

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

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.

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.

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

$$M_{(\text{Jérôme})} = 11,8$$

$$m_{(\text{Jérôme})} = 12$$

$$E_{(\text{Jérôme})} = 14$$

$$Q_1_{(\text{Jérôme})} = 6$$

$$Q_3_{(\text{Jérôme})} = 17$$

$$E_Q_{(\text{Jérôme})} = 11$$

$$M_{(\text{Bertrand})} = 11,8$$

$$m_{(\text{Bertrand})} = 12,5$$

$$E_{(\text{Bertrand})} = 5$$

$$Q_1_{(\text{Bertrand})} = 12$$

$$Q_3_{(\text{Bertrand})} = 14$$

$$E_Q_{(\text{Bertrand})} = 2$$

$$M_{(\text{Julie})} \approx 11,8$$

$$m_{(\text{Julie})} = 12$$

$$E_{(\text{Julie})} = 6$$

$$Q_1_{(\text{Julie})} = 10$$

$$Q_3_{(\text{Julie})} = 13$$

$$E_Q_{(\text{Julie})} = 3$$

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.

Dire que Jérôme à 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 2^{nde}5, 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.

Bonus : petit point sur la médiane et les quartiles

Médiane

Définition

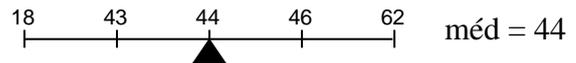
Soit une série statistique d'effectif total n , rangée par ordre croissant.

On appelle médiane la valeur "du milieu". On dit qu'elle partage la série en deux moitiés : il y a autant de valeurs en dessous qu'au dessus.

Pour déterminer son rang, il y a 2 cas :

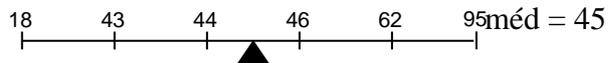
- si n est impair : la médiane est la valeur de rang

$$\frac{n+1}{2}$$



- si n est pair : nous prendrons la demi-somme des deux valeurs dont les rangs entourent le nombre

$$\frac{n+1}{2}$$



Remarque :

Si les données ont été regroupées en classes, on ne peut déterminer la valeur exacte de la médiane. En revanche, on appellera classe médiane, la classe qui la contient (et permet donc d'en donner un encadrement).

Exemples

Données discrètes "en vrac"

10, 7, 12, 18, 16, 15, 5, 11, 11, 20, 15, 11, 18, 14

Ordonnons la série par ordre croissant : 5, 7, 10, 11, 11, 11, 12, 14, 15, 15, 16, 18, 18, 20

Il y a 14 termes or $\frac{14+1}{2} = 7,5$.

La médiane est donc la demi somme des 7^{ème} et 8^{ème} termes : méd = $\frac{12 + 14}{2} = 13$

Avec un tableau d'effectifs

valeurs	1	2	3	4	5	6
effectifs	6	11	25	19	15	5
effectifs cumulés	6	17	42	61	76	81

Attention, il faut bien interpréter cette dernière ligne : Les données qui valent 3 ont un rang compris entre 18 et 42 inclus

L'effectif total est de 81 or $\frac{81+1}{2} = 41$.

La médiane est donc le 41^{ème} terme : méd = 3

Avec des données réparties par classes

classe	[0 ; 2[[2 ; 4[[4 ; 6[[6 ; 8]
fréquence	10%	38%	45%	7%
fréquence cumulée	10%	48%	93%	100%

48% des valeurs sont strictement inférieures à 4

Et 93% des valeurs sont strictement inférieures à 6

La classe médiane est donc la classe [4 ; 6[

On peut donc en déduire l'encadrement suivant $4 < \text{méd} < 6$

III. Quartiles

Définitions :

Le premier quartile Q_1 est la plus petite des valeurs de la série telle qu'au moins 25% de la population ait sa valeur inférieure ou égale à Q_1 .

Le troisième quartile Q_3 est la plus petite des valeurs de la série telle qu'au moins 75% de la population ait sa valeur inférieure ou égale à Q_3 .

Méthode pour trouver Q_1 et Q_3 .

Pour une population d'effectif n , $\frac{n}{4}$ (ou $\frac{3n}{4}$) si il est un entiers nous donne le rang de Q_1 (ou Q_3), dans le cas contraire on prendra la valeur de l'élément du rang immédiatement supérieur.

Exemple

Avec un tableau d'effectifs

valeurs	1	2	3	4	5	6
effectifs	6	11	25	19	15	5
effectifs cumulés	6	17	42	61	76	81

Ici $n = 81$

Donc $\frac{n}{4} = 20,25$ il va donc falloir que je prenne la valeur de l'individu de rang 21, donc $Q_1 = 3$

$\frac{3n}{4} = 60,75$ L'individu de rang 61 a pour rang 5 donc $Q_3 = 4$

Exercice 1

Une documentaliste a relevé, sur un trimestre, les emprunts effectués par les élèves de Seconde.

Ce tableau présente les résultats obtenus.

Nombre d'emprunts	0	1	2	3	4	5	6	7	8
Effectif	39	30	36	23	20	22	18	10	11

- Donner le tableau des effectifs cumulés croissants.
- Combien d'élèves ont emprunté trois livres au plus ?
- Calculer la fréquence cumulée croissante au millième près de la valeur 4 et l'interpréter.
- Représenter la série des effectifs par un diagramme en bâtons.

Exercice 2

Dans une classe de Seconde, les notes au dernier devoir de mathématiques sont représentées avec un histogramme des effectifs.



- A l'aide de l'effectif de la classe $[10;12[$, expliquer la légende du graphique qui indique une zone représentant une fréquence de 5 %.
- Le professeur annonce qu'environ 95 % des élèves ont obtenu moins de 14. La valeur exacte est-elle supérieure ou inférieure à 95 % ?
- Estimer la moyenne de ces notes, en supposant que la distribution est régulière dans chaque classe.
- On admet que la médiane de cette série est 8. Nina affirme : « Au moins 50 % des élèves ont une note inférieure ou égale à 8 ». Qu'en pensez-vous ?
- Dans quelles classes se trouvent les premier et troisième quartiles de la série ?

Exercice 3

Lors d'une séance en groupe, des élèves ont effectué un test comportant 100 questions.

Voici le nombre de réponses exactes pour chacun des élèves :

18; 36; 44; 46; 54; 54; 55; 56; 62; 70; 73; 74; 74.

- Déterminer la moyenne (à l'unité près) et la médiane de cette série.
- Déterminer les quartiles et les caractéristiques de dispersion de cette série.

Exercice 4

L'anaconda géant est le plus grand serpent d'Amérique du Sud.

Dans le tableau ci-dessous, on a relevé la taille de 100 femelles adultes.

Taille (en m)	4	5	6	7	8	9
Effectif	7	22	14	20	19	18

Déterminer la moyenne, la médiane, les quartiles et les caractéristiques de dispersion de cette série.

Exercice 5

Voici les tailles de 96 requins blancs.

Taille (en m)	1,5	2	2,5	3	3,5	4	4,5
Effectif	7	10	24	32	18	4	1

Reprendre les questions de l'exercice 7 pour cette série.

Exercice 6

On considère l'algorithme ci-contre.

1. Appliquer cet algorithme, avec chacune des listes ci-dessous.

a) 2; 3; 5; 25; 32; 48

b) 5; 8; 16; 27; 54; 69; 84

2. Quel est le rôle de cet algorithme ?

Variables :	n est un nombre entier naturel $(x_1; x_2; \dots; x_n)$ est une liste ordonnée de nombres réels
Entrées :	Saisir n, x_1, x_2, \dots, x_n
Traitement et sortie :	Si n est divisible par 2 alors Affecter à i la valeur $n/2$ Afficher $(x_i + x_{i+1})/2$ sinon Affecter à i la valeur $(n-1)/2$ Afficher x_{i+1} Fin Si

Exercice 7

Dans une rue passante de Salvador de Bahia (Brésil), on a mesuré le niveau de bruit en décibels (db) émis par 20 véhicules pris au hasard.

54,8; 55,0; 57,7; 59,6; 60,1

61,2; 62,0; 63,1; 63,5; 64,2

65,2; 65,4; 65,9; 66,0; 67,6

68,1; 69,5; 70,6; 71,5; 73,4

a) Regrouper les données par classes d'amplitude 5 db, à partir de 50 db.

b) Représenter la série par un histogramme.

c) Construire la courbe des fréquences cumulées croissantes.

d) Quelle est la classe de plus grande fréquence ? Interpréter dans le contexte de l'exercice.

Exercice 8

Un distributeur automatique de café propose des expressos.

Une pesée portant sur 30 expressos a donné les masses suivantes (en grammes) de café utilisé.

81	82	85	83	83	82	87	84	85	84
84	81	83	86	84	80	80	79	87	85
81	82	85	87	79	80	86	89	83	89

a) Reproduire et compléter le tableau par des valeurs approchées au centième près des fréquences et par les fréquences cumulées croissantes (FCC).

Masse (en g)	[79; 82[[82; 85[[85; 88[[88; 91[
Fréquence				
FCC				

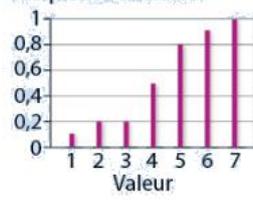
b) Représenter la courbe des fréquences cumulées.

c) Recopier et compléter par lecture graphique : « 75 % des expressos contiennent moins de ... g de café ».

Exercice 9

Expliquer pourquoi ce graphique ne peut pas correspondre à l'écran de la calculatrice.

Fréquence cumulée



```
Stats 1-Var
↑n=17
minX=1
Q1=3
Med=5
Q3=6
maxX=7
```

Utiliser Excel pour faire des statistiques :

Pour faire une légende variable :

les parties fixes s'écrivent entre guillemets, la concaténation se fait avec l'esperluète (le symbole &).

Pour faire dénombrer les cellules vérifiant un ou plusieurs critères

Il faudra taper =nb.si.ens(cellules à explorer ; condition 1 à écrire entre guillemet ; cellules à explorer ; condition 2 ;)

Pour indiquer les cellules à explorer on peut taper l'information à la main de la manière suivante : référence cellule supérieure gauche de la zone à couvrir : référence cellule inférieure droite de la zone à couvrir , par exemple :

B8 :D27 si on veut indiquer toutes les valeurs dans la zone allant de la ligne 8 à la 27 et allant de la colonne B à la D. Comme condition on indiquera par exemple « >=8 » si on a envie de sélectionner uniquement des cellules contenant des valeurs supérieures à 8.

Par exemple

Pour créer des valeurs aléatoires entre deux bornes :

=valeur minimale + ENT(ALEA()*(valeur maximale – valeur minimale))+1

=ALEA.ENTRE.BORNES(valeur minimale;valeur maximale)

Recherche de la médiane

Si on a un paquet de valeurs non classées il faudra utiliser la fonction mediane(comme suit :

« =MEDIANE(valeurs dont on cherche la médiane) »

Sinon il va falloir bricoler en utilisant les effectifs cumulés croissants

fonctions de base :

pour arrondir à l'entier supérieur (pratique pour les quartiles) : arrondi.sup(valeur, 0)

pour faire la somme de plusieurs valeurs : somme(valeurs)

Ressources :

<https://www.cours-gratuit.com/tutoriel-excel/tutoriel-excel-comment-compter-le-nombre-de-cellule-comprise-entre-deux-valeurs>

pour faire un polygone d'effectif ou de fréquence cumulées croissantes :

<https://www.youtube.com/watch?v=VaRUSQmwN8I>