

Chapitre 9

Statistiques

I. Vocabulaire

Population : Une population est un ensemble de personnes ou d'objets, appelés **individus**, définis par une propriété commune.

Exemple : les habitants d'un pays, les automobiles fabriquées en 2010,...

Caractère : Pour une population choisie, on peut étudier un caractère de ses individus.

Exemple : on peut étudier le caractère « taille » des élèves d'un lycée.

Caractère quantitatif : Un caractère est dit quantitatif lorsqu'il est possible de le mesurer en associant un nombre à chaque individu. Un caractère quantitatif est aussi appelé **variable**.

Exemple : l'âge, la taille, le nombre de frères et sœurs, ...

- Un caractère quantitatif est dit **continu** lorsque les nombres qui le mesurent peuvent prendre, à priori, toutes les valeurs d'un intervalle.

Exemple : le poids, la taille, la durée de vie d'un moteur, ...

- Il est **discret** dans le cas contraire.

Exemple : l'année de naissance, le nombre d'enfants par famille, ...

Caractère qualitatif : On appelle ainsi tout caractère non quantitatif.

Exemple : la couleur des yeux.

Exemple :

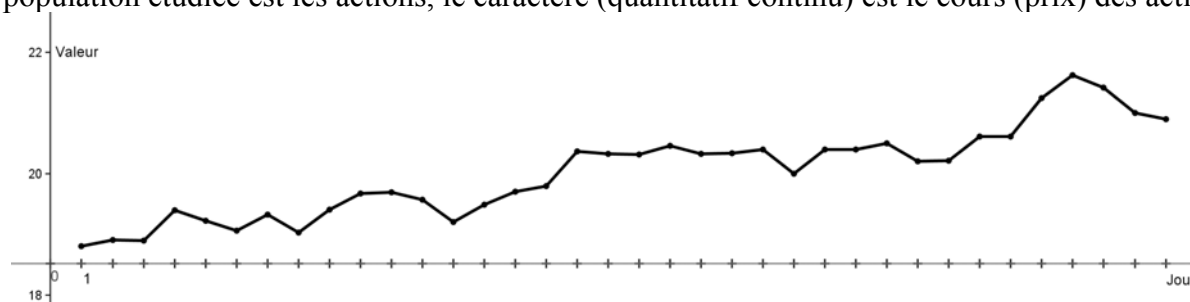
Le tableau suivant donne le relevé des cours d'une action sur 36 jours consécutifs de bourse.

jour	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
cours (en €)	18,8	18,91	18,9	19,4	19,22	19,06	19,33	19,03	19,41	19,67	19,69	19,57	19,2	19,49	19,7	19,8	20,37	20,33

jour	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
cours (en €)	20,32	20,46	20,33	20,34	20,4	20	20,4	20,4	20,5	20,2	20,21	20,61	20,61	21,25	21,63	21,42	21	20,9

Pour cette série statistique :

La population étudiée est les actions, le caractère (quantitatif continu) est le cours (prix) des actions.



II. Paramètres d'une série

Pour résumer une **série quantitative**, on utilise des **caractéristiques de position** (moyenne, médiane, quartiles, ...) et des **caractéristiques de dispersion** (étendue, écart interquartile, ...)

1) Caractéristiques de position

Moyenne

Définition :

Pour déterminer la **moyenne** d'une série prenant les valeurs x_1, x_2, \dots, x_n on utilise la formule :

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i$$

Exemple :

Pour la série étudiée, on a donc :

$$\bar{x} = \frac{18,8 + 18,91 + 19,9 + \dots + 21,42 + 21 + 20,9}{36} \simeq 20,02$$

Sur les 36 derniers jours le cours moyen de l'action est de 20,02 €.

Médiane et quartiles

Définitions :

- La **médiane** d'une série statistique est le nombre **Me** tel que :
 - 50% au moins des individus ont une valeur du caractère inférieure ou égale à **Me**.
 - 50% au moins des individus ont une valeur supérieure ou égale à **Me**.
- Le **premier quartile** est la plus petite valeur **Q₁** de la liste telle qu'au moins **un quart** des valeurs de la liste sont **inférieures ou égales à Q₁**.
- Le **troisième quartile** est la plus petite valeur **Q₃** de la liste telle qu'au moins les **trois quarts** des valeurs de la liste sont **inférieures ou égales à Q₃**.

Exemple :

Pour la série de l'exemple précédent :

18,8 18,9 18,91 19,03 19,06 19,2 19,22 19,33 19,4 19,41 19,49 19,57 19,67 19,69 19,7 19,8 20 20,2
20,21 20,32 20,33 20,33 20,34 20,37 20,4 20,4 20,4 20,46 20,5 20,61 20,61 20,9 21 21,25 21,42 21,63

- Donc la médiane vaut :

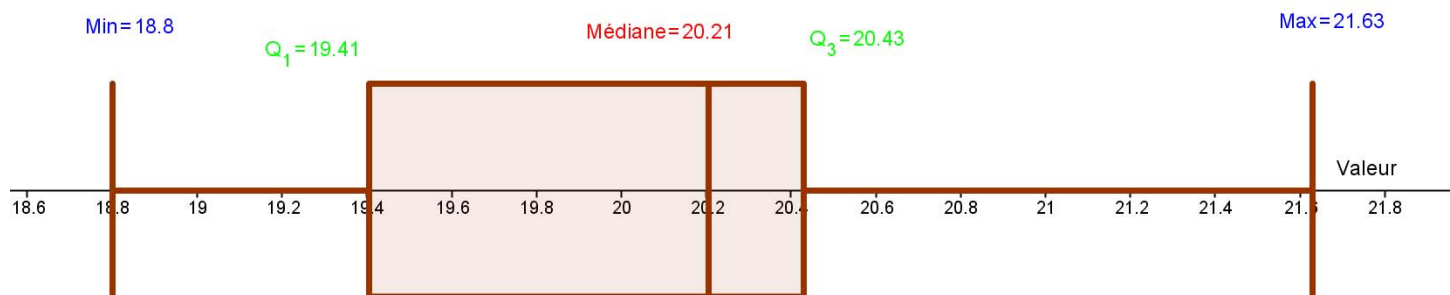
$$Me = \frac{20,2 + 20,21}{2} = 20,205$$

donc le cours de l'action était inférieur à 20,205 € durant 18 jours.

- **Q₁** : $\frac{36}{4} = 9$
donc le 1er quartile vaut 19,4 €.
- **Q₃** : $9 \times 3 = 27$
donc le 3ème quartile vaut 20,4 €.

Définition :

Un **diagramme en boîte** (ou **boîte à moustaches**) est une représentation graphique qui résume le caractère quantitatif étudié par les valeurs extrêmes, la médiane, les quartiles.

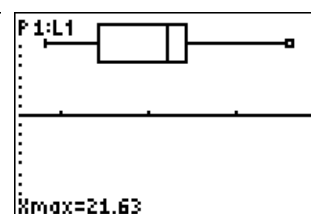
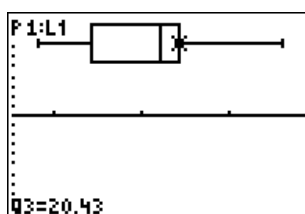
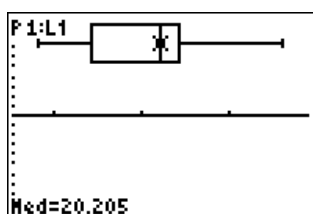
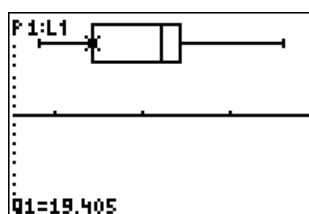
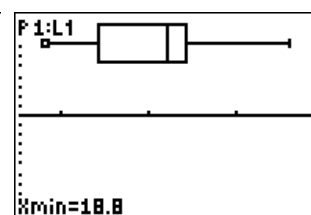


Calculatrice :

L1	L2	L3	1
18.91			
18.9			
19.4			
19.22			
19.06			
19.33			
L1(1)=18.8			

```
moyenne(L1)
20.02388889
médiane(L1)
20.205
```

```
Graph1 Graph2 Graph3
Type: L1 NAff
ListeX:L1
Effectifs:1
```



Remarque :

Les modes de calcul des quartiles diffèrent d'un logiciel à l'autre.

De ce fait, les résultats obtenus seront souvent différents des résultats attendus.

Pour la calculatrice, Q_1 est la médiane des valeurs comprises entre Xmin et Med et Q_3 est la médiane des valeurs comprises entre Med et Xmax.

2) Caractéristiques de dispersion

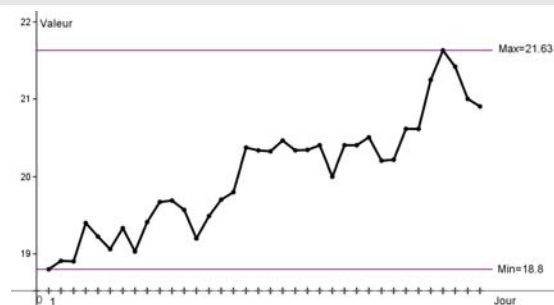
Étendue

Définition :

L'**étendue** est la différence entre les valeurs extrêmes de la série.

Exemple :

Pour la série étudiée l'étendue est 2,83 €.



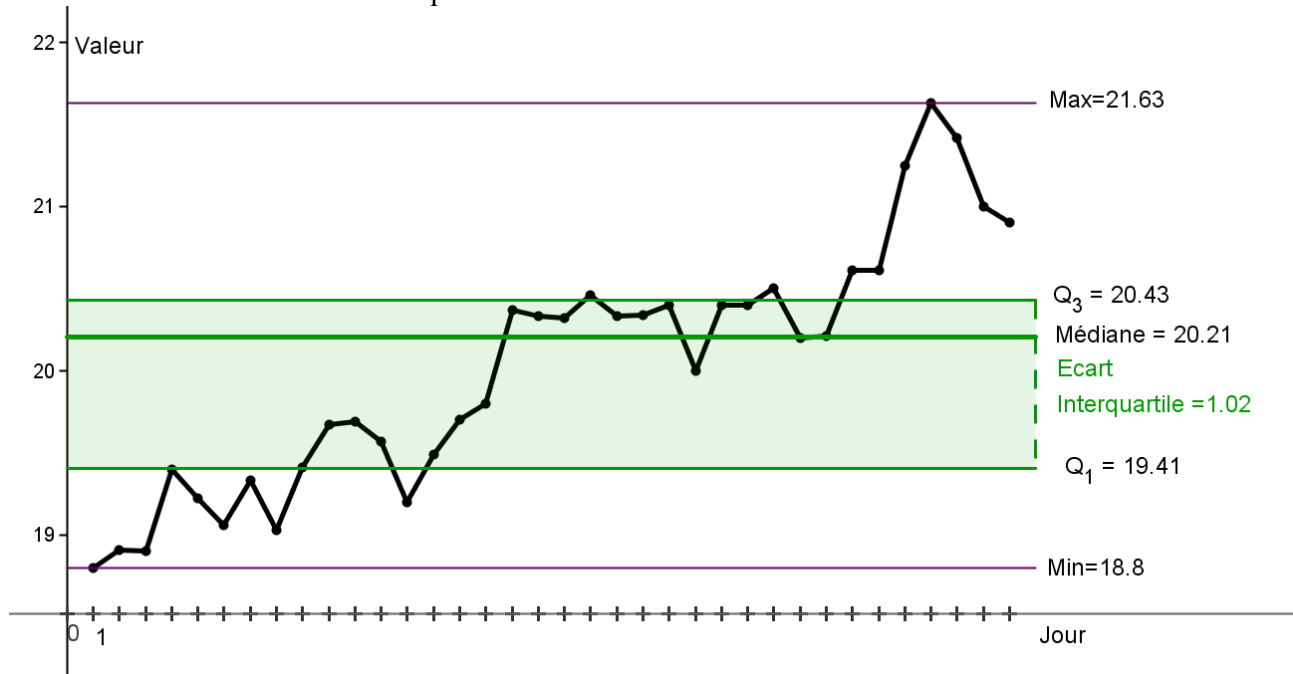
Écart interquartile

Définition :

L'**écart interquartile** est la différence entre les 3^{ème} et 1^{er} quartiles ($Q_3 - Q_1$).

Exemple :

Pour la série étudiée l'écart interquartile est 1 €.



Remarque :

L'écart interquartile ne prend pas en compte les valeurs extérieures à l'intervalle interquartile. Pour prendre ces valeurs en compte, on utilise une autre mesure de dispersion : l'**écart type**.

Écart type

Définitions :

- La **variance V** de la série prenant les valeurs x_1, x_2, \dots, x_n et de moyenne \bar{x} est définie par :

$$V = \frac{1}{n} \left[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2 \right] = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

- L'**écart type s** est alors défini comme la racine carrée de la variance :

$$s = \sqrt{V}$$

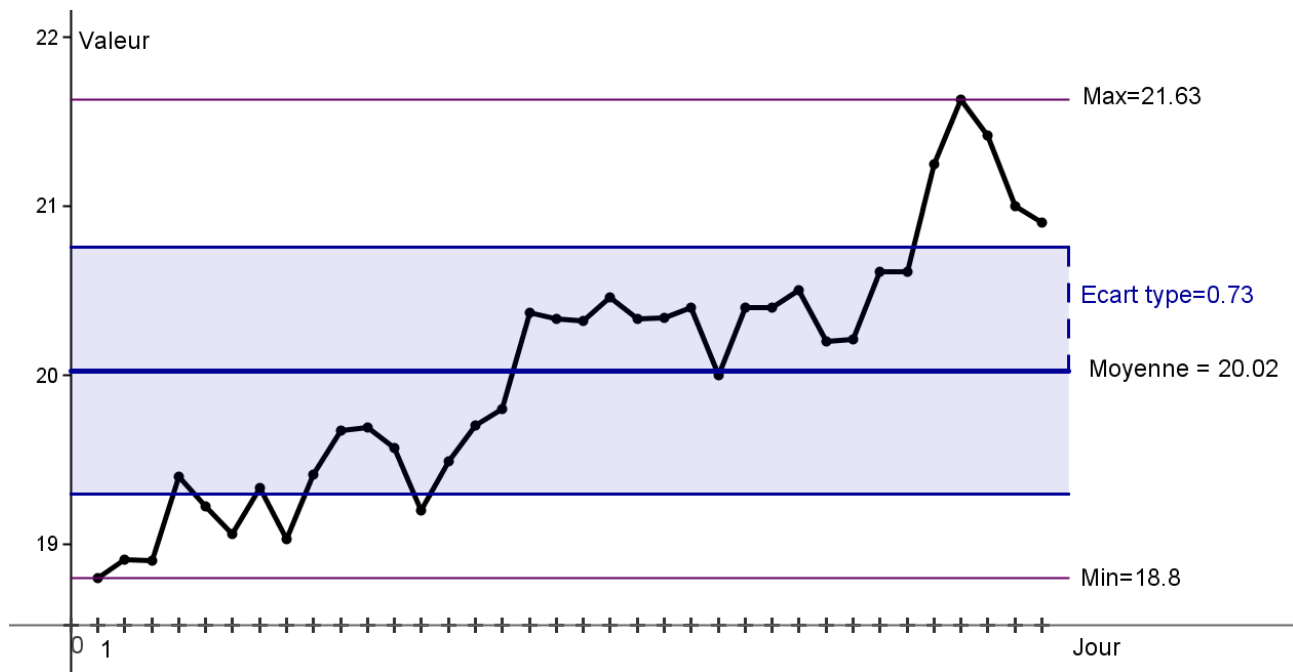
Exemple :

Pour la série étudiée on a :

$$V = \frac{(18,8 - 20,02)^2 + (18,91 - 20,02)^2 + \dots + (21 - 20,02)^2 + (20,9 - 20,02)^2}{36} \simeq 0,549$$

et $s = \sqrt{V} \simeq 0,74$.

Donc l'écart type est de 0,74 €.



Remarques :

- L'**écart type** est une quantité réelle positive qui mesure la dispersion autour de la moyenne.
- Pratiquement, c'est l'**écart type**, racine carrée de la variance, qui est utilisé car il possède les mêmes dimensions que le caractère.

Calculatrice :

L1	L2	L3	1
18.91	-----	-----	
18.9			
19.4			
19.22			
19.06			
19.33			
L1()=18.8			

```

variance(L1)
.5490244444
ecart-type(L1)
.7409618374

```

Pour obtenir un résumé statistique :

L1	L2	L3	1
18.91	-----	-----	
18.9			
19.4			
19.22			
19.06			
19.33			
L1()=18.8			

```

Stats 1-Var
x̄=20.02388889
Σx=720.86
Σx²=14453.6364
Sx=.7409618374
σx=.7305982244
↓n=36

```

```

Stats 1-Var
↑n=36
minX=18.8
Q1=19.405
Med=20.205
Q3=20.43
maxX=21.63

```

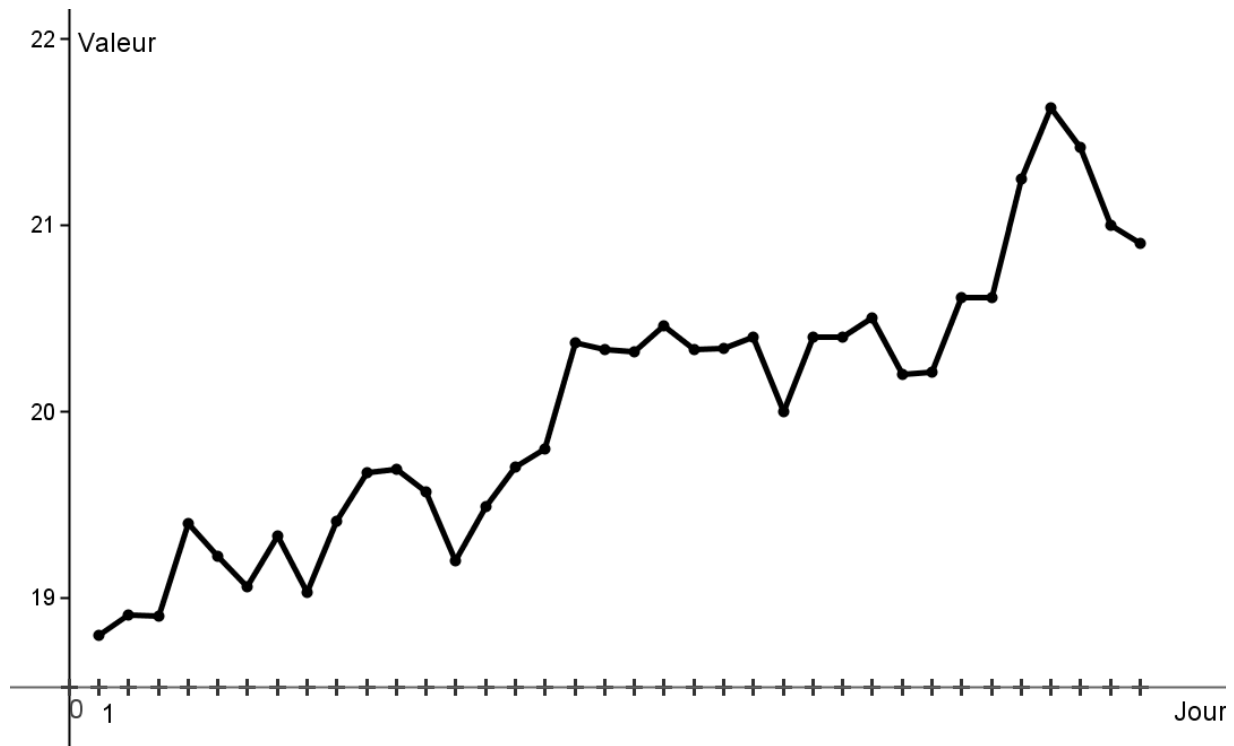
III. Série chronologique

Définition :

On appelle **série chronologique** une série statistique telle que les valeurs de la série sont observées à des intervalles de temps égaux (heures, mois, trimestre, années,...).

Exemple :

La série étudiée est donc bien une série chronologique.



Lissage par moyennes mobiles

Définition :

On appelle **moyenne mobile centrée d'ordre k** , à la date i , la moyenne arithmétique de l'observation x_i et des $2k$ observations qui l'encadrent.

Exemple :

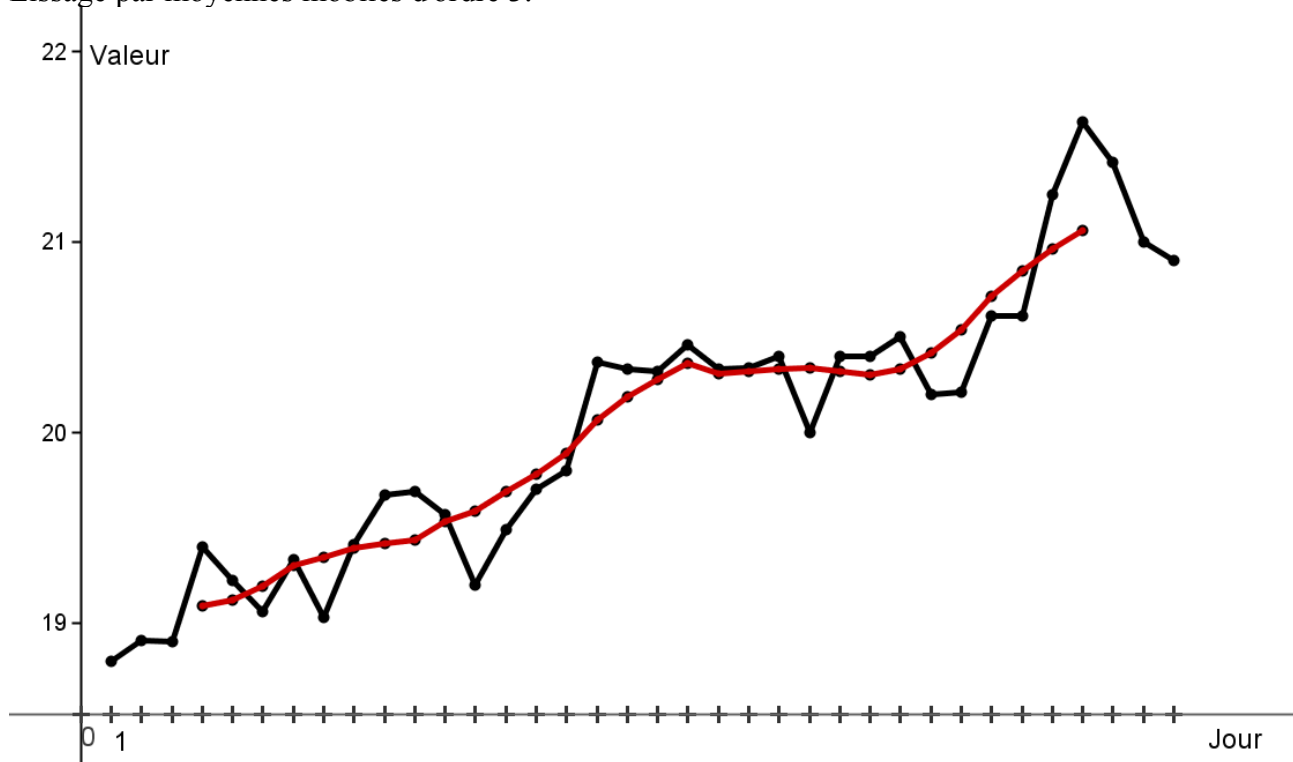
Pour la série étudiée, la série des moyennes mobiles d'ordre 3 est :

jour	1	2	3	4	5	6	...
cours (en €)				$\frac{18,8+18,91+18,9+19,4+19,22+19,06+19,33}{7}$ $\approx 19,09$	$\frac{18,91+18,9+19,4+19,22+19,06+19,33+19,03}{7}$ $\approx 19,12$	$\frac{18,9+19,4+19,22+19,06+19,33+19,03+19,41}{7}$ $\approx 19,19$...

jour	...	31	32	33	34	35	36
cours (en €)	...	$\frac{20,2+20,21+20,61+20,61+21,25+21,63+21,42}{7}$ $\approx 20,84$	$\frac{20,21+20,61+20,61+21,25+21,63+21,42+21,06}{7}$ $\approx 20,96$	$\frac{20,61+20,61+21,25+21,63+21,42+21,06+20,9}{7}$ $\approx 21,06$			

Exemple :

Lissage par moyennes mobiles d'ordre 3.



Remarque :

La série des moyennes mobiles permet de lisser la série chronologique initiale, en « gommant » les irrégularités.