I] Traitement des données simples

a) Effectifs et fréquences

En statistiques (analyse et représentation des données à des fins descriptives ou prévisionnelles), les données sont parfois fournies de façon brute, par la liste de toutes les valeurs obtenues.

Voici l'exemple d'une série de 40 mesures des diamètres de tubes PVC fabriqués par une usine :

12,5 – 12,4 – 12,5 – 12,4 – 12,5 – 12,5 – 12,6 – 12,4 – 12,5 – 12,5
12,5 – 12,6 – 12,7 – 12,5 – 12,5 – 12,6 – 12,6 – 12,5 – 12,4 – 12,4
12,5 – 12,5 – 12,5 – 12,6 – 12,5 – 12,4 – 12,6 – 12,4 – 12,5 – 12,4
12,5 – 12,4 – 12,5 – 12,5 – 12,5 – 12,5 – 12,4 – 12,5 – 12,4 – 12,5.

Évidemment ici, le premier travail sera de construire le tableau des effectifs, en comptant méthodiquement les valeurs identiques.

Valeurs des diamètres en *cm*	12,4	12,5	12,6	12,7	total
Effectifs (nombre de valeurs)	11	22	6	1	40

<u>Définitions</u> : L'*effectif* d'une valeur est le nombre d'individus (objets, personnes) qui présentent cette valeur dans la population considérée. L'*effectif total* est le nombre d'individus composant cette population.

Le quotient $f=\frac{effectif}{effectif\ total}$ est appelé *fréquence* de la valeur considérée.

Dans notre exemple on peut ainsi calculer les fréquences relatives à chaque valeur.

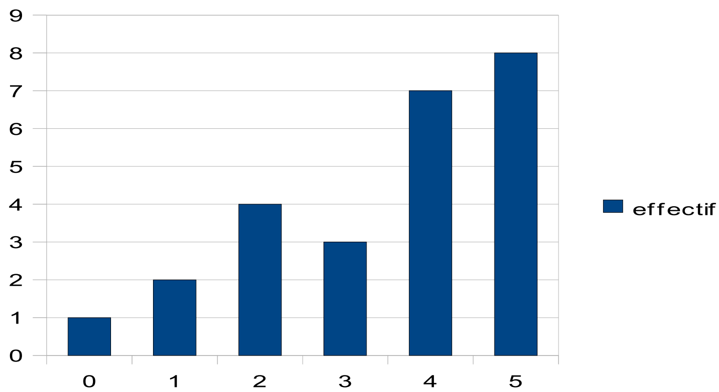
Le nombre obtenu est compris entre 0 et 1 (car l'effectif est toujours inférieur ou égal à l'effectif total) et il est souvent exprimé sous la forme d'un pourcentage : $f\,(en\ pourcentage)=\frac{effectif}{effectif\ total}\times 100\,\%$.

Valeurs des diamètres en *cm*	12,4	12,5	12,6	12,7	total
Fréquences (en %)	11÷40×100=27,5	55	15	2,5	100

<u>Propriété</u> : le total des fréquences des valeurs d'une série statistique est toujours égal à 1 (100%).

En ce qui concerne les données, elles sont parfois fournies par un graphique comme un *histogramme* ou un *diagramme circulaire* qui correspondent finalement à des tableaux d'effectifs ou de fréquences déguisés par la représentation en rectangles (histogrammes) ou en secteurs angulaires.

Voici donc un exemple où le tableau des effectifs se déduit de la lecture d'un graphique : il s'agit des notes (sur 5) obtenues à un DM.



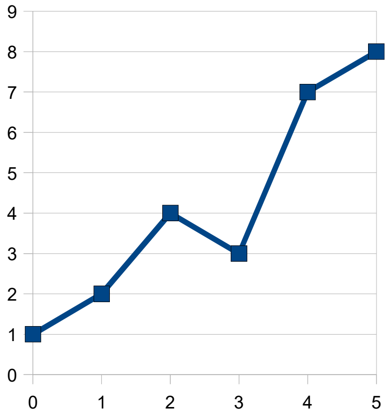
La lecture de ce graphique conduit au tableau des effectifs ou des fréquences suivant :

Note	0	1	2	3	4	5	total
Effectifs	1	2	4	3	7	8	25
Fréquences (en %)	4	8	16	12	28	20	100

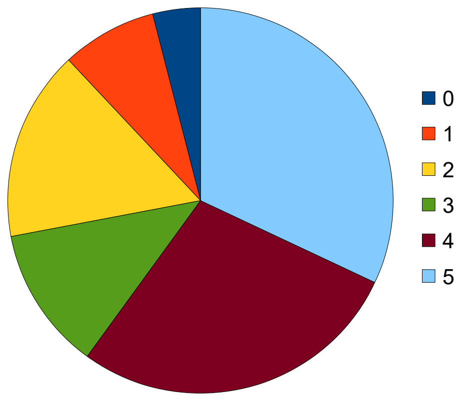
La représentation graphique des données est une étape importante en statistiques, car elle permet de ***communiquer*** des informations chiffrées d'une façon visuelle et globale. On aurait pu présenter autrement les informations concernant cette série de notes, par exemple avec un *diagramme circulaire* (un secteur représente une valeur) ou avec *une courbe* (une ligne montre l'évolution des effectifs).

La confection de tels diagrammes nécessite parfois l'emploi d'un coefficient de proportionnalité, comme pour le calcul de l'angle du secteur angulaire qui est proportionnel à l'effectif (ou la fréquence) de la valeur à représenter : l'effectif total correspond alors à 360° (un tour).

Note au DM (sur 5)	0	1	2	3	4	5	total
Angle du secteur	0,04×360=14°	29°	58°	43°	101°	72°	≈360°

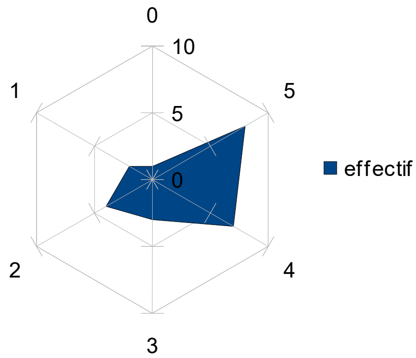


Remarque : du fait des arrondis, la somme des angles peut ne pas faire toujours 360°, mais elle doit se situer aux alentours de 360 : 358 ou 361 par exemple. De la même manière, la somme des fréquences peut ne pas toujours être égale à 100%. Par contre, les effectifs étant des nombres entiers qui ne sont pas arrondis, le total des effectifs devrait être connu exactement.



b) Utilisation d'un tableur

Les graphiques sont très faciles à faire avec un tableur.
Le tableur peut servir à beaucoup d'autres choses que la confection des graphiques (sa vocation première est de calculer) mais accessoirement il est très utile pour ce travail de mise en forme.
Pour information, j'utilise le tableur de OpenOffice (scalc qui enregistre des fichiers appelés « classeur » avec l'extension .odt) qui est libre et gratuit, téléchargeable sur internet en même temps qu'un traitement de texte et autres logiciels de bureautique, très performant et régulièrement mis à jour au même titre que d'autres ensembles logiciels mieux connus mais payants comme Excel de la suite Office de Microsoft.



Rappelons comment taper une formule pour ne pas changer une valeur lors du recopiage de la formule (en tirant sur la poignée en bas à droite de la cellule sélectionnée) : il faut insérer un $ devant la référence de ligne (un chiffre) ou de colonne (une lettre) qui ne doit pas être incrémentée.
Par exemple ici, j'ai tapé B2/$H2*360 pour le calcul de l'angle en B3. La référence $H2 indique au logiciel qu'il ne doit pas modifier la colonne quand je lui demanderai de recopier la formule en tirant sur la poignée noire (le petit carré en bas à droite de la cellule seléctionnée) : en C3 on trouvera C2/$H2*360.

B3 =B2/$H2*360

	A	B	C	D	E	F	G	H
1	note	0	1	2	3	4	5	total
2	effectif	1	2	4	3	7	8	25
3	angle	14	29	58	43	101	115	360

NB : Quand on utilise un tableur, le calcul de l'angle est évidemment inutile puisque le tableur calcule lui-même l'angle dont il a besoin pour tracer le graphique (voir ce graphique en haut à droite de la page).

II] Traitement des données regroupée par classe

a) Regroupement en classes

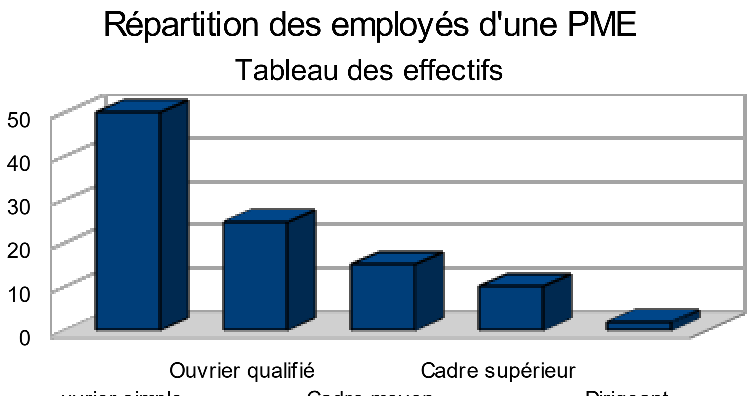
On peut être amené à regrouper les valeurs à l'intérieur de *classes* (groupe de valeurs proches) comme dans l'exemple suivant qui donne les effectifs et salaires des employés d'une PME (Petite et Moyenne Entreprises) :

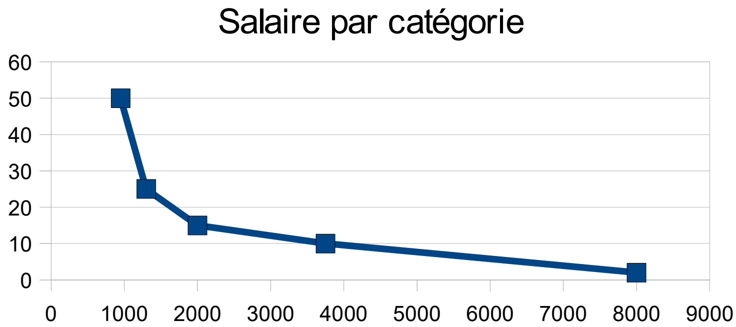
Catégorie	Ouvrier simple	Ouvrier qualifié	Cadre moyen	Cadre supérieur	Dirigeant
Effectif	50	25	15	10	2
Salaire	de 800 à 1100 €	de 1100 à 1500 €	de 1500 à 2500 €	de 2500 à 5000 €	de 5000 à 11000 €

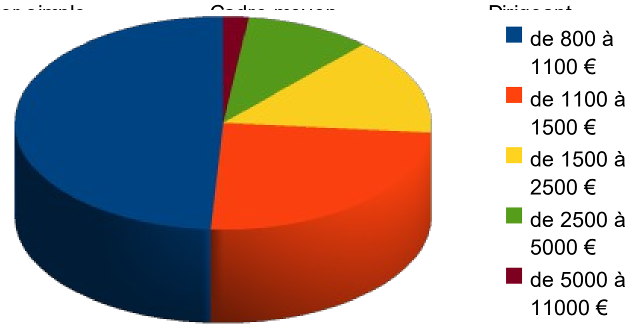
La *valeur statistique* étudiée ici est le *salaire* des employés. Chaque catégorie nominale (désignée par un nom) comme « Ouvrier simple » ou « Dirigeant » correspond à une fourchette de salaires, un intervalle à l'intérieur duquel sont situés les salaires des différents individus composant la classe considérée.

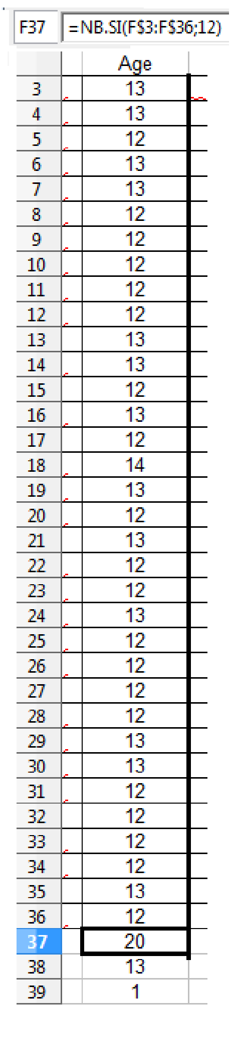
On peut se demander à quelle catégorie nominale appartient un employé qui gagnerait 1100 €, car cette valeur se situe comme limite supérieure de la catégorie « Ouvrier simple » et comme limite inférieure de la catégorie « Ouvrier qualifié ».
Pour éviter ce genre de problème de localisation, on dira que la classe « Ouvrier simple » concerne les employés ayant un salaire allant de 800 € inclus à 1100 € exclu. On notera cela $800 \leq salaire < 1100$ ou avec une notation spéciale [800 ; 1100[qui signifie la même chose (le crochet est tourné vers l'intérieur de l'intervalle pour une valeur incluse, tourné vers l'extérieur pour une valeur exclue).

Représentation graphique des données :
Que ce soit un histogramme ou un diagramme circulaire, on doit être en mesure de tirer l'information utile concernant les valeurs extrêmes des classes ainsi que les effectifs.
La légende et le titre prennent toute leur importance du fait que l'on doit comprendre sans ambiguïté ce qui a été représenté.









F37	=NB.SI(F$3:F$36;12)
	Age
3	13
4	13
5	12
6	13
7	13
8	12
9	12
10	12
11	12
12	12
13	13
14	13
15	12
16	13
17	12
18	14
19	13
20	12
21	13
22	12
23	12
24	13
25	12
26	12
27	12
28	12
29	13
30	13
31	12
32	12
33	12
34	12
35	13
36	12
37	20
38	13
39	1

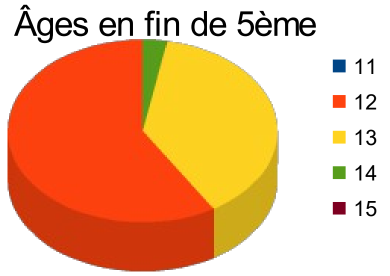
Remarque :
Les points de la courbe ci-dessus correspondent aux salaires moyens touchés par les employés d'une catégorie. Un « Ouvrier simple » gagnant entre 800 et 1100 €, on a considéré qu'il gagnait 950 € ($\frac{800+1100}{2}=950$).
Le regroupement sous forme de classes fait perdre l'information des valeurs individuelles. On se fait une idée du salaire d'un individus d'une classe en lui attribuant la valeur moyenne de sa classe.

b) Utilisation des fréquences

Les fréquences sont, en fait, des fractions calculées à partir des effectifs.
Avec la formule $f=\frac{effectif}{effectif\ total}\times 100$, on utilise des fractions de dénominateur 100 : des pourcentages. Mais on pourrait tout aussi bien utiliser des fractions simplifiées (dont le dénominateur est le plus petit entier possible) comme dans l'exemple suivant qui donne la répartition des élèves d'une classe de 5$^{\text{ème}}$ de collège selon leur âge en juin.
Les données brutes sont celles-ci :
13-13-12-13-13-12-12-12-12-12-13-13-12-13-12-14-13-
-12-13-12-12-13-12-12-12-12-13-13-12-12-12-12-13-12
On obtient le tableau des effectifs ci-dessous où les fréquences ont été exprimées sous la forme d'une fraction irréductible. On peut dire que 10 élèves sur 17 ont 12 ans. Cela revient au même de dire 20 élèves sur 34 ont 12 ans.
La fréquence décimale semble plus précise 10/17 ≈ 0,5882352941 mais elle apporte une information moins lisible et presque ridicule avec sa très grande précision. On lui préfèrera une valeur moins précise comme 0,59 et on l'exprimera en pourcntage : 59%.
Cela revient à dire que près de 59 élèves sur 100 ont 12 ans dans cette classe (c'est évidemment une proportion, il n'y a que 34 élèves dans cette classe).

âge fin 5ème	11	12	13	14	15	Total
effectif	0	20	13	1	0	34
fréquence	0	10/17	13/34	1/34	0	1

Remarque sur l'utilisation du tableur :
Les effectifs par classe d'âge de cette population sont aisés à calculer.
Mais le tableur peut être une aide précieuse pour effectuer des décomptes plus laborieux (lorsqu'en effectuant l'opération manuellement on n'obtient pas le total attendu, il faut tout recompter...).
Dans l'extrait de tableur ci-dessus (à gauche), j'ai sommé les données qui vérifiaient une propriété (que la valeur soit égale à 12 pour la ligne 37), en utilisant la fonction NB.SI() qui requiert dans les parenthèses une plage de données suivie de ; et de la valeur concernée. Dans la cellule F37 j'ai calculé l'effectif des élèves ayant 12 ans en tapant la formule =NB.SI(F$3:F$36;12) qui signifie :



NomBre de valeurs dans la plage allant de F3 à F36 où les valeurs sont égales à 12

L'utilisation de $ avant le numéro de ligne permet de recopier la formule en ne changeant pas ces références de la plage, dans une autre ligne. Pour l'âge 13 par exemple, j'ai juste recopié la formule en tirant sur la poignée, et en changeant le 12 en 13 dans la cellule F38 : =NB.SI(F$3:F$36;13).

Les fréquences exprimées en pourcentages permettent de :

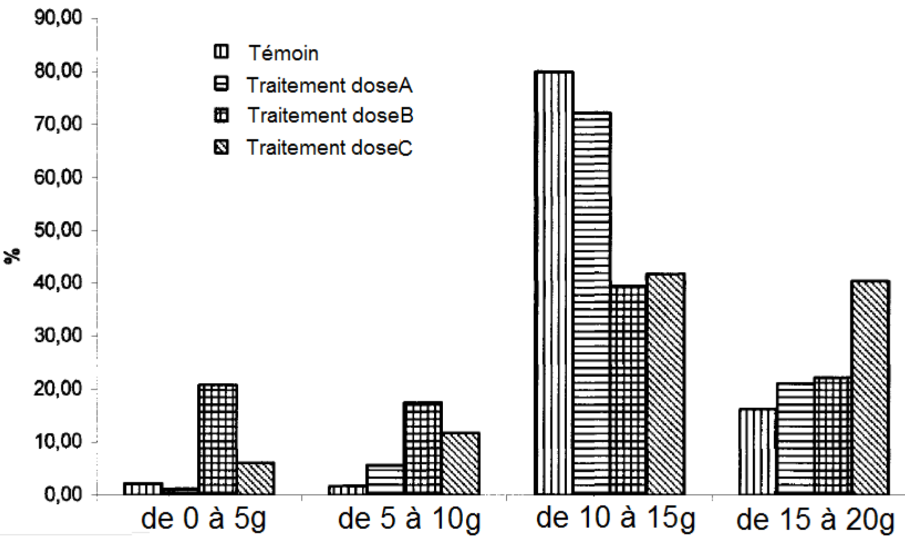
- Comparer entre elles différentes populations
- Comparer une même population en différents moments ou en différents lieux.

Exemple

Comparons par exemple quatre groupes de rats de laboratoire, après une expérience stressante, avec une préparation particulière pour chaque population (injection d'un traitement contre l'anxiété à plus ou moins forte dose). La *variable statistique* étudiée est la *perte de poids* des animaux après l'expérience.

Le graphique ci-dessous donne les résultats pour les quatre groupes.

On peut comparer directement les résultats sur ce graphique car ils sont exprimés en pourcentages.



Le traitement à la dose A n'apporte pas un grand changement dans la réponse au stress (la perte de poids consécutive à cette expérience est principalement de 10 à 15g (près de 80% pour le groupe témoin, des animaux qui n'ont pas été traité, et juste 75% pour le groupe traité à la dose A)
Les animaux traités aux doses B et C montrent davantage d'indifférence au stress (faibles pertes de poids) mais aussi davantage de réponses excessives (pertes de poids dépassant 15g).

III] <u>Moyennes</u>

Quel élève de 4ème ne sait pas déjà ce qu'est une moyenne et comment la calculer ?
L'approche du conseil de classe est souvent l'occasion du calcul de la moyenne trimestrielle pour chaque matière et de la moyenne globale (toutes matières confondues) et à chaque évaluation, le professeur donne la moyenne de la classe et l'élève peut évaluer l'écart de sa note par rapport à cette moyenne.
Ces notions prennent ainsi, périodiquement, une grande importance dans le vécu subjectif d'un élève, mais ce n'est pas pour cette raison, que toutes les finesses de ce concept central sont connues des élèves. Quel est le rôle et les effets des coefficients parfois attribués à certaines notes ? Comment modifier une moyenne avec une note supplémentaire (comment par exemple augmenter sa moyenne d'un point) ? Quel est l'effet d'un zéro dans une moyenne ? Autant de questions régulièrement évoquées comme si une part de mystère s'ajoutait à l'angoisse compréhensible de ces situations.

a) <u>Moyenne simple</u>

On effectue une moyenne simple lorsqu'on dispose de toutes les valeurs d'une grandeur.
La formule est simple : il s'agit d'ajouter toutes les valeurs et de diviser le total par le nombre de valeurs.

Exemple

Lors d'un séjour linguistique à Londres, j'ai dépensé 25€ le premier jour, 30€ le second et 95€ le dernier jour. Combien ai-je dépensé par jour ? Je calcule pour cela le total de mes dépenses : 10+30+95=135.
J'ai dépensé 135€ en 3 jours. Mes dépenses ont été, en moyenne, de 45€ par jour.

Une **formule littérale** pour la moyenne simple.
Admettons que l'on utilise la notation indicielle suivante :

- x_i est la $i^{\text{ème}}$ valeur de la grandeur étudiée.
- l'indice i est un entier qui va de 1 à n, le nombre de valeurs.

Avec ces notations, nos n valeurs se notent donc $x_1, x_2, x_3, x_4, \ldots x_n$.
La moyenne de ces n valeurs est notée, quant-à elle, $\overline{x}$.
Cette moyenne sera donc calculée grâce à la formule $\overline{x} = (x_1+x_2+x_3+x_4+\ldots+x_n) \div n$

Avec la notation fractionnaire $\overline{x}=\dfrac{x_1+x_2+x_3+x_4+\ldots+x_n}{n}$.

Une autre notation, plus concise mais aussi plus délicate à manier (surtout au début) est celle qui utilise le symbole $\sum$ qui se lit *sigma* (la lettre 's' en grec) et qui veut dire 'somme'.

On note les valeurs de début et de fin pour l'indice i de la façon suivante : $\sum_{i=1}^{n} x_i = x_1+x_2+x_3+x_4+\ldots+x_n$.

Ainsi on a, pour finir, la notation suivante pour le calcul de la moyenne : $\overline{x}=\dfrac{1}{n}\times\sum_{i=1}^{n} x_i$.

Une **tableau** pour le calcul de la moyenne simple.

	A	B	C
1	i	NOM	X_i
2	1	André	14
3	2	Bernard	16
4	3	Christine	12
5	4	Dorothée	10
6	5	Elsa	18
7	6	France	15
8	7	Garance	19
9	8	Hélène	17
10	9	Inès	9
11	10	Jules	5
12	11	Killian	12
13	12	Laurent	17
14	13	Mohamed	16
15	14	Nelly	11
16	15	Origène	20
17	16	Philippe	13
18		Moyenne $\overline{x}$ =	**14**

Ce tableau comporte deux lignes (ou deux colonnes), une pour les valeurs de la variable x_i et une autre pour l'indice i. Lorsqu'il y a peu de valeurs (lorsque n est petit), on peut se passer de la ligne correspondant à l'indice. Parfois, on écrira des libellés à la place de l'indice, ou dans une 3$^{\text{ème}}$ ligne.
Par exemple, les notes du trimestre en maths sont données par le tableau ci-contre à gauche où les éléments sont répartis en colonnes.
Je n'ai pas fait figurer la formule qui permet de calculer la moyenne de cette série de notes. Celle-ci est cachée dans la définition du contenu de la cellule[1] C18. On y a écrit : **=MOYENNE(C2:C17).**

X_i	cumuls
14	14
16	30
12	42
10	52
18	70
15	85
19	104
17	121
9	130
5	135
12	147
17	164
16	180
11	191
20	211
13	224
14	**14**

En réalité, le calcul peut être décomposé en cumulant les notes et en divisant le total (224) de toutes les notes cumulées par le nombre de notes (16).

Ainsi $\overline{x}=\dfrac{14+16+12+10+18+15+19+17+9+5+12+17+16+11+20+13}{16}=\dfrac{224}{16}=14$.

La moyenne en maths ce trimestre est de 14.
Dans le tableau de doite, j'ai ajouté une colonne donnant le cumul progressif des notes.

1 Le vocabulaire courant pour les tableurs est supposé connu des élèves qui l'ont étudié en cours de technologie. Néanmoins, rappelons que la cellule est une case de la feuille d'un classeur (les documents que l'on manipule avec un tableur du genre 'scalc', le tableur de OpenOffice) et qu'elle est repérée par une lettre indiquant la colonne et un chiffre indiquant la ligne.

Le tableau ci-dessous donne les moyennes obtenues par les élèves dans quatre matières différentes.

i	NOM	X_i	Y_i	Z_i	T_i	Moyennes
1	André	14	12	18	11	**13,75**
2	Bernard	16	13	17	12	**14,5**
3	Christine	12	12	12	12	**12**
4	Dorothée	10	12	9	9	**10**
5	Elsa	18	19	17	16	**17,5**
6	France	15	14	18	10	**14,25**
7	Garance	19	15	19	13	**16,5**
8	Hélène	17	18	16	14	**16,25**
9	Inès	9	12	11	10	**10,5**
10	Jules	5	7	12	3	**6,75**
11	Killian	12	17	16	15	**15**
12	Laurent	17	10	ABS	14	**13,67**
13	Mohamed	16	12	15	13	**14**
14	Nelly	11	13	15	10	**12,25**
15	Origène	20	11	17	18	**16,5**
16	Philippe	13	15	14	16	**14,5**
	Moyennes =	**14**	**13,25**	**15,07**	**12,25**	**13,64**

Nous pouvons y lire trois sortes de moyennes :

- *Les moyennes par matière* : $\overline{x}=14$ pour la moyenne des notes notées x_i (disons pour les maths), $\overline{y}=13{,}25$ pour la moyenne des notes notées y_i (disons pour la SVT), $\overline{z}=15{,}07$ pour la moyenne des notes notées z_i (disons pour la physique/chimie) et $\overline{t}=12{,}25$ pour la moyenne des notes notées t_i (disons pour la technologie). Ces moyennes ont été calculées comme on l'a indiqué précédemment, en divisant le total des notes d'une matière par le nombre de notes.
  Les formules sont $\overline{x}=\frac{1}{n}\times\sum_{i=1}^{n} x_i$, $\overline{y}=\frac{1}{n}\times\sum_{i=1}^{n} y_i$, $\overline{z}=\frac{1}{n}\times\sum_{i=1}^{n} z_i$ et $\overline{t}=\frac{1}{n}\times\sum_{i=1}^{n} t_i$.
- *Les moyennes par élèves* : André a une moyenne de 13,75. La meilleure moyenne est celle de Garance (16,5) et la plus faible est celle de Jules (6,75). Si l'on veut calculer ces moyennes à la main, il faut additionner les quatre notes et les diviser par 4.
  Ainsi la moyenne de Garance est $\overline{G}=\frac{x_7+y_7+z_7+t_7}{4}=\frac{19+15+19+13}{4}=\frac{66}{4}=16{,}5$,
  celle de Jules est $\overline{J}=\frac{x_{10}+y_{10}+z_{10}+t_{10}}{4}=\frac{5+7+12+3}{4}=\frac{27}{4}=6{,}75$.
- *La moyenne des moyennes,* notons la $\overline{Q}$ =13,64.
  Celle-ci peut être calculée de deux façons différentes :
  - en faisant la moyenne des moyennes par matières
  $$\overline{Q}=\frac{\overline{x}+\overline{y}+\overline{z}+\overline{t}}{4}=\frac{14+13{,}25+15{,}07+12{,}25}{4}=\frac{54{,}57}{4}=13{,}6425$$
  - en faisant la moyenne des moyennes par élèves
  $$\overline{Q}=\frac{\overline{A}+\overline{B}+\overline{C}+...+\overline{P}}{16}=\frac{13{,}75+14{,}5+12+...+14{,}5}{16}=\frac{217{,}92}{16}=13{,}62 .$$

Remarque
il y a une absence de note pour Laurent dans la matière Z (la physique/chimie).
Cette absence n'a pas été comptabilisée dans les notes pour faire la moyenne $\overline{z}$. La formule donnée plus haut a été légèrement adaptée pour tenir compte de cette absence, on a divisé le total par 15, et non par 16 (ce qui aurait pour effet de comptabiliser l'absence comme un 0) : $\overline{z}=\frac{226}{15}=15{,}0666...\approx 15{,}07$. Si l'on avait compté l'absence comme un 0 cela aurait baissé la moyenne de près de 1 point (15,07−14,125=0,945) car alors $\overline{z}=\frac{226}{16}=14{,}125$.
Cette absence n'a pas non plus été comptabilisée dans la moyenne de Laurent $\overline{L}$. La formule a, là aussi, été adaptée pour tenir compte de cette absence, on a divisé le total par 3 (et non par 4) : $\overline{L}=\frac{17+10+14}{3}=\frac{41}{3}=13{,}666...\approx 13{,}67$. Si l'on avait compté l'absence comme un 0 cela aurait baissé la

moyenne de près de 3,5 points (16,67−10,25=3,42) car alors $\overline{L}=\frac{41}{4}=10{,}25$.

Pour ce qui est de la moyenne des moyennes $\overline{Q}$, celle-ci est affectée par l'absence de Laurent, car on ne trouve plus le même résultat si l'on effectue la moyenne des moyennes par matières $\overline{Q}=13{,}6425$ ou si l'on effectue la moyenne des moyennes par élèves $\overline{Q}=13{,}62$. La différence est minime ici, mais elle existe et elle pourrait être amplifiée davantage si les résultats de Laurent avaient été davantage éloignés de la moyenne (par exemple, s'ils étaient supérieurs) tandis que les notes de la matière Z avaient été éloignées dans le sens contraire (s'ils étaient inférieurs).

i	NOM	X_i	Y_i	Z_i	T_i	Moyennes	Moyennes
1	André	14	12	18	11	**13,75**	**13,75**
2	Bernard	16	13	17	12	**14,5**	**14,5**
3	Christine	12	12	12	12	**12**	**12**
4	Dorothée	10	12	9	9	**10**	**10**
5	Elsa	18	19	17	16	**17,5**	**17,5**
6	France	15	14	18	10	**14,25**	**14,25**
7	Garance	19	15	19	13	**16,5**	**16,5**
8	Hélène	17	18	16	14	**16,25**	**16,25**
9	Inès	9	12	11	10	**10,5**	**10,5**
10	Jules	5	7	12	3	**6,75**	**6,75**
11	Killian	12	17	16	15	**15**	**15**
12	Laurent	17	10	ABS	14	**13,67**	**10,25**
13	Mohamed	16	12	15	13	**14**	**14**
14	Nelly	11	13	15	10	**12,25**	**12,25**
15	Origène	20	11	17	18	**16,5**	**16,5**
16	Philippe	13	15	14	16	**14,5**	**14,5**
	Moyennes =	**14**	**13,25**	**15,07**	**12,25**	**13,6198**	**13,4063**
	Moyennes =	**14**	**13,25**	**14,13**	**12,25**	**13,6417**	**13,4063**

On voit dans le tableau ci-dessus, les deux sortes de moyennes que l'on peut calculer, dans les deux options (absence non comptabilisée, ou absence comptée comme 0). Dans la dernière option, l'absence étant comptée comme une note, la moyenne des moyennes est identique pour les deux sortes de moyennes (colonne de droite, moyenne égale à 13,4063) alors que pour la 1$^{\text{ère}}$ option, cette moyenne des moyennes est légèrement plus élevée quand on fait la moyenne par matières (car la moyenne de la matière Z est plus élevée que la moyenne des matières) plutôt que par élèves (car la moyenne de Laurent est plus basse, proche de la moyenne de la classe).

Représentation graphique de la répartition des données

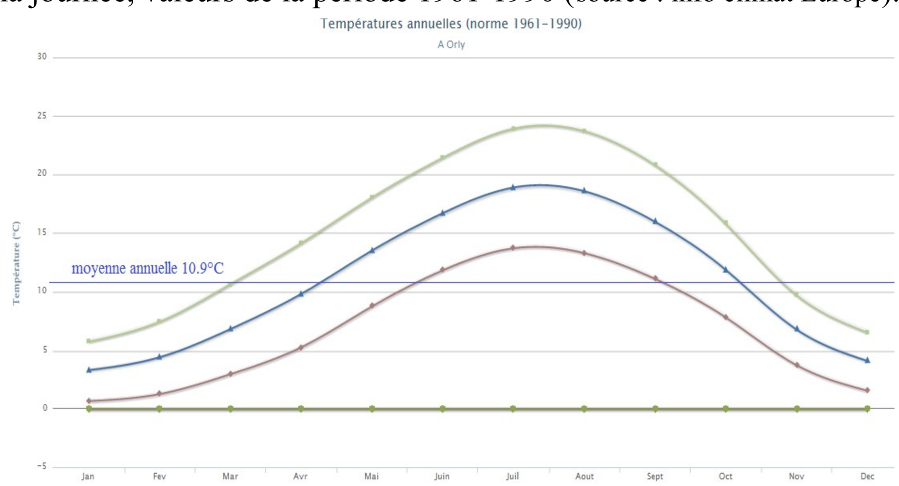
La moyenne peut apparaître dans un graphique de différentes façons.

Si les indices et les valeurs sont réparties selon un système d'axes gradués : par exemple i est disposé en abscisse et x_i est disposé en ordonnée. La moyenne peut être matérialisée par une droite horizontale autour de laquelle fluctuent (varient) les données.

Exemple :

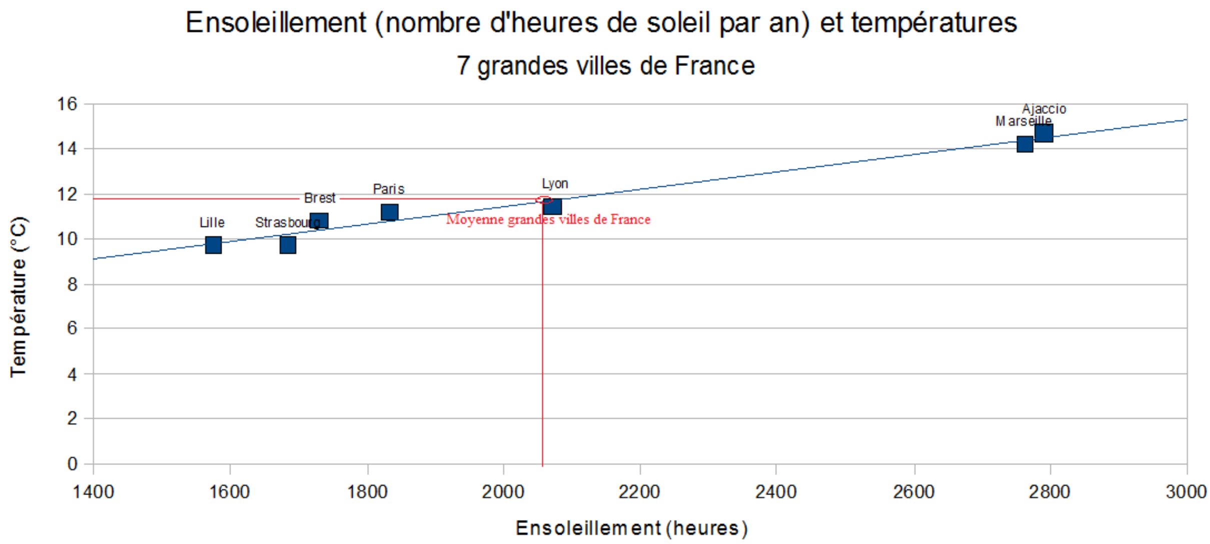
Températures à Orly selon l'heure de la journée, valeurs de la période 1961-1990 (source : info climat Europe).

Sur ce graphique, trois courbes représentent les températures moyennes quotidiennes (minimas, moyenne et maxima). Ce sont des moyennes sur la période de 30 ans considérée. La température moyenne quotidienne représente une autre sorte de moyenne : la moyenne des températures pendant la journée. Nous avons ajouté une 3$^{\text{ème}}$ moyenne : la moyenne annuelle des températures moyennes quotidiennes (la droite d'ordonnée 10,9°C).



Autre exemple, toujours dans le domaine de la météorologie :
La durée d'ensoleillement et la température moyenne dans sept grandes villes de France.
Ces deux grandeurs sont reliées et il est intéressant de les observer conjointement pour mesurer leur relation. La moyenne pour ces villes est approximativement représentée par la ville de Lyon.

	Ajaccio	Lyon	Marseille	Brest	Lille	Paris	Strasbourg	moyenne
Ensoleillement	2790	2072	2763	1729	1574	1833	1685	2064
Température	14,7	11,4	14,2	10,8	9,7	11,2	9,7	11,7



b) <u>Moyenne pondérée</u>
On effectue une moyenne pondérée lorsqu'on attribue un « poids » aux différentes valeurs observées.
Une valeur aura un poids d'autant plus important qu'elle va compter dans le calcul de la moyenne.
La formule pour calculer la moyenne va donc tenir compte de ces poids, en plus des valeurs elles-mêmes.

Exemple introductifs
Supposons que nous voulions calculer la moyenne des âges des élèves d'une classe.
Nous disposons pour cela de la liste des âges des élèves. Bien sûr, il est possible dans ce cas, d'effectuer directement un calcul de moyenne simple, comme nous l'avons décrit dans la partie a). Mais, comme il y a peu de valeurs différentes (seulement 11, 12, 13 et 14), nous pouvons regrouper celles-ci dans un tableau où, pour chaque valeur des A_i, on indique l'effectif n_i, c'est-à-dire le nombre d'individus qui possèdent la caractéristique A_i.
Pour le calcul de la moyenne, on va calculer le total des âges en multipliant chaque valeur A_i par son « poids », c'est-à-dire son effectif.

La formule utilisée ici peut se résumer à cette opération : $\overline{A}=\frac{1\times 11+5\times 12+9\times 13+1\times 14}{1+5+9+1}=\frac{202}{16}=12{,}63$

Cette façon de calculer une moyenne « pondérée » peut être notée $\overline{A}=\frac{\sum_{i=1}^{n} n_i A_i}{\sum_{i=1}^{n} n_i}$, l'écriture de cette formule est simplifiée généralement, car la présence des valeurs de début et de fin pour l'indice n'apporte pas grand chose ici et donc, on peut noter $\overline{A}=\frac{\sum n_i A_i}{\sum n_i}$.

Remarque
Les « poids » ou coefficients de la pondération peuvent être des effectifs ou des fréquences (rapports des effectifs d'une valeur sur l'effectif total). Si dans la formule, on affecte la fréquence f_i à la place de l'effectif n_i, il n'y a alors plus besoin de diviser par l'effectif total car la somme des fréquences vaut 1 :

$$\overline{A}=\frac{\sum n_i A_i}{\sum n_i}=\frac{n_1 A_1+n_2 A_2+\ldots+n_n A_n}{n}=\frac{n_1 A_1}{n}+\frac{n_2 A_2}{n}+\ldots+\frac{n_n A_n}{n}=f_1 A_1+f_2 A_2+\ldots+f_n A_n=\sum f_i A_i$$

i	NOM	A_i
1	André	13
2	Bernard	13
3	Christine	12
4	Dorothée	13
5	Elsa	12
6	France	13
7	Garance	11
8	Hélène	13
9	Inès	13
10	Jules	12
11	Killian	12
12	Laurent	14
13	Mohamed	13
14	Nelly	13
15	Origène	13
16	Philippe	12
	$\overline{A}$ =	12,63

Ages A_i	Effectifs n_i	Sommes A_in_i
10	0	0
11	1	11
12	5	60
13	9	117
14	1	14
15	0	0
Total	16	202
	Moyenne $\overline{A}$ =	12,63

Exemple

- 40% des élèves de la classe sont enfant unique (donc ont 0 frère ou sœur),
- 33% des élèves ont 1 frère ou 1 sœur,
- 20% ont deux frères ou sœurs,
- le restant des élèves ont 3 frères ou sœurs.

Combien y a t-il de frères ou sœurs en moyenne ?
Réponse : 0,94 en moyenne.
Il suffit de calculer $\overline{x}=0{,}33\times1+0{,}2\times2+(1-(0{,}4+0{,}33+0{,}2))\times3=0{,}33+0{,}4+0{,}21=0{,}94$.
On peut aussi utiliser un tableau

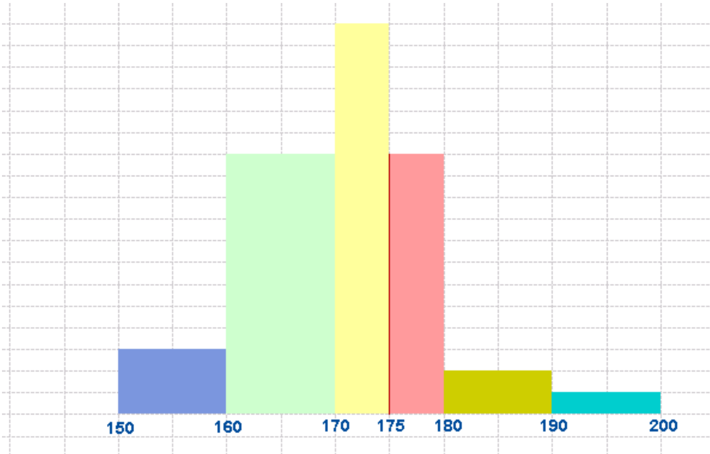
Fréquence f_i	nombre de frères ou sœurs x_i	f_ix_i
0,4	0	0,00
0,33	1	0,33
0,2	2	0,40
0,07	3	0,21
	Moyenne $\overline{x}$	0,94

Valeurs regroupées en classes

Les séries statistiques sont parfois regroupées en classe pour synthétiser les données. Dans ces cas, il y a un effectif associé à chaque classe (ou une fréquence) et les moyennes sont alors calculées en prenant pour chaque classe la valeur centrale ou « centre de la classe » qui n'est autre que la moyenne des extrémités de l'intervalle formé par les valeurs possibles dans l'intervalle.

Exemple

On a interrogé les élèves d'une classe sur leur taille. Les résultats sont regroupés dans six classes et sont donnés par l'histogramme ci-contre à partir duquel une simple lecture graphique permet de retrouver le tableau des effectifs qui indique également les intervalles des tailles pour chaque classe.
Les centres de classe sont calculés en effectuant juste la moyenne entre les bornes de ces intervalles (par exemple, 155 *cm* correspond à la moyenne entre 150 cm et 160 cm qui sont les bornes ou extrémités de l'intervalle[2] [150-160[).



2 Les crochets dans [150-160[indiquent que 150 est contenu dans l'intervalle alors que 160 ne l'est pas. Une taille qui serait contenue dans cet intervalle vérifierait l'encadrement 150≤taille<160. Si on veut inclure 160 dans l'intervalle, il faut noter [150-160], mais ce genre d'intervalle ne convient pas aux classes statistiques car dans les classes consécutives [150-160] et [160-170], la valeur 160 serait contenu 2 fois, ce qui ne se peut pas dans une répartition.

La moyenne des tailles de la classe est alors calculée avec la formule des moyennes pondérées :

$$\bar{x}=\frac{\sum n_i x_i}{n}=\frac{465+1980+...+195}{33}=\frac{5627,5}{33}=170,53$$

Tailles (en cm)	[150-160[	[160-170[	[170-175[	[175-180[	[180-190[	[190-200[	Totaux
Effectifs n_i	3	12	9	6	2	1	33
Centres de classe x_i	155	165	172,5	177,5	185	195	-
Sommes $x_i n_i$	465	1980	1552,5	1065	370	195	5627,5
Moyenne	-	-	-	-	-	-	170,53

<u>Coefficients de pondération</u>

Dans certains cas, les différentes valeurs sont associées à des coefficients qui ont pour fonction de rendre plus importants les « poids » de certaines valeurs relativement à d'autres valeurs de moindre importance. Ainsi, à l'examen du BAC, en série S, jusqu'en 2012, les coefficients de pondération sont donnés par le tableau ci-dessous sauf pour les épreuves anticipées de français (passées en 1$^{\text{ère}}$ de coefficient 2 pour l'écrit et 2 pour l'oral) et les épreuves facultatives :

Liste des épreuves	Coefficients	Nature de l'épreuve	Durée
3. Mathématiques	7 ou 9	écrite	4 heures
4. Physique-chimie	6 ou 8	écrite et pratique	3 heures 30 et 1 heure
5. Sciences de la vie et de la Terre	6 ou 8	écrite et pratique	3 heures 30 et 1 heure
6. Histoire-géographie	3	écrite	4 heures
7. LV 1 étrangère	3	écrite	3 heures
8. LV 2 étrangère ou régionale	2	écrite	2 heures
9. Philosophie	3	écrite	4 heures
10. Éducation physique et sportive	2	CCF (contrôle continu)	

Un candidat ayant choisi les mathématiques comme spécialité aura un coefficient de 9 pour l'épreuve de mathématiques, 6 pour la physique-chimie, 6 pour la SVT, etc.
La moyenne d'un candidat s'effectuera selon la formule des moyennes pondérées, exactement comme si il s'agissait d'un effectif de 9 notes de maths, 6 notes de physique-chimie, etc. chaque note ayant alors le même coefficient (moyenne simple).
Voici, par exemple, les notes de Simon qui a passé le Bac S spécialité maths en 2011 :

Liste des épreuves	Coeff. n_i	Note x_i	$x_i n_i$
1. Français écrit	2	15	30
2. Français oral	2	12	24
3. Mathématiques	9	17	153
4. Physique-chimie	6	14	84
5. Sciences de la vie et de la Terre	6	13	78
6. Histoire-géographie	3	11	33
7. LV 1 étrangère	3	15	45
8. LV 2 étrangère ou régionale	2	12	24
9. Philosophie	3	15	45
10. Éducation physique et sportive	2	17	34
Totaux, moyenne	38	14,47	550

La moyenne de Simon est calculée comme s'il avait 38 notes (la somme de tous les coefficients) selon la formule de la moyenne pondérée :

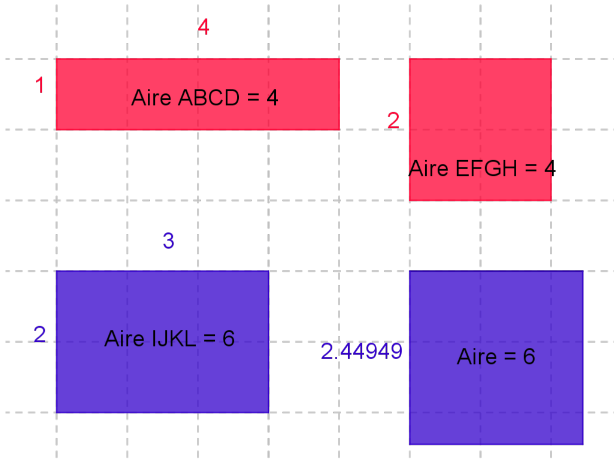
$$\bar{x}=\frac{\sum n_i x_i}{n}=\frac{30+24+...+34}{38}=\frac{550}{38}=14,47$$

c) *Complément (hors programme) : les autres moyennes*

Dans certaines situations, on peut penser à calculer une moyenne comme on vient de le faire (moyenne arithmétique) alors qu'il s'agit d'un autre type de moyenne.

i) Moyenne géométrique

Si on a un rectangle de longueur a et de largeur b, l'aire du rectangle est obtenu par la formule $a \times b$. Un carré qui aurait même aire que ce rectangle, aurait pour côté $\sqrt{a \times b}$.
Il s'agit d'une forme de moyenne, appelée *moyenne géométrique* de a et b.

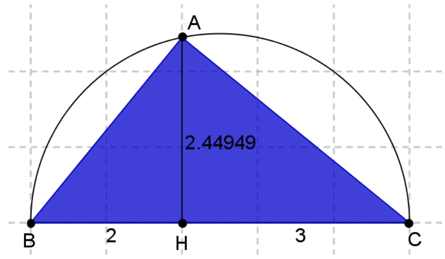


Le carré de côté 2 a la même aire que le rectangle de côtés 1 et 4, car 2 est la moyenne géométrique de 1 et 4 :

$$\sqrt{1 \times 4} = \sqrt{4} = 2$$

Le carré de côté $\sqrt{6} \approx 2{,}44949$ a la même aire que le rectangle de côtés 2 et 3, car $\sqrt{6}$ est la moyenne géométrique de 2 et 3 : $\sqrt{2 \times 3} = \sqrt{6}$.

Un autre moyen de construire géométriquement la moyenne géométrique de deux nombres a et b est indiqué sur la figure ci-dessous où la hauteur du triangle AH est égale à la moyenne géométrique $\sqrt{a \times b}$ de $a=BH$ par $b=CH$.



Avec $a=2$ et $b=3$, on retrouve la moyenne géométrique de 2 et 3 qui vaut $\sqrt{2 \times 3} = \sqrt{6} \approx 2{,}44949$. Pour construire A, il suffit de tracer le segment $[BC]$ de longueur $a+b$ contenant le point H tel que $BH=a$ et $CH=b$, puis le demi-cercle de diamètre $[BC]$ et la perpendiculaire à $[BC]$ passant par H. A est l'intersection du demi-cercle et de cette perpendiculaire. La justification de cette construction repose sur les angles des triangles ABH et CAH qui sont égaux.

Plus particulièrement, $\widehat{ABH} = \widehat{CAH}$ et $\widehat{BAH} = \widehat{ACH}$. Dans ces conditions, $\tan \widehat{ABH} = \frac{AH}{BH} = \frac{CH}{AH}$ et les produits en croix donnent l'égalité cherchée $AH \times AH = BH \times CH$ d'où $AH = \sqrt{BH \times CH}$.

La moyenne géométrique intervient dans une « suite géométrique », c'est-à-dire une suite de nombres construits à partir d'un premier nombre en multipliant à chaque fois par un même coefficient (appelé la raison de la suite). Par exemple, la suite de nombres 1, 3, 9, 27, 81, 972, etc. est une suite géométrique commençant par 1 dont la raison est 3. Dans toute suite géométrique, un nombre quelconque est la moyenne géométrique des nombres qui l'encadrent. Dans notre exemple, 3 est la moyenne géométrique de 1 et 9 (car $3 = \sqrt{1 \times 9}$), de même 9 est la moyenne géométrique de 3 et 27 (car $\sqrt{3 \times 27} = \sqrt{81} = 9$).

ii) Moyenne harmonique

Si, lors d'un trajet, vous roulez à 120 km/h pendant 1h, vous parcourez 120 km. Si vous roulez ensuite à 80 km/h pendant 1h, vous parcourez 80 km. En tout, vous avez roulé pendant 2 h et parcouru 200 km. Votre vitesse moyenne sur ce trajet a été de 100 km/h. C'est la moyenne arithmétique de 120 et 80.
Si, lors du même trajet, vous roulez à 120 km/h sur 100 km, ce parcours dure 100÷120 h, soit 50 minutes. Si vous roulez ensuite à 80 km/h sur les 100 km restants, cette fin du trajet parcours dure 100÷80 h, soit 1,25 h ou encore 75 minutes.

En tout, vous avez roulé pendant 125 minutes, soit $\frac{125}{60} = \frac{25}{12}$ h et parcouru 200 km.

Votre vitesse moyenne sur ce trajet a été de $200 \div \frac{25}{12} = \frac{200 \times 12}{25} = 8 \times 12 = 96$ km/h.

C'est la moyenne harmonique de 120 et 80.

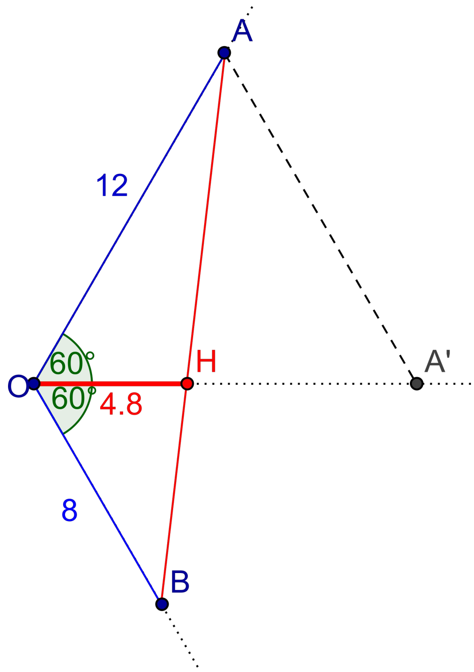
Faisons 1 km à v_1 km/h et 1 km à v_2 km/h. Nous mettons des temps en heures respectivement égaux à $\frac{1}{v_1}$ pour le premier et $\frac{1}{v_2}$ pour le second, soit un temps total égal à $\frac{1}{v_1} + \frac{1}{v_2}$ pour 2 km.

En divisant par 2, nous avons une durée moyenne au kilomètre : c'est l'inverse de la vitesse moyenne v. Ainsi l'inverse de la vitesse moyenne v est la moyenne (arithmétique) des inverses de v_1 et v_2 si v est la moyenne harmonique de v_1 et v_2.

On a donc $\frac{1}{v}=\frac{\frac{1}{v_1}+\frac{1}{v_2}}{2}$, cette formule se transforme pour obtenir directement $v=\frac{2\,v_1\,v_2}{v_1+v_2}$.

Ainsi, avec v_1=120 et v_2=80, on trouve $v=\frac{2\times 120\times 80}{120+80}=\frac{19200}{200}=\frac{192}{2}=96$.

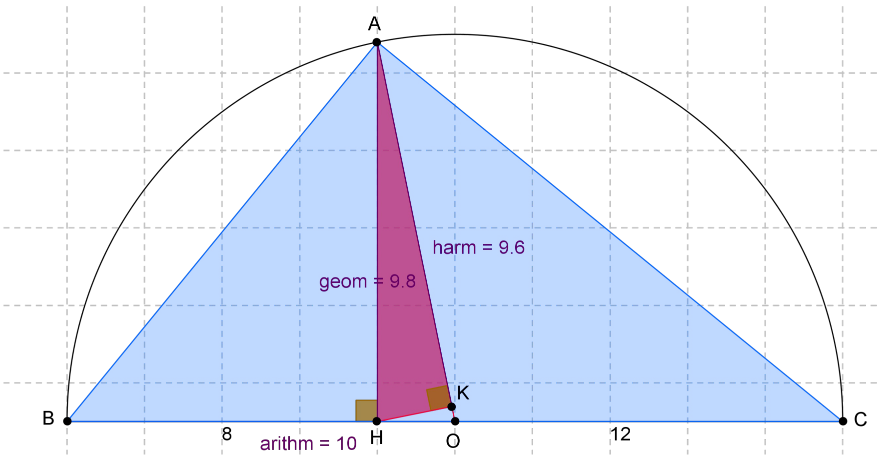
Un moyen de construire géométriquement la moyenne harmonique de deux nombres a et b est indiqué sur la figure ci-contre où [OA), [OB) et [OH) sont des demi-droites de même origine formant deux angles adjacents de 60°. On place A sur [OA) de manière à avoir OA=a et on place B sur [OB) de manière à avoir OB=b. Le segment [AB] coupe la bissectrice de l'angle $\widehat{AOB}$ en H tel que OH est la moitié de la moyenne harmonique de a et b. Sur notre figure, on voit que la moyenne harmonique de 8 et 12 est le double de 4,8 c'est-à-dire 9,6.



Pour justifier cette construction, construisons A' sur [OH) tel que (AA')//(OB). Dans ces conditions, on peut appliquer le théorème de Thalès et on a : $\frac{AH}{HB}=\frac{HA'}{HO}=\frac{AA'}{OB}=\frac{a}{b}$.

On en déduit que $\frac{HA'}{HO}=\frac{a-HO}{HO}=\frac{a}{HO}-1=\frac{a}{b}$ et donc, en divisant par a, on obtient $\frac{1}{HO}-\frac{1}{a}=\frac{1}{b}$, soit $\frac{1}{HO}=\frac{1}{a}+\frac{1}{b}$.

Une dernière construction géométrique (attribuée à Archimède) donnant les trois moyennes étudiées jusqu'ici :



On a déjà vu cette construction qui donne la moyenne géométrique AH des nombres BH et CH, la moyenne arithmétique étant BO=AO=CO.
Le pied de la hauteur issue de H dans le triangle AOH est un un point K tel que KA est la moyenne harmonique de BH et CH.
Avec 8 et 12, on trouve une moyenne géométrique de $\sqrt{8\times 12}=\sqrt{96}\approx 8{,}797958971$ et on retrouve la moyenne harmonique égale à 9,6.
La moyenne harmonique intervient dans une « suite harmonique », c'est-à-dire une suite de nombres construits à partir des inverses des nombres composants une suite arithmétique. La suite des inverses des entiers $1,\frac{1}{2},\frac{1}{3},\frac{1}{4},\frac{1}{5},\frac{1}{6},etc.$ est un exemple de suite harmonique construite à partir de la suite arithmétique des entiers. Dans toute suite harmonique, un nombre quelconque est la moyenne harmonique des nombres qui l'encadrent.

Dans notre exemple, $\frac{1}{3}$ est la moyenne géométrique de $\frac{1}{2}$ et $\frac{1}{4}$ car $\frac{2\times\frac{1}{2}\times\frac{1}{4}}{\frac{1}{2}+\frac{1}{4}}=\frac{\frac{1}{4}}{\frac{3}{4}}=\frac{1}{4}\times\frac{4}{3}=\frac{1}{3}$,

de même $\frac{1}{4}$ est la moyenne géométrique de $\frac{1}{3}$ et $\frac{1}{5}$ car $\frac{2\times\frac{1}{3}\times\frac{1}{5}}{\frac{1}{3}+\frac{1}{5}}=\frac{\frac{2}{15}}{\frac{8}{15}}=\frac{2}{15}\times\frac{15}{8}=\frac{2}{8}=\frac{1}{4}$.

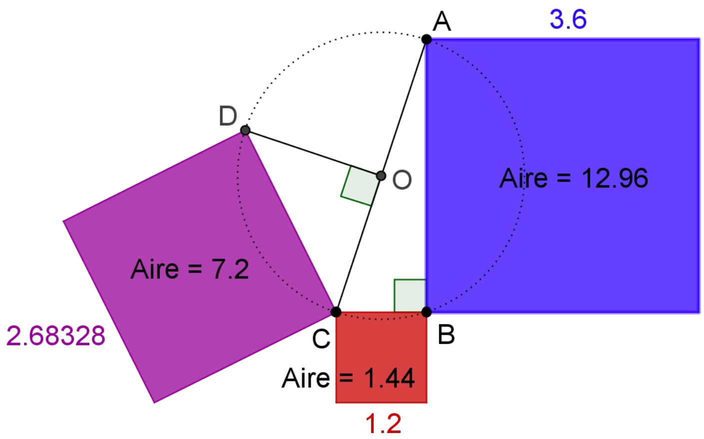
iii) Moyenne quadratique

Ce sera la dernière moyenne étudiée ici, mais il faut savoir qu'on peut en définir d'autres en prenant une fonction f croissante : la f-moyenne de a et b est un nombre m tel que $2f(m)=f(a)+f(b)$.
La moyenne quadratique correspond au cas où la fonction f est la fonction carrée.

Ainsi m est la moyenne quadratique de a et b si $2m^2=a^2+b^2$, c'est-à-dire si $m^2=\frac{a^2+b^2}{2}$, ou encore, a étant positif, $m=\sqrt{\frac{a^2+b^2}{2}}$. D'après cette définition, m est le côté d'un carré dont l'aire est la moyenne arithmétique des aires de deux carrés de côtés a et b.

Cela permet de caractériser l'hypoténuse c d'un triangle rectangle de cathètes a et b. D'après le théorème de Pythagore $c^2=a^2+b^2$ et donc $c^2=2\times\frac{a^2+b^2}{2}=2\times m^2$, d'où finalement $c=\sqrt{2}\times m$ (c est la diagonale d'un carré de côté m, moyenne quadratique de a et b).

Sur la figure ci-contre, à partir d'un triangle ABC rectangle en B, nous avons tracé un tel carré (en violet) : sa diagonale est égale à AC. Il est donc inscrit dans un cercle de diamètre $[AC]$, le point D est sur ce cercle et sur la médiatrice de $[AC]$. Nous avons construit le carré violet du côté de $[DC]$ où il n'y a pas le milieu O de $[AC]$ pour ne pas que nos carrés se chevauchent. L'aire du carré violet DC^2 est la moyenne des aires des 2 autres carrés ($12{,}96+1{,}44=2\times 7{,}2$).



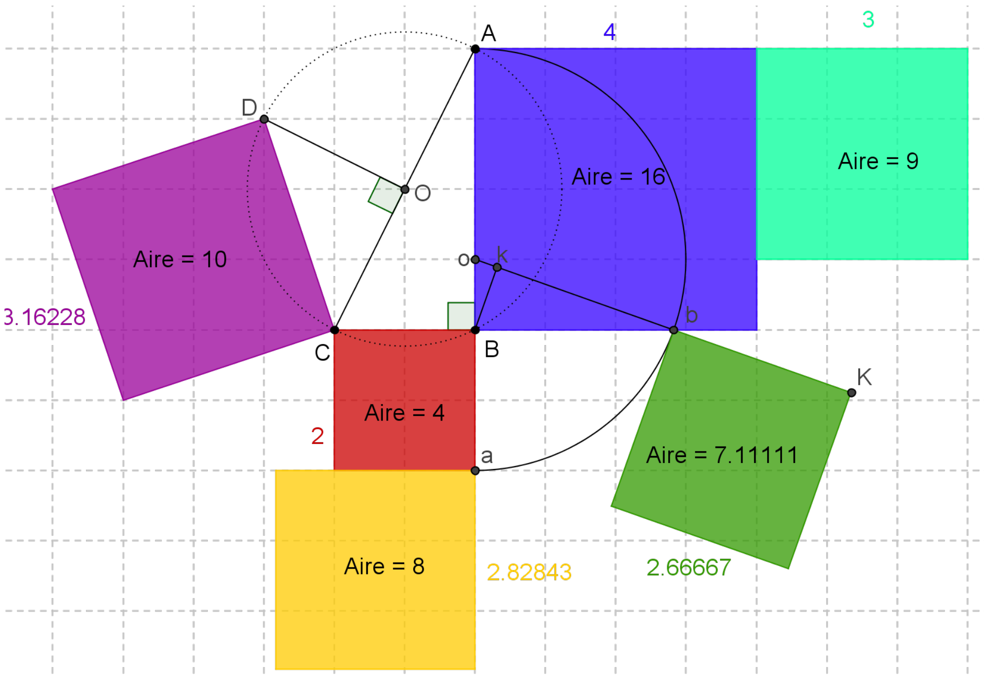
Une propriété générale sur les quatre moyennes définies ici : elles sont toujours dans un même ordre.

- La moyenne quadratique q est toujours la plus grande.
- La moyenne harmonique h est toujours la plus petite.
- Entre la géométrique g et l'arithmétique a, c'est toujours l'arithmétique qui est la plus grande.

Ainsi, nous avons pour tout couple de nombres (x ; y) leurs moyennes qui sont rangées dans cet ordre :

$$h < g < a < q$$

Nous avons illustré cela par une figure qui contient les quatre moyennes de AB et BC et les six carrés construits sur ces six nombres.
Les moyennes 'centrales' g et a sont représentées par des carrés qui suivent le quadrillage alors que les moyennes 'extrêmes' h et q sont représentées par des carrés obliques qui ne le suivent pas.
Ces constructions se déduisent naturellement des deux décrites plus haut.



IV] Médiane et étendue

La moyenne n'est pas le seul paramètre statistique calculé pour une série statistique.
La moyenne est un paramètre « central » dans la mesure où il donne une idée de la valeur d'un individus du centre de la répartition. D'autres paramètres centraux existent : la *médiane* et le *mode*.
Il existe aussi des paramètres statistiques dont le but n'est pas de caractériser un individus central. L'*étendue* d'une série (et aussi l'écart-type ou l'écart-moyen ou l'écart-interquartiles) permet d'évaluer la dispersion plus ou moins grande des valeurs de la répartition.

Exemple
On l'a dit, le premier intérêt des fréquences est de pouvoir comparer des séries qui ont des effectifs totaux différents. Ce qu'on compare alors, c'est la répartition des valeurs dans la série.
Sur l'exemple ci-dessous on dispose de deux séries de mesures faites par des observateurs différents :

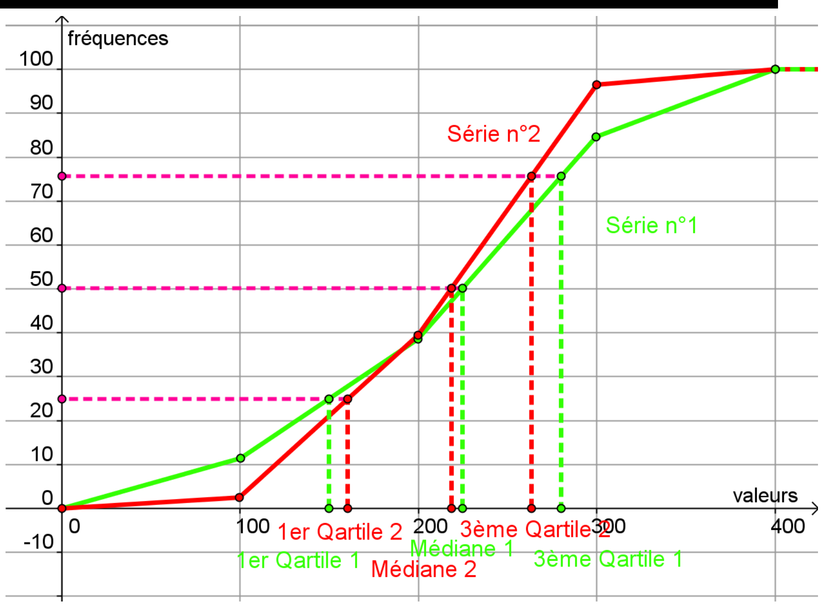
Valeurs	de 0 à 100	de 100 à 200	de 200 à 300	de 300 à 400	de 0 à 100	de 100 à 200	de 200 à 300	de 300 à 400
Effectifs	5	12	20	7	6	96	143	12
Fréquences	11,4%	27,3%	45,4%	15,9%	2,3%	37,4%	55,6%	4,7%

On constate qu'ici, les deux séries ont des profils assez différents :
l'allure de la répartition des valeurs est assez « espacée » pour la première série (on dit "dispersée") alors qu'elle est assez « concentrée » pour la deuxième. Cette caractéristique de plus ou moins grande concentration autour des valeurs centrales, qu'on qualifie de *dispersion* de la série, peut être estimée par un paramètre très simple à calculer : l'**étendue** de la série, qui est égale à la différence entre la valeur maximum et la valeur minimum. Sur notre exemple, l'étendue est difficile à estimer à cause du regroupement par classe (on a perdu l'information de la valeur de la plus petite et de la plus grande valeur), mais il est probable que l'étendue de la 2$^{\text{ème}}$ série soit plus petite que celle de la 1$^{\text{ère}}$, nous reviendrons sur ce paramètre dans un autre exemple.

La valeur qui partage la série en deux groupes de même effectif est ce qu'on appelle la **médiane**.
La médiane est une valeur centrale qui est généralement voisine de la moyenne mais rarement égale à celle-ci, car obtenue par un procédé différent. Pour l'obtenir, lorsqu'on cumule les effectifs (ou les fréquences), la médiane apparaît comme la valeur correspondant à la moitié de l'effectif total (ou à la fréquence cumulée de 50%).

valeurs	de 0 à 100	de 100 à 200	de 200 à 300	de 300 à 400	de 0 à 100	de 100 à 200	de 200 à 300	de 300 à 400
effectifs	5	12	20	7	6	96	143	12
Effectifs cumulés	5	17	37	44	6	102	245	257
fréquences	11,4%	27,3%	45,4%	15,9%	2,3%	37,4%	55,6%	4,7%
Fréq. cumulées	11,4%	38,7%	84,1%	100%	2,3%	39,7%	95,3%	100%

Une valeur cumulée s'interprète facilement : 84,1% des valeurs de la 1$^{\text{ère}}$ série sont inférieures à 300.
Le *polygone des fréquences cumulées* :
Cette courbe relie par des segments les points de coordonnées $(x_i ; fc_i)$ où fc_i est la fréquence cumulée des valeurs inférieures à la borne supérieure de la classe *i*.
Le point de coordonnées $(300 ; 0{,}841)$ correspond à la valeur de la 1$^{\text{ère}}$ série surlignée en bleu.
J'ai ainsi tracé les courbes polygonales des fréquences cumulées pour les deux séries : en vert la 1$^{\text{ère}}$ série ; en rouge la 2$^{\text{de}}$



Estimation de la valeur médiane :
On voit ici, en abscisse du point d'ordonnée 50%, que la médiane de la 1ère série se situe environ à 220, tandis que pour la 2ème série la médiane est un peu inférieure.
Si on veut *calculer* la médiane, il faut utiliser la proportionnalité, c'est-à-dire une formule qui se déduit du théorème de Thalès (demander une justification si nécessaire à votre professeur de maths préféré):

$$\text{médiane de la série 1} = 200+\frac{300-200}{\frac{84{,}1-38{,}7}{50-38{,}7}}=200+\frac{100}{4{,}017}\approx 225$$

NB : La moyenne, quant à elle se calcule avec la formule de la moyenne pondérée :

$$\text{moyenne 1} = \frac{50\times 5+150\times 9+250\times 20+350\times 7}{44}\approx 205$$

On constate que, pour la série 1, la moyenne est inférieure à la médiane. Ces deux paramètres centraux sont égaux lorsque la série des valeurs se disperse de la même façon avant et après la médiane. Dans le cas où la moyenne est inférieure à la médiane, il y a généralement un étalement des valeurs avant la médiane qui déplace la moyenne vers le bas. On parle dans ce cas, d'une dissymétrie de la répartition.

D'autres paramètres statistiques (hors programme) :
Deux autres paramètres statistiques ont été représentés sur le graphique : le 1er quartile et le 3ème quartile. Ces paramètres sont définis comme la médiane, sauf que <u>le 1er quartile est la valeur correspondant à un quart de l'effectif total</u> (ou à 25% de la fréquence cumulée) et <u>le 2ème quartile est la valeur correspondant à trois quarts de l'effectif total</u> (ou à 75% de la fréquence cumulée).
On voit sur le graphique que le 1er et le 3ème quartiles sont plus proches dans le cas de la série n°2 que dans celui de la série n°1, ce qui traduit le fait que cette série est plus "concentrée", moins "dispersée" que la 1ère : les valeurs se regroupent davantages autour des valeurs centrales (médiane ou moyenne). On calcule, pour mesurer cela ce qu'on appelle l'écart-interquartile qui mesure cette dispersion.

Autre exemple :
Je reprends les 40 mesures de diamètres de tubes PVC fabriqués par une usine (exemple de la partie Ia])

- Le paramètre *étendue*, est le plus simple à calculer.

Il mesure (grossièrement) la dispersion : cette série a une étendue égale à 12,7 − 12,4 = 0,3.
On peut commenter cette valeur, dire que l'étendue est faible par rapport aux valeurs centrales (moyenne ou médiane) qui se situent vers 12,5. La série des valeurs est donc "concentrée" autour des valeurs centrales. Mais tout est relatif! Il faudrait pouvoir comparer avec une autre série de valeurs semblables.

- Le paramètre *moyenne*, est également simple à calculer :

On utilise la formule. Le résultat peut éventuellement être obtenu grâce au tableur :

Valeurs des diamètres en *cm*	12,4	12,5	12,6	12,7	Totaux
Effectifs (nombre de valeurs)	11	22	6	1	40
	136,4	275	75,6	12,7	499,7
				Moyenne	12,4925

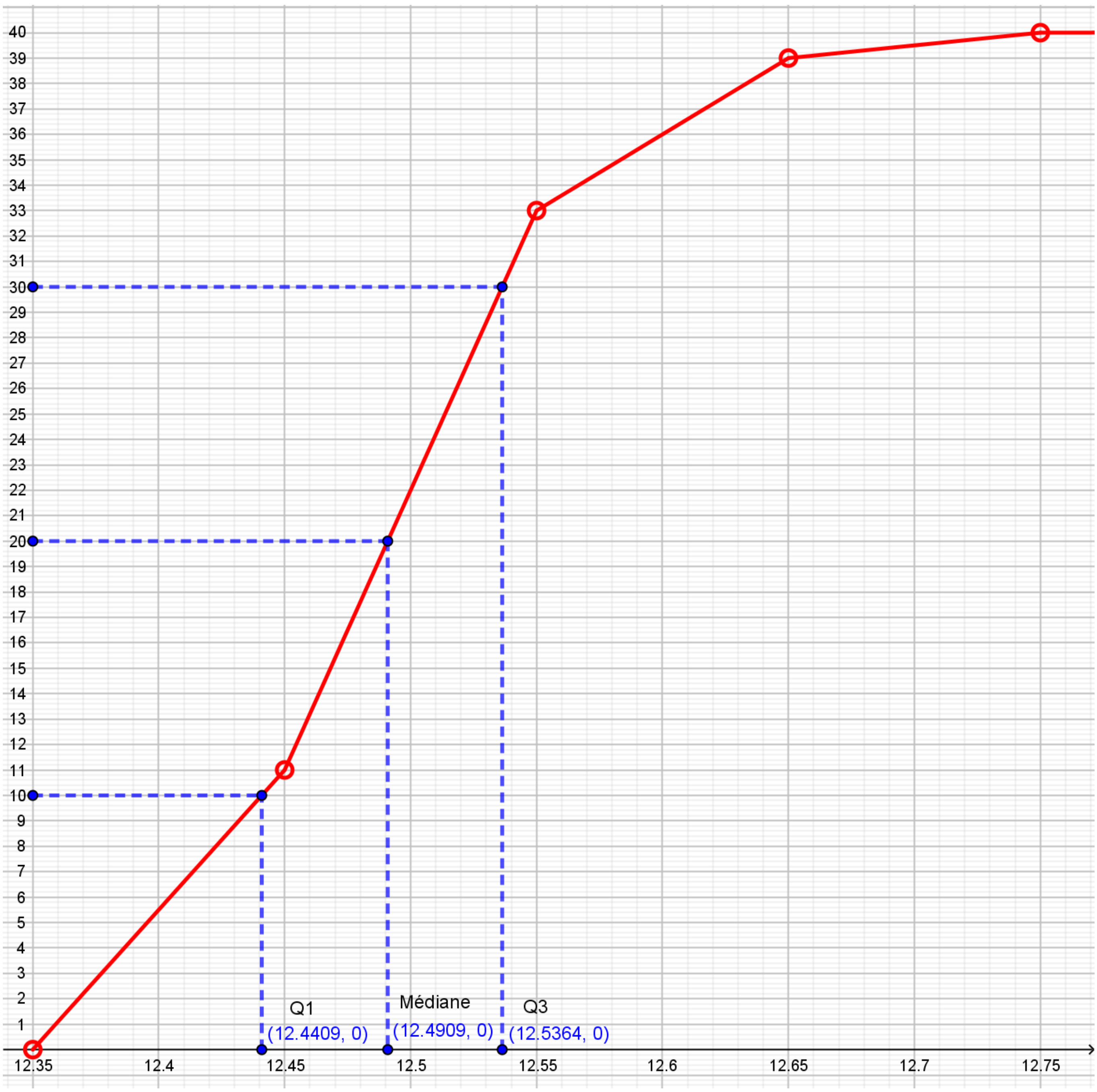
- Les paramètres médiane et écart interquartiles sont plus difficiles à obtenir puisqu'il faut disposer de la courbe des fréquences cumulées. Pour être plus précis ici, il faut comprendre qu'une mesure de 12,4 correspond en réalité à un intervalle qui va de 12,35 à 12,45.

	[12,35;12,45[	[12,45;12,55[	[12,55;12,65[	[12,65;12,75
effectif	11	22	6	1
effectifs cumulés	11	33	39	40

On peut donc maintenant considérer le tableau ci-dessus.
Celui-ci me sert à construire le polygone des effectifs cumulés (voir page suivante).
Je lis sur ce polygone : la **médiane** vaut 12,4909 soit sensiblement la même valeur que la moyenne. Un peu moins que la moyenne ce qui a tendance à montrer un léger étalement vers la droite.
Je lis également les quartiles, notés Q1 et Q3, qui me servent à calculer l'écart interquartile qui vaut

$$Q3-Q1= 12{,}5364-2{,}4409=0{,}0955$$

Dans cet intervalle interquartile se situe la moitié « centrale » de l'effectif.



Q1
(12.4409, 0)
Médiane
(12.4909, 0)
Q3
(12.5364, 0)