STATISTIQUES À UNE VARIABLE

I. Tableau des effectifs

POPULATION étudiée : Les élèves de la classe de 5e …

CARACTÈRE étudié : Usages d'Internet pour faire des recherches.

VALEURS DU CARACTERE :

EFFECTIF TOTAL : Le nombre d'individus de la population étudiée = 27

Usages d'Internet	***Effectif***
Plusieurs fois par jour	**2**
Environ une fois par jour	**7**
2 à 5 fois par semaine	**8**
Environ une fois par semaine	**6**
Une à trois fois par mois	**3**
Moins souvent	**1**
TOTAL	*27*

II. Fréquences

Vidéo **https://youtu.be/MwNV5eCBFrI**

On souhaite comparer les résultats de la classe à ceux réalisés lors d'une enquête nationale sur 1253 jeunes âgés de 15 à 24 ans.
Pour cela, les tableaux des effectifs ne sont pas adaptés car les effectifs totaux sont différents.

Enquête nationale :

Usages d'Internet	***Effectif***
Plusieurs fois par jour	551
Environ une fois par jour	276
2 à 5 fois par semaine	288
Environ une fois par semaine	100
Une à trois fois par mois	25
Moins souvent	13
TOTAL	***1253***

La fréquence qui met en rapport un effectif particulier avec l'effectif total nous permettra de comparer plus facilement les deux enquêtes.

$$\boxed{\text{Fréquence} = \frac{EFFECTIF}{EFFECTIF\ TOTAL}}$$

$\frac{2}{27} \approx 0{,}07$

$0{,}07 = \frac{7}{100} = 7\ \%$

Classe de 5^e … :

Usages d'Internet	**Effectif**	**Fréquence**	**Fréquence en %**
Plusieurs fois par jour	2	**0,07**	**7**
Environ une fois par jour	7	**0,26**	**26**
2 à 5 fois par semaine	8	**0,30**	**30**
Environ une fois par semaine	6	**0,22**	**22**
Une à trois fois par mois	3	**0,11**	**11**
Moins souvent	1	**0,04**	**4**
TOTAL	***27***	***1***	***100***

$\frac{551}{1253} \approx 0{,}44$

$0{,}44 = \frac{44}{100} = 44\ \%$

Enquête nationale :

Usages d'Internet	**Effectif**	**Fréquence**	**Fréquence en %**
Plusieurs fois par jour	551	**0,44**	**44**
Environ une fois par jour	276	**0,22**	**22**
2 à 5 fois par semaine	288	**0,23**	**23**
Environ une fois par semaine	100	**0,08**	**8**
Une à trois fois par mois	25	**0,02**	**2**
Moins souvent	13	**0,01**	**1**
TOTAL	***1253***	***1***	***100***

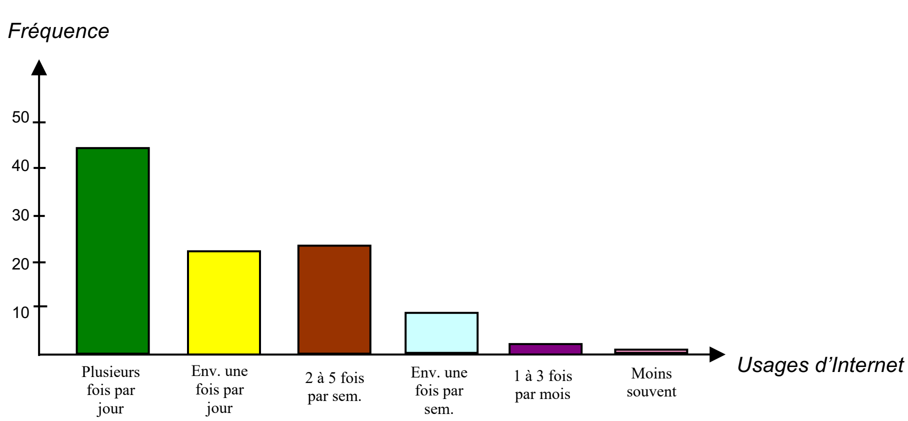
On peut maintenant comparer les deux populations.
On voit par exemple, que dans la classe, la proportion de jeunes utilisant Internet plusieurs fois par jour (7 %) est très faible par rapport au national (44 %).

III. Représentations graphiques

1) Diagramme en bâtons (ou à barres)

Vidéo **https://youtu.be/CR4lSAfho5A**
Vidéo **https://youtu.be/NZnhF5VDy04**

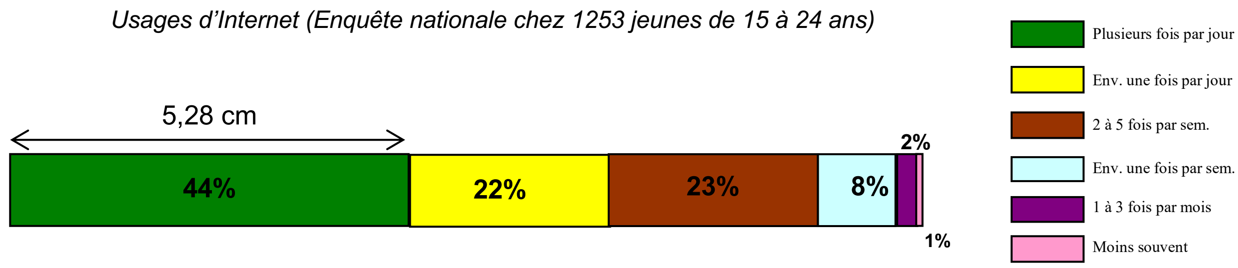
Usages d'Internet (Enquête nationale chez 1253 jeunes de 15 à 24 ans)



2) Diagramme à bandes

La totalité des fréquences est représentée par une bande rectangulaire de longueur 12 cm.
La valeur « Plusieurs fois par jour » est représentée par une bande (verte) de longueur
$\frac{44}{100}$ x 12 = 5,28 cm.
En effet, la valeur « Plusieurs fois par jour » correspond à 44 % du tout, soit 44 % de 12.
On fait de même pour calculer la longueur des autres bandes.

Usages d'Internet (Enquête nationale chez 1253 jeunes de 15 à 24 ans)



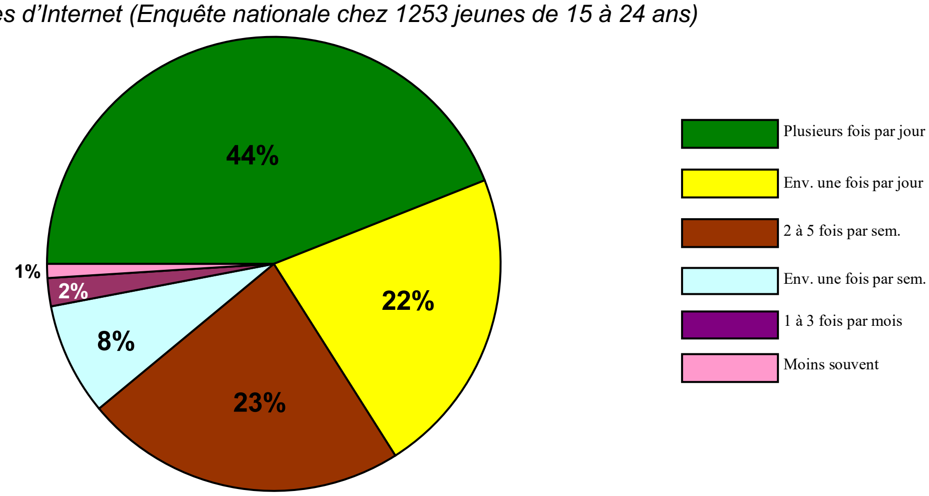
3) Diagramme circulaire ou « camembert »

Vidéo https://youtu.be/gpCY_3zq3bk

La totalité des fréquences est représentée par un disque (secteur de mesure 360°).
La valeur « Plusieurs fois par jour » est représentée par un secteur circulaire (vert) d'angle :
$\frac{44}{100}$ x 360 = 158,4°.
En effet, la valeur « Plusieurs fois par jour » correspond à 44 % du tout, soit 44 % de 360°.

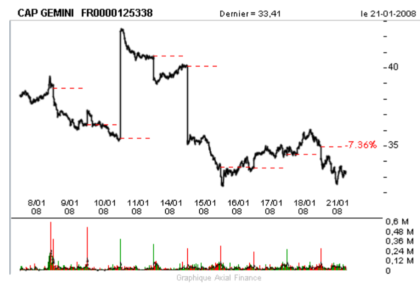
On fait de même pour calculer l'ouverture des autres secteurs.

Usages d'Internet (Enquête nationale chez 1253 jeunes de 15 à 24 ans)



IV. Moyenne, médiane

Voici les dernières notes obtenues par 3 élèves :



Victor : 4 ; 6 ; 18 ; 7 ; 17 ; 12 ; 12 ; 18
Nadir : 13 ; 13 ; 12 ; 10 ; 12 ; 3 ; 14 ; 12 ; 14 ; 15
Julie : 15 ; 9 ; 14 ; 13 ; 10 ; 12 ; 12 ; 11 ; 10

1) Moyenne

Vidéo https://youtu.be/a-RRUIS_CR8
Vidéo https://youtu.be/U1NamiLxBal

$M_{(Victor)}$ = (4 + 6 + 18 + 7 + 17 + 12 + 12 + 18) : 8 ≈ 11,8
$M_{(Nadir)}$ = (13 + 13 + 12 + 10 + 12 + 3 + 14 + 12 + 14 + 15) : 10 = 11,8
$M_{(Julie)}$ = (15 + 9 +14 + 13 + 10 + 12 + 12 + 11 + 10) : 9 ≈ 11,8

La **moyenne** est une caractéristique de position.

Méthode : Calculer une moyenne pondérée

Supposons qu'on attribue des coefficients aux notes de Victor :

Note	4	6	18	7	17	12	12	18
Coefficient	1	1	4	2	4	2	4	2

Calculer alors la moyenne pondérée des notes de Victor.

$$M'_{(Victor)} = \frac{1\times4+1\times6+4\times18+2\times7+4\times17+2\times12+4\times12+2\times18}{1+1+4+2+4+2+4+2} = \frac{272}{20} = 13,6$$

Dans ce cas, la moyenne de Victor est égale à 13,6. Cette moyenne est nettement supérieure à la moyenne brute (sans coefficient). Cela s'explique par le fait que les grands coefficients vont à ses meilleures notes, et à l'inverse, les petits coefficients correspondent à ses notes les plus faibles.

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 $\bar{x}$ et est égale à $\bar{x} = \frac{n_1x_1+\cdots+n_kx_k}{n_1+\cdots+n_k}$ ..

2) Médiane

Méthode : Calculer une médiane

Vidéo https://youtu.be/kr90dXv0NFY

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

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

<u>*Jérôme :*</u>

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

$m_{(Jérôme)} = 12$

<u>*Bertrand :*</u>

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

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

<u>*Julie :*</u>

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

$m_{(Julie)} = 12$

<u>Définition :</u>
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.

V. <u>Étendue, quartiles</u>

1) <u>Étendue</u>

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

<u>Méthode :</u> Calculer une étendue

Vidéo https://youtu.be/PPXGOs2b4Ls

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

$E_{(Jérôme)} = 18 - 4 = 14$

$E_{(Bertrand)} = 15 - 10 = 5$
On considère que 10 est la plus petite valeur

car « 3 » est négligeable dans la série de Bertrand.
*On dit qu'on a **élagué** la série.*

$E_{(Julie)} = 15 – 9 = 6$

L'étendue est une caractéristique de dispersion.

2) 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 : Calculer les quartiles

Vidéo https://youtu.be/Yjh-9nMVmEw
Vidéo https://youtu.be/2jbpNjXMdSA
Vidéo https://youtu.be/ljsDK0ODwlw

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

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.

Jérôme : 4 6 7 12 12 17 18 18

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

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

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

L'écart interquartile est égal à $E_{Q(Jérôme)} = Q_{3(Jérôme)} – Q_{1(Jérôme)} = 17 – 6 = 11$

Bertrand : 3 10 12 12 12 13 13 14 14 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.

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

L'écart interquartile est égal à $E_{Q(Bertrand)} = Q_{3(Bertrand)} - Q_{1(Bertrand)} = 14 - 12 = 2$

Julie : 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.

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

L'écart interquartile est égal à $E_{Q(Julie)} = Q_{3(Julie)} - Q_{1(Julie)} = 13 - 10 = 3$

Les **quartiles** sont des caractéristiques de position.
L'écart interquartile est une caractéristique de dispersion.

3) Interprétations

$M_{(Jérôme)} = 11,8$	$m_{(Jérôme)} = 12$	$E_{(Jérôme)} = 14$	$Q_{1(Jérôme)} = 6$ $Q_{3(Jérôme)} = 17$ $E_{Q(Jérôme)} = 11$
$M_{(Bertrand)} = 11,8$	$m_{(Bertrand)} = 12,5$	$E_{(Bertrand)} = 5$	$Q_{1(Bertrand)} = 12$ $Q_{3(Bertrand)} = 14$ $E_{Q(Bertrand)} = 2$
$M_{(Julie)} \approx 11,8$	$m_{(Julie)} = 12$	$E_{(Julie)} = 6$	$Q_{1(Julie)} = 10$ $Q_{3(Julie)} = 13$ $E_{Q(Julie)} = 3$

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).

VI. Regroupement par classes, histogramme

Méthode : Regrouper les effectifs d'une série par classes et présenter les résultats dans un histogramme

Vidéo **https://youtu.be/Lv3qvDjW6_Q**
Vidéo **https://youtu.be/iRWmgqycx_0**
Vidéo **https://youtu.be/GWDDay-mdVA**
Vidéo **https://youtu.be/BJMLHFmTMcE**

On interroge les élèves d'une classe sur leur taille en cm.
Voici les résultats de l'enquête :

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

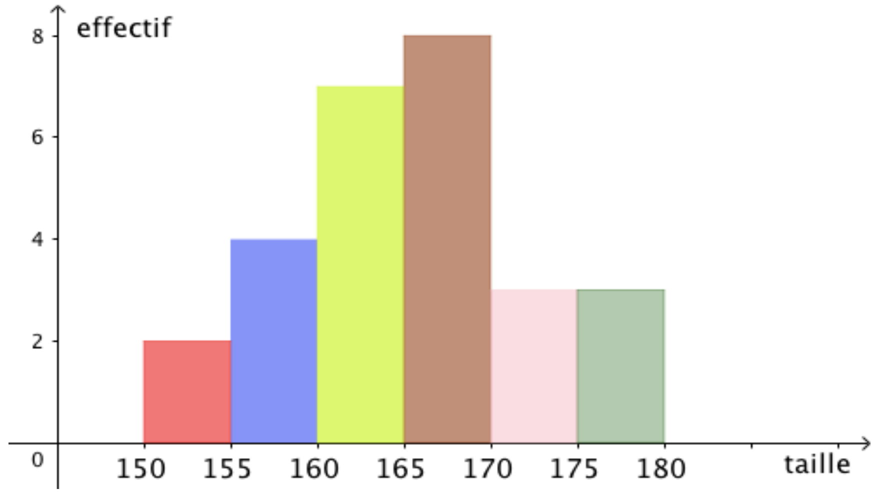
1) Calculer l'étendue de la série de tailles.
2) Regrouper les effectifs de cette série de tailles par classes de longueur 5 cm et présenter les résultats dans un histogramme.
3) Calculer les fréquences de chaque classe en % arrondies à l'unité.
4) a) Calculer la moyenne de la série après avoir centré les classes.
   b) Comparer le résultat précédent avec la moyenne exacte.

1) Étendue = Plus grande valeur – Plus petite valeur

Étendue des tailles = 179 – 151 = 28 cm

2) Regroupement de la série de tailles par classes de longueur 5 cm :

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*



3) Calcul des fréquences :

Tailles	150≤ t <155	155≤ t <160	160≤ t <165	165≤ t <170	170≤ t <175	175≤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.

4) Moyennes :

a) Calcul de la **moyenne en centrant les classes** :

Classes centrées	152,5	157,5	162,5	167,5	172,5	177,5
Effectifs	2	4	7	8	3	3

Il s'agit d'un calcul de moyenne pondéré :
(152,5 x 2 + 157,5 x 4 + 162,5 x 7 + 167,5 x 8 + 172,5 x 3 + 177,5 x 3) : 27
= 4462,5 : 27 $\approx$ 165,3 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
$\approx$ 164,6 cm
La méthode de calcul de moyenne en centrant les classes est assez fiable : 13 mm d'erreur.

VII. Variance, écart-type

Vidéo https://youtu.be/CiFoBkipJQk

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

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

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

Ainsi en reprenant l'exemple précédent des tailles, 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.

La **variance** et **l'écart-type** sont des caractéristiques de dispersion.