## I) Vocabulaire et définitions

**Une étude statistique** concerne une **population** (l'ensemble des éléments étudiés) et elle renseigne sur un **caractère** (ou variable) de cette population. **Les individus** sont les éléments de cette population.

Les valeurs prises par le caractère sont aussi appelées **les modalités**.

Le caractère étudié est dit **quantitatif** s'il prend des valeurs numériques (ex : âge, prix, notes….) et **qualitatif** si on ne peut pas le mesurer (ex : nationalité, couleur …).

*On distingue les caractères quantitatifs discrets qui ne peuvent prendre qu'un nombre fini de valeurs (notes…) et les caractères quantitatifs continus dont on regroupe les valeurs par intervalles ou par classes (tailles, durée…)*

| Valeurs du caractère $x_i$ | $x_1$ | $x_2$ | $x_3$ | … | … | $x_p$ | Total |
|---|---|---|---|---|---|---|---|
| Effectifs $n_i$ | $n_1$ | $n_2$ | $n_3$ | … | … | $n_p$ | $N$ |
| Effectif cumulé croissant **ECC** | $n_1$ | $n_2+n_1$ | $n_3+n_2+n_1$ | … | … | $N$ | |

Le nombre d'individus ($n_i$) d'une valeur est appelé **l'effectif**.

Le nombre total d'individus ($N$) de la population est appelé **l'effectif total**.

On appelle **l'effectif cumulé croissant** d'une valeur la somme des effectifs des valeurs inférieures ou égales à cette valeur. *(On écrira ECC en abrégé).*

**La fréquence $f_i$ d'une valeur est le rapport $\frac{n_i}{N}$**

$f_i$ est un nombre toujours compris entre 0 et 1 et souvent exprimé par un pourcentage.

**La somme des nombres $f_i$ est toujours égale à 1.**

On appelle **la fréquence cumulée croissante** d'une valeur la somme des fréquences des valeurs inférieures ou égales à cette valeur. *(On écrira FCC en abrégé).*

***Remarque :***

On définit de la même manière l'effectif cumulé décroissant (ECD ), ainsi que la fréquence cumulée décroissante. (FCD).

***Les deux exemples qui vont nous servir dans la construction du cours***

***Exercice A :*** Les notes sur 20 obtenues lors d'un devoir de mathématiques dans une classe de 2^de^ sont les suivantes : 10 ;8 ;11 ;9 ;12 ;10 ;8 ;10 ;7 ;9 ;10 ;11 ;12 ;10 ;8 ;9 ;10 ;9 ;10 ;11.

Population étudiée : *Clas se*     les élèves d'une classe     Individus :

Effectif total : *20*     les notes obtenues     Caractère *:*

- Série statistique définie par les effectifs :

| Valeurs du caractère (notes) $x_i$ | 7 | 8 | 9 | 10 | 11 | 12 | Total |
|---|---|---|---|---|---|---|---|
| Effectifs (Nbre d'élèves ayant la note) $n_i$ | *1* | *3* | *4* | *7* | *3* | *2* | 20 |
| Effectif cumulé croissant ***(ECC)*** | *11* | *4* | *8* | *15* | *18* | *20* | |

Combien d'élèves ont eu une note inférieure ou égale à 9 ? ..............8..................................

- Série statistique définie par les fréquences en pourcentage :

| Valeurs du caractère (notes) $x_i$ | 7 | 8 | 9 | 10 | 11 | 12 | Total |
|---|---|---|---|---|---|---|---|
| Fréquences en % $f_i$ | *5%* | *15%* | *20%* | *35%* | *15%* | *10%* | |
| Fréquences cumulées croissantes ***(FCC)*** | *5%* | *20%* | *40%* | *75%* | *90%* | *100* | |

***Exercice B :*** On donne le temps passé devant la télévision par 34 élèves d'une classe pendant une journée.

Population étudiée : *Classe* Individus :

Effectif total : *34* caractère :

- Série statistique définie par des classes (ou intervalles) d'amplitudes différentes :

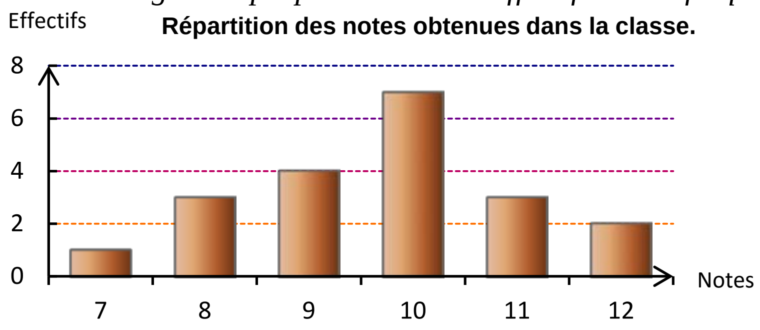
| Temps en minutes | [0 ; 15 [ | [15 ; 30 [ | [30 ; 60[ | [60 ; 120[ | [120 ; 180[ | Total |
|---|---|---|---|---|---|---|
| Nombre d'élèves | 7 | 5 | 8 | 10 | 4 | |
| Fréquences en % | *21%* | *15%* | *23%* | *29%* | *12%* | |
| **FCC** | *21%* | *36%* | *59%* | *88%* | *100%* | |
| **FCD** | *100%* | *79%* | *64%* | *41%* | *12%* | |

## ***II) Représentations graphiques***

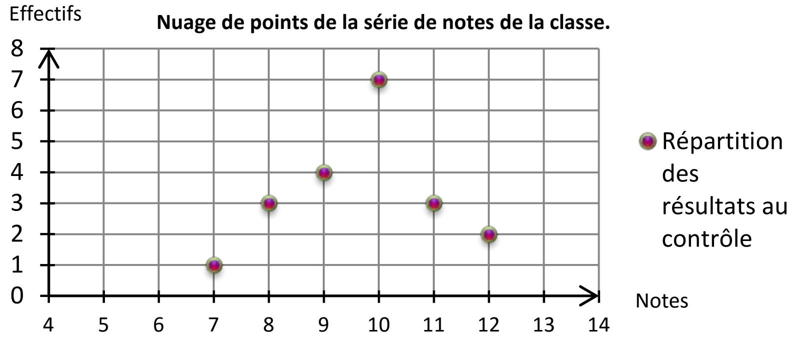
Selon le type de caractère, on utilise différentes représentations graphiques.

Pour la représentation graphique d'un caractère qualitatif ou d'un caractère quantitatif discret, on utilise le **diagramme en bâtons** *(On trace des segments proportionnels à l'effectif ou à la fréquence.)*

***Avec l'Exercice A :***

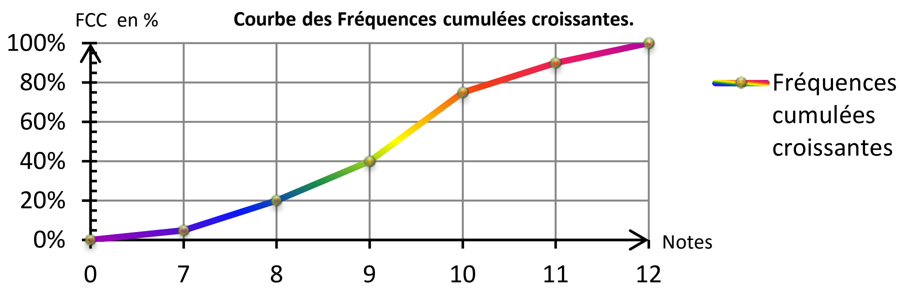


ou **le nuage de points**

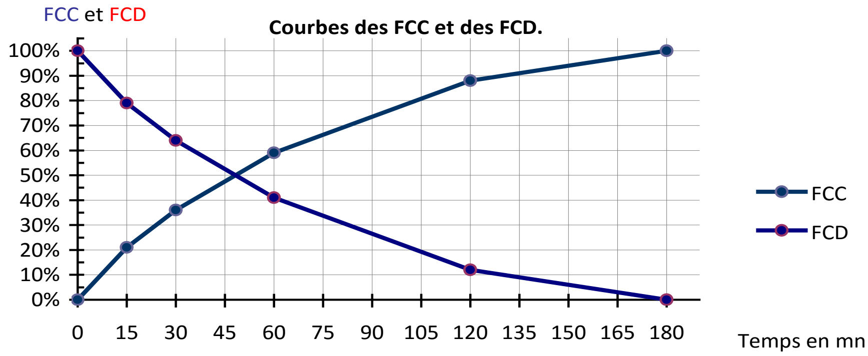



Pour la représentation graphique de tout type de caractère, on peut utiliser la courbe des fréquences cumulées croissantes ou décroissantes.

***Avec l'Exercice A :***



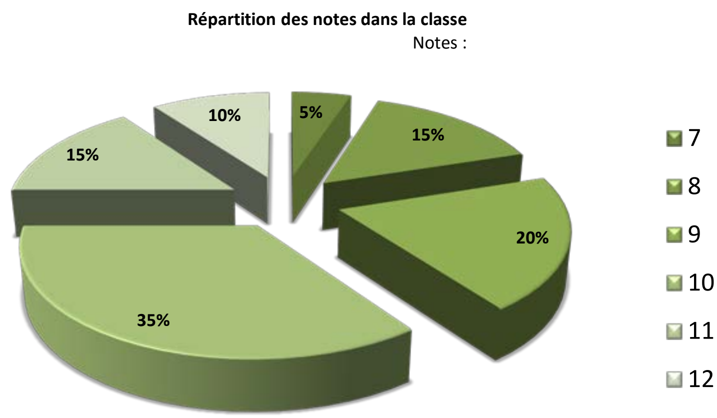

***Avec l'Exercice B :***




Pour la représentation graphique de tout type de caractère, on peut utiliser un **diagramme circulaire.**

***Avec l'Exercice A :***

| Notes | 7 | 8 | 9 | 10 | 11 | 12 | Total |
|---|---|---|---|---|---|---|---|
| Fréquences en % | *5%* | *15%* | *20%* | *35%* | *15%* | *10%* | *100%* |
| Angle du secteur en ° | *18°* | *54°* | *72°* | *126°* | *54°* | *36°* | *360°* |



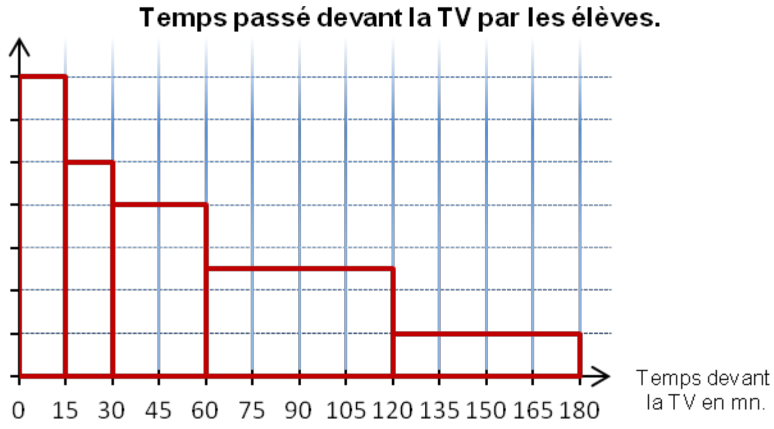
Pour la représentation graphique d'un caractère quantitatif continu, on utilise un **histogramme**.

*Dans un repère orthogonal on porte en abscisse les valeurs des bornes des intervalles, puis pour chaque intervalle on trace un rectangle dont* ***l'aire*** *est proportionnelle à l'effectif ou à la fréquence.*

*(En pratique, il est conseillé de commencer par construire un nouveau tableau donnant la largeur et l'aire de chaque rectangle. On peut alors facilement en déduire la hauteur.)*

***Avec l'Exercice B :*** Unité choisie : 1cm² représente un élève.

| Temps en minutes | [0 ; 15 [ | [15 ; 30 [ | [30 ; 60[ | [60 ; 120[ | [120 ; 180[ |
|---|---|---|---|---|---|
| Aire du rectangle en cm² | *7* | *5* | *8* | *10* | *4* |
| Amplitude de la classe | *15* | *15* | *30* | *60* | *60* |
| Largeur du rectangle en cm | *1* | *1* | *2* | *4* | *4* |
| Hauteur du rectangle en cm | *7* | *5* | *4* | *2,5* | *1* |




***Remarque :***

Dans la plupart des cas, les classes ont la même amplitude donc les rectangles ont la même largeur donc l'aire de chaque rectangle est proportionnelle à sa hauteur (donc à l'effectif).

*Une série statistique peut contenir de très nombreuses données. Il est donc nécessaire de trouver une façon de résumer ces données. On différencie les indicateurs de position et de dispersion.*

## III) Indicateurs de position

### a) La moyenne

**Définition**

**On considère une série statistique à caractère quantitatif discret dont les valeurs du caractère sont $x_1$, $x_2$, $x_3$, ..., $x_p$ et les effectifs associés sont $n_1, n_2, n_3, ..., n_p$ . $N$ représente l'effectif total.**

**La moyenne d'une série statistique se note $\overline{x}$ et $\overline{x} = \dfrac{n_1 x_1 + n_2 x_2 + n_3 x_3 + ... + n_p x_p}{N}$.**

*(On la note encore* $\overline{x} = \dfrac{\sum_{i=1}^{p} n_i x_i}{N}$ *. On lit « La somme de i =1 à p des $n_i$ $x_i$ le tout divisé par N ».)*

**Exemple :** Calculer la moyenne de la série statistique de **l'Exercice A.**

$$\overline{x} = \frac{1\times7 + 3\times8 + 4\times9 + 7\times10 + 3\times11 + 2\times12}{1 + 3 + 4 + 7 + 3 + 2} = \frac{194}{20} = 9{,}7$$ La moyenne est de 9,7.

On peut calculer la moyenne à partir des fréquences

**Propriété : Si on note $f_i = \dfrac{n_i}{N}$ la fréquence de la valeur $x_i$ alors : $\overline{x} = f_1 x_1 + f_2 x_2 + f_3 x_3 + ..... + f_p x_p$**

*(On peut aussi noter* $\sum_{i=1}^{p} f_i x_i$ *)*

**Preuve :** $\overline{x} = \dfrac{n_1 x_1 + n_2 x_2 + n_3 x_3 + ... + n_p x_p}{N}$

$$\overline{x} = \frac{n_1 x_1}{N} + \frac{n_2 x_2}{N} + \frac{n_3 x_3}{N} + \ldots + \frac{n_p x_p}{N}$$

$$\overline{x} = \frac{n_1}{N} x_1 + \frac{n_2}{N} x_2 + \frac{n_3}{N} x_3 + \ldots\ldots + \frac{n_p}{N} x_p = f_1 x_1 + f_2 x_2 + f_3 x_3 + \ldots.. + f_p x_p$$

**Exemple :** Calculer la moyenne de la série statistique de **l'Exercice A.**

$$\overline{x} = 0{,}05\times 7 + 0{,}15\times 8 + 0{,}2\times 9 + 0{,}35\times 10 + 0{,}15\times 11 + 0{,}1\times 12 = 9{,}7$$

**Remarque :** Pour calculer la moyenne pour un caractère quantitatif continu, on se ramène aux formules précédentes en remplaçant chaque classe par **son centre**.

*Le centre d'une classe [a ; b[ est le nombre* $\dfrac{a + b}{2}$ .

*(On peut réécrire la définition et la propriété précédentes : On considère une série statistique à caractère* *quantitatif continu dont les centres des classes sont* $c_1, c_2, c_3, \ldots,$ *et les effectifs associés sont :* $n_1, n_2, n_3, \ldots$

*la moyenne est :* $\overline{x} = \dfrac{n_1 c_1 + n_2 c_2 + c_3 x_3 + \ldots + c_p x_p}{N}$ *, avec N l'effectif total.*

*On la note encore* $\overline{x} = \dfrac{\sum\limits_{i=1}^{p} n_i c_i}{N}$. *Avec les fréquences :* $\overline{x} = \sum\limits_{i=1}^{p} f_i c_i$ *.)*

***Exemple :*** Calculer la moyenne de la série statistique de ***l'ExerciceB.***

| Temps en minutes | [0 ; 15 [ | [15 ; 30 [ | [30 ; 60[ | [60 ; 120[ | [120 ; 180[ | Total |
|---|---|---|---|---|---|---|
| Effectif | 7 | 5 | 8 | 10 | 4 | 34 |
| Centre des classes | 7,5 | 22,5 | 45 | 90 | 150 | |

Donc $\overline{x} = \dfrac{7\times7{,}5 + 5\times22{,}5 + 8\times45 + 10\times90 + 4\times150}{34} = \dfrac{2025}{34} \approx 59{,}55$

Les élèves regardent la TV en moyenne 59,56 min *(soit environ 59 min et 34 s par jour)*

## b) La médiane

***Définition : La médiane est le nombre noté Me tel que 50% au moins des individus ont une valeur du caractère inférieure ou égale à Me et 50% au moins des individus ont une valeur du caractère supérieure ou égale à Me.***

Pour un caractère quantitatif discret, on range les valeurs de la série dans l'ordre croissant et on peut utiliser le tableau des effectifs cumulés croissants pour la trouver.

- Si l'effectif total est de taille $2n+1$ (impair), *Me* est la valeur de rang $n + 1$ *(située au milieu),*

- Si l'effectif total est de taille $2n$ (pair), *Me* est la demi somme des termes de rangs $n$ et $n+1$ *(la moyenne des 2 valeurs situées au milieu).*

***Exemple :*** Déterminer la médiane de la série statistique de ***l'Exercice A.***

La médiane partage les 20 notes en deux groupes de 10. Elle se trouve entre la 10$^{\text{ième}}$ et 11$^{\text{ème}}$ valeur, donc entre 10 et 10. Donc la médiane est 10.

***Exemple :*** Voici l'âge des salariés dans une entreprise 30-28-47-30-44-50-60-26-29-37-30-29-58-59-28

Pour déterminer la médiane, les valeurs doivent être ordonnées :

26-28-28-29-29-30-30-30-37-44-47-50-58-59-60

La médiane partage les 15 valeurs en deux groupes de 7. La médiane est la 8$^{\text{ième}}$ valeur.    Donc Me = 30

***Remarque :*** Pour un caractère quantitatif continu, on utilise les fréquences cumulées. La médiane correspond à la valeur du caractère ayant une FCC (ou FCD) de 0,5.

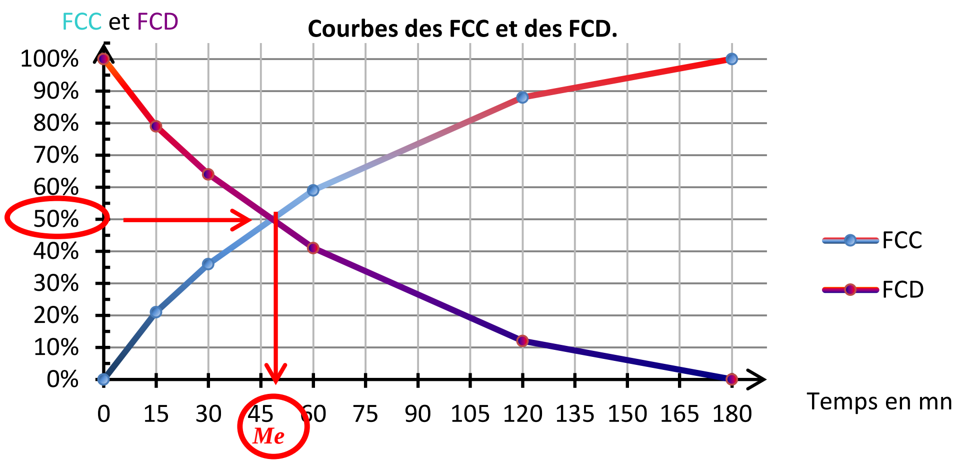
*(Ou la moitié de l'effectif total pour un ECC ou ECD).*

Pour déterminer une valeur de façon plus précise, on utilise la courbe des fréquences cumulées croissantes *(ou celles des effectifs cumulés croissantes, ou encore les FCD ou les ECD).*

***Exemple :*** Dans ***l'ExerciceB*** déterminer dans quelle classe se trouve la médiane.

Puis déterminer graphiquement une valeur de la médiane de cette série statistique.

0251662336




En traçant un trait horizontal à l'ordonnée 50% (ou en ayant tracé simultanément les courbes des FCC et FCD), on lit graphiquement une valeur approchée de la médiane *Me*.

Ici, on lit : $Me \approx 48$.

## c) Quartiles

***Définition :***

• Le premier quartile se note $Q_1$ . C'est la plus petite valeur de la série telle qu'au moins un quart des valeurs (ou 25%) lui soient inférieures ou égales.

• Le troisième quartile se note $Q_3$ . C'est la plus petite valeur de la série telle qu'au moins trois quarts (ou 75%) des valeurs lui soient inférieures ou égales.

***Remarque :*** $Q_1$ et $Q_3$ sont des valeurs prises par la série, mais *Me* pas toujours.

Pour un caractère quantitatif discret, on range les valeurs de la série dans l'ordre croissant et on commence par chercher le RANG du quartile et on peut utiliser le tableau des effectifs cumulés croissants.

- On calcule $\frac{N}{4}$ et on prend l'entier supérieur ou égal au résultat qui correspondra au rang (à la position) de la valeur cherchée, $Q_1$.

- On calcule $\frac{3N}{4}$ et on prend l'entier supérieur ou égal au résultat qui correspondra au rang (à la position) de la valeur cherchée, $Q_3$ .

***Exemple :*** Dans ***l'Exercice A***, il y a 20 valeurs. Pour déterminer $Q_1$ , on cherche d'abord son rang.

Rang de $Q_1 = \frac{20}{4} = 5$ C'est la 5ème valeur, donc $Q_1 = 9$ .

Rang de $Q_3 = \frac{3\times 20}{4} = 15$ C'est la 15ème valeur, donc $Q_3 = 10$ .

***Remarque :*** Dans le cas d'un caractère continu, on peut d'abord déterminer dans quelles classes se trouvent $Q_1$ et $Q_3$.

Pour plus de précision on utilise la courbe des FCC.

Cela revient à faire une lecture graphique en partant du quart et/ou de trois-quarts de l'effectif sur l'axe des ordonnées, ou encore de 25% et/ou de 75% pour les fréquences sur l'axe des ordonnées.

***Exemple :*** Lire graphiquement les valeurs de $Q_1$ et de $Q_3$ sur le graphique précèdent.

Ici, on lit : $Q_1 \approx 19$ et $Q_3 \approx 93$

## ***IV) Indicateurs de dispersion***

Ils permettent de mesurer la façon dont les valeurs sont réparties autour de la médiane et de la moyenne.

**L'Etendue : C'est la différence entre la plus grande et la plus petite valeur.**

**L'Ecart interquartile : $Q_3 - Q_1$**

***Exemple :*** Dans ***l'Exercice A***, l'étendue est $12 - 7 = 5$

L'écart interquartile est : $Q_3 - Q_1 = 10 - 9 = 1$.

L'intervalle interquartile est : [9 ; 10]

## ***V) Résumé d'une série statistique***

On peut résumer les indicateurs par un diagramme en boîte : il s'agit d' un graphique permettant de résumer un caractère quantitatif (discret ou continu) par ses valeurs extrêmes et ses quartiles rapidement.

① Une droite graduée : La graduation va de la valeur minimale à la valeur maximale de la série.

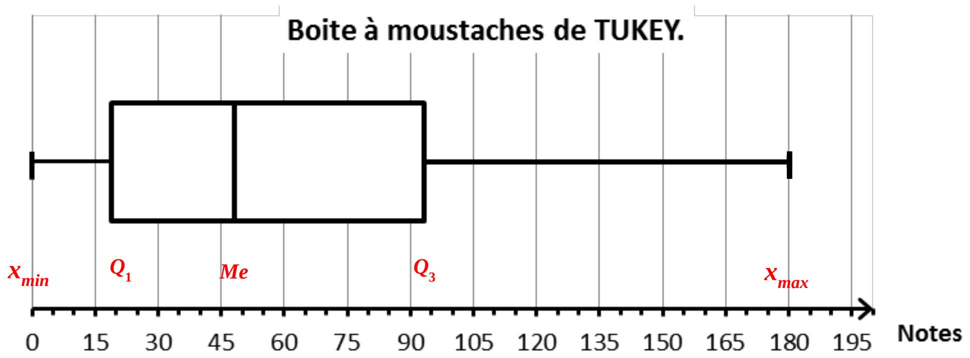
② L'interquartile (défini par $Q_3 - Q_1$): il donne les "dimensions" de la boîte, où $Q_1$ et $Q_3$ sont respectivement le premier et le troisième quartile.

③ La médiane : On partage ce rectangle par un segment vertical en deux parties disjointes au niveau de la médiane *Me*. (La médiane peut être égale à $Q_1$ ou $Q_3$).

④ Les extrémités ou "moustaches" : Ce sont les extrémités gauche et droite du graphique.

L'extrémité gauche est définie par la valeur minimale de la série

L'extrémité droite est définie par la valeur maximale de la série.




2516664322516654082516643842516633602516623336

***Exemple***

Pour la série **3 ; 3 ; 5 ; 6 ; 7 ; 10 ; 10 ; 11 ; 12 ; 12 ; 13 ; 17** :