17 (les quartiles).

II. Cas de pondération d'une série statistique

1) Série statistique

Tailles des élèves de 2^{nde}5 en cm :

174 – 160 – 161 – 166 – 177 – 172 – 157 – 175 – 162 – 169 – 160 – 165 – 170 – 152 – 168 – 156 – 163 – 167 – 169 – 158 – 164 – 151 – 162 – 166 – 156 – 165 – 179

2) Regroupement par classe

Regrouper cette série de tailles par classes de longueur 5 cm et calculer les fréquences arrondies au centième :

Tailles	$150 \leq t < 155$	$155 \leq t < 160$	$160 \leq t < 165$	$165 \leq t < 170$	$170 \leq t < 175$	$175 \leq t < 180$
Effectifs						
Fréquences						

L'effectif total est 27.

Méthode 5 : comment réaliser un Histogrammes

Faire un histogramme en suivant la répartition proposée dans le tableau

Remarque :

Dans un histogramme, l'aire des rectangles est proportionnelle à l'effectif (ou à la fréquence). Pour illustrer ça, on va dessiner à côté de l'histogramme précédent un autre représentant les mêmes données mais en fusionnant les deux premières classes.

3) Moyenne pondérée

Classes centrées x_i						
Effectifs n_i						

Il s'agit d'un calcul de moyenne pondéré car des effectifs différents n_i sont associés à chaque valeur x_i .

Définition :

La **moyenne** d'une série statistique dont les valeurs sont x_1, x_2, \dots, x_k et les effectifs correspondants n_1, n_2, \dots, n_k est notée \bar{x} et est égale à $\bar{x} = \frac{n_1x_1 + \dots + n_kx_k}{n_1 + \dots + n_k}$..

Ainsi dans l'exemple :

$$\bar{x} = (2 \times 152 + 4 \times 157 + 7 \times 162 + 8 \times 167 + 3 \times 172 + 3 \times 177) : 27 = 4449 : 27$$

$$\approx 164,8 \text{ cm}$$

Remarque : Calcul de la moyenne exacte

$$(174 + 160 + 161 + 166 + 177 + 172 + 157 + 175 + 162 + 169 + 160 + 165 + 170 + 152 + 168 + 156 + 163 + 167 + 169 + 158 + 164 + 151 + 162 + 166 + 156 + 165 + 179) : 27$$

$$= 4444 : 27 \approx 164,6 \text{ cm}$$

La méthode de calcul de moyenne en centrant les classes est très fiable (ici : 2 mm d'erreur)

4) Linéarité de la moyenne

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

Exemple :

	Moyennes			
x_i	4	7	-2	$(4 + 7 - 2) : 3 = 3$
$2x_i - 5$	3	9	-9	$(3 + 9 - 9) : 3 = 1$

On a $\bar{x} = 3$, et donc en appliquant la propriété, la moyenne de la série $2x_i - 5$ est égale à : $2\bar{x} - 5 = 2 \times 3 - 5 = 1$. On retrouve bien le résultat calculé directement dans le tableau.

5) Variance, écart-type

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

Ainsi en reprenant l'exemple des tailles de la 2nde5, la variance est égale à :

$$V = \frac{2 \times (152 - 164,8)^2 + 4 \times (157 - 164,8)^2 + 7 \times (162 - 164,8)^2 + 8 \times (167 - 164,8)^2 + 3 \times (172 - 164,8)^2 + 3 \times (177 - 164,8)^2}{27}$$

$$\approx 46,914$$

$$\sigma \approx \sqrt{46,914} \approx 6,85$$

L'écart-type possède la même unité que les valeurs de la série.

Ainsi pour la série étudiée, l'écart-type est environ égal à 6,85 cm.

Remarque :

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