

Statistique Descriptive

Professeur : Jean-Louis MONINO

Contact : monino@wanadoo.fr ou jean-louis.monino@univ-montp1.fr

Toute recommandation par mail devra être accompagnée d'un objet pour qu'elle puisse être lue

Bibliographie optionnel : *Statistique descriptive*, Jean-Louis MONINO, Edition DUNOD

<http://www.dunod.com/livre-dunod-9782100513239-td-statistique-descriptive.html>

Le cours de Statistique descriptive sur PORTICES (**Attention**, il ne s'agit pas des diapos que J-L Monino diffuse en cours !!):

Chapitre 1 : <http://www.portices.fr/StatistiqueC1>

Chapitre 2 : <http://www.portices.fr/StatistiqueC2>

Chapitre 3 : <http://www.portices.fr/StatistiqueC3>

Chapitre 4 : <http://www.portices.fr/StatistiqueC4>

Chapitre 5 : <http://www.portices.fr/StatistiqueC5>

Chapitre I

Le vocabulaire

Un peu d'histoire ...

En France, on dénombre la population à un peu plus de 65.000.000 d'habitants. De tout temps, un problème a été soulevé : Comment calculer la force d'une nation ?

Les premières méthodes consistaient à compter un par un le nombre d'hommes valides, donc pour évaluer en quelque sorte, la force de travail disponible. Il n'était pas encore impossible d'effectuer ce genre de recensement car la population de l'époque n'était pas aussi conséquente que celle d'aujourd'hui. L'idée fut prise aussi de calculer la production de chacun et de leur revenu...

Cependant avec le temps, il fut de plus en plus compliqué d'évaluer au cas par cas chaque individu. Cette contrainte à pousser les statisticiens à changer de méthodes. Les écoles Française, Anglaise et Russe ont démontré que connaître la vie de chacun n'était pas si pertinent. Ils ont mis en place le sondage partiel.

A. Définitions et Introduction à la Statistique

La statistique recouvre un ensemble de méthodes qui contribuent à la connaissance, sous formes de résumés qui sont souvent numériques, d'un ensemble analysable que l'on doit étudier.

Le terme statistique a un double sens en français :

- Il désigne la totalité des données numériques (par exemple : Les revenus des agriculteurs français) ou non numérique dans un ensemble.
- Il désigne également un ensemble cohérent de méthodes scientifiques (par exemple : le calcul du revenu moyen des agriculteurs français) qui permettent de rassembler l'ensemble des données numériques et non numériques.

On dénombre aussi de genre de Statistique :

- La Statistique Descriptive, que nous étudierons dans ce cours
- Et la Statistique Mathématique, qui utilise la Statistique Descriptive pour le calcul de probabilités.

Lorsque la Statistique est mise au pluriel, on parle alors de données statistiques, de tableaux, de valeurs où la Statistique est appliquée.

Un tableau type : Le Tableau Individus-Caractères

Numéro	Sexe	Âge	Type d'occupation	Nombre d'actifs	Nombre de voitures
1	Homme	25	Locataire	3	0
2	Homme	35	Propriétaire	4	2
3	Femme	26	Locataire	3	1
4	Homme	38	Propriétaire	1	2
5	Femme	34	Locataire	3	2
6	Homme	33	Locataire	1	3
7	Homme	41	Propriétaire	3	1
8	Homme	64	Locataire	3	3
9	Homme	52	Propriétaire	1	2
10	Femme	39	Propriétaire	3	2
11	Homme	34	Propriétaire	3	2
12	Femme	27	Locataire	4	1
13	Femme	31	Propriétaire	4	1
14	Homme	38	Propriétaire	2	1
15	Femme	64	Propriétaire	2	1

Ici, l'unité n'est pas précisée car on la sous entend dans les informations divulguées. Cependant, elle est importante dans une majeure partie des tableaux statistiques, ne jamais l'oublier.

- Les données comme l'âge donnent des valeurs quantitatives continues en entier réel.
- Les données comme le nombre de voitures ou le nombre d'actifs donnent des valeurs quantitatives discrètes en entier naturel car indivisibles et toujours positives (Il n'est pas possible de recenser des demis voitures ou de posséder un nombre négatif de voiture...)

Ce genre de tableau sera étudié plus en détail à partir du chapitre 2

Un tableau type : La répartition entre l'âge et les CSP dans la région nantaise en 1999

CSP \ Âge	Agriculteurs exploitants	Artisans, Commerçants	Cadres supérieurs	Professions intermédiaires	Employés	Ouvriers	Retraités	Etudiants	Demandeurs d'emplois	Autres	Non réponse	Total
De 20 à 39 ans	12	121	417	781	1173	903		525	257	148	9	4346
De 40 à 59 ans	37	225	446	559	701	572	140	1	269	280	4	3234
De 60 à 74 ans	0	6	9	6	5	1	1207	0	0	171	3	1408
Plus de 75 ans	0	0	1	0	0	0	554	0	0	49	0	604
Non réponse	0	1	0	1	3	0	3	0	1		0	9
Total	49	353	873	1347	1882	1476	1904	526	527	648	16	9601

La classification des données permet une aération des statistiques tout en le rendant plus pertinente

Ce genre de tableau sera étudié dans le chapitre 3

B. Les données et calculs principaux utilisés en Statistique

Les opérateurs

L'opérateur somme sert à contracter l'écriture d'une somme de nombre. On peut par exemple écrire :

$$x_1 + x_2 + x_3 + x_4 = \sum_{i=1}^{i=4} x_i = \sum_{i=1}^4 x_i$$

$$x_1 + \dots + x_i + \dots + x_n = \sum_{i=1}^{i=n} x_i$$

Les variable est les indices sont dits muets.

Les opérateurs permettent la compression des écritures ici en terme général, toutes lettres utilisées dans un opérateur sont en minuscule.

Les caractères et variables

- A tout caractère...
- X variable statistique
 - Variable quantitative (Donnée calculable, exemple : Âge, ...)
 - Variable continue (Donnée divisible, nombres réels)
 - Variable discrète (Donnée indivisible, nombres entiers)
 - Variable qualitative (Donnée non calculable, exemple : Hommes, femmes, ...)
 - Valeur ordinale (Donnée classable)
 - Valeur nominative (Donnée non classable)

Variable statistique, modalités et effectifs absolus

Toutes ces variables sont repérées par une lettre majuscule.

Tout caractère (ou question) se transforme en variable, qui va donner des modalités et des effectifs.

Effectif = Fréquence absolue

Effectif cumulée

La première valeur est l'effectif premier, le 2^{ème} effectif cumulé est la somme du premier effectif et du second, le 3^{ème} est le premier + le deuxième + le nouvel effectif.

Fréquence ou Fréquence relative

La fréquence relative est le résultat de la division de l'effectif n_i sur l'effectif total N . Elle se note f_i .

Fréquence relative cumulée

Elle se calcule de la même façon que l'effectif cumulé. Cependant, la variable n'est plus n_i mais f_i . La particularité de la fonction est qu'elle est croissante, positive et bornée.

Récapitulatif

Valeur de la variable x	Effectif n	Effectif cumulé N	Fréquence relative f	Fréquence relative cumulée F
x_1	n_1	$N_1 = n_1$	$f_1 = \frac{n_1}{N}$	$F_1 = f_1$
...
x_i	n_h	$N_h = \sum_{i=n_1}^{n_h} i$	$f_h = \frac{n_h}{N}$	$F_h = \sum_{i=f_1}^{f_h} i$
...
x_n	n_r	$N_r = \sum_{i=n_1}^{n_r} i = N$	$f_n = \frac{n_r}{N}$	$F_n = \sum_{i=f_1}^{f_n} i = 1$
Total	N		1	

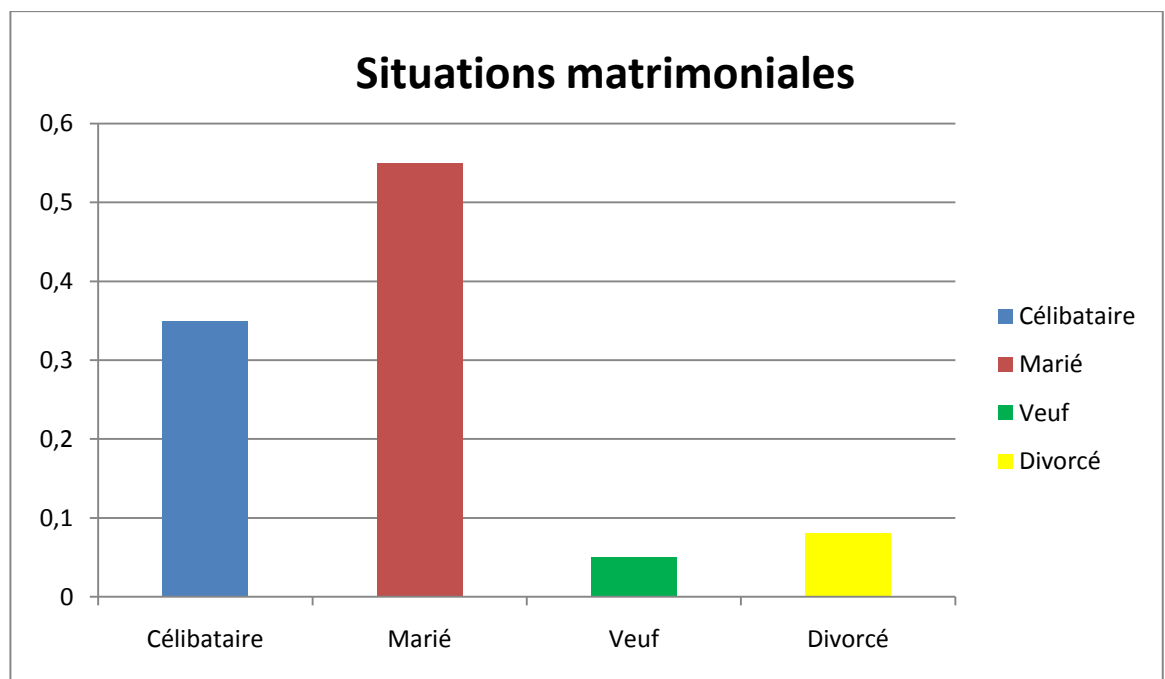
Exercice – Variable discrète

Valeurs x_i	Effectifs n_i	Effectifs cumulés (croissants) N_i	Fréquences relatives f_i	Fréquences relatives cumulés (croissantes) F_i
0	10	10	$\frac{10}{50}$	$\frac{10}{50}$
1	9	10 + 9	$\frac{9}{50}$	$\frac{10}{50} + \frac{9}{50} = \frac{19}{50}$
2	13	10 + 9 + 13	$\frac{13}{50}$	$\frac{10}{50} + \frac{9}{50} + \frac{13}{50} = \frac{32}{50}$
3	7	10 + 9 + 13 + 7	$\frac{7}{50}$	$\frac{10}{50} + \frac{9}{50} + \frac{13}{50} + \frac{7}{50} = \frac{39}{50}$
4	5	10 + 9 + 13 + 7 + 5	$\frac{5}{50}$	$\frac{10}{50} + \frac{9}{50} + \frac{13}{50} + \frac{7}{50} + \frac{5}{50} = \frac{44}{50}$
5	4	10 + 9 + 13 + 7 + 5 + 4	$\frac{4}{50}$	$\frac{10}{50} + \frac{9}{50} + \frac{13}{50} + \frac{7}{50} + \frac{5}{50} + \frac{4}{50} = \frac{48}{50}$
6	2	10 + 9 + 13 + 7 + 5 + 4 + 2 = 50	$\frac{2}{50}$	$\frac{10}{50} + \frac{9}{50} + \frac{13}{50} + \frac{7}{50} + \frac{5}{50} + \frac{4}{50} + \frac{2}{50} = \frac{50}{50} = 1$
Total	50		1	

Valeurs x_i	Effectifs n_i	Effectifs relatifs cumulés (décroissants) N_i	Fréquences relatives cumulés (décroissantes) F_i
0	10	50	$\frac{50}{50} = 1$
1	9	50 - 9	$\frac{50}{50} - \frac{9}{50} = \frac{41}{50}$
2	13	50 - 9 - 13	$\frac{50}{50} - \frac{9}{50} - \frac{13}{50} = \frac{28}{50}$
3	7	50 - 9 - 13 - 7	$\frac{50}{50} - \frac{9}{50} - \frac{13}{50} - \frac{7}{50} = \frac{21}{50}$
4	5	50 - 9 - 13 - 7 - 5	$\frac{50}{50} - \frac{9}{50} - \frac{13}{50} - \frac{7}{50} - \frac{5}{50} = \frac{16}{50}$
5	4	50 - 9 - 13 - 7 - 5 - 4	$\frac{50}{50} - \frac{9}{50} - \frac{13}{50} - \frac{7}{50} - \frac{5}{50} - \frac{4}{50} = \frac{12}{50}$
6	2	50 - 9 - 13 - 7 - 5 - 4 - 2 = 10	$\frac{50}{50} - \frac{9}{50} - \frac{13}{50} - \frac{7}{50} - \frac{5}{50} - \frac{4}{50} - \frac{2}{50} = \frac{10}{50}$
Total	50		

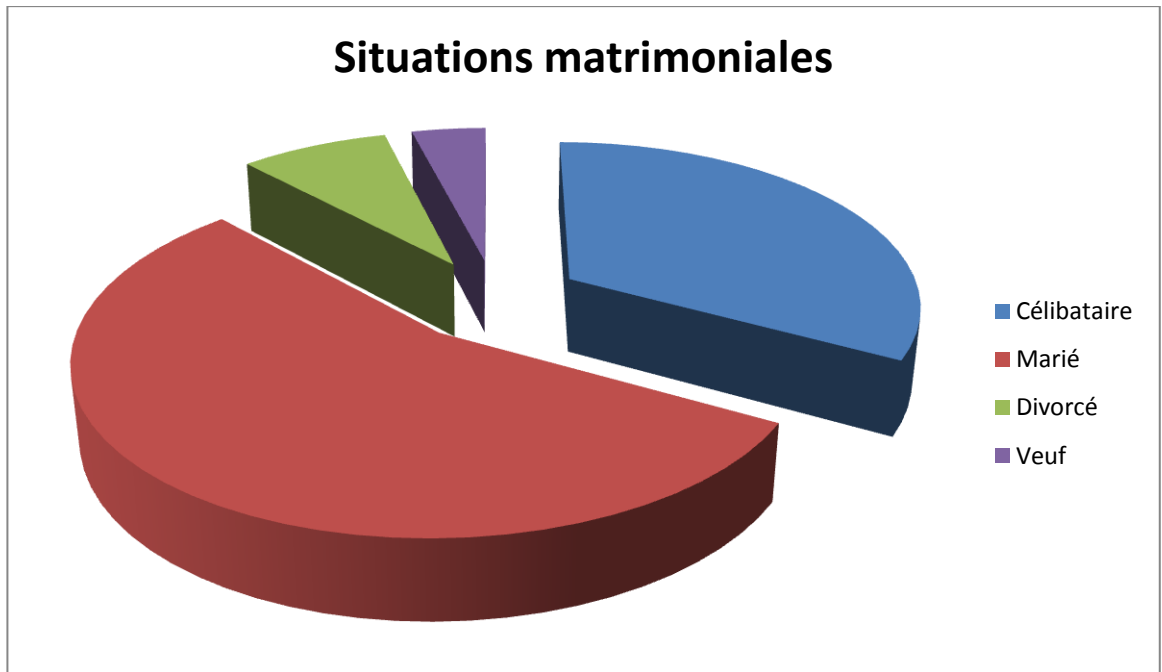
C. Les représentations graphiques communes

Diagramme en tuyaux d'orgues



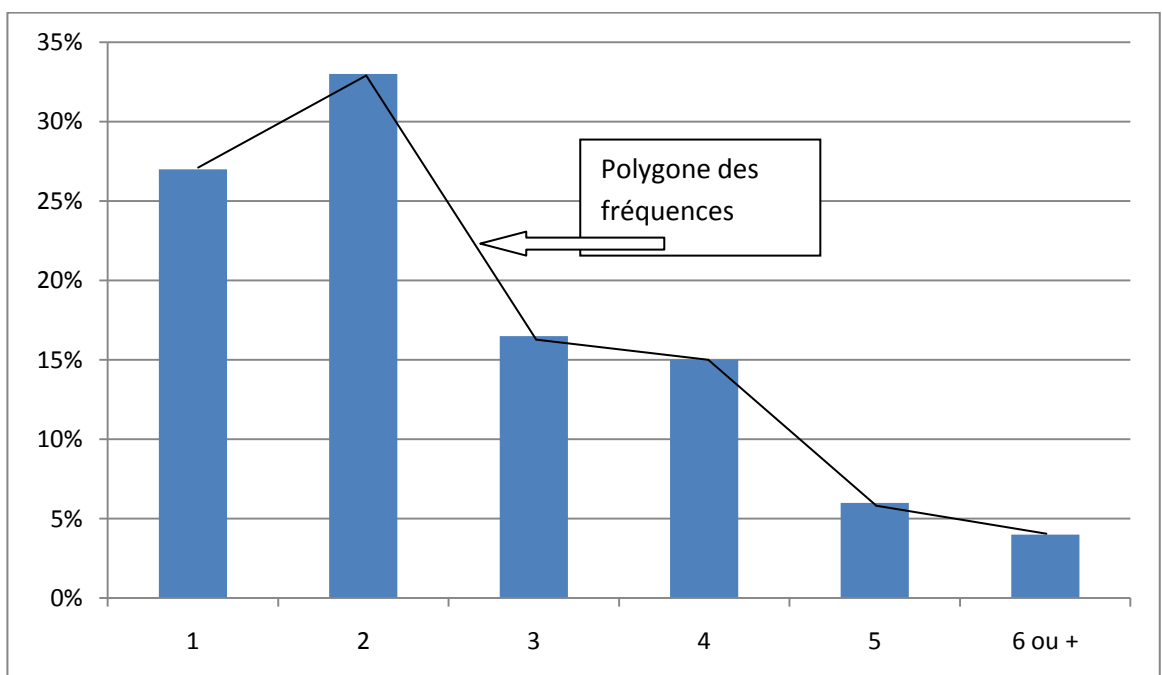
Dans ce type de représentation, les rectangles (ou tuyaux) ont pour base une modalité et comme hauteur l'effectif (ou la fréquence). La base de chacun des rectangles (base sur l'axe des abscisses) ne possède aucune signification numérique (variable qualitative)

Diagramme en secteurs



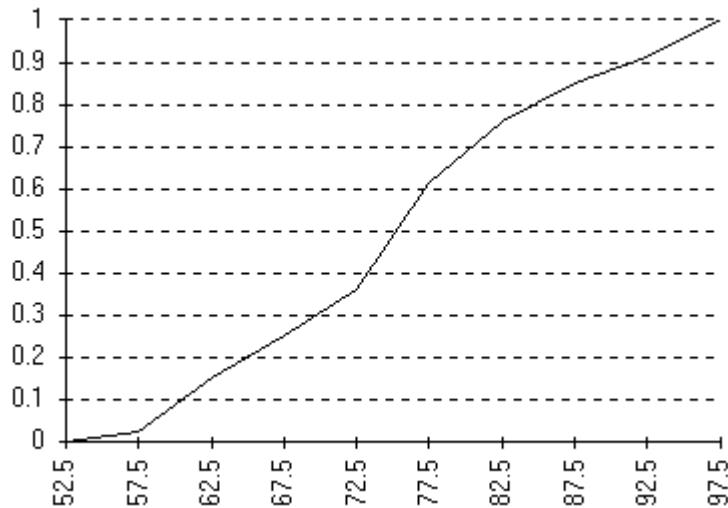
Dans ce type de représentation, nous utilisons un disque (plus communément appelé camembert) ; chacune des modalités est représentée par un secteur qui est proportionnel à l'effectif (ou à la fréquence)

Le polygone des fréquences



Comme vous pouvez le remarquer, le polygone des fréquences est, au final, une courbe insérée dans un diagramme. Chacun des points de la courbe représentée se situe toujours au centre de sa colonne (du diagramme) respective.

Diagramme cumulatif



Comme vous pouvez le constater, le diagramme cumulatif est monotone, et cela est toujours le cas qu'il soit croissant ou décroissant. Pour cause, il agit de la même façon que pour les valeurs de l'effectif cumulatif : c'est une addition ou une soustraction des valeurs précédentes. La représentation ci-dessus est effectuée avec des variables continues, sinon, il s'agit d'un graphique dit « en escalier ».

Regroupement en classe

L'avantage du regroupement en classe est qu'il permet de regrouper des séries de données au sein même d'un intervalle (appelé classe). Prenons par exemple une série de nombres {3 ; 3,5 ; 4,1 ; 5,2 ; 6} et mettons en place une classe 1.

Les nombres de la série de nombres vont rentrer dans la classe 1 qui sera pondéré par une borne inférieure (par exemple : 1) et une borne supérieure (par exemple : 7)

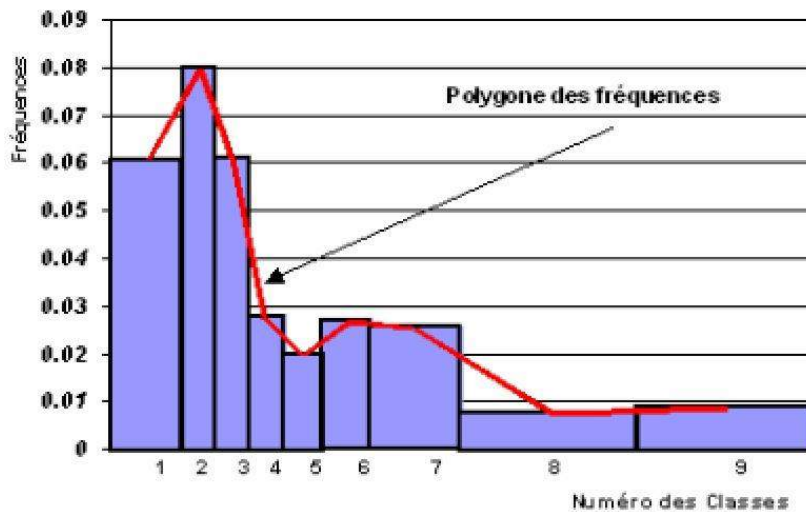
Pour définir l'amplitude d'une classe, notée a_i , il suffit de soustraire la borne inférieure à la borne supérieure :

$$a_i = 7 - 1 = 6$$

Pour calculer le centre de la classe, notée (dans les histogrammes notamment) $x_{i\text{centre}}$, il faut faire la somme de la borne inférieure et de la borne supérieure, puis diviser le résultat par 2

$$x_{i\text{centre}} = \frac{1 + 7}{2} = 4$$

L'histogramme & polygone des fréquences



Graphique venant du cours

L'histogramme est la représentation graphique très proche en apparence au diagramme en tuyaux d'orgues, sauf qu'elle représente une distribution statistique (x_i, n_i) ou (x_i, f_i) ou $(x_i, f_i\%)$ groupée en classe, où chaque classe est représentée par un rectangle de base proportionnelle à l'amplitude et de surface proportionnelle à l'effectif (il est donc rare d'avoir tous les rectangles de l'histogramme d'une largeur uniforme). Il est donc important de ne pas confondre ces deux représentations graphiques.

Le mode correspond au maximum en mathématique. C'est un effectif qui est encadré par 2 effectifs qui lui sont inférieurs. Le mode est sujet à la modification de la classification par classe.

Série chronologique

On peut appeler Suite chronologique une analyse statistique effectuée sur des variables dans une période t donnée. Elles servent essentiellement à tenter de faire de la prévision. Il s'agit souvent une variable quantitative discrète.

Plus de précisions en fin de document... (Il est **probable** (et non pas sûr) que cela tombe au partiel de stat !)

FIN DU PREMIER CHAPITRE

Des QCM pour tester vos connaissances sont à votre disposition sur le site Portices de J-L Monino

([Voir en début de cours](#) ou directement [ici](#))

Chapitre 2

Les principaux résumés de la statistique

A. Les résumés des positions et des valeurs

Les propriétés des indicateurs – Les conditions de Yule

Les distributions statistiques à une variable sont représentées par un petit nombre d'indicateurs (résumés numériques) qui doivent être représentatifs de la distribution statistique. Il est souhaitable que les paramètres ou résumés numériques possèdent certaines propriétés, appelées Conditions de Yule (du nom de son inventeur : George Udny Yule) :

- Être défini de manière objective
- Dépendre de toutes les observations
- Avoir une signification concrète
- Être facilement calculables et interprétables
- Être peu sensibles aux fluctuations d'échantillonnages
- Se prêter aisément aux calculs algébriques

Médiane

La médiane X_m d'une distribution statistique est la valeur de la variable qui partage l'effectif total de la distribution en deux parties égales, telle que la première moitié des observations soit inférieure (ou égale) à X_m et la seconde moitié soit supérieure (ou égale) à X_m

Si $(x_i, F(x))$ est la distribution des fréquences cumulées d'une variable statistique, alors la médiane est donnée par l'équation suivante :

$$F(X_m) = \frac{1}{2}$$

Si la variable est continue, on effectue une interpolation à l'intérieur de la classe médiane.

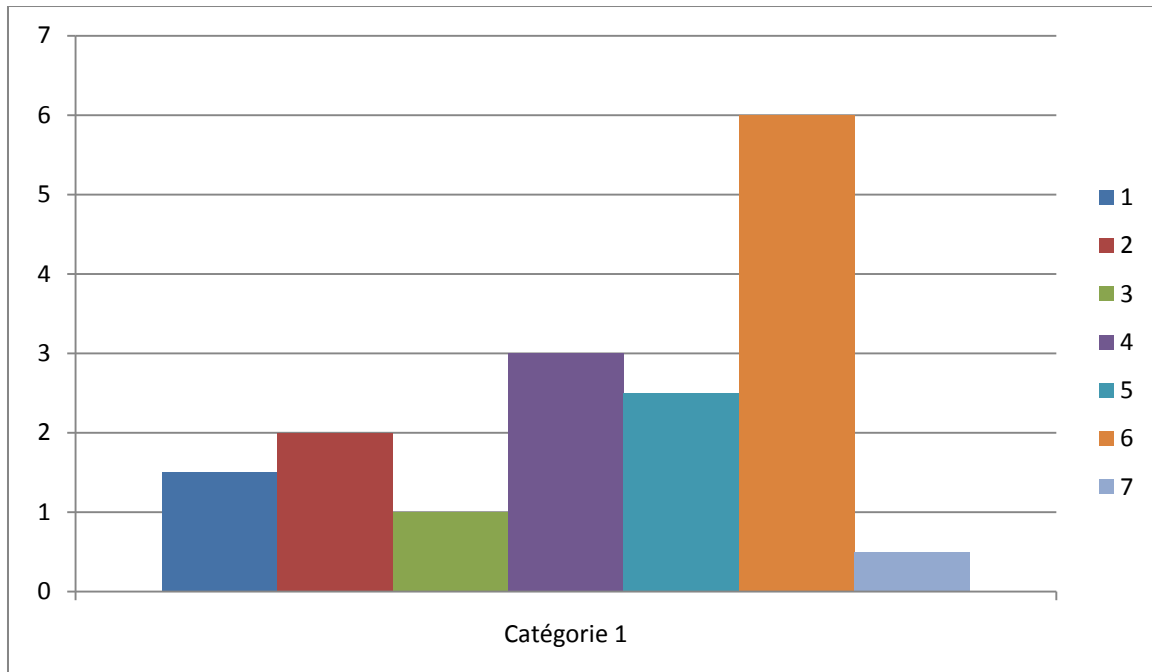
Mode

Dans la statistique, il existe deux types de classes modales :

La classe modale absolue : Il s'agit de la valeur de la variable pour laquelle l'effectif est le plus élevé

La *classe modale relative* : il s'agit de la valeur de la variable pour laquelle l'effectif (ou la fréquence) est encadré par des valeurs qui lui sont inférieures. La nuance avec le mode absolu, c'est qu'il peut y avoir plusieurs modes relatifs à contrario du mode absolu où il est unique.

Exemple :



Il y a ici 3 classes modales relatives et 1 classe modale absolue

La classe modale absolue est représentée par la colonne 6

La première classe modale relative est représentée par la colonne 2, entourée de la valeur 1 et de la valeur 3, possédant des effectifs plus « petits ».

Il en va de même pour les deux autres classes modales relatives (4 et 6)

Remarque : Lorsque les variables sont regroupés en classe, on appelle classe modale le groupement de valeur de la variable (classe) pour lesquelles les effectifs sont encadrés par des classes plus « petites ».

Cependant, il est **parfois** utile de remplacer la notion de classe modale par la notion de mode initiale, **à vos risques et périls** cependant : une classe modale reste une classe modale ; nous ne pouvons donc pas affirmer la valeur du mode (ou cas exceptionnel) par un simple coup d'œil sur le diagramme!!, un calcul est nécessaire à cela que je décris ci dessous.

Pour se faire en bons termes, on effectue une interpolation linéaire à l'intérieur de la classe modale ; Le mode se calcule de la manière suivante :

$$\text{Borne inférieure de la classe modale} + \text{Amplitude de classe} \left(\frac{d_1}{d_1 + d_2} \right)$$

Où $d_1 =$ Différence entre l'effectif de la classe modale et celle de la classe précédente

$d_2 =$ Différence entre l'effectif de la classe modale et celle de la classe supérieure

De manière totalement algébrique le calcul aurait donc cette tête :

$$X_m = x_{i-1} + a_i \left(\frac{d_1}{d_1 + d_2} \right)$$

Formule que vous pouvez retrouver, page 50/113 sur le site Portices de M. Monino (cf. début de cours)

Les quartiles, les déciles et les centiles

Il s'agit de valeurs d'une série ou d'une distribution statistique rangée dans un ordre particulier (croissant ou décroissant), partageant l'effectif total en plusieurs parties égales. Les quartiles sont dénombrables au nombre de 3 (Q_1, Q_2, Q_3 : Q_1 représentant le premier quart, Q_2 , le second, et Q_3 , le troisième), les déciles, au nombre de 9 (puisque qu'il divise la série en 10 parties égales) et les centiles, au nombre de 99 (puisque'il divise la série en 100 parties égales)

Les moyennes

Moyenne arithmétique

La moyenne arithmétique est la moyenne la plus connue et la plus utilisée. Pour la réaliser, il faut faire la somme de tous les effectifs de la variable statistique et la diviser par le nombre de ces dernières. Mathématiquement parlant, ça donnerait ceci :

Soit $x_1, \dots, x_i, \dots, x_r$ les observations numériques d'une variable statistique X et soit les effectifs respectifs $n_1, \dots, n_i, \dots, n_r$ de ces r valeurs numériques.

Sous la forme de calcul, le résultat serait :

$$\bar{X} = \frac{n_1 x_1 + \dots + n_i x_i + \dots + n_r x_r}{n_1 + \dots + n_i + \dots + n_r}$$

$$\bar{X} = \frac{\sum_{i=1}^{i=r} n_i x_i}{N}$$

Moyenne géométrique

La moyenne géométrique d'une série statistique quantitative discrète positive non nulle est définie telle que son logarithme est la moyenne arithmétique des logarithmes des valeurs discrètes positives non nulles de la distribution.

Autrement dit :

$$\log(\bar{x}) = \frac{n_1 \log(x_1) + \dots + n_i \log(x_i) + \dots + n_r \log(x_r)}{n_1 + \dots + n_i + \dots + n_r}$$

$$\log(\bar{x}) = \frac{\sum_{i=1}^{i=r} n_i \log(x_i)}{N} = \sum_{i=1}^{i=r} f_i \log(x_i)$$

Moyenne harmonique

La moyenne harmonique est l'inverse de la moyenne arithmétique de l'inverse des termes.

Sous la forme de calcul, ça donnerait ceci :

$$\frac{1}{H} = \frac{n_1 \frac{1}{x_1} + \dots + n_i \frac{1}{x_i} + \dots + n_r \frac{1}{x_r}}{n_1 + \dots + n_i + \dots + n_r}$$

Que l'on peut donc écrire de la manière suivante :

$$H = \frac{n_1 + \dots + n_i + \dots + n_r}{n_1 \frac{1}{x_1} + \dots + n_i \frac{1}{x_i} + \dots + n_r \frac{1}{x_r}}$$

$$H = \frac{1}{N} \sum_{i=1}^{i=r} \frac{n_i}{x_i}$$

Moyenne quadratique

La moyenne quadratique est la racine carrée de la moyenne du carré des effectifs d'une variable statistique, défini mathématiquement de la façon suivante :

$$Q^2 = \frac{n_1 x_1^2 + \dots + n_i x_i^2 + \dots + n_r x_r^2}{n_1 + \dots + n_i + \dots + n_r}$$

$$Q^2 = \frac{1}{N} \sum_{i=1}^{i=r} n_i x_i^2 = \sum_{i=1}^{i=r} f_i x_i^2$$

$$\text{Donc : } Q = \sqrt{\sum_{i=1}^{i=r} f_i x_i^2}$$

B. Dispersion et concentration

Variance

On peut interpréter la variance comme la moyenne des carrés des écarts à la moyenne ou, dit plus simplement, la différence entre la moyenne des carrés et le carré des moyennes, mathématiquement interprété de la façon suivante :

$$S_X^2 = \frac{1}{N} \sum_i^{r-1} n_i x_i^2 + \bar{X}^2 = \frac{1}{N} \sum_i^{r-1} n_i (x_i - \bar{X})^2 = \sum_i^{r-1} f_i (x_i - \bar{X})^2$$

Ecart-type

L'écart-type n'est autre que la racine carré de la variance, il se notera donc mathématiquement de la façon suivante :

$$S_X = \sqrt{\frac{1}{N} \sum_i^{r-1} n_i (x_i - \bar{X})^2}$$

Moments non centré d'ordre s

On appelle un moment non centré d'ordre s la moyenne des effectifs de la variable statistique à la puissance s, mathématiquement représenté de la façon suivante :

$$m_s = \frac{1}{N} \sum_i n_i x_i^s$$

Avec les moments non centrés, on peut générer les 4 moyennes

Moment centré d'ordre s

Les 4 principaux moments centrés utilisé en Stat descriptive sont :

$$\mu_2 = m_2 - m_1^2$$

$$\mu_3 = m_3 - 3m_1m_2 + 2m_1^3$$

$$\mu_4 = m_4 - 4m_1m_3 + 6m_1^2m_2 - 3m_1^4$$

Ils seront probablement les seuls moments centrés demandés lors des partiels.

Remarque : Le moment centré d'ordre 2 μ_2 correspond à l'expression de la variance

Médiale

Se calcule de la même façon que la médiane.

On compare la médiane (construite à partir des n_i) à la médiale (construite à partir des $n_i x_i$)

Si $(x_i, Pseudo F_i(x))$ est la distribution des fréquences cumulées d'une variable statistique, alors la médiale est par l'équation :

$$Pseudo F(X_{ML}) = \frac{1}{2}$$

$$X_{ML} = x_{i-1} + a_i \left(\frac{0,5 - pF_{i-1}}{pf_i} \right)$$

Effectifs	« Pseudo » fonctions relative	« Pseudo » fonctions cumulatives (croissantes) La médiale
$x_i \cdot n_1$	$\frac{x_i \cdot n_1}{\sum_i n_i \cdot x_i} = pf_1$	$pf_1 = pF_1$
...
$x_i \cdot n_i$	$\frac{x_i \cdot n_i}{\sum_i n_i \cdot x_i} = pf_i$	$pf_1 + pf_2 + pf_3 + \dots + pf_i = pF_i$
...		...
$x_i \cdot n_r$	$\frac{x_i \cdot n_r}{\sum_i n_i \cdot x_i} = pf_r$	$pf_1 + \dots + pf_i + \dots + pf_r = pF_r$
$\sum_i n_i \cdot x_i$		

Courbe de Lorenz

Aire de concentration correspond à l'aire de l'intérieur de ce qu'il se trouve entre les deux courbes
(max : 0,5)

Somme des 4 trapèzes appartenant aux 4 classes

0.5-aire des trapèzes = aire de concentration

Indice de (concentration de) Gini = $1 - \sum_i [F_i(nx) + F_{i-1}(nx)] * f_i(x)$

Dernier TD de stat : Les formules à savoir pour les séries chronologiques

✚ Moyenne :

$$\bar{Y} = \frac{1}{p} \sum_{j=1}^p y_{ij}$$

✚ Variance :

$$S_y^2 = \frac{1}{p} \sum_j (y_{ij} - y)^2$$

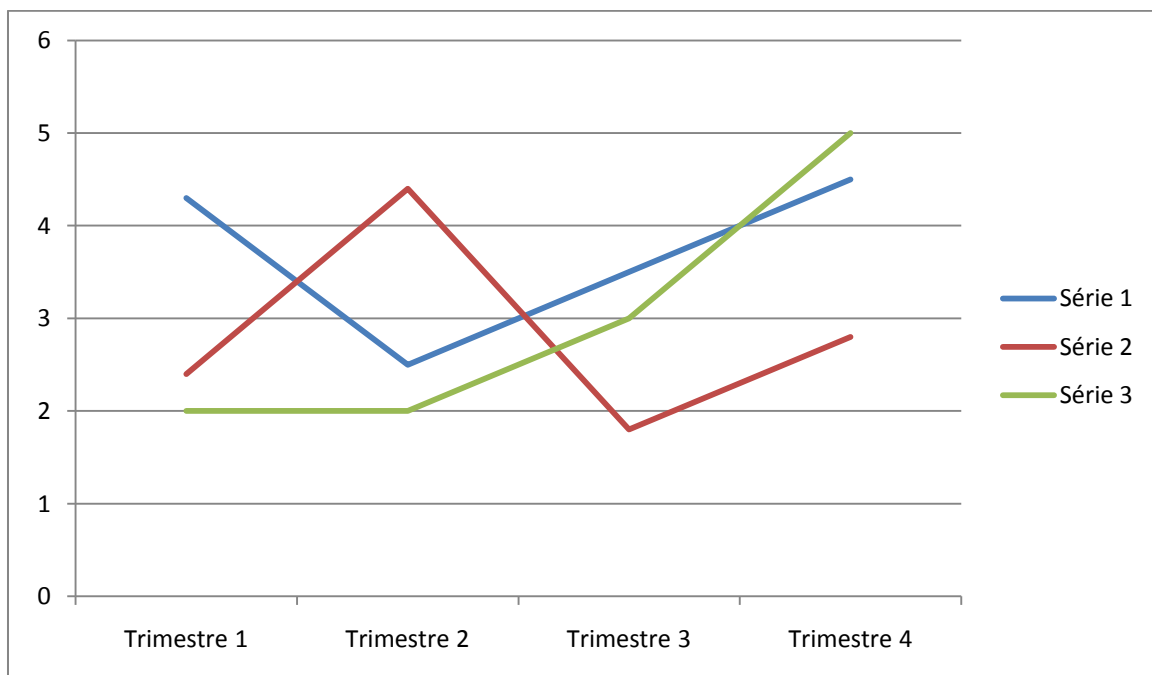
✚ Ecart Type :

$$S_y = \sqrt{S_y^2} = \sqrt{\frac{1}{p} \sum_j (y_{ij} - y)^2}$$

✚ Profil Temporel :

Représentation graphique de toutes les années dans un même graphique (y = valeur, x = fraction de période)

Exemple :



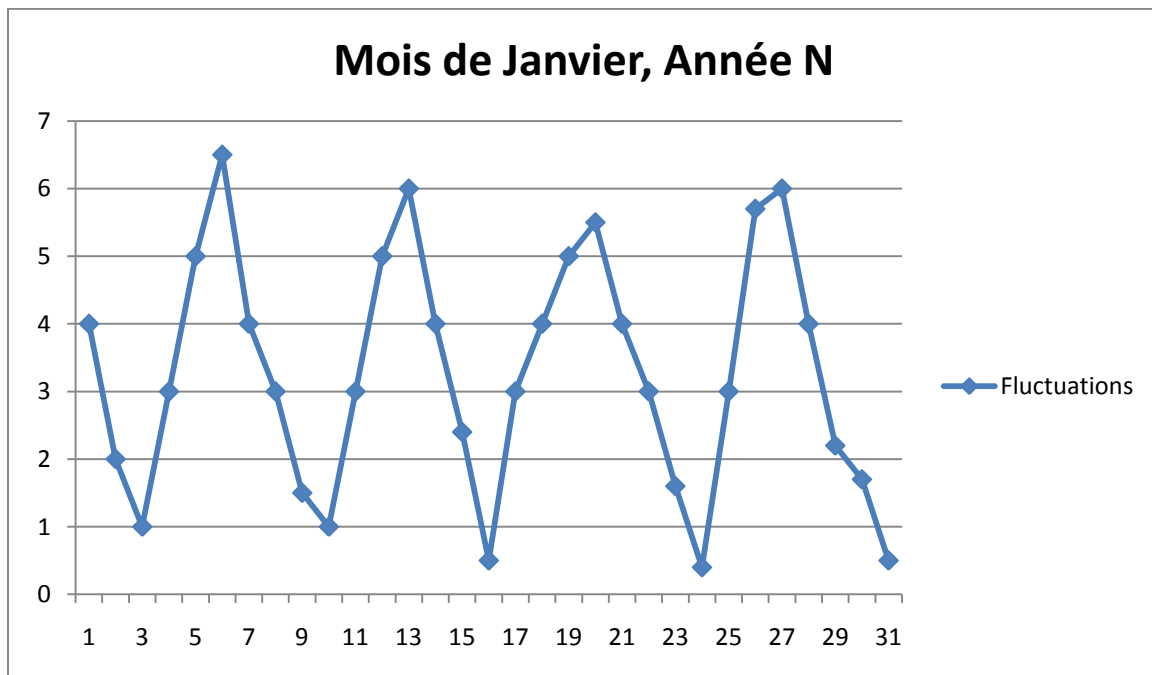
Période choisie : 1 année, Type de division : Trimestrielle

Il existe 3 types de composantes dans le profil temporel :

- La composante :

C'est une courbe dont les données sont celles relevées division de période par division de période

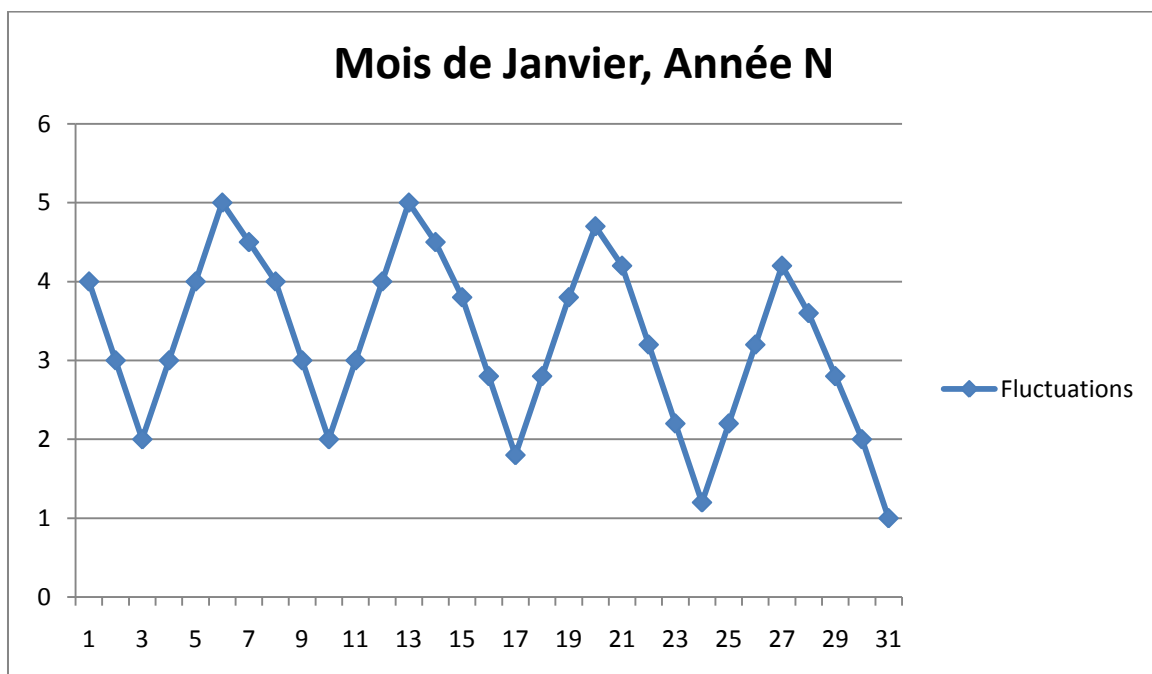
Exemple :



- La composante saisonnière :

C'est un arrondissement de la composante précédente afin de la rendre moins abrupte et donc plus lisible et compréhensible.

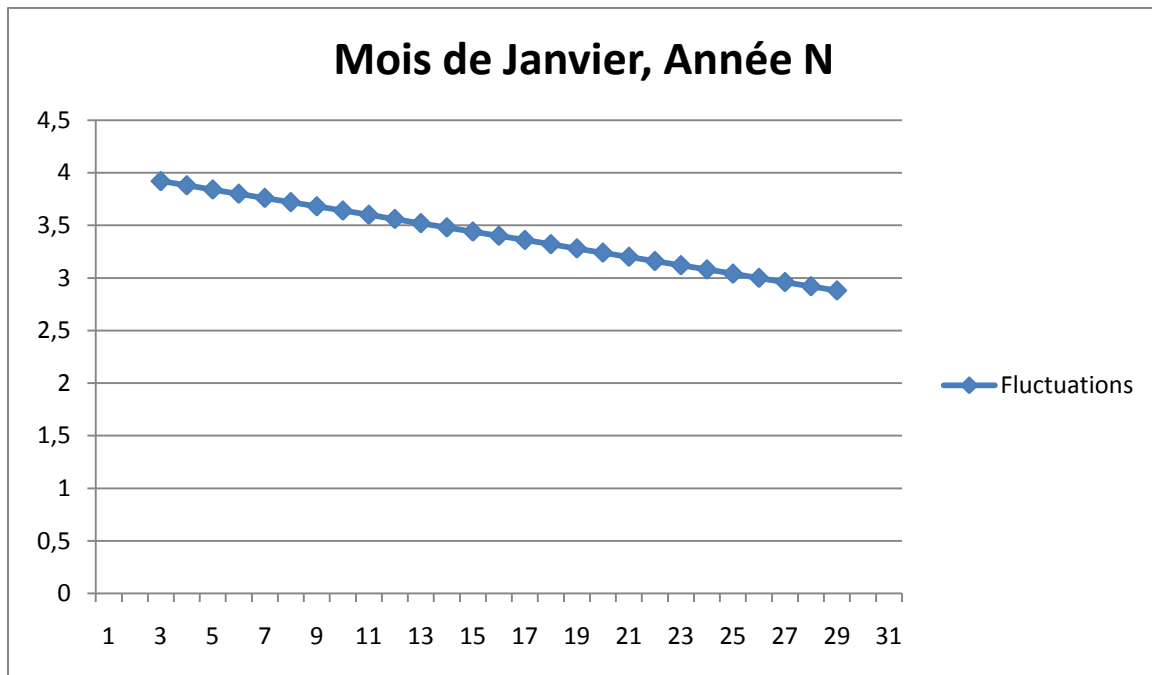
Exemple :



- La composante tendancielle :

On rend la composante en fonction quasi-affine. On peut donc distinguer de la réelle évolution d'une valeur pendant la période

Exemple :



Le choix de la division de la période implique la parité de cette même division, ce qui joue sur les calculs qui vont déterminer la courbe. On dit qu'une période est paire lorsque le nombre de division de la période est lui-même pair (Ex : Trimestre : 4 divisions de la période [année]). Réciproquement, on dit qu'une période est impaire lorsque le nombre de division de la période est impair (Ex : Jour : 7 division de la période [semaine])

➤ En période paire (Périodicité = $2K$)

▶ Moyenne mobile (ou glissante) (Profil Temporel)

Ce calcul permet de rendre la composante plus lisible en la transformant en composante tendancielle

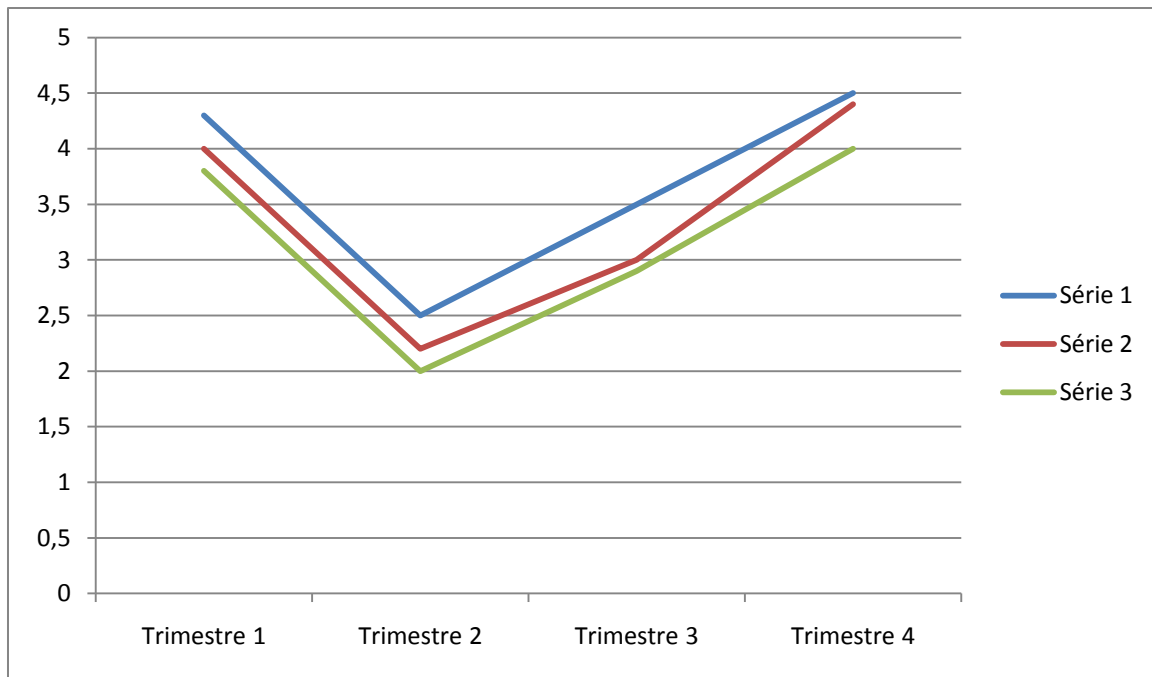
$$mm_t^p(y_t) = x_t = \frac{1}{p} \left[\sum_{i=-k+1}^{i=k-1} y_{t+i} + \frac{1}{2}y_{t-k} + \frac{1}{2}y_{t+k} \right]$$

La moyenne mobile n'existe pas si $t - k$ ou $t + k$ sortent du domaine de définition

▶ Le modèle additif (Profil temporel)

Courbes du profil temporel relativement parallèles et régulières

Exemple :



❖ Variation saisonnière (Modèle additif – Profil temporel)

La variation permet de déterminer l'agrégat superflu qui rend la composante plus abrupte du fait des normes saisonnières qui l'influe

$$d_{ij} = y_t - x_t$$

❖ Coefficient saisonnier

$$S_j = \frac{1}{p-1} \sum_{j=1}^{p-1} d_{ij}$$

❖ Coefficient saisonnier corrigé des variations saisonnières

$$S_j^* = S_j - \left(\frac{1}{p} \sum_{j=1}^p S_j \right)$$

❖ Série corrigé des variations saisonnières

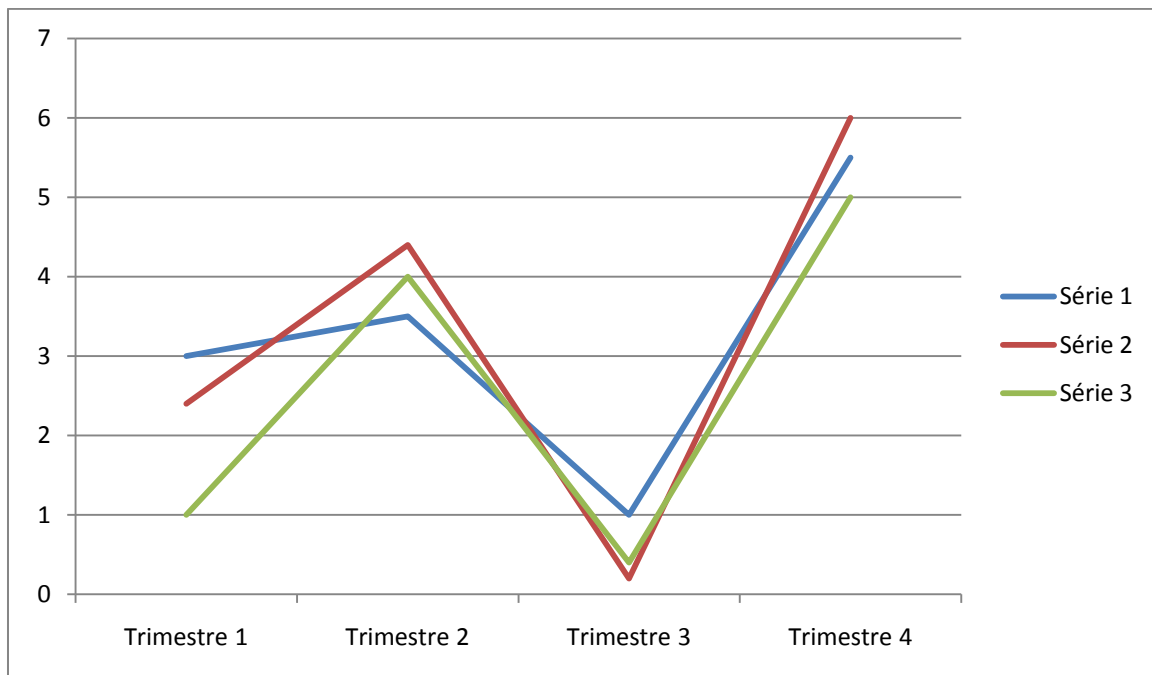
Cette série détermine la composante saisonnière

$$y_t^{CVS} = y_t - S_j^*$$

▶ Le modèle multiplicatif (Profil Temporel)

Courbe (Série) du profil temporel dont les variations sont irrégulières et de plus en plus importantes

Exemple :



❖ Rapport saisonnier (Modèle multiplicatif – Profil Temporel)

$$r_{ij} = \frac{y_t}{x_t}$$

❖ Coefficient saisonnier

S_j : Moyenne des r_{ij}

❖ Coefficient saisonnier corrigé des variations saisonnières

$$S_j^* = \frac{S_j}{\frac{1}{p} \sum_j S_j}$$

❖ Série corrigée des variations saisonnières

$$y_t^{cvs} = \frac{y_t}{S_j^*}$$