Mme GUESMIA AZIZA Jan. 2013

Statistiques partie n°1

Lycée Nafta 2ème Economie et Services

Caractéristique de position d'une série statistique

1) Séries statistiques

Voici les séries de notes obtenues par 3 élèves :

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

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

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

2) Moyennes

$M\ (Jihad) = (4 + 6 + 18 + 7 + 17 + 12 + 12 + 18) : 8 \approx 11{,}8$

$M\ (Bilal) = (13 + 13 + 12 + 10 + 12 + 3 + 14 + 12 + 14 + 15) : 10 = 11{,}8$

$M(Hamza) = (15 + 9 + 14 + 13 + 10 + 12 + 12 + 11 + 10) : 9 \approx 11{,}8$

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*.

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

Jihad : $\underbrace{4 \quad 6 \quad 7 \quad 12}_{\text{4 données}} \quad \underbrace{12 \quad 17 \quad 18 \quad 18}_{\text{4 données}}$

$m_{(Jihad)} = 12$

Bilal: $\underbrace{3 \quad 10 \quad 12 \quad 12 \quad 12}_{\text{5 données}} \quad \underbrace{13 \quad 13 \quad 14 \quad 14 \quad 15}_{\text{5 données}}$

$m_{(Bilal)} = (12 + 13) : 2 = 12{,}5$

Hamza: $\underbrace{9 \quad 10 \quad 10 \quad 11}_{\text{4 données}} \quad 12 \quad \underbrace{12 \quad 13 \quad 14 \quad 15}_{\text{4 données}}$

$m_{(Hamza)} = 12$

La médiane est une caractéristique de position.

I. 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.

En reprenant les données de l'exemple étudié au paragraphe I. :

$E_{(Jihad)} = 18 - 4 = 14$ $E_{(Bilal)} = 15 - 3 = 12$ $E_{(Hamza)} = 15 - 9 = 6$

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.

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.

Jihad : 4 6 7 12 12 17 18 18

$\frac{1}{4}$ x 8 = 2, le premier quartile est la 2e donnée de la série ordonnée.

$\frac{3}{4}$ x 8 = 6, le troisième quartile est la 6e donnée de la série ordonnée.

$Q_{1(Jihad)} = 6$ $Q_{3(Jihad)} = 17$

Bilal: 3 10 12 12 12 12 13 14 15 15

$\frac{1}{4}$ x 10 = 2.5, le premier quartile est la 3e donnée de la série ordonnée.

$\frac{3}{4}$ x 10 = 7.5, le troisième quartile est la 8e donnée de la série ordonnée.

$Q_{1(Bilal)} = 12$ $Q_{3(Bilal)} = 14$

Hamza: 9 10 10 11 12 12 13 14 15

$\frac{1}{4}$ x 9 = 2.25, le premier quartile est la 3e donnée de la série ordonnée.

$\frac{3}{4}$ x 9 = 6.75, le troisième quartile est la 7e donnée de la série ordonnée.

$Q_{1}(Hamza) = 10$ $Q_{3}(Hamza) = 13$

Les quartiles sont des caractéristiques de dispersion.

3) Interprétations

$M(Jihad) = 11,8$ $m(Jihad) = 12$ $E(Jihad) = 14$ $Q_{1}(Jihad) = 6$ $Q_{3}(Jihad) = 17$

$M(Bilal) = 11,8$ $m(Bilal) = 12,5$ $E(Bilal) = 5$ $Q_{1}(Bilal) = 12$ $Q_{3}(Bilal) = 14$

$M(Hamza) \approx 11,8$ $m(Hamza) = 12$ $E(Hamza) = 6$ $Q_{1}(Hamza) = 10$ $Q_{3}(Hamza) = 13$

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

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

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

II. Effectifs cumulés et fréquences cumulées

1) Série statistique

Tailles des élèves de 9ème de base 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

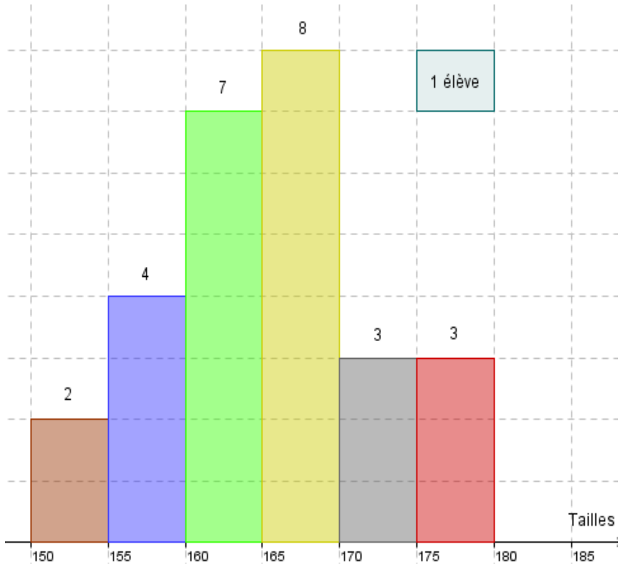
2) Regroupement par classe

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	$\frac{2}{27} x100 = 7$	15	26	30	11	11

L'effectif total est 27.

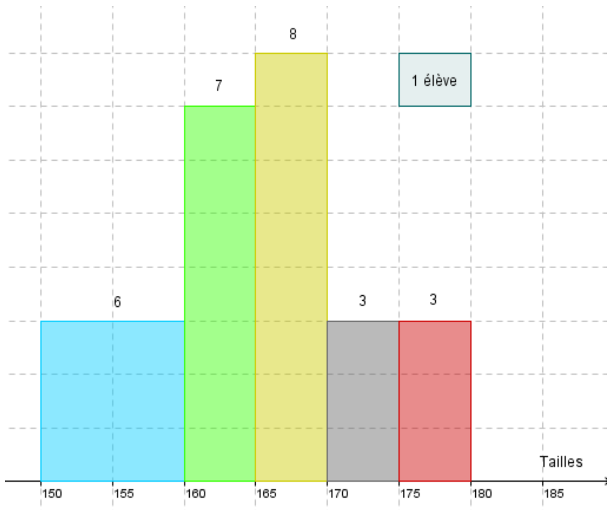
HISTOGRAMME DES EFFECTIFS DES TAILLES



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) Moyenne :

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, …, n_k est notée $\overline{x}$ et est égale à $\overline{x} = \dfrac{n_1x_1 + \ldots + n_kx_k}{n_1 + \ldots + n_k}$.

Ainsi dans l'exemple :

La moyenne est

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

$\bar{x}$ = 4449 : 27

$\bar{x} \approx$ 164,8 cm

b) Calcul de la moyenne exacte :

$\bar{x}$ = (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

$\bar{x}$ = 4444 : 27

$\bar{x} \approx$ 164,6 cm

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

4) 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 ?

Réponse : *78%*

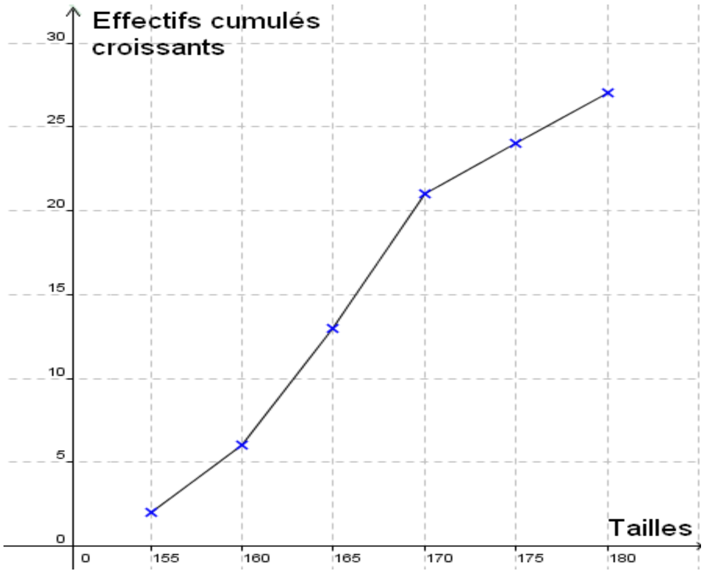
Quel est le pourcentage d'élèves dont la taille est comprise entre 160 cm et 170 cm?

Réponse *: 78 – 22 = 56%*

Combien y a-t-il d'élèves dont la taille est supérieure à 165 cm ?

Réponse : *27 - 13 = 14.*

Polygone des effectifs cumulés croissants :



III. Echantillonnage

1) Notion d'échantillon

Exemple :

Si, sur l'ensemble des cartes à puce produites par une entreprise en une semaine, on en prélève 200, on dit que cet ensemble de 200 cartes à puce constitue un **échantillon de taille 200** de la population de toutes les cartes à puce produites en une semaine.

Définition :

Un échantillon de taille n est constitué des résultats de n répétitions indépendantes de la même expérience sur l'ensemble des personnes ou objets sur lesquels porte l'étude statistique (la population).

Un échantillon issu d'une population est donc l'ensemble de quelques éléments de cette population.

2) Intervalle de fluctuation

On suppose que 22% des cartes à puce produites par l'entreprise sont défectueuses.

La **proportion effective** *p* est donc égale à 22%.

On prélève un échantillon de taille 200 parmi cette production et on compte le nombre de cartes à puce défectueuses parmi cette échantillon. Ce nombre est égal à 41.

Dans ce cas, la **fréquence observée** *f* est égale à $\frac{41}{200} = 0,205$.