STATISTIQUES
& PROBABILITÉS

1

Pourcentages & Statistiques

- Un étudiant dit à son maître : « *Vous m'enseignez à me battre, mais vous parlez de paix. Comment conciliez-vous les deux ?* »
- Le maître a répondu : « *Il vaut mieux être un guerrier dans un jardin qu'être un jardinier dans une guerre.* »

Bouddha

Introduction

Le monde dans lequel nous vivons regorge d'informations chiffrées : proportions, pourcentages, taux d'évolution, etc. Il est donc essentiel de bien comprendre et maîtriser ces informations ! Un exemple très simple (et utile dans la vie de tous les jours !) :

> *« Imaginons que je sois en train de faire les soldes. Un jean m'intéresse et il est à 120€ avec une remise de 50%. Le vendeur me dit « Attendez la prochaine démarque, il sera soldé à 30% en plus des 50%. »*

A votre avis, quelle est donc la remise totale ? 80% ? Pas du tout ! La remise totale est de 65% ... Nous verrons dans ce chapitre comment cela est possible avec ce que l'on appelle les taux d'évolutions successifs. Un autre exemple tout aussi déroutant :

Je suis un instagrammeur influent : 500 000 followers. Un jour, le badbuzz : je prétends que La Terre est plate et que personne n'a été sur La Lune. Internet me ridiculise et je perds 40% de mes abonnés... Les marques qui me sponsorisent font pression sur moi en me disant de m'excuser, de m'expliquer et de regagner mes 500 000 followers.

A votre avis, quel est le pourcentage à appliquer pour réatteindre les 500 000 followers ? 40% ? Encore une fois non ... Il faut environ 67%. Nous verrons également dans ce chapitre comment cela est possible avec ce que l'on appelle les taux d'évolutions réciproques.

Certains individus jouent du fait qu'un grand nombre de personnes ne comprennent pas les subtilités et pièges des pourcentages pour les flouer... N'en faites pas partie !

Concernant les statistiques, nous pouvons partir d'un constat très simple. Prenons trois groupes différents de trois élèves :

- Dans le premier groupe, tout le monde a eu 10.
- Dans la deuxième, le premier a eu 0, le second 10 et le troisième 20.
- Dans la troisième, le premier a eu 9, le second 10 et le troisième 11.

Que se passe-t-il si on calcule la moyenne de ces trois groupes ? Hé oui ! C'est la même moyenne alors que les notes sont radicalement différentes ! Il nous faut donc des indicateurs complémentaire pour mesurer cette dispersion des données autour de la moyenne.

Dans ce chapitre, nous allons revenir sur la notion de *proportion* et *pourcentage*. Puis, dans un second temps, sur les notions de *moyenne pondérée* et de *médiane*. Ces deux indicateurs sont dits *de position*. Nous étudierons deux nouveaux indicateurs dits *de dispersion* : les *quartiles* et l'*écart type* qui nous permettrons de résoudre l'absence d'informations de la seule moyenne.

Les statistiques, dans le sens populaire du terme, traitent à l'aide des mathématiques l'étude de groupe d'une population. En statistique descriptive, on se contente de décrire un échantillon à partir de grandeurs comme la moyenne, la médiane, l'écart type, la proportion, la corrélation, etc. C'est souvent la technique qui est utilisée dans les recensements.
Dans un sens plus large, la théorie statistique est utilisée en recherche dans un but inférentiel. Le but de l'inférence statistique est de dégager le portrait d'une population donnée, à partir de l'image plus ou moins floue constituée à l'aide d'un échantillon issu de cette population.
Dans un autre ordre d'idées, il existe aussi la statistique « mathématique » où le défi est de trouver des estimateurs judicieux (non biaisés et efficaces). L'analyse des propriétés mathématiques de ces estimateurs sont au cœur du travail du mathématicien spécialiste de la statistique.

Auto-évaluation diagnotisque

coucou@coquillagesetpoincare.fr
pour toute(s) question(s) / remarque(s).

1 Prendre 35% d'une quantité, c'est le multiplier par

ⓐ 35 ⓑ $\frac{35}{100}$ ⓒ 35×

2 Dans un lycée de 800 élèves : 25% des élèves sont en Seconde ; 45% des élèves de Seconde sont des filles. La part des filles de Seconde dans le lycée est :

ⓐ 11,25%. ⓑ 20%. ⓒ 70%.

3 Augmenter de 17 % une quantité, cela revient à multiplier cette quantité par

ⓐ 17 ⓑ 1,17 ⓒ $\frac{17}{100}$

4 Diminuer de 23 % une quantité, cela revient à multiplier cette quantité par

ⓐ 0,77 ⓑ 0,23 ⓒ $\frac{17}{100}$

5 « *Imaginons que je sois en train de faire les soldes. Un jean m'intéresse ! Il est à 120€ et en plus une remise de 50% lui sera appliqué. Le vendeur me dit « Attendez la prochaine démarque, il sera encore soldé à 30% en plus des 50%.* »
A votre avis, quelle est donc la remise totale ?

ⓐ 80% ⓑ 20% ⓒ 65%

6 « *Je suis un instagrammeur influent : 500 000 followers. Un jour, le badbuzz et je perds 40% de mes abonnés...* »
A votre avis, quel est le pourcentage à appliquer pour ré-atteindre les 500 000 followers ?

ⓐ 53% ⓑ 40% ⓒ 67%

Pour les questions 7 à 9, on considère la série statistique suivante

Valeur	6	9	13	15	16
Effectif	2	1	3	2	1

7 La médiane de cette série statistique vaut

ⓐ 11, 8 ⓑ 13 ⓒ 3

8 La moyenne de cette série statistique vaut

ⓐ 11, 8 ⓑ 10 ⓒ 13

9 Si je multiplie toutes les valeurs par 2, puis que j'ajoute 7 points (à toutes les valeurs). La moyenne de cette nouvelle série statistique vaut

ⓐ 33 ⓑ 27 ⓒ 30, 6

➤➤➤ Voir solutions p. ??

Cours

1. Autour des pourcentages

A. Proportion

Remarque : Soit t un nombre positif. On rappelle que prendre $t\%$ d'une quantité, c'est le multiplier par $\frac{t}{100}$.

Exemple

- Pour calculer 15% de 36 , on effectue le calcul :

$$\frac{15}{100} \times 36 = \frac{540}{100} = 5,4$$

- Pour calculer 15% de 36 , on effectue le calcul :

$$0,15 \times 36 = 5,4$$

- Pour calculer 15% de 36 km, on effectue le calcul :

$$0,15 \times 36 \text{ km} = 5,4 \text{ km}$$

Remarque : Quelques correspondances à bien connaître :

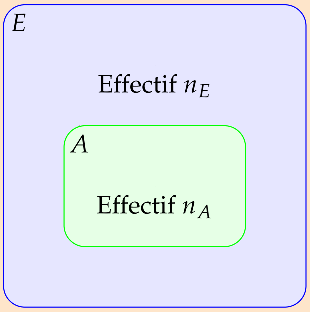
- prendre 25% de ... = prendre un quart de = $\frac{1}{4} \times$ ...
- prendre 50% de ... = prendre la moitié de = $\frac{1}{2} \times$ ...
- prendre 75% de ... = prendre trois quarts de = $\frac{3}{4} \times$ ...

DÉFINITION

Soient E un ensemble de référence non vide et n_E le nombre d'éléments de E.
Soient A une partie de E et n_A le nombre d'éléments de A.

La **proportion** p de A dans E est le réel défini par

$$p = \frac{n_A}{n_E}$$

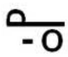


Remarque : On peut exprimer une proportion sous forme décimale, sous forme de fraction ou encore sous forme de pourcentage.

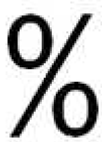
Attention : Attention à ne pas confondre la notion de ratio et celle de proportion !
Prenons un exemple pour observer la différence entre la proportion et le ratio :
Dans une classe de 6ème de 30 élèves, il y a 12 garçons et 18 filles.
La proportion de garçons est de $\frac{12}{30}$ et celle de filles $\frac{18}{30}$.
Quand on parle de ratio, on dit qu'il y a dans la classe 12 garçons pour 18 filles.
On dit que le ration est de 12 : 18. On peut aussi dire qu'il est de 2 : 3 car $\frac{12}{18} = \frac{2}{3}$.

Un peu d'histoire : À l'origine, les traités mathématiques en latin n'étaient pas notés à l'aide de chiffres et de symboles, mais uniquement en mots. Ainsi, l'expression de la fraction 1/100 s'écrivait *unu per cento*.
Plus tard, vers 1425, cette écriture fut simplifiée, en plaçant un P couché sur le *cento*.
Dès 1650, les traités abrégèrent également cento, ne gardant que le o final, ce qui donnait une forme presque similaire au % actuel, avec une barre horizontale au lieu de diagonale.
Dès le début XVIIIe siècle, le % gardera sa forme actuelle

La notation "%" au XVe siècle, abréviation de per cento.



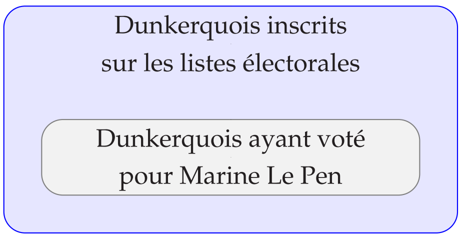
La notation "%" au XVIIe siècle, il ne reste que le o de cento.



La notation "%" dès le XVIIIe siècle, notez la barre diagonale.

Exemple

À Dunkerque, à l'élection présidentielle de 2017, sur les 62852 inscrits sur les listes électorales, 28648 personnes ont voté pour Marine Le Pen. La proportion de votants pour Marine Le Pen est donc $p = \dfrac{28648}{62852} = 0,4558$.
Il y a donc $45,58\%$ des dunkerquois inscrits sur les listes électorales qui ont voté pour Marine Le Pen.



Remarque :

- L'eau recouvre plus de 72% de la surface de la Terre mais ne représente que 0.006% de sa masse.
- La probabilité de mourir assassiné est bien plus élevée que celle de mourir d'un accident d'avion (la première est de 1 sur 20 000 et la seconde, de 1 sur 25 millions).

PROPOSITION

Pour tout ensemble A contenu dans un ensemble non vide E, on a :

$$0 \leq p \leq 1$$

PREUVE

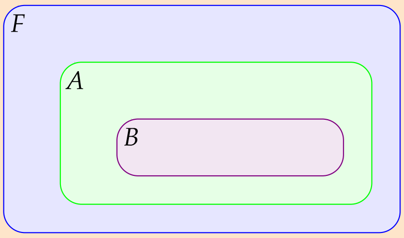
- On a $n_A \geq 0$ et $n_E > 0$.
  Or le quotient de deux réels positifs et encore positif, d'où $\dfrac{n_A}{n_E} \geq 0$, et donc $p \geq 0$.
- A étant une partie de E, on a $n_A \leq n_E$ d'où $\dfrac{n_A}{n_E} \leq 1$ et donc $p \leq 1$.

■

B. Proportion de proportion

PROPOSITION

Soient F un ensemble de référence non vide, A une partie non vide de F et B une partie de A.
Si p_1 est la proportion de A dans F, et si p_2 est la proportion de B dans A alors la proportion de B dans F est $p = p_1 \times p_2$.



PREUVE Soient F un ensemble de référence non vide et n_F le nombre d'éléments de F. A une partie non vide de F contenant n_A éléments et B une partie de A contenant n_B éléments.
On a donc $n_A \neq 0$ et $n_B \neq 0$.

- Si p_1 est la proportion de A dans F, alors $p_1 = \dfrac{n_A}{n_F}$.
- Si p_2 est la proportion de B dans A, alors $p_2 = \dfrac{n_B}{n_A}$.
- Si p est la proportion de B dans F, alors $p = \dfrac{n_B}{n_F}$

Par ailleurs,

$$p_1 \times p_2 = \frac{n_A}{n_F} \times \frac{n_B}{n_A} = \frac{n_A \times n_B}{n_F \times n_A} = \frac{\cancel{n_A} \times n_B}{n_F \times \cancel{n_A}} = \frac{n_B}{n_F} = p$$

■

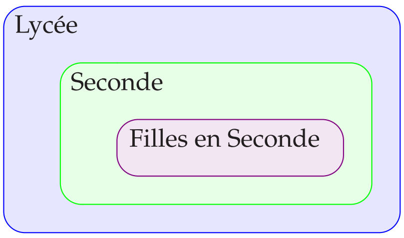
ATTENTION : Pour utiliser la formule précédente ($p = p_1 \times p_2$), on utilise des proportions sous forme décimale ou sous forme de fraction mais **JAMAIS** sous forme de pourcentage !

REMARQUE : On parle aussi de **proportions échelonnées**.

Exemple

Dans un lycée de 800 élèves : 25% des élèves sont en Seconde ; 45% des élèves de Seconde sont des filles.
La part des filles de Seconde dans le lycée est :

$$\frac{25}{100} \times \frac{45}{100} = \frac{1125}{10000} = \frac{11,25}{100} = 11,25\%$$



REMARQUE : Il faut toujours bien préciser l'ensemble de référence, c'est-à-dire par rapport à quel ensemble on calcule la proportion.

C. Variation absolue et variation relative

Dans tout ce qui va suivre, on va noter V_D une **v**aleur de **d**épart pour une quantité quelconque qui va varier pour atteindre une **v**aleur d'**a**rrivée V_A.

DÉFINITION

La **variation absolue ΔV** est donnée par :

$$\Delta V = V_A - V_D$$

La **variation relative** (ou **taux d'évolution**) t est le quotient de la différence entre V_A et V_D par V_D. Elle est donnée par

$$t = \frac{V_A - V_D}{V_D}$$

Remarque : Les variations absolues et relatives peuvent être positives ou négatives : si la quantité augmente, les variations absolue et relative sont positives, sinon elles sont négatives.

Exemple

Nabilla avait 350 neurones (le nombre total de neurones du cerveau humain est estimé de 86 à 100 milliards.) avant les Anges de la Téléréalité 4. A la fin de l'émission, elle n'en possédait plus que 46. On a $V_D = 350$ et $V_A = 46$ et ainsi

$$V_A - V_D = 46 - 350 = -304$$

La variation absolue est donc de -304.
Par ailleurs, on a

$$\frac{V_A - V_D}{V_D} = \frac{46 - 350}{350} \simeq -0,87$$

Le taux d'évolution (en pourcentage) est de -87% (soit donc une perte de 87% de sa matière grise).

D. Coefficient multiplicateur

PROPOSITION

Soit t le taux d'évolution qui permet à une quantité de passer de V_D à V_A. On a alors :

$$V_A = (1+t)V_D$$

PREUVE D'après la définition du taux d'évolution t, pour $V_D \neq 0$, on a $t = \dfrac{V_A - V_D}{V_D}$. D'où

$$t = \frac{V_A - V_D}{V_D} \Leftrightarrow t \times V_D = V_A - V_D \Leftrightarrow t \times V_D + V_D = V_A$$

$$\Leftrightarrow (t+1) \times V_D = V_A$$

■

DÉFINITION

$1+t$ est appelé **coefficient multiplicateur** associé au taux d'évolution t.
On pourra le note CM.

PROPOSITION

- Dans le cas d'une baisse, t est négatif et CM est un réel compris entre 0 et 1.
- Dans le cas d'une augmentation, t est positif et CM est un réel supérieur à 1.

PREUVE D'après la définition du taux d'évolution t, pour $V_D \neq 0$, on a $t = \dfrac{V_A - V_D}{V_D}$.

- (On ne démontrera que $CM \leq 1$ Dans le cas d'une baisse, on a

$$V_A \leq V_D \Leftrightarrow V_A - V_D \leq \cancel{V_D} - \cancel{V_D} \Leftrightarrow \frac{V_A - V_D}{V_D} \leq \frac{0}{V_D} \Leftrightarrow \underbrace{1 + \frac{V_A - V_D}{V_D}}_{=CM} \leq 1 \Leftrightarrow CM \leq 1$$

- Dans le cas d'une hausse, on a

$$V_A \geq V_D \Leftrightarrow V_A - V_D \geq \cancel{V_D} - \cancel{V_D} \Leftrightarrow \frac{V_A - V_D}{V_D} \geq \frac{0}{V_D} \Leftrightarrow \underbrace{1 + \frac{V_A - V_D}{V_D}}_{=CM} \geq 1 \Leftrightarrow CM \geq 1$$

■

Exemple

- Calculons le coefficient multiplicateur associé à une augmentation de 18% :

$$k = 1 + \frac{18}{100} = 1 + 0,18 = 1,18$$

  Le coefficient multiplicateur associé à l'augmentation de 18% est 1,18.
- Trouvons la variation en pourcentage correspondant au coefficient de 0,85.
  $k = 0,85$ donc $t = k - 1 = 0,85 - 1 = -0,15$ soit -15%.
  Comme $k \leq 1$, on a une diminution de 15%.

Exemple

- Augmenter de 17% une quantité, cela revient à multiplier cette quantité par $1 + \frac{17}{100}$. Par conséquent, si on augmente 51 de 17%, cela donne :

$$51 \times \left(1 + \frac{17}{100}\right) = 51 \times (1 + 0,17) = 51 \times 1,17 = 59,67$$

- Réduire une quantité de 56%, cela revient à multiplier cette quantité par $1 - \frac{56}{100}$. Par conséquent, si on réduit 38 de 56%, cela donne :

$$38 \times \left(1 - \frac{56}{100}\right) = 38 \times (1 - 0,56) = 38 \times 0,44 = 16,72$$

E. Evolutions successives

DÉFINITION

Lorsqu'une quantité subit des **évolutions successives** t_1, t_2, ..., t_n de sa valeur, elle subit alors une **évolution globale** t.

REMARQUE : L'ordre dans lequel les évolutions successives sont appliquées n'a pas d'importance.

PROPOSITION

Le coefficient multiplicateur global CM associé à l'évolution t est le produit des coefficients multiplicateurs $CM_1, CM_2, \ldots, CM_n$ associés respectivement aux évolutions $t_1, t_2, \ldots, t_n$. On a donc :

$$CM = CM_1 \times CM_2 \times \cdots \times CM_n$$

ATTENTION : Comme on l'a vu en introduction, le taux d'évolution globale **N'EST PAS** égal à la somme des taux d'évolutions successifs !

REMARQUE : Une petite astuce pour ne jamais faire l'erreur d'additionner les pourcentages ! Raisonner avec des « chiffres ronds » et retenez surtout l'exemple suivant : un article coûte 100€ et il y a deux diminutions successives de 50%. Si on additionnait les pourcentages, cela voudrait dire que l'on aurait une diminution globale de 100% et donc que notre article serait gratuit ! Ce qui n'est pas le cas...
En effet, 100€ diminué de 50%, ça donne 50€ et ces derniers 50€ diminué de 50% donne 25€.
D'ailleurs petit question : quelle est la diminution globale correspondant à deux diminutions successives de 50% ?

Exemple

Reprenons notre exemple introductif :

> *Imaginons que je sois en train de faire les soldes. Un jean m'intéresse ! Il est à 120€ mais avec une remise de 50%. Le vendeur me dit « Attendez la prochaine démarque, il sera encore soldé à 30% en plus des 50%. »*

A votre avis, quelle est donc la remise totale ?

On a $t_1 = \dfrac{50}{100} = 0,5$ et $t_2 = \dfrac{30}{100} = 0,3$ et ainsi $CM_1 = 1 + t_1 = 1 + (-0,5) = 0,5$
Par ailleurs, $CM_2 = 1 + t_2 = 1 + (-0,3) = 0,7$. (Il y a des signes « - » car il s'agit d'une diminution !)
On calcule donc le coefficient multiplicateur global CM associé au taux d'évolution global t :

$$CM = CM_1 \times CM_2 = 0,5 \times 0,7 = 0,35$$

D'où $t = 0,35 - 1 = -0,65$, soit une diminution globale de 65%.

REMARQUE : Le terme « *coefficient multiplicateur* » porte très bien son nom. Non seulement, c'est le coeffcient par lequel on multiplie notre quantité de départ, mais en plus de cela, les coefficients multiplicateurs se ... multiplient entre eux dans le cadre des évolutions successives.

PÉDAGOGIE 1 Représentation graphique des évolutions successives

Les évolutions successives posent toujours quelques problèmes au moment de réaliser les calculs. Le plus simple étant de faire un schéma afin d'avoir tous les éléments au clair.

Heuristique

Réalisons un schéma avec trois évolutions successives :

$$\begin{tikzcd} \boxed{\text{Prix initial}} \arrow[r, bend left, "\times CM_1"] & \boxed{\text{Prix après la } 1^{ere} \text{ évolution}} \arrow[r, bend left, "\times CM_2"] & \boxed{\text{Prix après la } 2^{eme} \text{ évolution}} \arrow[r, bend left, "\times CM_3"] & \boxed{\text{Prix final}} \end{tikzcd}$$

Exemple

La population d'une ville de 45 304 habitants augmente de 5% puis diminue de 10% l'année suivante.

1. Calculer le taux d'évolution global sur les deux années.

On a $t_1 = \dfrac{5}{100} = 0,05$ et $t_2 = -\dfrac{10}{100} = -0,1$ et ainsi $CM_1 = 1 + t_1 = 1 + 0,05 = 1,05$ et $CM_2 = 1 + t_2 = 1 + (-0,1) = 0,9$.
On calcule donc le coefficient multiplicateur global CM associé au taux d'évolution global t :

$$CM = CM_1 \times CM_2 = 1,05 \times 0,9 \simeq 0,945$$

D'où $t = 0,945 - 1 = -0,055$, soit une diminution globale de $5,5\%$.

2. Calculer le nombre d'habitants après ces évolutions. Arrondir à l'entier le plus proche.

Pour cela, nous pouvons faire les deux calculs suivants représentés sur le schéma :

$$\begin{tikzcd} \boxed{45\,304} \arrow[r, bend left, "\times 1{,}05"] \arrow[rr, bend right, "\times 0{,}945"'] & \boxed{\simeq 47\,569} \arrow[r, bend left, "\times 0{,}9"] & \boxed{\simeq 42\,812} \end{tikzcd}$$

F. Evolutions réciproques

DÉFINITION

Une quantité non nulle V_D subit une évolution de taux t et devient égale à une quantité V_A.
Le **taux réciproque** de t est le taux t' permettant de passer de V_A à V_D.

Exemple

Un article coûte 50€ et subit une baisse de 20%, il coûte donc 40€. Il faut une augmentation de 25% pour revenir au prix initial de 50€.
On a donc ici $t = -20\%$ et le taux réciproque est $t' = +25\%$.

PROPOSITION

Le coefficient multiplicateur réciproque CM' associé à l'évolution réciproque t' est l'inverse du coefficient multiplicateur non nul CM associé à l'évolution de départ t. On a :

$$CM' = \frac{1}{CM}$$

PREUVE Soient t et t' deux évolutions successives non nulles telles que la valeur d'arrivée soit la même que la valeur de départ. Le coefficient multiplicateur global est 1. On notre CM et CM' les coefficients multiplicateurs respectifs des évolutions t et t'. On a alors

$$CM \times CM' = 1 \quad \Leftrightarrow \quad CM' = \frac{1}{CM}$$

■

Exemple

Reprenons notre exemple introductif :

> *Je suis un instagrammeur influent : 500 000 followers. Un jour, le badbuzz : je prétends que La Terre est plate et que personne n'a été sur La Lune. Internet me ridiculise et je perds 40% de mes abonnés... Les marques qui me sponsorisent font pression sur moi en me disant de m'excuser, de m'expliquer et de regagner mes 500 000 followers.*

A votre avis, quel est le pourcentage à appliquer pour ré-atteindre les 500 000 followers ?

On calcule le coefficient multiplicateur CM correspondant à la baisse de 40%. On a ainsi

$$CM = 1 - \frac{40}{100} = 1 - 0,4 = 0,6$$

. Le coefficient multiplicateur réciproque CM' associé à l'évolution réciproque t' est ainsi

$$CM' = \frac{1}{CM} = \frac{1}{0,6} \simeq 1,67$$

L'évolution réciproque est donc $t' = CM' - 1 = 1,67 - 1 = 0,67$, soit une augmentation de 67% pour ré-atteindre les 500 000 followers.

PÉDAGOGIE 2 Représentation graphique des évolutions réciproques

Tout comme les évolutions successives, les évolutions réciproques posent quelques problèmes au moment de réaliser les calculs. Le plus simple étant là aussi de faire un schéma similaire aux évolutions successives afin d'avoir tous les éléments au clair.

Heuristique

Réalisons un schéma avec une évolution réciproque :

$$\begin{tikzcd} \boxed{\text{Prix initial}} \arrow[r, bend left, "\times\, CM"] & \boxed{\text{Prix après la } 1^{eme} \text{ évolution}} \arrow[r, bend left, "\times\, CM'"] & \boxed{\text{Prix initial}} \end{tikzcd}$$

Exemple

Suite à son passage en machine à laver, un pull a rétréci de 7%. En utilisant des astuces pour récupérer sa taille d'origine, de quel pourcentage (à 0,1% près) doit-il alors s'agrandir ?

Un petit schéma pour comprendre ce qu'il se passe :

$$\begin{tikzcd} \boxed{\text{Pull normal}} \arrow[r, bend left, "\searrow 7\%"] & \boxed{\text{Pull rétréci}} \arrow[r, bend left, "\nearrow\ ?"] & \boxed{\text{Pull normal}} \end{tikzcd}$$

On calcule le coefficient multiplicateur CM correspondant à la baisse de 7%. On a ainsi

$$CM = 1 - \frac{7}{100} = 1 - 0,07 = 0,93$$

. Le coefficient multiplicateur réciproque CM' associé à l'évolution réciproque t' est ainsi

$$CM' = \frac{1}{cm} = \frac{1}{0,93} \simeq 1,075$$

L'évolution réciproque est donc t'=$CM'$$-1 = 1,075 - 1 = 0,075$, soit une augmentation de $7,5\%$ pour que le pull retrouve sa taille normale.

$$\begin{tikzcd} \boxed{\text{Pull normal}} \arrow[r, bend left, "\times 0{,}93"] \arrow[rr, bend right, "\times 1"'] & \boxed{\text{Pull rétréci}} \arrow[r, bend left, "\times\ 1{,}075"] & \boxed{\text{Pull normal}} \end{tikzcd}$$

2. Autour des statistiques

A. Premier indicateur de position : la médiane

DÉFINITION

La **médiane** M_e d'une série statistique de n valeurs **rangées par ordre croissant** est :
- la valeur centrale si n est impair,
- la moyenne des deux valeurs centrales si n est pair.

REMARQUE :
- On a au moins 50% des valeurs qui sont supérieures ou égales à M_e.
- Dans le cas où l'effectif total est un nombre pair, il est possible que la médiane ne soit pas une valeur de la série.

ATTENTION : Ne pas oublier de ranger les nombres par ordre croissant !

Exemple

Cherchons la médiane de la série statistique suivante

2 ; 10 ; 5 ; 30 ; 7 ; 50 ; 8 ; 40 ; 9 ; 70 ; 12 ; 200 ; 15 ; 50 ; 10 ; 5 ; 30 ; 7 ; 50 ; 8 ; 40 ; 9 ; 70 ; 12 ; 200 ; 15 ; 50 ; 10 ; 5 ; 30 ; 7 ; 50 ; 8 ; 40 ; 9 ; 70 ; 12 ; 200 ; 15 ; 50 ; 10 ; 5 ; 30 ; 7 ; 50 ; 8 ; 40 ; 9 ; 70 ; 12 ; 200 ; 15 ; 50

On range les données par ordre croissant :

2 ; 5 ; 5 ; 5 ; 5 ; 7 ; 7 ; 7 ; 7 ; 8 ; 8 ; 8 ; 8 ; 9 ; 9 ; 9 ; 9 ; 10 ; 10 ; 10 ; 10 ; 12 ; 12 ; 12 ; 15 ; 15 ; 15 ; 15 ; 30 ; 30 ; 30 ; 30 ; 40 ; 40 ; 40 ; 40 ; 50 ; 50 ; 50 ; 50 ; 50 ; 50 ; 50 ; 50 ; 70 ; 70 ; 70 ; 200 ; 200 ; 200 ; 200 ; 7 012.

L'effectif total de la série est 52. Or, $52 = 26 + 26$. La 26e donnée est 15. La 27e donnée est 15. Donc la médiane de la série est 15.

Exemple

Cherchons la médiane de la série statistique suivante :

Valeur	2	5	7	8	9	10	12	15	20	30
Effectif	1	1	1	1	1	3	1	2	1	3

L'effectif total de la série est 15. Or, $15 = 7 + 1 + 7$. La médiane de la série est la 8e donnée. Donc la médiane de la série est 10.

ATTENTION : Quand vous listez les valeurs d'une série statistique comme dans l'exemple précédent, il est important de choisir le bon caractère pour séparer les valeurs.
Lorsqu'il n'y a pas d'ambiguïté (par exemple si l'on ne travaille pas avec des nombres décimaux), on peut séparer les valeurs par une virgule.
Cependant, il vaut mieux éviter ce type de séparateur et de prendre la bonne habitude de toujours utiliser le point-virgule...

B. Deuxième indicateur de position : la moyenne pondérée

DÉFINITION

La **moyenne pondérée** $\overline{x}$ de la série statistique de valeurs $x_1, x_2, ..., x_p$ et d'effectifs respectifs $n_1, n_2, ..., n_p$ est le nombre :

Valeur	x_1	x_2	...	x_p
Effectif	n_1	n_2	...	n_p

$$\overline{x} = \frac{n_1 \times x_1 + n_2 \times x_2 + ... + n_p \times x_p}{n_1 + n_2 + ... + n_p}$$

Remarque :

- L'effectif total est $n = n_1 + n_2 + ... + n_p$.
- La moyenne est plus sensible aux valeurs extrêmes que la médiane.

Notation : La moyenne peut aussi être notée à l'aide de son initiale m, M ou avec la lettre grecque correspondante μ.

Exemple

Calcul de la moyenne pondérée de la série suivante

Valeur	6	9	13	15	16
Effectif	2	1	3	2	1

$$\overline{x} = \frac{6 \times 2 + 9 \times 1 + 13 \times 3 + 15 \times 2 + 16 \times 1}{2 + 1 + 3 + 2 + 1} = \frac{106}{9} \simeq 11,8$$

La moyenne des valeurs pondérées par les effectifs est environ de $11,8$.

Exemple

Calcul de la moyenne pondérée de la série suivante.

Valeur	150	155	160	165	170	175	180	185	190
Effectif	25	23	30	50	40	18	10	3	1

La somme des données de la série est :

$$25 \times 150 \text{ cm} + 23 \times 155 \text{ cm} + \ldots + 3 \times 185 \text{ cm} + 190 \text{ cm} = 32860 \text{ cm}$$

L'effectif total de la série est :

$$25 + 23 + 30 + 50 + 40 + 18 + 10 + 3 + 1 = 200$$

Donc la moyenne de la série est égale à :

$$\frac{32860 \text{ cm}}{200} = 164,3 \text{ cm}$$

Exemple

Calcul de la moyenne pondérée de la série suivante. Il s'agit de notes obtenues par une classe de 25 élèves en mathématiques. Les notes sont regroupées en « intervalles » ou « classes ».

Notes	$[2;8[$	$[8;14[$	$[14;20[$
Effectif	12	8	5

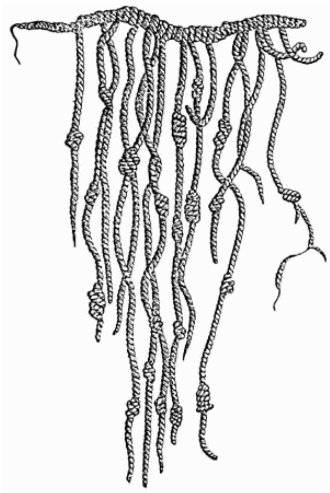
Pour calculer la moyenne générale, on prend la note moyenne de chaque classe (le **centre**) et on l'affecte du nombre d'élèves correspondant :

$$\overline{x} = \frac{5 \times 12 + 11 \times 8 + 17 \times 5}{12 + 8 + 5} = \frac{233}{25} = 9,32$$

La moyenne des notes pondérées par les effectifs est de $9,32$.

Un peu d'histoire :

Quipu, quipou, khipu ou quipo, signifie « noeud » et « compte » en quechua. Le terme désigne aujourd'hui les objets qu'utilisait l'administration inca pour le recensement des données statistiques concernant l'économie et la société de l'empire. En l'absence d'écriture, l'administration figurait les entiers naturels à l'aide de successions de nœuds le long de cordelettes de diverses couleurs fixées à une corde : l'ensemble constituait un quipu.



Il est toutefois possible qu'une partie des quipus ait véhiculé une information d'un autre type, notamment des mots-clefs comme payé ou vendu, voire de véritables textes.

Les quipus constituent un système original de consignation de données qui a été développé très tôt dans le Pérou ancien. En effet, certains pourraient dater de quelques milliers d'années comme ceux découverts par Ruth Shady sur le site de la civilisation de Caral remontant à 4 500 ans.

PROPOSITION : Linéarité de la moyenne

Soient a et b deux nombres réels.
Si une série de valeurs $\{x_i\}_{1 \leq i \leq p}$ a pour moyenne $\overline{x}$, alors la série de valeurs $\{ax_i + b\}_{1 \leq i \leq p}$ a pour moyenne $a\overline{x} + b$.

Exemple

- Le professeur décide d'ajouter un point à toutes les copies, alors la moyenne sera augmentée de 1 point.
- Le professeur a noté sur 10 et trouve une moyenne de 6 sur 10. Pour ramener les notes sur 20, il multiplie toutes les notes par 2 et la moyenne sera aussi multipliée par 2, on trouve 12 sur 20 de moyenne.

PÉDAGOGIE 3 Prévision de la moyenne par des modifications linéaires

Il ne s'agit pas d'une proposition particulièrement difficile ; elle est même plutôt naturelle quand on analyse l'exemple précédent.

Cependant pour démontrer cette proposition, il faut surtout comprendre ce que l'on raconte ! On passe de ce tableau :

Valeur	x_1	x_2	...	x_p
Effectif	n_1	n_2	...	n_p

à ce tableau

Valeur	ax_1+b	ax_2+b	...	ax_p+b
Effectif	n_1	n_2	...	n_p

En gros, on multiplie **toutes** les valeurs par a, puis on ajoute b. Le proposition dit que la moyenne subit la même transformation !

Heuristique

On sait que la moyenne de la série $\{x_i\}_{1\leq i\leq p}$ est :

$$\overline{x}=\frac{n_1\times x_1+n_2\times x_2+...+n_p\times x_p}{n_1+n_2+...+n_p}$$

Calculons maintenant la moyenne de la série de valeurs $\{ax_i+b\}_{1\leq i\leq p}$ et notons la $\overline{x'}$. On a :

$$\begin{aligned}
\overline{x'} &= \frac{n_1\times(ax_1+b)+n_2\times(ax_2+b)+...+n_p\times(ax_p+b)}{n_1+n_2+...+n_p}\\
&= \frac{n_1\times ax_1+n_1\times b+n_2\times ax_2+n_2\times b+...+n_p\times ax_p+n_p\times b}{n_1+n_2+...+n_p}\\
&= \frac{n_1\times ax_1+n_2\times ax_2+...+n_p\times ax_p+n_1\times b+n_2\times b+...+n_p\times b}{n_1+n_2+...+n_p}\\
&= \frac{n_1\times ax_1+n_2\times ax_2+...+n_p\times ax_p}{n_1+n_2+...+n_p}+\frac{n_1\times b+n_2\times b+...+n_p\times b}{n_1+n_2+...+n_p}\\
&= a\underbrace{\frac{n_1\times x_1+n_2\times x_2+...+n_p\times x_p}{n_1+n_2+...+n_p}}_{=\overline{x}}+b\underbrace{\frac{n_1+n_2+...+n_p}{n_1+n_2+...+n_p}}_{=1}\\
&= a\overline{x}+b
\end{aligned}$$

■

PÉDAGOGIE 4 Moyenne *VS* Médiane

Il ne faut pas confondre la médiane et sa cousine la moyenne. Cette dernière est égale à la valeur totale des montants mesurés, divisée par le nombre d'individus. Contrairement à la médiane, la moyenne est très influencée par les valeurs extrêmes et, dans certains cas, moins représentative.
Pour le comprendre, prenons un exemple.

Heuristique

- **Le calcul absurde, les dix souris et l'éléphant.** Prenons un cas absurde. Nous allons faire la moyenne de la masse de 20 souris qui pèsent chacune 30 grammes et de 1 un éléphant qui pèse 5 tonnes soit 5 millions de grammes. Pour calculer cette moyenne nous faisons $\dfrac{20 \times 30 + 5000000}{21}$, car il y a 21 animaux, soit 238 123 grammes par animaux, soit 238 kilos par animaux ce qui est une aberration pour une souris, même une grosse souris.

  On pourrait faire remarquer que mon exemple est mauvais, car j'ai mélangé des espèces différentes et c'est vrai, mais l'exemple montre que **lorsqu'il y a quelque chose de trop gros ou de trop petit, la moyenne n'a plus de sens, on a donc besoin d'une nouvelle valeur mathématique, la médiane.**

  La médiane de nos souris et de notre éléphant :

  $$30 \ ; \ 30 \ ; \ 30 \ ; \ 30 \ \ldots \ 5\,000\,000$$

  La médiane est de 30 c'est à dire que notre éléphant a été retiré naturellement du paquet.

- **Salaire médian ou salaire moyen ?**

  On comprend donc que dans certains cas la médiane a bien plus de sens que la moyenne et notamment pour le salaire des Français. Si je cherche à calculer le salaire moyen des Français, je prends tous les Français qui travaillent, j'ajoute leur salaire et je divise par le nombre de français qui travaillent. Cela signifie que les quelques milliardaires que nous avons en France, vont faire grossir le salaire moyen comme l'éléphant la masse des souris. Ces gens-là correspondent à des valeurs extrêmes et ne sont donc pas représentatifs. Si en outre, on calcule le salaire médian, on va trier du plus petit salaire jusqu'au plus gros éliminant d'un côté les très petits salaires, de l'autre les très gros pour conserver une valeur sans les extrêmes.

  Par conséquent, le salaire moyen est toujours plus élevé que le salaire médian, car les riches rajoutent leur poids dans la balance, si on veut comparer son salaire par rapport au reste des Français, il vaut mieux choisir le salaire médian que moyen.

PROPOSITION : Moyenne et sous-groupes

On étudie un caractère dans une population partagée en deux sous-groupes d'effectifs respectifs n_1 et n_2.
Si la moyenne du caractère est $\overline{x_1}$ dans le premier sous-groupe et $\overline{x_2}$ dans le second, alors la moyenne du caractère dans la population entière est :

$$\overline{x} = \frac{n_1 \times \overline{x_1} + n_2 \times \overline{x_2}}{n_1 + n_2}$$

Exemple

Deux sous-groupes de sportifs se rencontrent.
Dans le premier sous-groupe, il y a 20 sportifs et le temps moyen pour courir un 100 mètres est de 11 secondes.
Dans le deuxième sous-groupe, il y a 30 sportifs et le temps moyen est de 12 secondes.
- Dans le groupe de 50 sportifs, le temps moyen m pour courir un 100 mètres est :

$$m = \frac{20 \times 11 + 30 \times 12}{50} = 11,6$$

Un troisième sous-groupe de 10 personnes rejoint ces 50 sportifs. Dans ce sous-groupe, le temps moyen est de 13 s.
Dans le groupe des 60 sportifs, le temps moyen m' pour courir un 100 mètres est :

$$m' = \frac{50 \times 11,6 + 10 \times 13}{60} = 11,8$$

C. Premier indicateur de dispersion : l'étendue

DÉFINITION

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

Exemple

Calcul de l'étendue de la série suivante :

$$24; 7; 1; 9; 46; 15$$

La plus grande valeur est 46 et la plus petite est 1.
$46 - 1 = 45$, donc l'étendue de la série est égale à 45.

Exemple

Valeur	2	5	7	8	9	12	15
Effectif	10	30	50	40	70	200	50

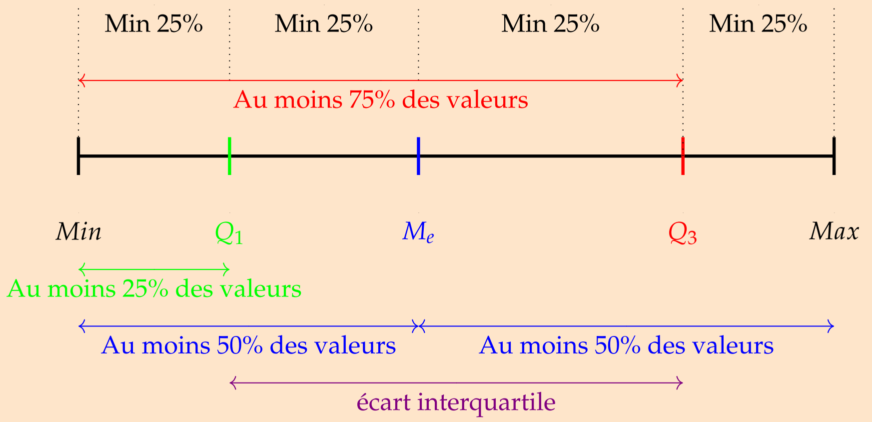
L'étendue de la série est égale à $15 - 2 = 13$.

REMARQUE : L'étendue ne dépend pas de la taille de la série, mais uniquement des valeurs extrêmes.

D. Deuxième indicateur de dispersion : les quartiles et écart inter-quartile

DÉFINITION

• Le **premier quartile**, noté Q_1, est la plus petite valeur de la série telle qu'au moins 25% des valeurs soient inférieures ou égales à Q_1.
• Le **troisième quartile**, noté Q_3, est la plus petite valeur de la série telle qu'au moins 75% des valeurs soient inférieures ou égales à Q_3.
• L'**écart interquartile** est la différence $Q_3 - Q_1$.
• L'**intervalle interquartile** est l'intervalle $[Q_1 ; Q_3]$.



REMARQUE :

- Le **rang** du 1^{er} quartile d'une série de n valeurs est le plus entier supérieur ou égal à $\frac{1}{4}n$.
- Le **rang** du 3^{e} quartile d'une série de n valeurs est le plus entier supérieur ou égal à $\frac{3}{4}n$.

Exemple

On considère la série statistique suivante dont l'effectif total est $n = 10$:

$$2,5 ; 4,8 ; 2 ; 5 ; 5 ; 1 ; 8 ; 9 ; 7 ; 4$$

- D'abord, on range la série dans l'ordre croissant, ce qui donne :

$$1 ; 2 ; 2,5 ; 4 ; 4,8 ; 5 ; 5 ; 7 ; 8 ; 9$$

- Calcul de Q_1 : $\frac{1}{4}n = \frac{1}{4} \times 10 = \frac{10}{4} = 2,5$
  On remplace $2,5$ par 3 (on arrondit à l'entier supérieur). Donc le 1^{er} quartile Q_1 est la 3^{e} valeur, c'est-à-dire $2,5$. $Q_1 = 2,5$
- Calcul de Q_3 : $\frac{3}{4}n = \frac{3}{4} \times 10 = \frac{30}{4} = 7,5$
  On remplace $7,5$ par 8 (on arrondit à l'entier supérieur). Donc le 3^{e} quartile Q_3 est la 8^{e} valeur, c'est-à-dire 7. $Q_3 = 7$.

ATTENTION :

- Ne pas oublier de ranger les nombres par ordre croissant !
- En notant n l'effectif total, on calcule $\frac{n}{4}$ et $\frac{3n}{4}$, puis il faut arrondir à l'entier supérieur.
- Les calculs de $\frac{n}{4}$ et $\frac{3n}{4}$ donnent les positions des valeurs dans la liste et non les valeurs !

PRENONS DE LA HAUTEUR : Les quartiles font partie d'une famille beaucoup plus grande d'indicateurs de dispersion : **les quantiles**.

En statistiques et en théorie des probabilités, les quantiles sont les valeurs qui divisent un jeu de données en intervalles de même probabilité égale. Il y a donc un quantile de moins que le nombre de groupes créés. Par exemple, les quartiles sont les trois quantiles qui divisent un ensemble de données en quatre groupes de même probabilité. La médiane quant à elle est le quantile qui sépare le jeu de données en deux groupes de même probabilité.

Les quantiles des multiples du dixième sont des **déciles**. Ils sont d'usage fréquent en géologie minière, en hydrologie, ainsi que dans nombre de statistiques médicales.
Le décile est donc calculé en tant que 10-quantile :

- le seuil du 1er décile sépare les données entre les 10 % inférieurs et le reste des données
- le seuil du 9e décile sépare les 90 % inférieurs des données des 10 % supérieurs.

Par exemple, ce tableau présente le niveau de vie par an en France métropolitaine, en 2007, réparti en déciles. En 2007, le revenu moyen des ménages était de 29 696 €/an.

Tranches de niveau de vie	Limites des tranches (déciles) en €/an	Niveau de vie moyen dans la tranche en €/an
10 % avaient moins de	10 012 (D1, 1er décile)	7 698 (A)
10 % avaient entre	10 012 et 12 403 (D2)	11 253
10 % avaient entre	12 403 et 14 363 (D3)	13 391
10 % avaient entre	14 363 et 16 201 (D4)	15 297
10 % avaient entre	16 201 et 18 165 (D5)	17 132
10 % avaient entre	18 165 et 20 316 (D6)	19 220
10 % avaient entre	20 316 et 22 969 (D7)	21 565
10 % avaient entre	22 969 et 26 624 (D8)	24 698
10 % avaient entre	26 624 et 33 896 (D9, 9e décile)	29 768
10 % avaient plus de	33 896 (D9)	50 778 (B)
Rapport D9/D1 et B/A	3,39	6,6

- 10 % des Français avaient un niveau de vie inférieur à 10 012 € durant l'année. Ces 10 % de Français avaient, en moyenne, un niveau de vie de 7 698 €. 10 % des Français avaient un revenu compris entre 10 012 € et 12 403 €. Ces Français avaient un niveau de vie moyen de 11 253 €. 10 % des Français avaient un revenu supérieur à 33 896 € (supérieur au 9e décile), et avaient, en moyenne, un niveau de vie de 50 778 €.
- Les 10 % des Français les plus riches avaient, en moyenne, un niveau de vie 6,6 fois supérieur aux 10 % les Français les plus pauvres. Le plancher au-dessus duquel on appartenait aux 10 % les Français les plus riches (le 9e décile, D9) était 3,39 fois supérieur à celui en dessous duquel on appartenait aux 10 % les Français les plus pauvres (le 1er décile, D1).

3. Les représentations

A. Autour des pourcentages

En ce qui concerne les pourcentages, on utilise très souvent un tableau de proportionnalité.

On peut cependant aussi utiliser les fonctions affines (et plus précisément les fonction linéaires) pour représenter une sitation de proportionnalité.

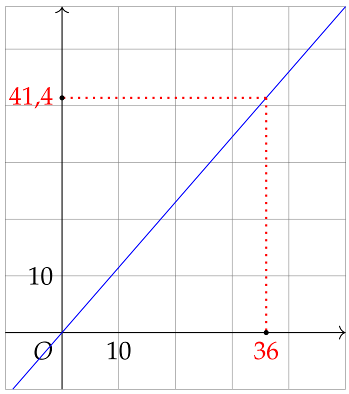
> **REMARQUE :** On rappelle qu'une fonction f définie sur $\mathbb{R}$ est dite **affine** lorsqu'il existe deux réels m et p tels que, pour tout $x \in \mathbb{R}$, $f(x) = mx + p$.
> - m s'appelle le **coefficient directeur** (ou **la pente**).
> - p s'appelle l'**ordonnée à l'origine**.
> - Lorsque $p = 0$, la fonction f est **linéaire**.

En reprenant l'exemple ci-dessus, on obtiendrait la fonction linéaire f suivante, puis on calculerait l'image de 36 par f :

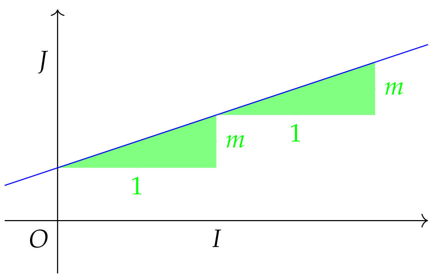
$$f : x \xrightarrow{\times 1{,}15} 1,15x$$

$$f : 36 \xrightarrow{\times 1{,}15} 41,4$$

Ce qui nous donnerait la représentation graphique ci-contre :



> **REMARQUE :** En particulier, si la différence des abscisses est de 1, alors m (qui représente l'augmentation ou la diminution via le coefficient multiplicateur) est égal à la différence des ordonnées.



B. Autour des statistiques

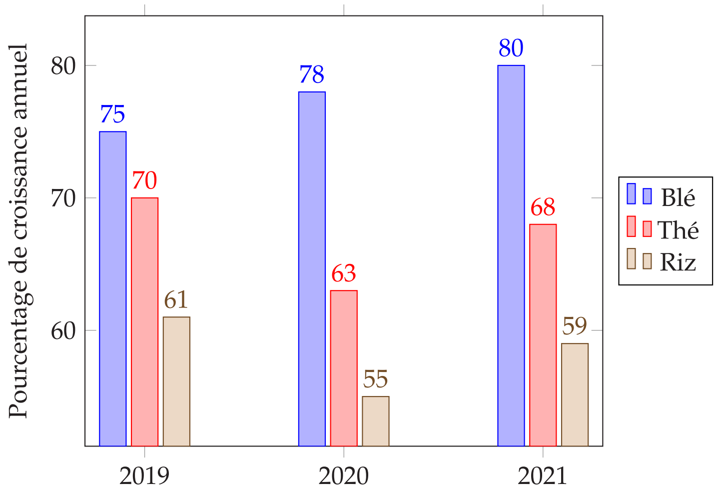
DÉFINITION : Le diagramme à barres

Un **diagramme à barres** (ou en barres), aussi appelé **diagramme à bâtons** (ou en bâtons), est un graphique représentant des variables avec des barres rectangulaires (verticalement ou horizontalement) avec des hauteurs proportionnelles aux valeurs qu'elles représentent.

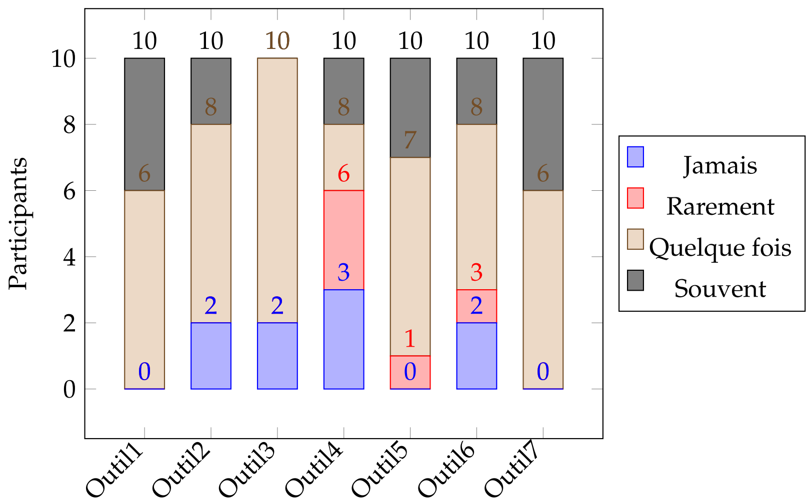
Exemple

Voici des données statistiques et le diagramme à barres correspondant :

Matières	Blé	Thé	Riz
Quantité en 2019 (en %)	75	70	6
Quantité en 2020 (en %)	70	63	6
Quantité en 2021 (en %)	61	55	59



REMARQUE : Il existe une variante que l'on appelle **diagramme à barres empilées**.
Dans l'exemple ci-dessous, on a demandé à 10 personnes leur habitudes de consommation concernant 10 outils numériques.



DÉFINITION : L'histogramme

Un **histogramme** est un moyen de représenter une série statistique dont le caractère est quantitatif continu. Si la série statistique est donnée par les classes ($[a_i, a_{i+1}[$), il est constitué par des rectangles dont la base est le segment $[a_i, a_{i+1}[$ (sur l'axe des réels) et l'aire est proportionnelle à l'effectif de la classe.

REMARQUE : Il faut bien noter que c'est l'aire qui doit être proportionnelle à l'effectif de la classe et non la hauteur elle-même. Si toutes les classes ont la même étendue, cela n'a pas d'importance.
Sinon, il faut procéder de la façon suivante : on note ni l'effectif de la classe $[a_i, a_{i+1}[$. On choisit un rapport de proportionnalité k. La hauteur du rectangle de base $[a_i, a_{i+1}[$ sera $\dfrac{k \times n_i}{a_{i+1} - a_i}$.

Exemple

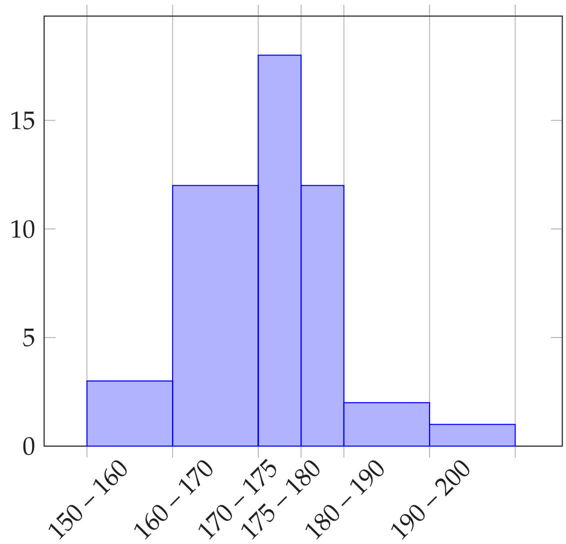
On a demandé la taille des élèves dans une classe de 33 élèves. On obtient les résultats suivants :

Taille (en cm) :	$150 - 160$	$160 - 170$	$170 - 175$	$175 - 180$	$180 - 190$	$190 - 200$
Effectif :	3	12	9	6	2	1

En choisissant un rapport de proportionnalité $k = 10$.

- La hauteur du rectangle de base $[150, 160[$ sera $\dfrac{10 \times 3}{160 - 150} = 3$.
- La hauteur du rectangle de base $[160, 170[$ sera $\dfrac{10 \times 12}{170 - 160} = 12$.
- La hauteur du rectangle de base $[170, 175[$ sera $\dfrac{9 \times 12}{175 - 170} = 18$.
- On réitère l'opération pour tous les intervalles...

L'histogramme correspondant est donc :



DÉFINITION : Le diagramme circulaire

Un **diagramme circulaire** ou **diagramme en secteurs** est un type de graphique utilisé en statistiques. Il permet de représenter un petit nombre de valeurs (ou de classes) par des angles proportionnels à la fréquence (ou l'effectif) de ces valeurs.

REMARQUE : Il est aussi appelé **camembert** et **diagramme en pointes de tarte**.

Exemple

Aude est une candidate qui prépare sérieusement son Concours de Recrutement de Professeur des Écoles (CRPE). Elle décide de lister, par thème, dans un tableau la quantité de notions qu'elle ne maîtrise pas. Voici son tableau :

Matières	Arithmétique	Géométrie	Algèbre	Algorithmique
Nombre de notions	36	72	144	108

Pour mieux se rendre compte de la répartition des notions, elle décide de faire un diagramme circulaire. Elle doit « convertir » ces données en secteurs d'angle.

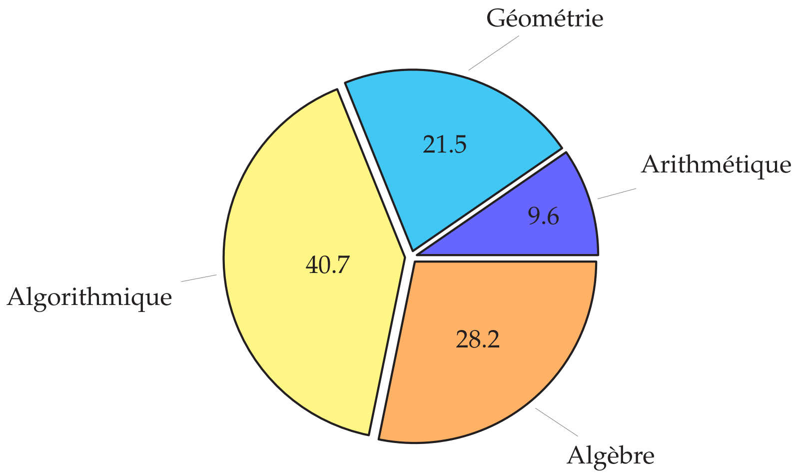
Matières	Arithmétique	Géométrie	Algèbre	Algorithmique	**Total**
Nombre de notions	17	38	72	50	177
Pourcentage	9,6	21,5	40,7	28,2	100
Secteur d'angle	17	38	72	50	177

Pour déterminer les secteurs d'angles, Aude a tout simplement réalisé un produit en croix. Le nombre total de notions, à savoir 177, correspond à l'angle total, à savoir 360°.

Par exemple, pour l'angle correspondant à la géométrie, on a le calcul suivant :

$$\text{Angle recherché} = \frac{38 \times 360}{177} \simeq 21,5$$

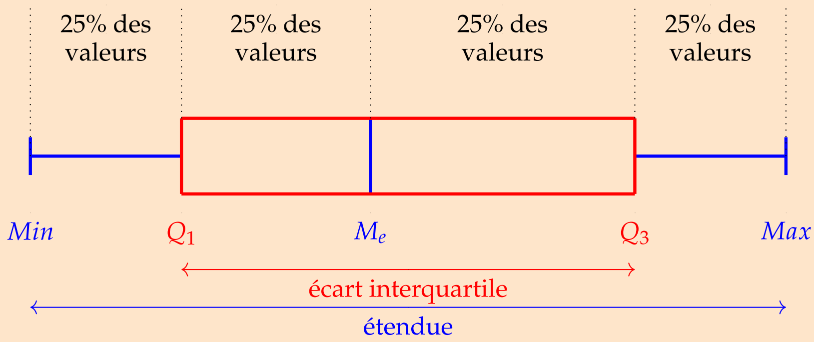
Nombre de notions	**Angle correspondant**
38	Angle recherché
177	360



REMARQUE : À cause d'erreurs d'arrondis, il peut arriver que la somme des pourcentages ne donne pas 100, tout comme la somme des angles au centre ne donne pas 360°.

DÉFINITION : La boîte à moustache

Il est possible de résumer, sous la forme d'un graphique, l'information fournie par l'étendue, ainsi que par la médiane, les deux quartiles et les intervalles qui les séparent. Ce graphique porte le nom de **boîte à moustaches**, ou encore de **boîte à pattes** ou **diagramme en boîte** (*boxplot* en anglais).



REMARQUE : On parle aussi de **boîte de Tukey** (de son inventeur, John Tukey, en 1977).

Exemple

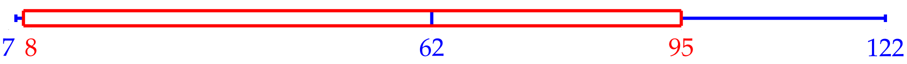
Réalisons la boîte à moustache de la série suivante :

$$73; 122; 73; 8; 95; 43; 64; 44; 32; 111; 115; 7; 7; 8; 111; 60$$

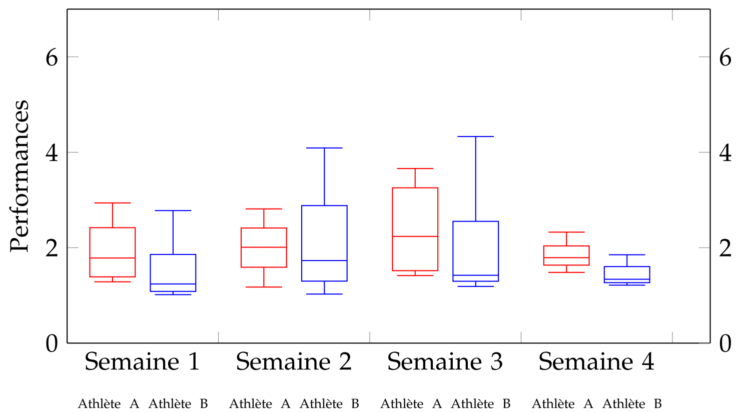
On trouve :

$$Min = 7 \ ; \ Q_1 = 8 \ ; \ Me = 62 \ ; \ Q_3 = 95 \ ; \ Max = 122$$

On a donc



REMARQUE : À noter qu'une boîte à moustache seule n'est pas vraiment très intéressante...
On les utilise souvent pour comparer plusieurs séries statistiques car c'est un moyen rapide de figurer le profil essentiel de séries statistiques quantitatives.
Par exemple, dans le graphique ci-dessous, nous pouvons comparer rapidement les performances deux athlètes sur 4 semaines.



DÉFINITION : Diagramme de Kiviat

Le **diagramme de Kiviat**, ou **diagramme en toile d'araignée** sert à représenter sur un plan en deux dimensions au moins trois ensembles de données multivariées. Chaque axe, qui part d'un même point, représente une caractéristique quantifiée. Est ainsi facilitée une analyse détaillée de plusieurs objets, ainsi que leur comparaison générale (comparaison des surfaces) ou point par point. Ce type de diagramme n'est utile que si les axes sont correctement normés selon l'importance donnée à chaque caractéristique.

REMARQUE :

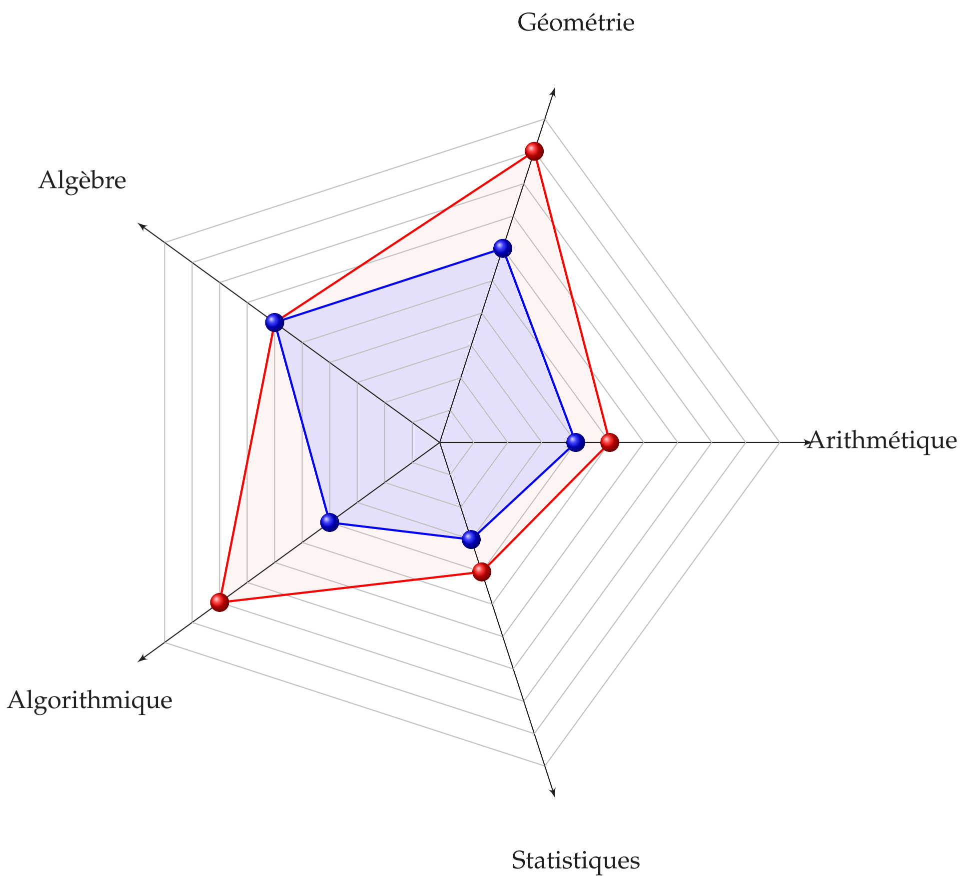
- Il est aussi appelé **diagramme en radar** ou encore **diagramme en étoile**.
- Il fut créé en 1877 par le statisticien allemand Georg von Mayr.

Exemple

Considérons les données suivantes

Matières	Arithmétique	Géométrie	Algèbre	Algorithmique	Statistiques
Notes de l'élève A	5	9	6	8	4
Notes de l'élève B	4	6	6	4	3

Alors le diagramme de Kiviat est



Je teste mes connaissances

À la fin de ce chapitre, je dois être capable de :

- Calculer des proportions et des pourcentages.
- Manipuler les pourcentages de pourcentages.
- Utiliser les coefficients multiplicateurs.
- Calculer des taux d'évolutions.
- Connaissant deux taux d'évolutions successifs, déterminer le taux d'évolution global.
- Connaissant un taux d'évolution, déterminer le taux d'évolution réciproque.
- Déterminer une moyenne pondérée et utiliser les propriétés de la moyenne.
- Déterminer une médiane et les quartiles Q_1 et Q_3.
- Choisir, calculer et interpréter des les indicateurs de position et de dispersion.
- Comparer des séries statistiques en utilisant les indicateurs de position et de dispersion ou des données sous forme de graphiques.

QCM d'auto-évaluation

coucou@coquillagesetpoincare.fr
pour toute(s) question(s) / remarque(s).

Voici un QCM d'auto-évaluation pour vous tester. Vous avez quelques questions reprenant l'ensemble des notions abordées dans ce cours.

10 Quel coefficient multiplicateur correspond à une augmentation de 37% ?

ⓐ 0,37 ⓑ 1,37 ⓒ 0,63

11 Quel coefficient multiplicateur correspond à une diminution de 58% ?

ⓐ 0,58 ⓑ 1,58 ⓒ 0,42

12 Une augmentation de 20% suivie d'une diminution de 65% correspondent à une diminution globale de :

ⓐ 58% ⓑ 45% ⓒ 85%

Pour les questions 13 à 15, on considère la série statistique suivante

Valeur	12	14	20	25	43	47
Effectif	5	7	14	5	2	32

13 La médiane et la moyenne de cette série statistique valent respectivement

ⓐ $M_e = 22,5$ et $\overline{x} \simeq 33,12$ ⓑ $M_e = 43$ et $\overline{x} \simeq 26,83$ ⓒ $M_e = 43$ et $\overline{x} \simeq 33,12$

14 L'étendue de cette série statistique vaut

ⓐ 35 ⓑ 30 ⓒ 47

15 Les quartiles de cette série statistique valent

ⓐ $Q_1 = 12$ et $Q_3 = 47$ ⓑ $Q_1 = 27$ et $Q_3 = 43$ ⓒ $Q_1 = 20$ et $Q_3 = 47$

S'entraîner

Proportion

16 Ouvrez la cage aux oiseaux CORRIGÉ

Dans une réserve de protection d'oiseaux, il y a 315 pics noir, ce qui représente 21 % du nombre total d'oiseaux.

Quel est le nombre d'oiseaux de cette réserve ?

17 En avant la musique CORRIGÉ

Parmi les 600 spectacteurs d'un concert, 240 ont moins de 18 ans.

Calculer la proportion des personnes mineures dans le public en pourcentage.

18 En avant la musique II CORRIGÉ

Parmi les 600 spectacteurs d'un concert, 300 ont moins de 18 ans.

Calculer la proportion des personnes mineures dans le public en pourcentage.

19 Le cadeau commun CORRIGÉ

Le cadeau commun que nous souhaitons faire à Timothée coûte 90 €. Je participe à hauteur de 20 % du prix total.

Combien ai-je donné pour le cadeau de Timothée ?

Proportion de proportion

20 Pour démarrer CORRIGÉ

Dans un lycée, la proportion des élèves de seconde qui sont passés en première est égale à 0,92.
Parmi ces élèves passés en première, la proportion des élèves qui ont choisi d'aller en première STMG est égale à 0,35.

Calculer la proportion d'élèves ayant choisi d'aller en première STMG parmi les élèves qui étaient en seconde. Exprimer cette proportion en pourcentage.

21 Comme chiens et chats CORRIGÉ

Un refuge pour animaux accueille 120 chiens et chats mâles et femelles :

- 30% sont des chats ;
- parmi les chiens, 25% sont des mâles ;
- il y a 32 mâles.

Compléter le tableau ci-contre

	Chiens	Chats	Total
Mâles			
Femelles			
Total			120

Coefficient multiplicateur

22 Le point de départ CORRIGÉ

Soit $V_D = 48$. Dans chacun des cas suivants, déterminer la valeur d'arrivée V_A :

1) après une hausse de 37% ;
2) après une baisse de 11%.

23 Coefficient multiplicateur associé CORRIGÉ

Pour chacun des taux d'évolution suivants, déterminer le coefficient multiplicateur associé.

1) Augmentation de :
   a) 24%
   b) 0,04%
   c) 350%
2) Diminution de :
   a) 45%
   b) 0,008%
   c) 87,01%

24 La meilleure promotion CORRIGÉ

Deux offres sont proposées pour une bouteille de lessive.
Laquelle est la plus intéressante pour le client ?

25 Récapitulons un peu CORRIGÉ

Compléter le tableau ci-dessous

Valeur de départ	Valeur d'arrivée	Coefficient multiplicateur	Evolution
185			Augmentation de 20%
	$22,95$		Diminution de 15%
1075		1,002	
240	180		
	38,01	0,21	

Evolutions successives

26 Dans tous les sens... CORRIGÉ

Dans chacun des cas suivants, déterminer le taux d'évolution global, arrondi à $0,01\%$ près si nécessaire.

1) Lors d'une augmentation de 45% puis d'une augmentation de 70%.
2) Lors d'une diminution de $13,4\%$ puis d'une augmentation de $5,8\%$.
3) Lors d'une diminution de $28,47\%$ puis d'une diminution de $37,61\%$.

27 Prenons la route CORRIGÉ

De juin à août, le temps perdu dans les embouteillages à Paris durant les heures de pointes diminue en moyenne de 80% puis augmente de 275% en septembre pour atteindre 15 secondes perdues par kilomètre parcouru (*Source : Etude V-Traffic*).

Combien de temps est perdu en moyenne par kilomètre par un automobiliste en juin ?

28 Toujours avec des voitures... CORRIGÉ

En 2015, il y a eu 17 268 immatriculations de voitures électriques. Ce nombre à augmenter de 44,26% en deux ans dont 25,96% la première année (*Source : Etude Fiches-auto.fr*).

Déterminer le taux d'évolution du nombre d'immatriculations de voitures électriques de 2016 à 2017. Arrondir à 0,01% près.

Evolutions réciproques

29 Pour démarrer CORRIGÉ

Dans chacun des cas suivants, déterminer le taux d'évolution réciproque, arrondi à $0,01\%$ près si nécessaire.

1) d'une augmentation de 21%
2) d'une augmentation de 150%
3) d'une diminution de 43%
4) d'une diminution de $4,7\%$.

30 Une très gentille commerçante CORRIGÉ

Après une augmentation de ses prix de 11,3%, puis de $5,7\%$, une commerçante souhaite récompenser un client fidèle en lui accordant une remise telle qu'elle compense ses deux dernières augmentations.

Déterminer le pourcentage de remise que doit effectuer la commerçante.

Statistiques

31 Pour démarrer

CORRIGÉ

On considère la série statistique suivante :

Valeur	1	3	4	5	10	13
Effectif	2	3	1	4	5	5

1) Déterminer :
   a) La médiane
   b) La moyenne
   c) L'étendue
   d) Le premier et le troisième quartile, ainsi que l'écart interquartile.
2) Représenter cette série sous la forme d'une boîte à moustache.

32 La lutte des classes

CORRIGÉ

Un groupe industriel comporte deux entreprises A et B. La répartition des salaires des employés suivant l'entreprise et leur catégorie professionnelle est donnée dans le tableau suivant :

Site A	Cadres	Ouvriers
Salaire moyen en €	1400	1000
Effectif	500	400
Site B	**Cadres**	**Ouvriers**
Salaire moyen en €	1500	1100
Effectif	200	800

1) Pour chaque catégorie (cadre et ouvrier), indiquer quelle entreprise « paie le mieux ».
2) Calculer la moyenne des salaires de chaque usine.
3) Quel paradoxe apparent y a-t-il dans les résultats des deux premières questions. Comment s'explique-t-il ?

33 Bien calculer sa moyenne

CORRIGÉ

Lison-Shan a relevé les notes qu'elle a obtenues cette année en mathématiques (les notes sont toutes sur 20) :

- Premier trimestre : 9 ;11 et 13 ;
- Deuxième trimestre : 12 ;15 ;14 et 16 ;
- Troisième trimestre : 10 et 8 ;

Lison-Shan avait prévenu ses parents que sa moyenne annuelle serait de 12. Malheureusement, à la réception de son bulletin, cette moyenne annuelle n'est que de 11,4.

1) Pourquoi Lison-Shan s'est-elle trompée ?
2) Comment expliquer cet écart ?

34 Brest *VS* Moscou

CORRIGÉ

Le tableau fournit les températures (en °C) moyennes mensuelles à Brest et à Moscou.

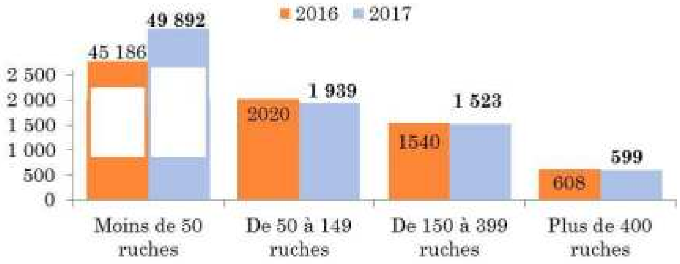
Mois	J	F	M	A	M	J	J	A	S	O	N	D
Brest	9,1	9,4	11	12,5	15,6	18,1	20,4	20,6	18,7	15,3	11,9	10
Moscou	-6,3	-4,2	1,5	10,4	18,4	21,7	23,1	21,5	15,4	8,2	1,1	-3,5

1) Pour chacune des séries, déterminer :
   a) La médiane
   b) La moyenne
   c) L'étendue
   d) Le premier et le troisième quartile, ainsi que l'écart interquartile
2) Représenter ces deux séries sous la forme d'une boîte à moustache.
3) Les affirmations suivantes sont-elles exactes ? Justifier.
   a) « Il fait toujours plus froid à Moscou qu'à Brest. »
   b) « En moyenne sur un an, il fait bien 6 degrés de plus à Brest qu'à Moscou. »
   c) « La température médiane de Brest dépasse de 4,6 degrés celle de Moscou. »

Vu au CRPE

35 **Groupement 5 - CRPE 2021** CRPE

Voici un graphique présentant le nombre d'apiculteurs en France métropolitaine :



Source : Observatoire de la production de miel et gelée royale FranceAgriMer 2018 d'après la déclaration de la DGAL. Extrait de « France-Agricole-FranceAgriMer-MIEL-2018-Observatoire miel et GR 2017 »

1) **a)** Donner le nombre d'apiculteurs qui possédaient entre 150 ruches et 399 ruches en France métropolitaine en 2017.

b) Donner le nombre total d'apiculteurs en 2017 en France métropolitaine.

c) Donner le nombre d'apiculteurs qui possédaient entre 50 ruches et 399 ruches en France métropolitaine en 2016.

2) **a)** Expliquer pourquoi la partie du diagramme concernant les apiculteurs possédant moins de 50 ruches n'est pas représenté de la même façon que les autres parties.

b) Pour les apiculteurs ayant moins de 50 ruches, le pourcentage d'augmentation entre 2016 et 2017 a été de 10,4 %. Calculer le nombre d'apiculteurs en 2016 dans cette catégorie.

3) Calculer le pourcentage d'évolution du nombre d'apiculteurs possédant au moins 150 ruches de 2016 à 2017, on arrondira le résultat au dixième d'unité de pourcentage.

36 **Groupement 4 - CRPE 2021** CRPE

Dans une classe de 28 élèves, on souhaite comparer les tailles des filles et des garçons. Voici les données dont on dispose :

- Taille des treize filles en centimètre :

149	155	161	142	167	163	157	150	165	152	161	159	160

- Taille des garçons : Toutes les tailles sont des nombres entiers de centimètres et il n'y a pas deux garçons qui ont la même taille.
- On connait également les indicateurs suivants : Étendue : 29 cm Moyenne : 159 cm Médiane : 161 cm

1) Calculer l'écart, en centimètre, entre la taille moyenne des filles et la taille moyenne des garçons.

2) Dans la classe, l'élève de plus petite taille mesure 140 cm. Quelle est la taille de l'élève le plus grand ?

3) Dans la classe, combien d'élèves mesurent 162 cm ou plus ?

4) Calculer la taille moyenne, en centimètre, arrondie au millimètre des élèves de cette classe.

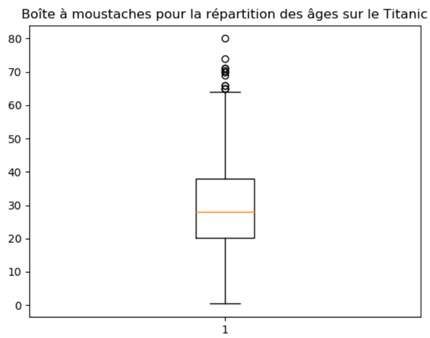
Les statistiques à grande échelle et Python

Sur Internet, il est possible de trouver le fichier *titanic.csv* (et plein d'autres fichiers de ce type !). Il s'agit d'un fichier recensant plusieurs informations sur les survivants et morts du Titanic : s'ils ont survécu ou non, leur nom, prénoms, âge, etc.
J'ai pu trouvé un fichier recensant 887 personnes, et j'ai voulu tracé une boîte à moustache me donnant la répartition des âges des personnes étant sur le Titanic.
A l'aide du module **pandas** qui permet de lire des fichiers *.csv* (plus précisément, n'importe quelles informations se trouvant dedans !) et du module **matplotlib**, il possible de réaliser un tel diagramme sans faire le relevé soi-même à la main ! Voici le script et son résultat :

```
import matplotlib.pyplot as plt
import pandas

data=pandas.read_csv('titanic.csv',sep=',')
agetitanic=data.Age

plt.boxplot(agetitanic)
plt.title("Boite a moustaches pour la repartition des ages sur le Titanic")
plt.show()
```



Le *gerrymandering* ou comment « truquer » des élections...

Le *gerrymandering* est un terme politique nord-américain pour désigner le découpage des circonscriptions électorales ayant pour objectif de donner l'avantage à un parti, un candidat, ou un groupe donné.