STATISTIQUES

En italien, « stato » désigne l'état. Ce mot à donné « statista » pour « homme d'état ».
En 1670, le mot est devenu en latin « statisticus » pour signifier ce qui est relatif à l'état.
Les statistiques ont en effet d'abord désigné l'étude des faits sociaux relatifs à l'état.

I. Caractéristique de position d'une série statistique

1) Séries statistiques

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) Moyennes

Méthode : Calculer une moyenne

Vidéo **https://youtu.be/88_16UbkdZM**

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

La moyenne est une caractéristique de position.

3) Médianes

Définition :
La médiane m est une valeur de la série telle que la moitié de l'effectif ait des valeurs inférieures à m, l'autre moitié des valeurs supérieures à m.

Méthode : Calculer une médiane

Vidéo **https://youtu.be/kr90dXv0NFY**

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

La médiane est une caractéristique de position.

II. Caractéristiques de dispersion d'une série statistique

1) Etendue

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

Méthode : Calculer une étendue

Vidéo **https://youtu.be/PPXGOs2b4Ls**

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

L'étendue est une caractéristique de dispersion.

2) Quartiles

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.

Méthode : Calculer les quartiles

Vidéo https://youtu.be/Yjh-9nMVmEw
Vidéo https://youtu.be/2jbpNjXMdSA

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

Les quartiles sont des caractéristiques de dispersion.

3) Interprétations

$M_{(Jérôme)} = 11{,}8$ $m_{(Jérôme)} = 12$ $E_{(Jérôme)} = 14$ $Q_{1(Jérôme)} = 6$
$Q_{3(Jérôme)} = 17$

$M_{(Bertrand)} = 11{,}8$ $m_{(Bertrand)} = 12{,}5$ $E_{(Bertrand)} = 5$ $Q_{1(Bertrand)} = 12$
$Q_{3(Bertrand)} = 14$

$M_{(Julie)} \approx 11{,}8$ $m_{(Julie)} = 12$ $E_{(Julie)} = 6$ $Q_{1(Julie)} = 10$
$Q_{3(Julie)} = 13$

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.

III. Effectifs cumulés et fréquences cumulées

Vidéo https://youtu.be/zJ625zpPTds

1) Série statistique

Tailles des élèves de 2nde5 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

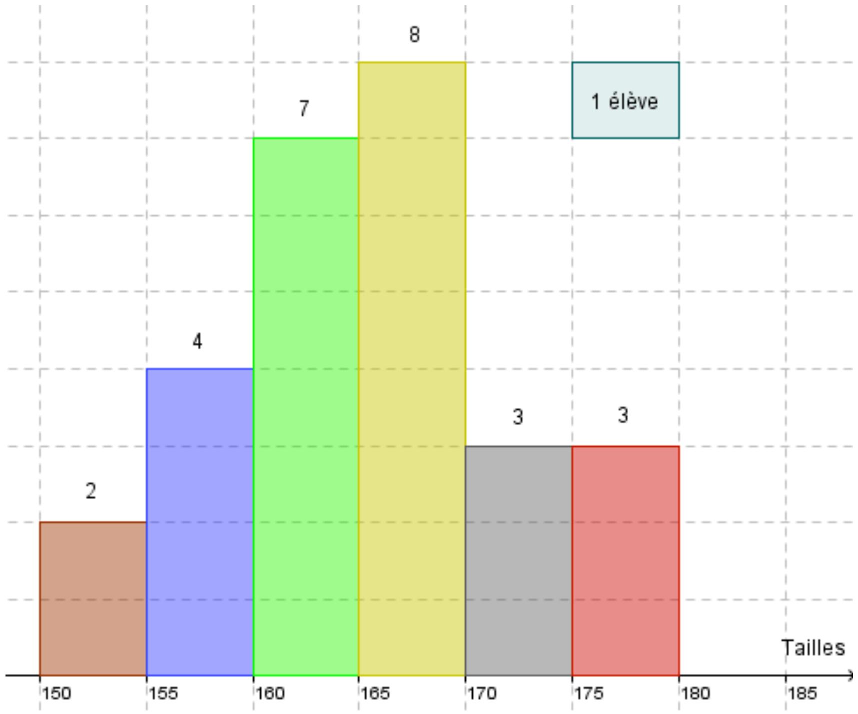
2) Regroupement par classe

Regrouper cette série de tailles par classes de longueur 5 cm et calculer les fréquences en % arrondies à l'unité :

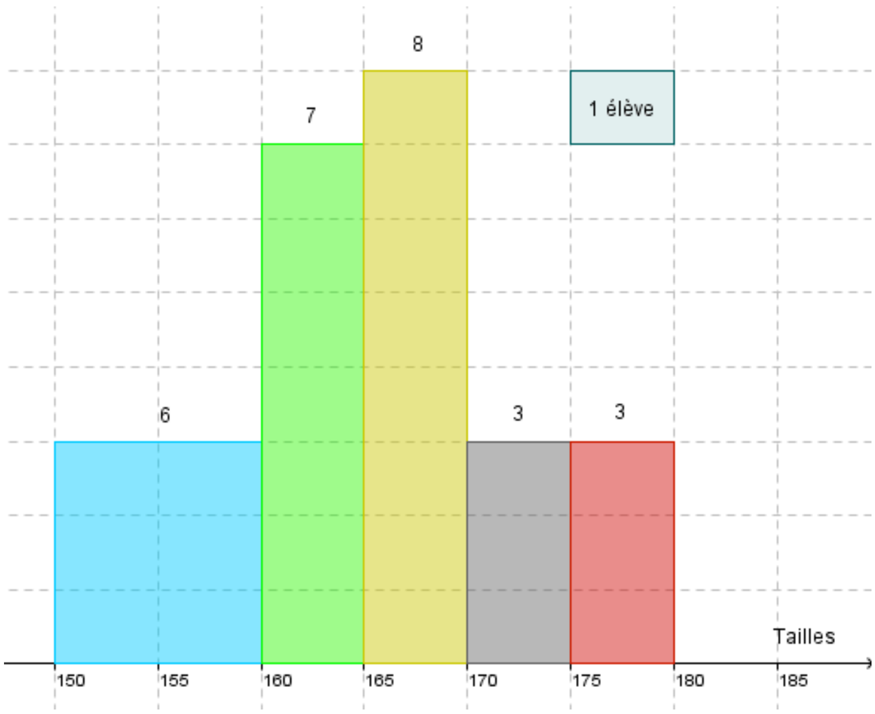
Tailles	$150 \leq t < 155$	$155 \leq t < 160$	$160 \leq t < 165$	$165 \leq t < 170$	$170 \leq t < 175$	$175 \leq t < 180$
Effectifs	2	4	7	8	3	3
Fréquences	$\frac{2}{27} \times 100$ $= 7$	15	26	30	11	11

L'effectif total est 27.

HISTOGRAMME DES EFFECTIFS DES TAILLES



Remarque :
Dans un histogramme, l'aire des rectangles est proportionnelle à l'effectif (ou à la fréquence).
Ainsi, dans l'exemple, en regroupant les deux premières classes, on obtiendrait la représentation suivante :



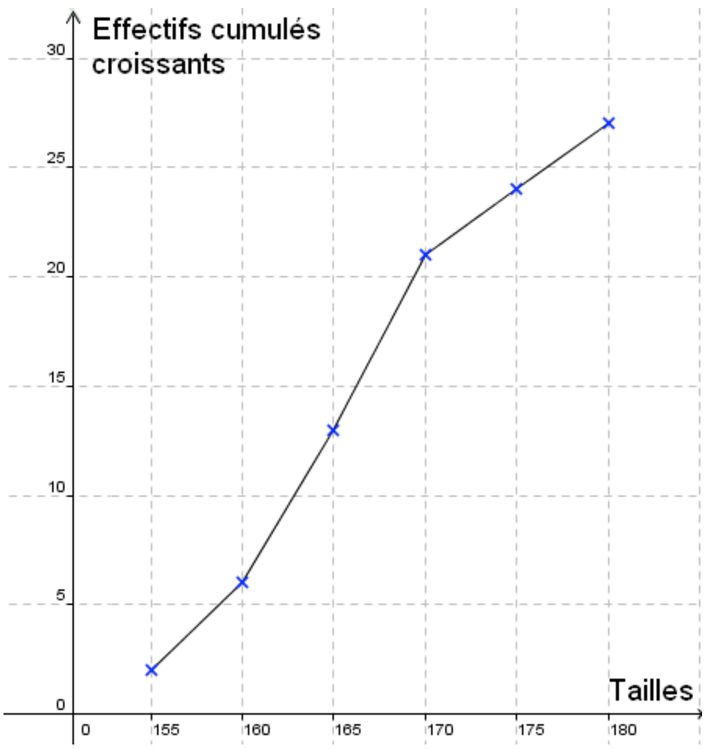
3) Cumuls:

Tailles	t <155	t <160	t <165	t <170	t <175	t <180
Effectifs cumulés croissants	*2*	*6*	*13*	*21*	*24*	*27*
Fréquences cumulées croissantes en %	*7*	*22*	*48*	*78*	*89*	*100*

Quel est le pourcentage d'élèves dont la taille est inférieure à 170 cm ? *78%*
Quel est le pourcentage d'élèves dont la taille est comprise entre 160 cm et 170 cm?
78 – 22 = 56%
Combien y a-t-il d'élèves dont la taille est supérieure à 165 cm ? *27 - 13 = 14*

Polygone des effectifs cumulés croissants :

Vidéo https://youtu.be/DVN-4u6BCPY



4) Moyenne :

Vidéo https://youtu.be/88_16UbkdZM

a) Calcul de la moyenne en centrant les classes :

Classes centrées	*152*	*157*	*162*	*167*	*172*	*177*
Effectifs	*2*	*4*	*7*	*8*	*3*	*3*

Il s'agit d'un calcul de moyenne pondéré.

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

Ainsi dans l'exemple :
(152 x 2 + 157 x 4 + 162 x 7 + 167 x 8 + 172 x 3 + 177 x 3) : 27 = 4449 : 27 ≈ 164,8 cm

b) 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 ≈ 164,6 cm
La méthode de calcul de moyenne en centrant les classes est très fiable *(ici : 2 mm d'erreur)*