République Algérienne Démocratique et Populaire

Ministère de L'Enseignement Supérieur et de la Recherche Scientifique

Université Mouloud MAMMERI de Tizi Ouzou

Faculté des Sciences Economiques, Commerciales et des Sciences de Gestion

Département des Sciences Economiques

Polycopié de cours

Analyse des Sériés Temporelles

1ère année du cycle Master

Semestre 1 – Economie es Affaires

Réalisé par Dr. Gouraya BELBACHIR

Maître de conférences B

Email : gouraya.belbachir@ummto.dz

Année universitaire : 2024 – 2025

To access the online courses, please scan the **QR** code using the anonymous login

Identification du module :

<table>
<tr><td>Unité d’enseignement</td><td colspan="2">Méthodologique</td></tr>
<tr><td>Crédits</td><td colspan="2">5</td></tr>
<tr><td>Coefficient</td><td colspan="2">2</td></tr>
<tr><td rowspan="3">Volume horaire hebdomadaire</td><td>Cours</td><td>1h 30 min</td></tr>
<tr><td>TD</td><td>1h 30 min</td></tr>
<tr><td>Σ</td><td>3 h</td></tr>
<tr><td rowspan="2">Méthode d’évaluation</td><td>Cours</td><td>20 / 20</td></tr>
<tr><td>TD</td><td>20 / 20</td></tr>
<tr><td rowspan="3">Poids relatif d’évaluation</td><td>Cours</td><td>60 %</td></tr>
<tr><td>TD</td><td>40 %</td></tr>
<tr><td>Σ</td><td>100 %</td></tr>
<tr><td>Moyenne de Module</td><td colspan="2">= (Note EMD * 0,6) + (Note TD * 0,4)</td></tr>
<tr><td>Exemple</td><td colspan="2">Note EMD = 16 / 20
Note TD = 14 / 20
Moyenne de Module = (16 * 0,6) + (14 * 0,4)
Moyenne de Module = (9,6 + 5,6)
Moyenne de Module = 15,2 / 20</td></tr>
</table>

Objectifs de l’enseignement :

Le présent cours, conçu conformément au programme ministériel, vise à ce que l’étudiant acquiert des compétences en la matière, à savoir :

- Maitriser les concepts clés de l’Analyse des Séries Temporelles ;
- La présentation des techniques d’analyse des séries temporelles multivariées couramment utilisées dans les applications (séries stationnaires et séries non-stationnaires) ;
- Dans le cas stationnaire, il s’agit de l’étude des modèles VAR (vectoriels auto-régressifs) ;
- Dans le cas non-stationnaire le cours se concentre sur l’inférence en présence de racines unités et à l’estimation de relations de cointégration.
- Mettre en œuvre les méthodes d’estimation et de test ;
- S’initier à l’usage des logiciels statistiques utilitaires, notamment le logiciel de base EViews, pour la manipulation de données, l’analyse statistique et économétrique, ainsi que pour la réalisation de prévisions, de simulations et de présentations analytiques de données.

Les chapitres dispensés en cours magistral font l’objet chacun d’un traitement appliqué, sous forme de séries d’exercices, en séances de Travaux Dirigés (TD). Lors de ces dernières des éclaircissements et des informations supplémentaires, surtout d’ordre pratique, sont fournis aux étudiants.

Une bibliographie révisée et mise à jour est fournie, dans le but de permettre aux étudiants d’approfondir leurs connaissances.

Les chapitres du présent cours sont également publiés en ligne, sur la plateforme Moodle (Elearning) en accès anonyme et mon e-mail (**gouraya.belbachir@ummto.dz**) est mis à la disposition des étudiants pour toute question ou renseignement éventuel.

Plan du cours :

- Introduction générale

Chapitre 1 : Introduction à l'analyse de la relation entre deux variables quantitatives

Section 1 : La régression linéaire simple

Section 2 : Les méthodes d'ajustement

Section 3 : L'analyse de la Corrélation

Chapitre 2 : Définitions et présentation graphique des séries temporelles

Section 1 : Définition et objectifs des Séries Temporelles (ST)

Section 2 : La Représentation Graphique des Séries Temporelles (ST)

Section 3 : Les Composantes Fondamentales des Séries Temporelles (ST)

Chapitre 3 : Les Modèles de décomposition des séries temporelles

Section 1 : Le modèle additif

Section 2 : Le modèle multiplicatif

Section 3 : Test du modèle (modèle de Bays Ballot)

Chapitre 4 : Les Modèles d'Ajustement et de Lissage des Séries Temporelles

Section 1 : La Méthode des Moyennes Mobiles (MM)

Section 2 : La Méthode des Moyennes Mobiles Centrées (MMC)

Chapitre 5 : Mesure des Composantes d'une Série Temporelle

Section 1 : Détermination du Trend (Tt)

Section 2 : Détermination de la Composante Saisonnière (St)

Section 3 : Détermination de la Composante Cyclique (Ct)

Introduction générale

Introduction générale

La série temporelle, chronologique, dite aussi chronique, représente une suite d'observations réalisées dans le temps pour une variable quelconque. C'est donc une série bivariée où l'une des variables est le temps (t). C'est la suite des mesures d'une grandeur relevée à différentes périodes ou instants précis (t) (Les jours, les semaines, les mois, les trimestres, les semestres, les années ...) pris dans une période donnée (l'année par exemple). Le plus souvent, la variable étudiée est notée Y_i et le temps X_i ou t_i .

L'analyse des Séries Temporelles (ST) constitue une branche de l'économétrie dont l'objet est l'étude des variables au cours du temps (t). Parmi leurs principaux objectifs figurent la détermination de *tendances* au sein de ces séries ainsi que la *stabilité* des valeurs (et de leur variation) au cours du temps.

C'est dans ce sens qu'elles représentent un autre moyen dynamique cette fois et très puissant dans la description des populations observées ainsi que l'analyse des variables économiques qui nous intéressent au lieu d'arrêter avec une analyse statique comme auparavant.

En effet, les Séries Temporelles (ST) rendent les variables étudiées de plus en plus explicites en les décomposant en plusieurs composantes qui constituent les effets de leur variation. Ce qui peut, par conséquent, faciliter leur utilisation dans la pratique de l'économie notamment dans cadre de la prévision et la prise de décision.

Nous allons présenter d'abord l'intérêt et les définitions des Séries Temporelles (ST) afin d'en préciser certaines particularités. Ensuite, nous présenterons leurs modèles de décomposition pour appréhender leurs différentes composantes. Puis, nous exposerons les modèles d'ajustement et de lissage qui sont jugés nécessaires pour nettoyer les séries brutes des effets non désirables. Enfin, nous mettrons l'accent sur les méthodes de mesure de leurs composantes afin d'estimer leurs poids et procéder à la prévision des valeurs futures des séries observées.

Chapitre 1 :

Introduction à l’analyse de la relation entre deux variables quantitatives

Chapitre 1 : Introduction à l’analyse de la relation entre deux variables quantitatives

Dans tous les domaines socio-économiques, la recherche de la relation entre deux (ou plusieurs) variables est fondamentale. En statistique, on appelle cette analyse l’analyse de la régression. Lorsque celle-ci porte sur deux variables seulement, elle est dite régression simple, lorsqu’elle porte sur plus de deux variables, elle est dite régression multiple.

L’objet du présent chapitre est de faire découvrir à l’étudiant que dans le on peut mettre en évidence l'existence d'une relation linéaire significative entre deux caractères quantitatifs X et Y, on peut chercher à formaliser la relation moyenne qui unit ces deux variables à l'aide d'une des trois équations suivantes :

- $a.X + b.Y + c = 0$: équation de la droite moyenne liant les caractères X et Y ;
- $Y = a.X + b$: droite de régression de Y en fonction de X ;
- $X = a.Y + b$: droite de régression de X en fonction de Y.

Les trois équations proposées ci-dessus correspondent à trois droites différentes, trois résumés différents du nuage de points (X,Y). La différence entre les trois droites vient du fait que les trois équations proposées correspondent à trois objectifs différents.

Section 1 : La régression linéaire simple
Section 2 : Les méthodes d’ajustement
Section 3 : L’analyse de la Corrélation

Section 1 : La régression linéaire simple

Etudier la relation entre deux variables revient à étudier leur plus ou moins dépendance. Autrement dit, la régression simple consiste à expliquer les variations d'une variable, dite variable dépendante, endogène ou expliquée, généralement notée (y_i), par une autre variable dite indépendante, autonome, exogène ou explicative, généralement notée (x_i) : c'est-à-dire expliquer les variations de y_i par les variations de x_i.

La forme mathématique la plus simple permettant de relier deux variables est la forme linéaire de type « y = ax + b », c'est-à-dire une droite, d'où l'appellation linéaire.

On peut rencontrer trois types de relations linéaires simples :

- La relation linéaire positive, où les deux variables varient en même temps et dans le même sens (Figure 1),

- La relation linéaire négative, où les deux variables varient simultanément mais en sens inverse : l'une augmente pendant que l'autre diminue et vis-versa (Figure 2),

- L'absence de relation, où l'une des variables varie (la variable indépendante) pendant que l'autre demeure constante (Figure 3).

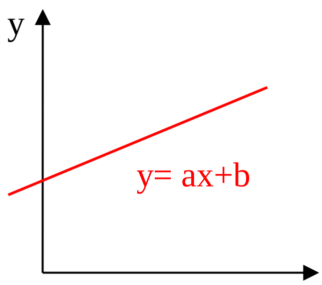


Figure 1

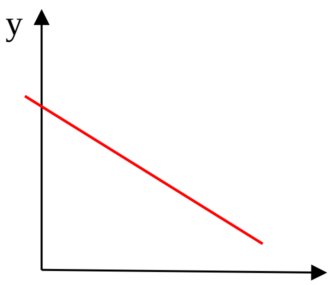


Figure 2

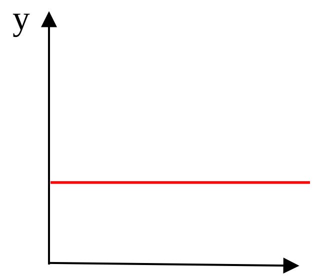


Figure 3

L'**ajustement** est l'un des moyens d'estimer la relation entre deux variables, c'est aussi le moyen le plus utilisé dans l'analyse de la régression. Il existe plusieurs méthodes d'ajustements.

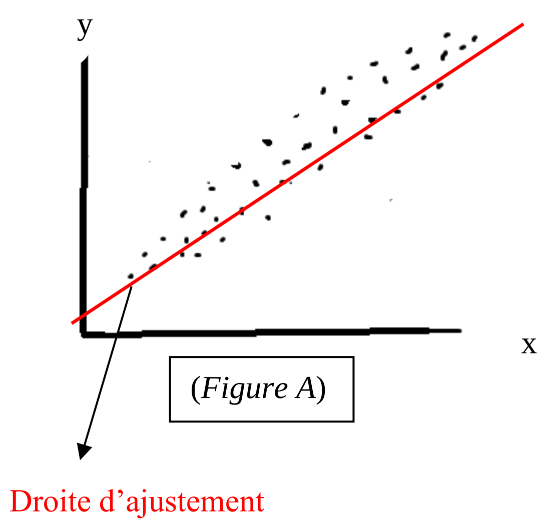
Section 2 : Les méthodes d'ajustement

Il existe en statistique plusieurs manières d'ajuster les données relatives aux deux caractères étudiés. Nous en exposons dans ce qui suit les plus usitées.

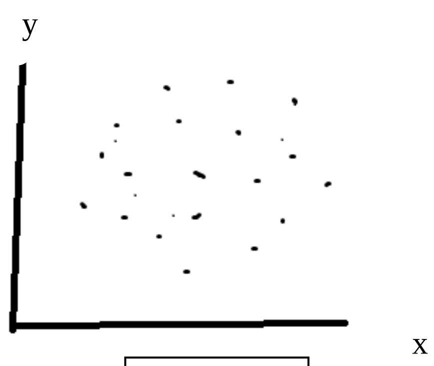
1. L'ajustement graphique :

Une autre manière d'exposer les distributions statistiques à deux caractères, après le tableau de contingence, est le graphique : on l'appelle dans ce cas le nuage de points, où les points représentent les couples (x_i ; y_i).

La répartition des points du nuage, ou bien l'allure de ce dernier, est la première étape qui permet de nous indiquer s'il y a (figure A) ou pas (figure B) relation entre les deux variables. Autrement dit, voir si les deux variables varient globalement dans le même sens et suivent la même tendance dans leurs variations. Lorsque une tendance à la liaison entre les deux variables se dégage, on s'intéressera alors de savoir si la forme de celle-ci peut être linéaire ou pas. C'est-à-dire regarder si le nuage de points suggère une droite ou une tendance à la linéarité (figure A). Lorsque c'est le cas, on cherchera alors une droite d'ajustement (appelée aussi droite de régression). C'est justement ce cas précis de la relation linéaire qui nous intéresse dans la présente section.



(Figure A)



(Figure B)

2. L'ajustement mécanique :

Il consiste à « lisser » (c'est-à-dire gommer ou effacer les variations du court terme) la série temporelle par les procédés ou les techniques des moyennes échelonnées et des moyennes mobiles. La technique des moyennes mobiles étant la plus appropriée et la plus utilisée dans le cas des séries temporelles, nous ne reviendrons donc que sur celle-ci.

Il consiste en la décomposition de la série des données (xi ; yi) en sous-séries ou groupes. Deux techniques principales sont généralement utilisées pour cela.

a) Ajustement par les moyennes échelonnées :

Cette technique consiste à remplacer des groupes de modalités, d'ordre t (ou de taille t), de la distribution bivariée par leurs médianes, pour la variable X, et par leurs moyennes arithmétiques pour la variable Y. Afin de faciliter la détermination des médianes, on retient un ordre t impair, généralement 3, et on prendra soin d'ordonner les modalités x_i par ordre croissant.

Exemple :

Soit la distribution bivariée suivante :

Xi	80	85	90	92	98	105
Yi	106	110	124	120	101	102

$\bar{Y}_1$ $\bar{Y}_2$

Déterminer les moyennes échelonnées d'ordre 3 (ou décomposer, ou ajuster, la série suivante par la méthode des moyennes échelonnées d'ordre 3).

Solution :

$\bar{Y}_1 = (y_1 + y_2 + y_3)/3 = 113,3 \Rightarrow X_2 = 85$ (Médiane)

$\bar{Y}_2 = (y_4 + y_5 + y_6)/3 = 107,7 \Rightarrow X_5 = 98$ (Médiane)

On obtient ainsi les couples de points (xi ; yi) = {(85 ; 113,3) ; (107,7 ; 98)} par lesquels on peut ajuster le nuage de points.

b) Ajustement par les moyennes mobiles :

Cette technique est surtout utilisée dans le cas de séries chronologiques où l'une des variables est le temps. Elle consiste à remplacer la série brute (informations collectées) ou les modalités par une série de moyennes et de médianes. Ce principe rappelle beaucoup celui des moyennes échelonnées. Il s'agit de décomposer la série en sous-séries ou groupes, d'ordre (ou de taille) t. On prendra les médianes pour la variable X et les moyennes pour la variable Y. Afin de faciliter la détermination des médianes, on retient un ordre t impair, généralement 3, et on prendra soin d'ordonner les modalités x_i par ordre croissant.

La formule des moyennes mobiles est simple :

Si Y = y ; y ; y ;…… ; y_t ,

Les moyennes mobiles, d'ordre 3, seront :

$$\mathbf{\overline{Y}_t = (y_{t-1} + y_t + y_{t+1}) /3}$$

$$\overline{Y}_2 = (y_1 + y_2 + y_3) /3$$

$$\overline{Y}_3 = (y_2 + y_3 + y_4) /3$$

Cette méthode fait perdre de l'information aux extrémités de la série, et ce, d'autant plus que le nombre de modalités observées est grand. Cette méthode est particulièrement intéressante pour corriger les variations saisonnières dans les séries chronologiques.

Exemple :

Reprendre les données de l'exemple 2 et déterminer les moyennes mobiles d'ordre 3 (ou décomposer ou ajuster la série en utilisant la méthode des moyennes mobiles d'ordre 3).

Solution :

$\overline{Y}_2 = (106 + 110 + 124) /3 = 113{,}3 \Rightarrow x_2 = 85$ (Médiane)

$\overline{Y}_3 = (110 + 124 + 120) /3 = 118 \Rightarrow x_3 = 90$ (Médiane)

$\overline{Y}_4 = (124 + 120 + 101) /3 = 115 \Rightarrow x_4 = 92$ (Médiane)

$\overline{Y}_5 = (120 + 101 + 102) /3 = 107{,}7 \Rightarrow x_5 = 98$ (Médiane)

On aura donc à représenter sur le graphique les quatre points suivants :

$(x_i ; y_i)$ = { (113,3 ; 85) ; (118 . 90) ; (115 ; 92) ; (107,7 ; 98) }.

3. L'ajustement analytique :

C'est la méthode la plus utilisée en statistique, en particulier pour des séries non chronologiques. Elle consiste à traiter le nuage de points obtenu par la méthode graphique en l'ajustant par une droite. En effet, lorsque le nuage de points suggère une droite ou une tendance à la linéarité (figure A), on cherchera alors une droite d'ajustement (appelée aussi droite de régression), parmi toutes les droites possibles, qui minimise les écarts entre les points situés sur la droite (notés y_c), qui sont des points estimés, et les points non situés sur la droite (notés y_i), qui sont des points observés. Pour ce faire, on fait recours au principe des moindres carrés ordinaires (MCO/OLS)*.

Le principe des moindres carrés stipule que la somme des carrés des écarts est minimum. Si on note nos écarts $\mathbf{e_i}$, on obtient :

$\mathbf{e_i} = (y_i - y_c)$, avec ; $\Sigma \mathbf{e_i}^2$ ---> Minimum.

L'ajustement analytique consiste donc à construire l'équation mathématique de la droite d'ajustement qui permet de minimiser ces écarts. C'est une équation simple de type y = ax + b permettant de relier les deux variables (xi et yi). Cette équation, construite à partir de données réelles (informations recueillies) d'une période donnée, permettra par la suite d'établir des prévisions pour la période à venir.

Il suffit alors de déterminer la valeur des paramètres **a** et **b** de l'équation telles que :

$\mathbf{e_i} = (y_i - y_c)$ ---> On pose $\Sigma e_i^2 = G$, on aura :

$G = [y_i - (ax + b)]^2$ = Minimum.

---

* MCO : Moindres Carrés Ordinaires
OLS : Ordinary Least Squares

Du point de vue mathématique, pour que la fonction G admette un minimum, il faut que sa dérivée première (dérivées partielles par rapport à a et à b) s'annule et que sa dérivée seconde soit positive :

$\frac{\partial' G}{\partial \ a} = 0$ et $\frac{\partial' G}{\partial \ b} = 0$ Avec $\frac{\partial'' G}{\partial a} > 0$ et $\frac{\partial'' G}{\partial b} > 0$

En déterminant ces dérivées partielles et après développement mathématique, on aura deux formules de **a** :

$$\mathbf{a} = \frac{\overline{XY} - \bar{X} \cdot \bar{Y}}{\overline{X^2} - (\bar{X})^2}$$

et $\mathbf{b} = \bar{Y} - a\,\bar{X}$

$$\mathbf{a} = \frac{\overline{XY} - \bar{X} \cdot \bar{Y}}{\overline{X^2} - (\bar{X})^2}$$

$$\mathbf{a} = \frac{\text{Cov}(x\,;\,y)}{V(x)} = \frac{(xi - \bar{X})(yi - \bar{Y})}{\overline{X^2} - (\bar{X})^2}$$

et $\mathbf{Y} = \mathbf{aX} + \mathbf{b} \Rightarrow \bar{Y} = a\,\bar{X} + b \Rightarrow b = \bar{Y} - a\,\bar{X}$

Section 3 : L'analyse de la Corrélation

En général, on rencontre trois situations concernant la relation entre deux variables [1] :

- Une relation totale, appelée aussi relation fonctionnelle, où les variations d'une variable sont expliquées exclusivement et totalement par les variations de l'autre. Dans ce cas les deux variables sont totalement dépendantes. Ce type de relation est très rare en pratique, notamment dans le domaine socio-économique ;

[1] Madjid HADJEM, **Polycopié Cours de Statistique I**, Semestre 1, 1ere Année Licence, 1 - Département des Sciences Commerciales, Faculté des Sciences Economique, Commerciale et des Sciences de Gestion, Université Mouloud MAMMERI de Tizi-Ouzou, 2023/2024, P : 100, https://dspace.ummto.dz/server/api/core/bitstreams/01256031-7855-41ea-ad3c-dbb7ab165058/content

- Absence de relation, signifiant que les variations d'une variable n'ont aucun effet sur les variations de l'autre. Les deux variables sont dans ce cas indépendantes ;

- Une relation relative, appelée aussi liaison ou dépendance partielle, où les variations d'une variable sont en grande partie, mais non en totalité, expliquées par celles de l'autre. C'est le cas le plus fréquent dans le domaine socio-économique. Une part des variations est laissée au hasard ou aux influences de l'environnement. La dépendance partielle suppose donc une partie certaine importante expliquée par les variations de l'une des variables ($y = f(x)$), et une partie incertaine ou aléatoire, peu importante et réduite non expliquée par les variations de l'une des variables mais par le hasard, on l'exprime souvent par le symbole Epsilon « $\boldsymbol{\varepsilon}$ » pour signifier son caractère négligeable dans la formulation mathématique ($y = f(x) + \varepsilon$).

Dans la relation linéaire simple, on parle d'analyse de la corrélation linéaire simple. Celle-ci mesure le degré ou l'intensité de la liaison entre deux variables. Elle consiste en la mesure d'un paramètre, appelé coefficient de corrélation linéaire de Pearson.

1. Le coefficient de corrélation :

Noté **r**, il est le rapport de la covariance entre X et Y au produit des écarts-types de X et Y. C'est un nombre sans dimension. On écrit :

$$\mathbf{r = Cov(x,y) / \delta_x.\delta_y}$$

ou bien :

$$\mathbf{r = a.(\delta x/\delta y)}$$

2. Interprétation de coefficient de corrélation [1] :

r varie entre -1 et 1: $\mathbf{r} \in [-1 ; 1]$.

Si :

a) r =1, les deux variables varient dans le même sens et la liaison est *totale* ou *fonctionnelle*. Cela signifie que la droite de régression (données estimées) s'ajuste parfaitement aux données réelles.

[1] Madjid HADJEM, op.cit., P : 101.

b) r = -1, les deux variables varient en sens inverse.

c) r = 0, pas de liaison entre les deux variables. La droite de régression est alors une droite de pente a = 0, soit parallèle à l'axe des abscisses, soit parallèle à l'axe des ordonnées : les variations d'une variable n'influence pas celles de l'autre.

d) r proche de 0, la relation entre les deux variables est faible.

e) r < 0, il existe une corrélation relative et inverse ou négative entre les deux variables, toute variation d'une variable entraine une variation en sens inverse de l'autre variable.

f) r > 0, il existe une corrélation *relative* ou *partielle* et positive entre les deux variables, toute variation d'une variable entraine une variation relative, dans le même sens, de l'autre variable. Dans ces deux derniers cas, une grande partie de la variation de Y est expliquée par les variations de X. L'autre partie, infime, est expliquée par le hasard.

g) r proche de ± 1, la relation est forte.

Il ne faut cependant pas oublier qu'on ne peut juger de la qualité de l'ajustement, comme on vient de le démontrer, à partir du seul examen du coefficient de corrélation. L'examen graphique et la signification des variables sont des compléments nécessaires au constat établi avec **r**.

Au-delà de la mesure de l'intensité de la relation entre les deux variables, on peut aussi mesurer la partie de Y expliquée par les variations de X. C'est à dire sur la totalité des variations de Y, la part qui est induite par les variations de X, soit :

% = variation expliquée / variation totale

Il existe un paramètre qui permet de mesurer cela, on l'appelle le **coefficient de détermination**, noté « $\mathbf{r^2}$ », et qui est exprimé en pourcentage**.** Il n'est rien d'autre que le carré du coefficient de corrélation. On écrit :

$$\mathbf{r^2 = (r)^2 . 100}$$

Ainsi, par exemple, si r = 0,6 $\Rightarrow$ $r^2 = 0,36 . 100 = 36\%$. Autrement dit, 36% de la variation totale de Y est due à sa relation avec X ou due à la variation de X. Ou bien encore, que 36% des variations de Y dépendent des variations de X.

Exemple :

Une agence de voyage a réalisé une étude pour savoir s'il existe une relation entre le prix moyen mensuel de billet (X en 1000 DA) et le nombre de clients qui ont réservé un voyage (Y). Les résultats sont donnes ci-après :

X	7	9	9.5	11	12.5	13	15	17	19	21
Y	90	110	130	150	170	190	210	230	250	270

1. Tracer le diagramme de dispersion du nombre de clients qui ont réservé un voyage en fonction du prix moyen de billet. Conclure quant à la nature de la relation.

2. Calculer un paramètre qui permet de mesurer l'intensité de la relation linéaire entre les deux variables étudiées. Interpréter votre résultat.

3. Déterminer l'équation de la droite donnant le nombre de client qui ont réservé un voyage en fonction du prix moyen de billet.

4. Déterminer le nombre de clients qui ont réservé pour des prix moyens de billet qui sont respectivement de 8000 DA et 10000 DA.

Solution :

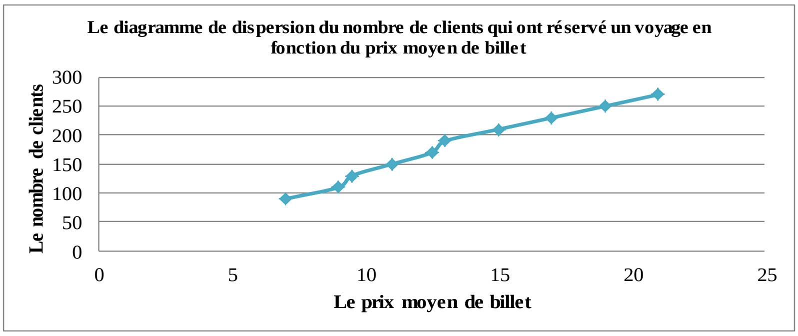
La population est représentée par dix (10) observations portant sur dix (10) clients d'une agence de voyage.

Alors que les variables utilisées sont définies par :

▪ **X** : « le prix moyen mensuel de billet en 1000 DA » comme variable exogène.

▪ **Y** : « le nombre de clients qui ont réservé un voyage suite à un prix affiché par l'agence » comme variable endogène.

1. Le diagramme de dispersion du nombre de clients qui ont réservé un voyage en fonction du prix moyen de billet est présenté comme suit :



On constate que le nombre de clients ayant réservé un voyage évolue positivement et linéairement au prix moyen de billet.

2. Le paramètre qui permet de **mesurer l'intensité de la relation** linéaire entre les deux variables étudiées est représenté par le **coefficient de corrélation** (r) définit comme suit :

$$r = \frac{Cov(x,y)}{\delta(x) * \delta(y)}$$

D’où le tableau et les résultats suivants :

	$\bar{X}$	$\bar{Y}$	$\overline{X^2}$	$\overline{Y^2}$	$\overline{XY}$
	X	**Y**	**X^2**	**Y^2**	**X.Y**
	7	90	49	8100	630
	9	110	81	12100	990
	9.5	130	90.25	16900	1235
	11	150	121	22500	1650
	12.5	170	156.25	28900	2125
	13	190	169	36100	2470
	15	210	225	44100	3150
	17	230	289	52900	3910
	19	250	361	62500	4750
	21	270	441	72900	5670
Σ	**134**	**1800**	**1982.5**	**357000**	**26580**

Avec :

$$r = \frac{Cov(x,y)}{\delta(x) * \delta(y)}$$

$$r = \frac{\overline{XY} - \overline{X} * \overline{Y}}{\sqrt{\overline{X^2} - (\overline{X})^2} * \sqrt{\overline{Y^2} - (\overline{Y})^2}}$$

- La moyenne arithmétique de X ($\overline{X}$) :

$\overline{X} = \frac{\sum X_i}{N}$ / Sachant que N = 10 (Nombre de situations)

$\overline{X} = \frac{134}{10}$

$\overline{X} = 13,4$

Le prix moyen des prix mensuels fixé pour les 10 clients (ou 10 derniers mois) pris dans cet échantillon est de 13400 DA.

- La moyenne arithmétique au carrée des X ($\overline{X})^2$:

$(\overline{X})^2 = (13,4)^2$

$(\overline{X})^2 = 179,56$

- La moyenne arithmétique des X au carrée ($\overline{X^2}$) :

$\overline{X^2} = \frac{\sum X_i^2}{N}$

$\overline{X^2} = \frac{1982,5}{10}$

$\overline{X^2} = 198,25$

- La moyenne arithmétique de Y ($\overline{Y}$) :

$\overline{Y} = \frac{\sum Y_i}{N}$

$\overline{Y} = \frac{1800}{10}$

$\overline{Y} = 180$

Le nombre moyen des clients ayant réservé un voyage sur les dix derniers mois est de 180 clients.

- La moyenne arithmétique au carrée des Y ($\overline{Y})^2$:

$(\overline{Y})^2 = (180)^2$

$(\overline{Y})^2 = 32400$

- La moyenne arithmétique des Y au carrée ($\overline{Y}^2$) :

$$\overline{Y}^2 = \frac{\sum Y_i^2}{N}$$

$$\overline{Y}^2 = \frac{357000}{10}$$

$$\overline{Y}^2 = 35700$$

- La moyenne arithmétique de $\overline{X} * \overline{Y}$:

$$\overline{X} * \overline{Y} = (13,4) * (180)$$

$$\overline{X} * \overline{Y} = 2412$$

- La moyenne arithmétique de XY :

$$XY = \frac{\sum X_i Y_i}{N}$$

$$XY = \frac{26580}{10}$$

$$XY = 2658$$

- Cov(x,y) :

$$Cov(x,y) = XY - \overline{X} * \overline{Y}$$

$$Cov(x,y) = 2658 – 2412$$

$$Cov(x,y) = 246$$

- L'écart type δ(x) :

$$\delta(x) = \sqrt{\ \overline{X}^2 - (\overline{X})^2}$$

$$\delta(x) = \sqrt{198,25 - 179,56}$$

$$\delta(x) = \sqrt{18,66}$$

$$\delta(x) = 4,32$$

En moyenne, le prix de billet s'écarte du prix moyen calculé sur les 10 derniers mois d'environ 4320 DA.

▪ L'écart type δ(y) :

$\delta(y) = \sqrt{\overline{Y^2} - (\overline{Y})^2}$

$\delta(y) = \sqrt{35700 - 32400}$

$\delta(y) = \sqrt{3300}$

$\delta(y) = 57{,}44 \approx 58$

En moyenne, le nombre des clients ayant réservé un voyage s'éloigne du nombre moyen des clients sur les dix derniers mois d'environ 58 clients.

▪ $r = \dfrac{\overline{XY} - \overline{X} * \overline{Y}}{\sqrt{\overline{X^2} - (\overline{X})^2} * \sqrt{\overline{Y^2} - (\overline{Y})^2}}$

$r = \dfrac{2658 - 2412}{4{,}32 * 57{,}44}$

$r = \dfrac{246}{248{,}14}$

$r = 0{,}99$

Le prix de billet et le nombre de clients sont dépendants positivement.

Enfin, on trouve :

Les deux variables observées dans cette population sont donc ***fortement*** et ***positivement corrélées*** puisque jusqu'à ***99%*** de la variabilité du nombre de clients est expliquée par la variabilité du prix moyen fixé par mois. Ainsi, il y a seulement ***1%*** des changements susceptibles sur le nombre de clients de cette agence qui est expliqué par ***les autres variables non observées***.

3. L'équation de la droite donnant le nombre de client qui ont réservé un voyage en fonction du prix moyen de billet est représentée par :

Y = aX + b ⇒ Dont les valeurs des deux paramètres ***a*** et ***b*** sont estimées par :

▪ $a = \dfrac{Cov(x,y)}{V(X)}$

$a = \dfrac{\overline{XY} - \overline{X} * \overline{Y}}{\overline{X^2} - (\overline{X})^2}$

$$a = \frac{2658 - 2412}{198{,}25 - 179{,}56}$$

$$a = \frac{246}{18{,}69}$$

a = 13,16

Ce qui signifie que lorsque le prix moyen par mois ***augmente*** de ***1%***, le nombre de clients ***augmente aussi*** d'environ ***13%***.

▪ b = $\bar{Y}$ - a $\bar{X}$

b = 180 – (13,16) * (13,4)

b = 180 – 176,34

b = 3,65

Ce qui représente le nombre constant de clients qui peut réserver un voyage ***indépendamment*** du prix moyen par mois du billet.

Ce qui donne, enfin, l'équation de la droite d'ajustement donnée par :

Y = aX + b

Y = 13,16X + 3,65

4. Sur la base de la droite précédente on peut donner les prévisions suivantes :

▪ Si le prix moyen de billet par mois est fixé à 8000 DA alors le nombre de clients qui peut réserver sera prévu à :

Y = (13,16) * (8) + 3,65

Y = 105,28 + 3,65

Y ≈ 109

Soit donc un nombre prévisionnel d'environ 109 clients.

▪ Si, par contre, le prix moyen de billet par mois est fixé à 10000 DA alors le nombre de clients qui peut réserver sera prévu à :

Y = (13,16) * (10) + 3,65

Y = 131,6 + 3,65

Y = 135,25

Y ≈ 135

Soit donc un nombre prévisionnel d'environ 135 clients.

Chapitre 2 :

Définitions et présentation graphique des séries temporelles

Chapitre 2 : Définitions et Présentation Graphique des Séries Temporelles

Les séries temporelles sont des ensembles de données chronologiques enregistrées à intervalles réguliers, qu'il s'agisse de données économiques, météorologiques, financières, ou autres. L'analyse des séries temporelles vise à comprendre les dynamiques sous-jacentes, à identifier les tendances, les cycles et les comportements saisonniers, et à prévoir les valeurs futures sur la base des observations passées.

Section 1 : Définition et objectifs des Séries Temporelles (ST)
Section 2 : La Représentation Graphique des Séries Temporelles (ST)
Section 3 : Les Composantes Fondamentales des Séries Temporelles (ST)

Section 1 : Définition et objectifs des Séries Temporelles

1. Définitions des Séries Temporelles

Une Série Temporelle (ST) est une suite ordonnée d'observations d'une grandeur chiffrée au cours du temps. Elle peut être définit comme une série statistique bivariée notée $(t\ , y_t)$.

Avec :

- **t :** Représente le temps (dates) t = 1, 2, 3, …, N / Sachant que N : Le nombre d'observations repérées par l'indice t.
- **y_t :** La variable numérique prenant ses valeurs dans le temps t.

Une Série Temporelle (ST) est donc toute suite d'observations correspondant à la même variable. Il peut s'agir de données :

a) Macroéconomiques : Le PIB d'un pays, le chômage, l'inflation, les exportations, la croissance, ... ;

b) Microéconomiques : Les ventes d'une entreprise donnée, le chiffre d'affaires (CA), le cours d'une action, nombre d'employés, le revenu d'un individu, ... ;

c) Démographiques : La taille moyenne des habitants, leur âge, le nombre d'enfants à la naissance, … ;

d) Politiques : Le nombre de votants, de voix reçues par un candidat, … ;

e) Santé : Le taux de sucre, électrocardiogrammes … ;

f) Métrologiques : La température, la pluviométrie, le nombre de jours de soleil, … .

Exemple :

Le nombre de ventes réalisées par l'entreprise ALPHA par trimestre est donné par le tableau suivant :

Trimestre / Année	1	2	3	4
2020	860	794	1338	1148
2021	1096	1021	1705	1505
2022	1436	1363	2319	2047

N = 12 observations

$t = 1, 2, 3, \ldots, 12$ Trimestres

$y_t : t = 1 \Rightarrow t = 12$

- $y_1 = 860 \Rightarrow$ Correspond à la première observation de la série (l'observation du premier trimestre de la première année) ;
- $y_2 = 794 \Rightarrow$ Correspond à la deuxième observation de la série (l'observation de deuxième trimestre de la première année) ;
- $y_7 = 1705 \Rightarrow$ Correspond à la septième observation de la série (l'observation de troisième trimestre de la deuxième année) ;
- $y_{10} = 1363 \Rightarrow$ Correspond à la dixième observation de la série (l'observation de deuxième trimestre de la troisième année).

2. Les Objectif des Séries Temporelles (ST) :

L'étude des Séries Temporelles (ST) répond aux objectifs suivants :

- Décrire l'évolution dans le temps (Compréhension du phénomène dans le temps) ;
- En permettre l'explication en guidant l'interprétation (Représentation simplifiée par un modèle) ;
- Permettre d'établir des prévisions futures (Obtenir des prévisions d'un phénomène à partir de passé) ;
- Etablir une Prévision (Il s'agit ici de supposer que les mêmes causes produisent les mêmes effets) ;
- Créer des liens entre les variables, afin d'établir des comparaisons ainsi que des corrélations entre eux.
- Par conséquent, on va pouvoir écarter certaines relations qui ne présentent aucun sens avec la série ou, au contraire, associer d'autres relations qui interagissent avec la série observée.

Section 2 : La Représentation Graphique des Séries Temporelles

On représente en général les Séries Temporelles (ST) sur des graphiques de valeurs (Ordonnées Y_t) en fonction du temps (Abscisses t) :

▪ Lorsqu'une série est stable autour de sa moyenne ($\bar{X}$), on parle de série stationnaire. Inversement on trouve aussi des séries non stationnaires ;

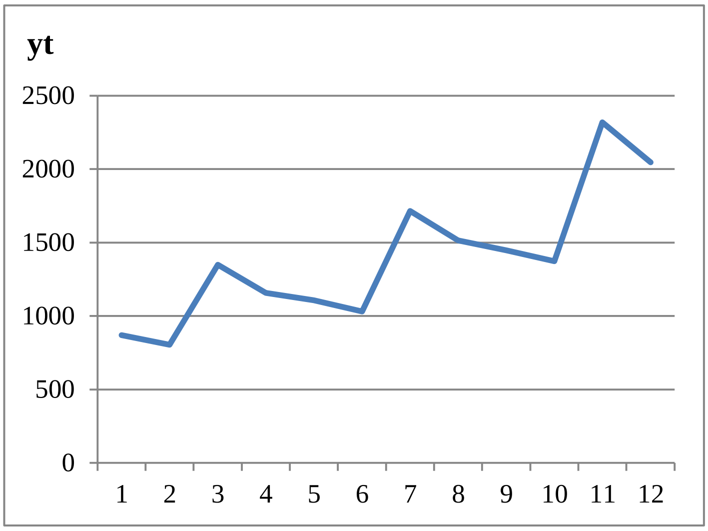
▪ Lorsqu'une série croit sur l'ensemble de l'échantillon et donc possède une moyenne qui n'est pas constante, on parle de Tendance ou Trend (T_t) ;

▪ Lorsqu'on observe des phénomènes qui se reproduisent à des périodes régulières on parle de phénomènes Saisonniers (S_t).

La série présentée dans le tableau précédent peut être représentée graphiquement en traçant la ligne brisée reliant les valeurs de y_t en fonction du temps t, telle que :

Trimestre \ Année	**2020**	**2021**	**2022**
1	860	1096	1436
2	794	1021	1363
3	1338	1705	2319
4	1148	1505	2047

Année	**t**	**y_t**
2020	1	860
	2	794
	3	1338
	4	1148
2021	5	1096
	6	1021
	7	1705
	8	1505
2022	9	1436
	10	1363
	11	2319
	12	2047



La lecture de la représentation graphique peut être plus enrichissante pour l'utilisateur dans la mesure où elle permet de détecter certaines des composantes de la série temporelle notamment la présence éventuelle de tendances de longue durée (Trend T_t), cycliques (C_t) ou parfois saisonnières (S_t) et mêmes accidentelles (ε_t) que nous exposerons par la suite.

Dans ce dernier graphe, par exemple, on peut facilement détecter la présence de :

▪ Tendances de longue durée (Trend T_t) vu l'évolution croissante des ventes de l'entreprise ALPHA dans le temps ;

▪ Variations saisonnières (S_t) étant donné les hauts et les bas enregistrés dans les ventes dans certaines périodes. Les 3º trimestres de chaque année présentent, par exemple, des piques correspondant aux ventes importantes alors que les 2º trimestres présentent des creux où les plus faibles ventes sont réalisées ;

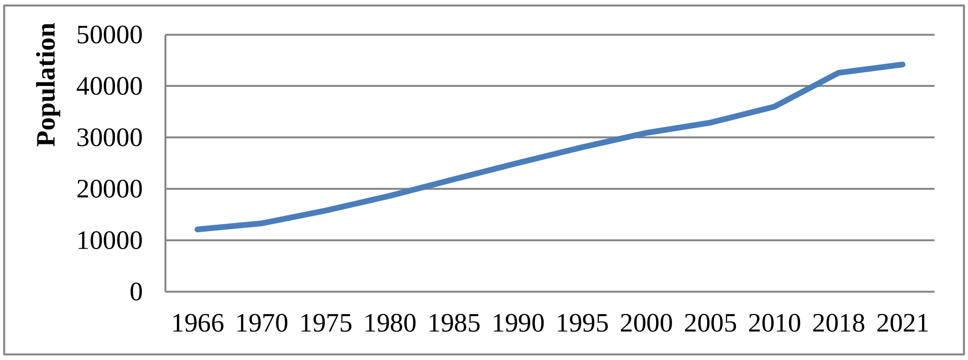
▪ Tendances cycliques (C_t) aussi puisque les mêmes mouvements périodiques sont pratiquement repris de la même façon sinon avec une amplitude élevée et ce sur la longue durée.

Ces diverses composantes peuvent éventuellement interagir différemment sur l'évolution globale de la série selon leur présence.

Afin de mieux éclaircir cette situation, nous présentons dans ce qui suit quelques autres exemples de graphes.

1° CAS : Les données relatives à la croissance de la population Algérienne 1966-2021 (en milliers)

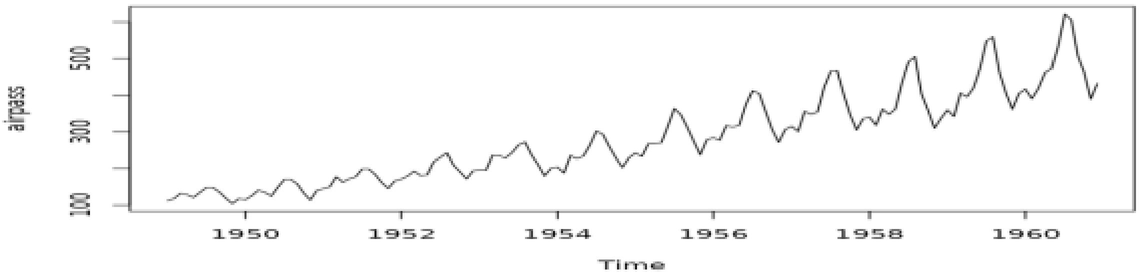
t	y_t
1966	12096
1970	13309
1975	15768
1980	18666
1985	21863
1990	25022
1995	28060
2000	30879
2005	32906
2010	35978
2018	42578
2021	44180



Qui montre la présence nette de tendance (T_t) mais pas de saisonnalité (S_t).

2° CAS : Monthly airline passenger numbers 1949 to 1960

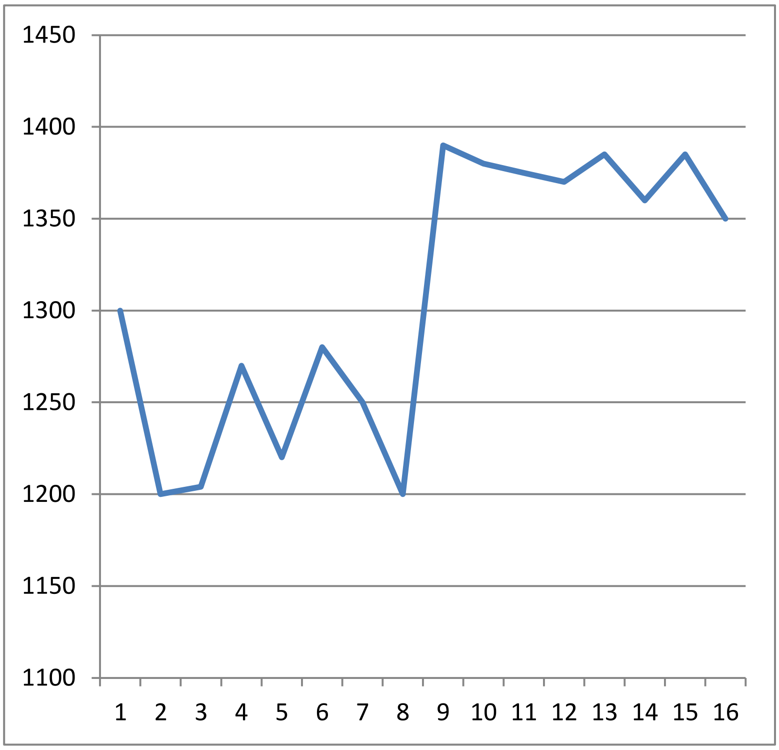
(Nombre mensuel de passagers aériens 1949-1960)



Qui présente aussi bien les effets de tendance (T_t) que les variations saisonnières (S_t) sur plusieurs années.

3° CAS : Les données concernant le cours journalier d'une action au cours d'un mois

t	y_t
1	1300
2	1200
3	1204
4	1270
5	1220
6	1280
7	1250
8	1200
9	1390
10	1380
11	1375
12	1370
13	1385
14	1360
15	1385
16	1350



L'observation de la courbe montre clairement l'absence de tendance (Trend T_t) et de saisonnalité. Seules les variations aléatoires qui peuvent être notées.

Section 3 : Les composantes fondamentales des series temporelles

Ce sont les éléments constitutifs de l'évolution globale ; on les appelle les « composantes » de la Série Temporelle (ST). Toute Série Temporelle (ST) peut être analysée comme étant la combinaison de différents mouvements qui sont les « composantes » de la série.

En général, une Série Temporelle (ST) y_t est la résultante de plusieurs composantes fondamentales, notamment :

1. La Tendance - Trend (T_t) :

Notée habituellement par (T_t), la *tendance* demeure la composante la plus fréquente pour une Série Temporelle (ST). Appelée également *Trend*, elle représente l'évolution à long terme de la série étudiée (ou du phénomène étudié). Elle traduit le comportement « moyen » de la série.

Par exemple : La croissance des ventes d'un centre commercial au cours des 12 mois de l'année.

a) Les différentes natures de la fonction de tendance :

Généralement, le trend (T_t) peut être de plusieurs natures : Linéaire, exponentiel, polynomial ou autre. Ceci diffère selon la nature des variables étudiées ainsi que la nature de leurs évolutions.

- **Linéaire :**

Formule : Le modèle de tendance linéaire est le suivant :

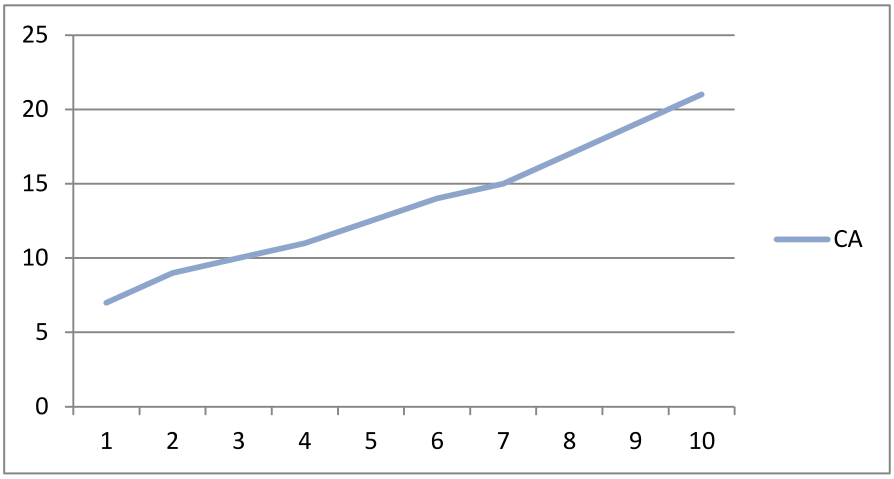
$y_t = \alpha + \beta\, t + \varepsilon_t$

Notation :

Terme	Description
y_t :	La variable numérique prenant ses valeurs dans le temps t.
α :	Constante
β :	Variation moyenne d'une période à la suivante
t :	Représente le temps (dates)
ε_t :	Terme d'erreur

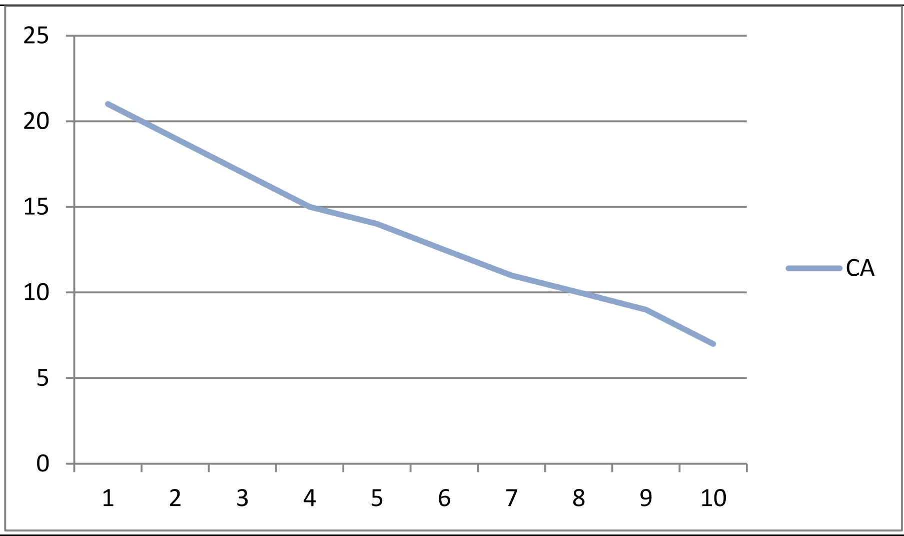
t	Durée (mois)
y_t	Chiffre d'affaires (CA)

t	**y_t**
1	7
2	9
3	10
4	11
5	12,5
6	14
7	15
8	17
9	19
10	21



t	Durée (mois)
y_t	Chiffre d'affaires (CA)

t	**y_t**
1	21
2	19
3	17
4	15
5	14
6	12,5
7	11
8	10
9	9
10	7



- **Croissance Exponentielle :**

Formule : Le modèle de tendance à croissance exponentielle prend en compte une croissance ou une décroissance exponentielle. Par exemple, un compte d'épargne peut présenter une croissance exponentielle.

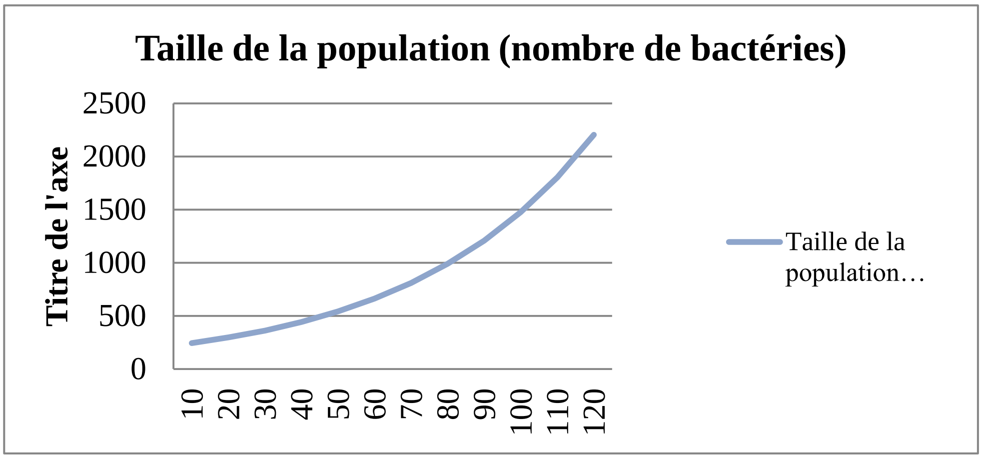
$y_t = \alpha * e^{\beta t}$

Notation :

Terme	Description
y_t :	La variable numérique prenant ses valeurs dans le temps t.
α :	Constante
β :	Coefficient
t :	Représente le temps (dates)

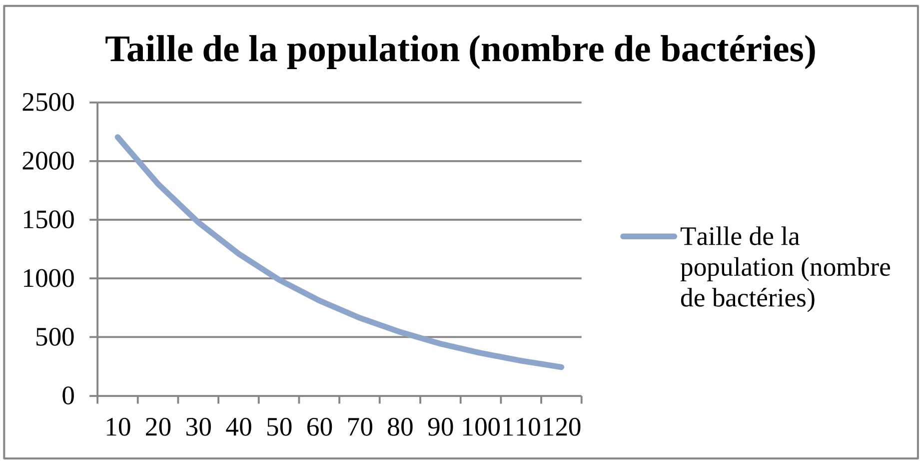
t	Durée (min)
y_t	Taille de la population (nombre de bactéries)

t	y_t
10	244
20	298
30	364
40	445
50	544
60	664
70	811
80	991
90	1210
100	1478
110	1805
120	2205



t	Durée (min)
y_t	Taille de la population (nombre de bactéries)

t	y_t
10	2205
20	1805
30	1478
40	1210
50	991
60	811
70	664
80	544
90	445
100	364
110	298
120	244



- **Quadratique :**

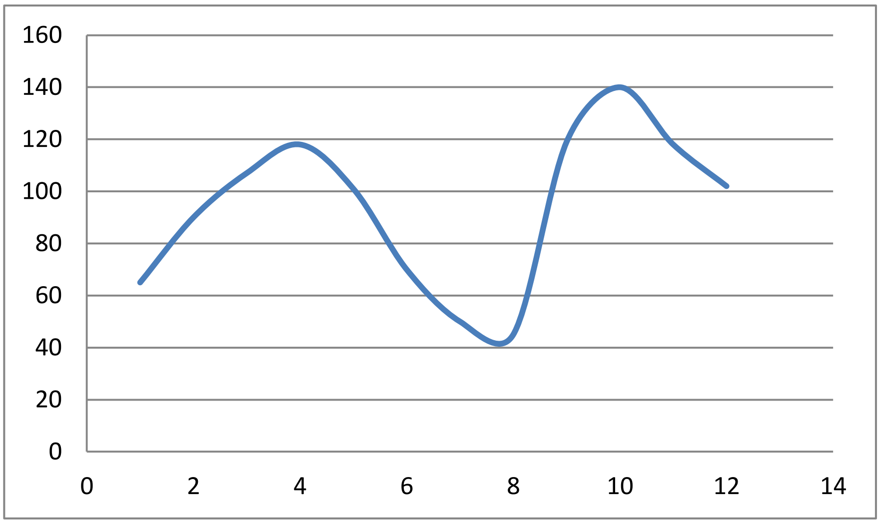
Formule : Le modèle de tendance quadratique, qui peut rendre compte d'une courbure simple dans les données, est le suivant :

$y_t = \alpha + \beta_1 t + \beta_2 t^2 + \varepsilon_t$

Notation

Terme	**Description**
$\mathbf{y_t}$ **:**	La variable numérique prenant ses valeurs dans le temps t.
α :	Constante
$\boldsymbol{\beta_1}$ **et** $\boldsymbol{\beta_2}$ **:**	Coefficients
t :	Représente le temps (dates)
$\boldsymbol{\varepsilon_t}$ **:**	Terme d'erreur

t	y_t
1	65
2	90
3	107
4	118
5	101
6	70
7	50
8	45
9	119
10	140
11	118
12	102



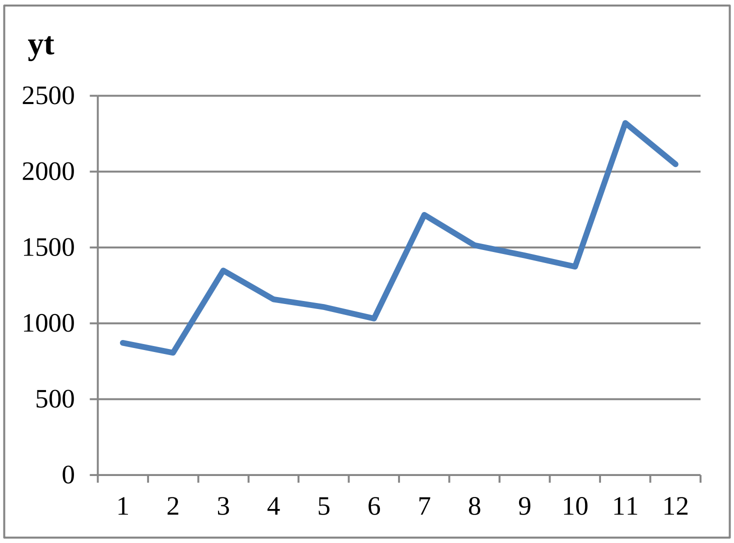
2. La Composante Saisonnière - Saisonnalité (S_t) :

Cette composante correspond à un phénomène qui se répète à intervalles de temps réguliers (périodique). D'où les variations saisonnières comme le cas des jours fériés par exemple.

On note généralement sa période par (P) qui représente les facteurs saisonniers engendrés par les mouvements saisonniers, tel que :

- **P = 2** $\Rightarrow$ Pour les séries ou les variations semestrielles (2 Semestres).
- **P = 4** $\Rightarrow$ Pour les séries ou les variations trimestrielles (4 Trimestres).
- **P = 12** $\Rightarrow$ Pour les séries ou les variations mensuelles (12 Mois).

Année	t	y_t
2020	1	860
	2	794
	3	1338
	4	1148
2021	5	1096
	6	1021
	7	1705
	8	1505
2022	9	1436
	10	1363
	11	2319
	12	2047



3. Les Mouvements Cycliques (C_t) :

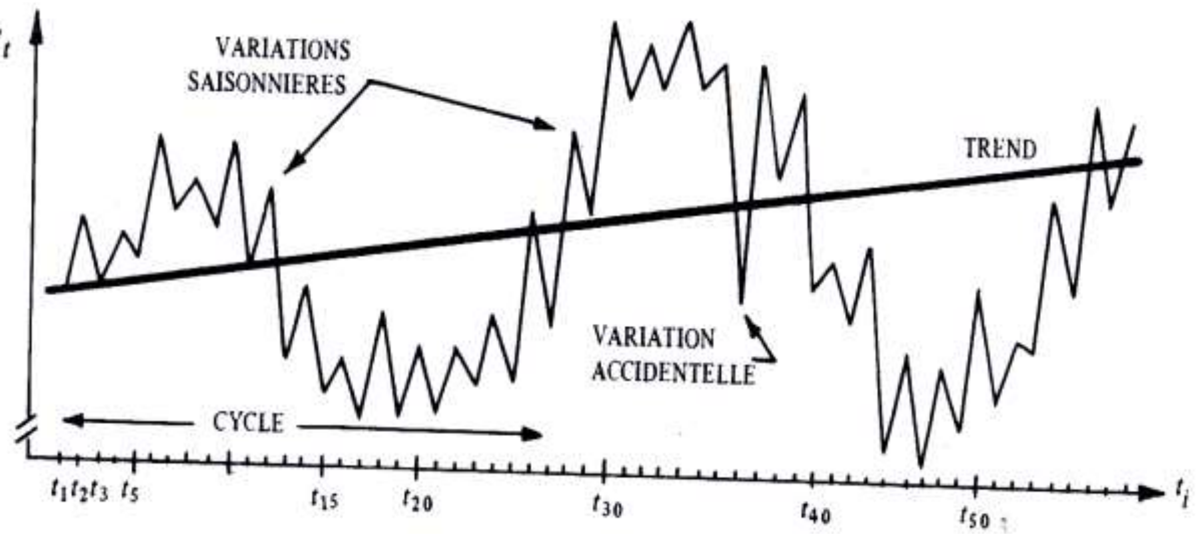
Ce sont les mouvements rencontrés sur une grande période autour du trend. Ces mouvements peuvent être périodiques. Ce qui est représenté par des variations ponctuelles de fortes amplitudes.

Exemple : L'expansion ou la récession (crise) économique de long terme.

4. La Composante Résiduelle - Résidu - Bruit (ε_t) / (R_t) :

Elle représente les fluctuations irrégulières ou les mouvements accidentels.
Elles sont, en général, de faible intensité et de nature aléatoire (imprévisibles). D'où le terme des aléas comme le cas des grèves, de guerre, des inondations, de crash financier, qui ont un caractère imprévisible et non contrôlable.

Ces fluctuations sont généralement gênantes dans le sens où elles influencent l'évolution de la série mais leur effet est faible dans le temps. D'où la nécessité de les enlever de la série brute avant toute analyse statistique de celle-ci.



Modèle-Type d'une Série Temporelle (ST)

Si l'on observe le schéma, ci-dessus, qui représente la représentation graphique « Idéal », « Typique » ou de « Référence » d'une série Temporelle, on peut déduire qu'il existe quatre (04) composantes dans une série Temporelle :

- La composante globale ou le Trend (T_t) ;
- La composante saisonnière (S_t) ;
- La composante cyclique (C_t) ;
- La composante accidentelle ou résiduelle (ε_t) / (R_t).

Ces quatre composantes forment le *modèle ou le schéma « Idéal »* de la série Temporelle.

Si la présentation d'une ST à travers sa définition et sa représentation graphique est utile dans le sens d'avoir une indication sur son évolution, elle est, malheureusement, souvent insuffisante pour l'analyse. Ce qui doit être aussi important c'est le fait de savoir comment une ST peut être décomposée afin de dissocier entre ses diverses composantes et mesurer par la suite leurs effets sur cette dernière. D'où, les modèles de décomposition qui permettent d'en préciser le contenu.

5. Quelques exemples pour bien comprendre[1] :

A quelles composantes d'une série temporelles correspondent les expressions suivantes ?

a) Incendie dans une usine entrainant un retard de 15 jours de la production.

b) Evolution moyenne du prix d'un produit entre 1993 et 2013.

c) Ventes d'un magasin pendant le mois de juin entre 1992 et 2012.

d) Ces mêmes ventes pendant le mois du ramadhan de 1996.

e) Une période entre deux crises.

f) Production mensuelle de blé en Algérie de 1992 à 2012.

Réponses :

a) Composante accidentelle.

b) Composante globale ou cyclique.

c) Composante saisonnière.

d) Composante accidentelle.

e) Composante cyclique.

f) Composante globale.

[1] Madjid HEDJEM, op.cit., P 147.

Chapitre 3 :
Les Modèles de décomposition des séries temporelles

Chapitre 3 : Les modèles de décomposition des séries temporelles

Nous présentons les modèles qui nous permettent généralement de décomposer la Série Temporelle (ST) à travers ses différentes composantes. Ce qui représente le fait de la décortiquer en plusieurs éléments la composant dans le but de séparer par la suite leurs effets éventuels.

Deux modèles peuvent être éventuellement utilisés dans la décomposition des éléments d'une Série Temporelle (ST) :

- Le modèle Additif ;
- Le modèle Multiplicatif.

Ces derniers doivent être utilisés au choix non subjectif de l'utilisateur étant donné l'existence de plusieurs critères sur lesquels on se base généralement et qui représentent les caractéristiques des modèles de décomposition.

Section 1 : Le modèle additif

Section 2 : Le modèle multiplicatif

Section 3 : Test du modèle (modèle de Bays Ballot)

Section 1 : Le modèle Additif

Qui se caractérise par le fait d'avoir :

1. Toutes les composantes de la Série Temporelle (ST) sont supposées être indépendantes et s'ajoutent notamment les unes aux autres dans leur influence de la série étudiée, telle que :

$$\mathbf{Y_t = T_t + S_t + C_t + \varepsilon_t}$$

$$\mathbf{Y_t = T_t + S_t + C_t + R_t}$$

2. L'amplitude (c.à.d. la variation entre les bornes ou les valeurs supérieure et inférieure) de la composante saisonnière (S_t) et du bruit ($\boldsymbol{\varepsilon_t}$) reste constante au cours du temps (Cf. le graphe ci- dessous). Cela signifie que ces composantes gardent pratiquement les mêmes effets et, en conséquence, la série temporelle va subir les mêmes fluctuations aussi.

3. Les mêmes composantes (S_t) et (ε_t) sont de faible influence mais aussi de moyenne nulle :

$\mathbf{\sum S_t = 0}$ $\quad / \, t \in \{1, \ldots, P\}$
et $\mathbf{\sum \varepsilon_t = 0}$ $\quad / \, t \in \{1, \ldots, N\}$

Ce qui signifie qu'elles sont centrées, leurs effets s'annulent en quelque sorte dans le temps et, par conséquent, toute l'information concernant la tendance est uniquement contenue dans le Trend (T_t).

4. Dans le cas d'un schéma additif la somme des coefficients doit être nulle.

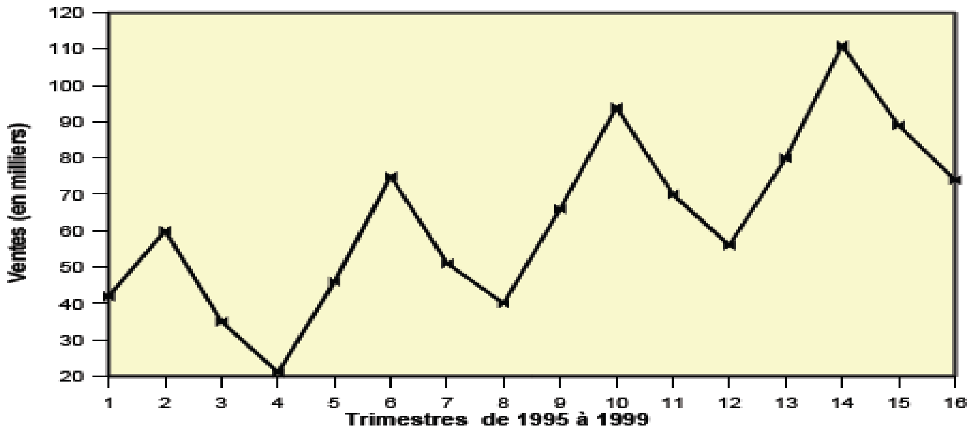
Soit S_j le $j^{ème}$ coefficient provisoire, on calcule la somme des coefficients :

$\mathbf{S = \sum S_j}$

- Si **S = 0,** les S_j sont les coefficients saisonniers définitifs ;
- Si **S ≠ 0,** les coefficients saisonniers sont normés afin que leur somme soit nulle.

Les coefficients définitifs sont donnés par : $\mathbf{S_j^* = S_j - \bar{S}}$

En termes de représentation graphique :



Comme nous l'avons souligné, à partir de la représentation graphique on peut facilement détecter à travers l'évolution de la série, ses différentes composantes les plus apparentes comme on peut extraire aussi une première analyse de description de la variable en question.

En effet, la comparaison par exemple des courbes joignant les extremums de la série nous permet d'avoir une idée sur l'amplitude notamment des composantes saisonnières et résiduelles. Comme cela est montré sur la représentation graphique, celle-ci parait en effet *constante* en représentant ainsi les mêmes fluctuations de la série dans le temps.

L'évolution de longue durée, ou le **Trend**, peut être également visualisée à travers la droite de centre et qu'on pourra certainement estimée par la suite.

Comme on peut aussi constater l'existence des effets cycliques (C_t) qui sont nettement identifiés sur ce graphe vu que sur le long terme, la série tend à l'augmentation et ce malgré les effets de saisonnalité qui font que certaines diminutions sont enregistrées (voir les piques positifs et négatifs de saisonnalité).

Ces différents constats peuvent être également faits sur le graphe suivant qui représente cette fois l'évolution trimestrielle des ventes d’un produit. Le modèle de décomposition doit être certainement aussi représenté par une addition des différentes composantes de la série.

Section 2 : Le Modèle Multiplicatif

Une des caractéristiques fondamentales du modèle additif reste celle liée aux mouvements *saisonniers* et *cycliques* qui sont supposés être indépendants du niveau Yt atteint sur le Trend. Si, par contre, on pense que les variations saisonnières et cycliques suivent l'évolution de la grandeur étudiée, dans ce cas, on peut adopter plutôt un modèle multiplicatif tel que la relation qui lie ses composantes est donnée par :

$\mathbf{Y_t = T_t * S_t * C_t * \varepsilon_t}$

$\mathbf{Y_t = T_t * S_t * C_t * R_t}$

On peut le rendre additif en utilisant le log :

$$\log \boldsymbol{Y} = \log \boldsymbol{T} + \log \boldsymbol{S} + \log \boldsymbol{C} + \log \boldsymbol{R}$$

Le modèle multiplicatif se caractérise généralement par :

1. Les composantes de la ST ne sont plus indépendantes mais elles sont supposées se *multiplier* cette fois-ci. D'où l'effet de multiplication entre elles et par suite la multiplication de leurs effets. Ce qui renforce l'évolution de la série.

2. L'amplitude de la composante saisonnière et du bruit n'est plus constante. Ce qui revient à supposer que les fluctuations de la série varient au cours du temps proportionnellement à la tendance (T_t).

3. Les composantes (S_t) et ($\boldsymbol{\varepsilon_t}$) auront une influence parfois notable et leurs moyennes ne sont plus nulles, telles que :

$\sum \mathbf{S_t = P}$ $\quad$ / t ∈ {1, …, P}

et $\sum \boldsymbol{\varepsilon_t} \mathbf{= 1}$ $\quad$ / t ∈ {1, …, N}

4. Dans le cas d'un schéma multiplicatif la moyenne des coefficients, pour une année donnée, doit être égale à 1.

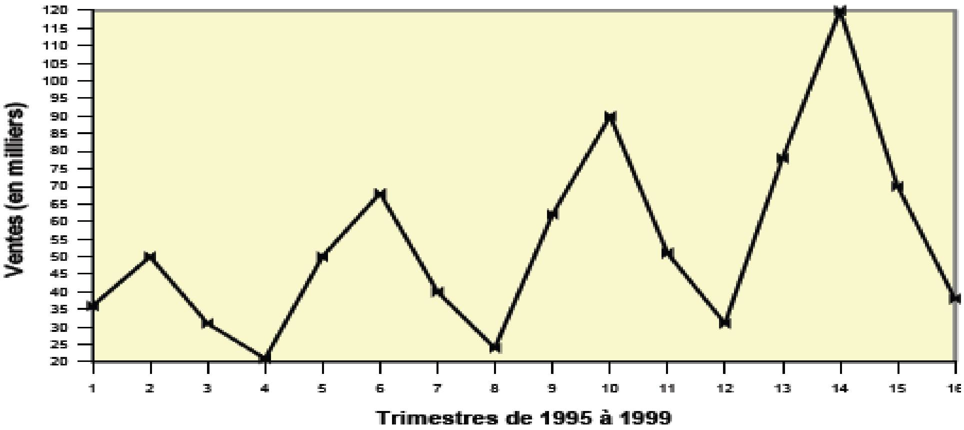
Soit S_j le $j^{\text{ème}}$ coefficient provisoire, on calcule la somme des coefficients :

$\mathbf{S = \sum S_j}$ avec P la période de la saisonnalité.

Les coefficients définitifs sont donnés par : $\mathbf{S_j^* = S_j / \bar{S}}$

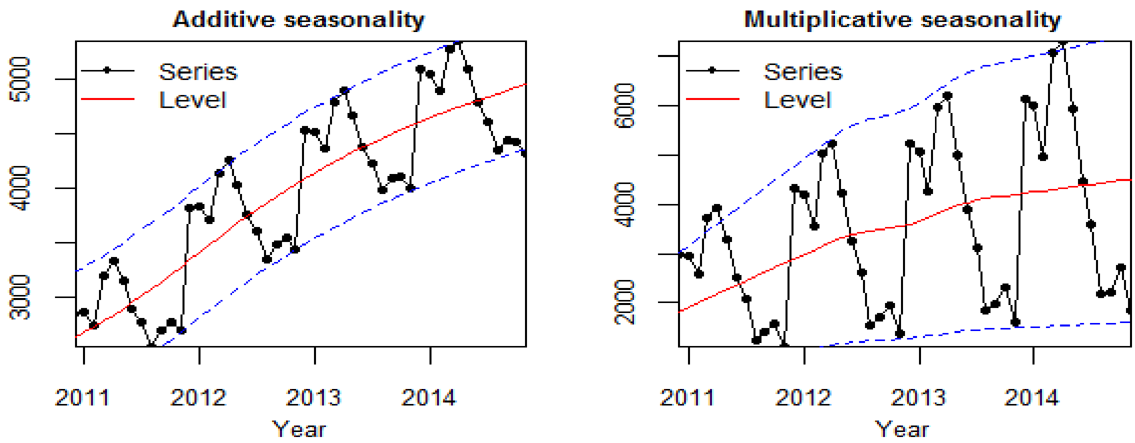
La méthode de dessaisonalisation doit être adaptée au type de saisonnalité : rigide ou souple.

Ces différentes caractéristiques sont facilement notables sur la représentation graphique suivante :



On s'aperçoit à partir du graphe ci-dessus, en effet, que l'amplitude des composantes saisonnières et résiduelles est variable (croissante dans ce cas), en révélant des variations différentes dans le temps. Ce qui signifie qu'elles ne sont plus faites dans la même cadence mais augmentent avec le temps. Le Trend est toujours représenté par la droite de centre marquant aussi une évolution soutenue en augmentant dans le temps. Les piques des variations saisonnières sont de plus en plus élevés.

Ces différentes variations que se soient de tendance ou saisonnières ou encore de résidus peuvent être encore mieux visualisées sur le graphe suivant :



Section 3 : Test du modèle (Modèle de Bays Ballot)

1. Définition de Test du modèle (Modèle de Bays Ballot) :

Il est fondé sur le calcul des moyennes et les écarts types par année :
Le modèle est additif : si la moyenne et l'écart type sont indépendants.
Le modèle est multiplicatif dans le cas contraire lorsque le nombre d'année est suffisant.

Nous pouvons estimer par les MCO/OLS les paramètres B_0 et B_1 de l'équation : $Sigma_i = B_0 + B_1 Moy_i + epsilon$
Avec **i** le nombre d'années i = 1, 2, …, N

Dans le cas où le coefficient B_1 est significativement différent de zéro (test de Student), alors on accepte l'hypothèse d'un modèle multiplicatif.
Dans le cas contraire nous retenons l'hypothèse d'un modèle additif.

2. Déroulement de test du modèle (Modèle de Bays Ballot) :

Test de Student :

$H_0 : \beta = 0 \iff$ Modèle additif

$H_1 : \beta \neq 0 \iff$ Modèle multiplicatif

$$T_c = \left| \frac{\hat{\beta} - 0}{\delta_{\hat{\beta}}} \right|$$

Si $T_c > t_{n-2}^{\propto/2}$ on accepte H_1

Si $T_c \leq t_{n-2}^{\propto/2}$ on accepte H_0

Exemple :

Soit la série X_i observé pendant 3 ans : P = 4, N = 3

Année Trimestre	**T1**	**T2**	**T3**	**T4**
2021	12	13	10	31
2022	8	10	11	29
2023	11	14	12	30

-Déterminer le modèle générateur de données pour la série X_t.

Solution :

Les résultats d'estimation sont donnés dans le tableau suivant :

Dependent Variable: ET
Method: Least Squares
Date: 10/13/24 Time: 20:58
Sample: 1 3
Included observations: 3

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	16.65486	3.897574	4.273135	0.1463
MOY	-0.478630	0.244385	**-1.958507**	0.3005

Déterminer le modèle générateur de données pour la série X_t

$H_0 : \beta = 0 \implies$ Modèle additif

$H_1 : \beta \neq 0 \implies$ Modèle multiplicatif

$$T_c = \left|\frac{\hat{\beta} - 0}{\delta_{\hat{\beta}}}\right| \leadsto \text{Student } (n-2)\text{dL}$$

Sous H_0 $T_c = \left|\frac{\hat{\beta}-0}{\delta_{\hat{\beta}}}\right| = \left|\frac{0,22}{\delta_{\hat{\beta}}}\right|$

$$t_{n-2}^{\propto/2} = t_{3-2}^{0,025} = t_1^{0,025} = 12,706$$

$t_c = 0,73 < t_1^{0,025} = 12,706$ On accepte $H_0 \Rightarrow$ (Modèle additif)

Chapitre 4 :
Les Modèles d'Ajustement et de Lissage des Séries Temporelles

Chapitre 4 :
Les Modèles d'Ajustement et de Lissage des Séries Temporelles

L'analyse des Séries Temporelles (ST) rend nécessaire la mise au point d'instruments spécifiques ayant pour objet de mettre en œuvre et de mesurer l'influence des mouvements saisonniers.

Ces instruments doivent permettre d'épurer **(nettoyer)** la Série Temporelle (ST) de l'effet perturbateur des fluctuations saisonnières, de calculer une série dite **désaisonnalisée**, ou encore **corrigée** des variations saisonnières plus facile à interpréter et plus utilisable si on veut prolonger le mouvement constaté dans l'avenir et formuler des prévisions d'évolution.

En effet, l'existence de ces effets non désirables influencent continuellement les variables économiques étudiées dans un sens qu'on ne peut pas les contrôler ni les reproduire. Ce qui explique que les Séries Temporelles doivent être ainsi nettoyées de ces derniers.

L'ajustement ou le **lissage** représente donc l'opération de **purifier** les ST des composantes saisonnières (S_t) et résiduelles (ε_t). Cette dessaisonalisation peut être décomposée en trois étapes essentielles :

1. Etape 1 : Estimer la Tendance Générale de la (ST)

C'est l'étape de la détermination du trend (T). On cherche à ce niveau à réduire les fluctuations et à dégager une évolution de long terme.

2. Etape 2 : Calculer les Valeurs du Mouvement Saisonnier (VMS) de la (ST)

IL s'agit à ce niveau de confronter les valeurs de la série brute et celle de la tendance estimée précédemment.

3. Etape 3 : Extraire de la Série Brute l'Influence du Mouvement Saisonnier
Afin d'obtenir la série corrigée des variations saisonnières.

Et pour ce faire, on dispose généralement de plusieurs techniques dont on cite dans ce cadre essentiellement :

-La Méthode des Moyennes Mobiles (MM) ;

-La Méthode des Moyennes Mobiles Centrées (MMC) ;

-La Méthode Exponentielle.

Le choix d'une méthode particulière doit se faire à partir de la nature de l'évolution des données (représentation graphique) c'est à dire le schéma de décomposition de la série temporelle selon s'il est **additif** ou **multiplicatif**.

Section 1 : La Méthode des Moyennes Mobiles (MM)

Section 2 : La Méthode des Moyennes Mobiles Centrées (MMC)

Section 1 : La Méthode des Moyennes Mobiles (MM)

La méthode des Moyennes Mobiles (MM) est la plus courante pour le lissage des Séries Temporelles (ST). Elle consiste à remplacer la série brute (informations collectées) ou les couples **(t_i ; y_t)** par une série de médianes (Me) et de moyennes ou par les couples **(t_i ; $\bar{y}_t$)** ou bien **(Me_t ; $\bar{y}_t$)**.

Il s'agit de décomposer la série en sous-séries ou groupes. On prendra les médianes pour la variable X ou t et les moyennes pour la variable Y. Afin de faciliter la détermination des médianes, on retient un ordre t impair, généralement 3, et on prendra soin d'ordonner les modalités xi par ordre croissant.

La méthode des Moyennes Mobiles (MM) permet de :

1. Lisser la ST en éliminant les mouvements saisonniers (S_t) et/ou accidentels (ε_t). Cela explique son utilisation quelque soit le modèle de décomposition additif ou multiplicatif ;

2. Mettre en évidence l'allure de la tendance (T_t) en supprimant la composante saisonnière (S_t) et en atténuant le bruit (ε_t). Ce qui signifie la possibilité de retrouver le trend (T_t) à partir d'une série brute en passant notamment par les moyennes mobiles (MM) dont l'ordre reste à préciser.

Les moyennes mobiles d'ordre **(p)** sont généralement notées par leurs initiales : MMp (t)

Avec :

t : Les dates moyennes de la ST ;

p : L'ordre de la Moyenne Mobile.

La formule générale de la Moyenne Mobile (MM) s'écrit :

$MM_p (t) = \bar{Y}_t = (y_{t-1} + y_t + y_{t+1}) / p$

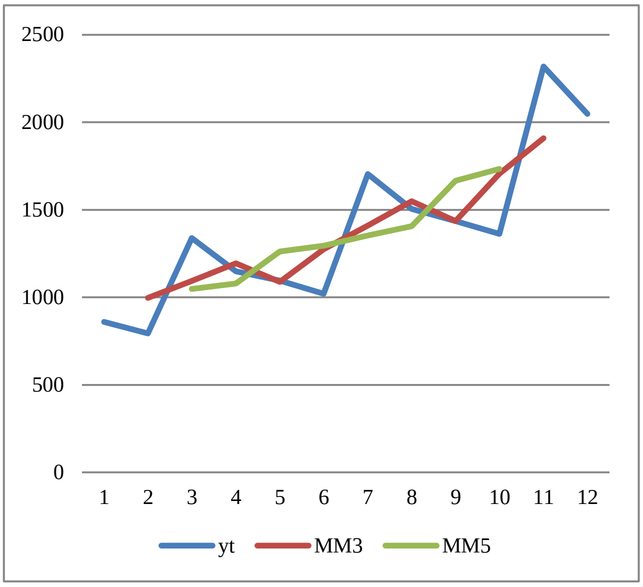
Si $p = 3 \Rightarrow MM_3 (t) = \bar{Y}_t = (y_{t-1} + y_t + y_{t+1}) / 3$

Si $p = 5 \Rightarrow MM_5 (t) = \bar{Y}_t = (y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2}) / 5$

Exemple :

On suppose maintenant calculer les Moyennes Mobiles (MM) d'ordre 3 et 5, on aura les nouvelles séries qui seront définies comme suit :

Année	T	y_t	MM_3	MM_5
2020	1	860	-	-
	2	794	997,33	-
	3	1338	1093,33	1047,20
	4	1148	1194,00	1079,40
2021	5	1096	1088,33	1261,60
	6	1021	1274,00	1295,00
	7	1705	1410,33	1352,60
	8	1505	1548,67	1406,00
2022	9	1436	1434,67	1665,60
	10	1363	1706,00	1734,00
	11	2319	1909,67	-
	12	2047	-	-



- Si $p = 3 \Rightarrow MM_3\ (t) = \overline{Y}_t = (y_{t-1} + y_t + y_{t+1}) / 3$
- Moyennes Mobiles MM_3 (t) :

$\overline{Y}_1 = (y_0 + y_1 + y_2) / 3 \Rightarrow \overline{Y}_1 =$ On ne peut pas le calculer car y_0 n'existe pas.

$\overline{Y}_2 = (y_1 + y_2 + y_3) / 3 \Rightarrow \overline{Y}_2 = (860 + 794 + 1338) / 3 \Rightarrow \overline{Y}_2 = 997{,}33$

$\overline{Y}_3 = (y_2 + y_3 + y_4) / 3 \Rightarrow \overline{Y}_3 = (794 + 1338 + 1148) / 3 \Rightarrow \overline{Y}_3 = 1093{,}33$

$\overline{Y}_4 = (y_3 + y_4 + y_5) / 3 \Rightarrow \overline{Y}_4 = (1338 + 1148 + 1096) / 3 \Rightarrow \overline{Y}_4 = 1194{,}00$

$\overline{Y}_5 = (y_4 + y_5 + y_6) / 3 \Rightarrow \overline{Y}_5 = (1148 + 1096 + 1021) / 3 \Rightarrow \overline{Y}_5 = 1088{,}33$

$\overline{Y}_6 = (y_5 + y_6 + y_7) / 3 \Rightarrow \overline{Y}_6 = (1096 + 1021 + 1705) / 3 \Rightarrow \overline{Y}_6 = 1274{,}00$

$\overline{Y}_7 = (y_6 + y_7 + y_8) / 3 \Rightarrow \overline{Y}_7 = (1021 + 1705 + 1505) / 3 \Rightarrow \overline{Y}_7 = 1410{,}33$

$\overline{Y}_8 = (y_7 + y_8 + y_9) / 3 \Rightarrow \overline{Y}_8 = (1705 + 1505 + 1436) / 3 \Rightarrow \overline{Y}_8 = 1548{,}67$

$\overline{Y}_9 = (y_8 + y_9 + y_{10}) / 3 \Rightarrow \overline{Y}_9 = (1505 + 1436 + 1363) / 3 \Rightarrow \overline{Y}_9 = 1434{,}67$

$\overline{Y}_{10} = (y_9 + y_{10} + y_{11}) / 3 \Rightarrow \overline{Y}_{10} = (1436 + 1363 + 2319) / 3 \Rightarrow \overline{Y}_{10} = 1706{,}00$

$\overline{Y}_{11} = (y_{10} + y_{11} + y_{12}) / 3 \Rightarrow \overline{Y}_{11} = (1363 + 2319 + 2047) / 3 \Rightarrow \overline{Y}_{11} = 1909{,}67$

$\overline{Y}_{12} = (y_{11} + y_{12} + y_{13}) / 3 \Rightarrow \overline{Y}_{12} =$ On ne peut pas le calculer car y_{13} n'existe pas.

Le premier couple de points sera donc (t_i ; $\overline{Y}_t$) = (t_2 ; $\overline{Y}_2$), sachant que t_2 est la médiane (Me).

- Si p = 5 $\Rightarrow$ MM_5 (t) = $\overline{Y}_t = (y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2}) / 5$
- Moyennes Mobiles MM_5 (t) :

$\overline{Y}_1 = (y_{-1} + y_0 + y_1 + y_2 + y_3) / 5 \Rightarrow \overline{Y}_1$ = On ne peut pas le calculer car y_{-1} et y_0 n'existent pas.

$\overline{Y}_2 = (y_0 + y_1 + y_2 + y_3 + y_4) / 5 \Rightarrow \overline{Y}_2$ = On ne peut pas le calculer car y_0 n'existe pas.

$\overline{Y}_3 = (y_1 + y_2 + y_3 + y_4 + y_5) / 5$

$\Rightarrow \overline{Y}_3 = (860 + 794 + 1338 + 1148 + 1096) / 5 \Rightarrow \overline{Y}_3 = 1047{,}20$

$\overline{Y}_4 = (y_2 + y_3 + y_4 + y_5 + y_6) / 5$

$\Rightarrow \overline{Y}_4 = (794 + 1338 + 1148 + 1096 + 1021) / 5 \Rightarrow \overline{Y}_4 = 1079{,}40$

$\overline{Y}_5 = (y_3 + y_4 + y_5 + y_6 + y_7) / 5$

$\Rightarrow \overline{Y}_5 = (1338 + 1148 + 1096 + 1021 + 1705) / 5 \Rightarrow \overline{Y}_5 = 1261{,}60$

$\overline{Y}_6 = (y_4 + y_5 + y_6 + y_7 + y_8) / 5$

$\Rightarrow \overline{Y}_6 = (1148 + 1096 + 1021 + 1705 + 1505) / 5 \Rightarrow \overline{Y}_6 = 1295{,}00$

$\overline{Y}_7 = (y_5 + y_6 + y_7 + y_8 + y_9) / 5$

$\Rightarrow \overline{Y}_7 = (1096 + 1021 + 1705 + 1505 + 1436) / 5 \Rightarrow \overline{Y}_7 = 1352{,}60$

$\overline{Y}_8 = (y_6 + y_7 + y_8 + y_9 + y_{10}) / 5$

$\Rightarrow \overline{Y}_8 = (1021 + 1705 + 1505 + 1436 + 1363) / 5 \Rightarrow \overline{Y}_8 = 1406{,}00$

$\overline{Y}_9 = (y_7 + y_8 + y_9 + y_{10} + y_{11}) / 5$

$\Rightarrow \overline{Y}_9 = (1705 + 1505 + 1436 + 1363 + 2319) / 5 \Rightarrow \overline{Y}_9 = 1665{,}60$

$\overline{Y}_{10} = (y_8 + y_9 + y_{10} + y_{11} + y_{12}) / 5$

$\Rightarrow \overline{Y}_{10} = (1505 + 1436 + 1363 + 2319 + 2047) / 5 \Rightarrow \overline{Y}_{10} = 1734{,}00$

$\overline{Y}_{11} = (y_9 + y_{10} + y_{11} + y_{12} + y_{13}) / 5 \Rightarrow \overline{Y}_{11}$ = On ne peut pas le calculer car y_{13} n'existe pas.

$\overline{Y}_{12} = (y_{10} + y_{11} + y_{12} + y_{13} + y_{14}) / 5 \Rightarrow \overline{Y}_{12}$ = On ne peut pas le calculer car y_{13} et y_{14} n'existent pas.

En termes de représentation graphique, on peut, constater que les séries obtenues à partir des Moyennes Mobiles (MM_3) et (MM_5) sont de plus en plus lissent et tendent, par conséquent, à se transformer en une droite. Ce qui montre l'effet de **purification** ou de **correction** de la série initiale ou brute (y_t) des effets non désirables comme celui des saisons (S_t) et des bruits ou les effets de perturbations (ε_t).

En effet, les nouvelles séries en Moyennes Mobiles (MM) sont moins variables que la série initiale ou brute (y_t) qui, parait-il, montre les effets des composantes saisonnières (S_t) et résiduelles (ε_t).

Section 2 : La Méthode des Moyennes Mobiles Centrées (MMC)

Les Moyennes Mobiles Centrées (MMC) sont définies exactement de la même façon que les (MM), en représentant aussi la suite des moyennes arithmétiques calculées sur les valeurs brutes de la série initiale (y_t), en les rapportant par rapport à leurs dates de milieu sauf que dans ce cas l'ordre désigné pour les moyennes mobiles est pair.

La méthode des Moyennes Mobiles Centrées (MMC) est donc proposée afin de pallier à l'inconvénient rencontré pour les Moyennes Mobiles (MM) d'ordre pair.

En effet, si l'on veut comparer la série lissée (obtenue à partir des Moyennes Mobiles (MM) d'ordre pair) à la série initiale, on a besoin d'avoir les valeurs pour les mêmes dates d'observations. Ce qui n'est pas le cas dans le cadre des MMS d'ordre pair.

Etant donné ce problème là et afin de lever l'ambigüité existante sur les interprétations, on est souvent conduit à une sorte de correction dans laquelle on procède dans le calcul par :

1. Calculer les moyennes sur (p + 1) valeurs au lieu de (p) qui est supposé être une valeur paire (comme p = 4 et on travaille sur p = 5) ;

2. Pondérer les deux valeurs extrêmes par (1/2) au lieu de 1 comme pour les autres valeurs restantes des moyennes à calculer ;

3. Diviser la somme par (p) au lieu de (p + 1) (comme 4 au lieu de 5).

Cette méthode consiste à calculer des moyennes mobiles d'ordre 2 aux moyennes mobiles déjà obtenues.

La formule générale de la Moyenne Mobile Centrée (MMC) s'écrit :

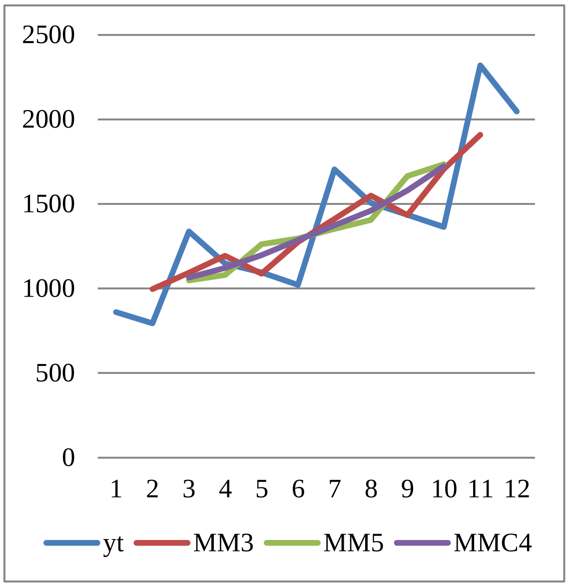
$$MM_p\ (t) = \bar{Y}_t = \left(\frac{y_{t-2}}{2} + y_{t-1} + y_t + y_{t+1} + \frac{y_{t+2}}{2}\right) / p$$

$$\text{Si } p = 4 \Rightarrow MM_4\ (t) = \bar{Y}_t = \left(\frac{y_{t-2}}{2} + y_{t-1} + y_t + y_{t+1} + \frac{y_{t+2}}{2}\right) / 4$$

$$\text{Si } p = 6 \Rightarrow MM_6\ (t) = \bar{Y}_t = \left(\frac{y_{t-3}}{2} + y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2} + \frac{y_{t+3}}{2}\right) / 6$$

On suppose maintenant calculer les Moyennes Mobiles Centrée (MMC) d'ordre 4, on aura la nouvelle série qui sera définie comme suit :

Année	t	y_t	MM_3	MM_5	MMC_4
2020	1	860	-	-	-
	2	794	997,33	-	-
	3	1338	1093,33	1047,20	1064,50
	4	1148	1194,00	1079,40	1122,38
2021	5	1096	1088,33	1261,60	1196,63
	6	1021	1274,00	1295,00	1287,13
	7	1705	1410,33	1352,60	1374,25
	8	1505	1548,67	1406,00	1459,50
2022	9	1436	1434,67	1665,60	1579,00
	10	1363	1706,00	1734,00	1723,50
	11	2319	1909,67	-	-
	12	2047	-	-	-



- $\text{Si } p = 4 \Rightarrow MM_4\ (t) = \bar{Y}_t = \left(\frac{y_{t-2}}{2} + y_{t-1} + y_t + y_{t+1} + \frac{y_{t+2}}{2}\right) / 4$

- Moyennes Mobiles Centrée MMC_4 (t) :

$$\bar{Y}_t = \left(\frac{y_{t-2}}{2} + y_{t-1} + y_t + y_{t+1} + \frac{y_{t+2}}{2}\right) / 4$$

$\bar{Y}_1 = \left(\frac{y_{-1}}{2} + y_0 + y_1 + y_2 + \frac{y_3}{2}\right) / 4 \Rightarrow \bar{Y}_1 =$ On ne peut pas le calculer car y_{-1} et y_0 n'existent pas.

$\bar{Y}_2 = \left(\frac{y_0}{2} + y_1 + y_2 + y_3 + \frac{y_4}{2}\right) / 4 \Rightarrow \bar{Y}_2 =$ On ne peut pas le calculer car y_0 n'existe pas.

$\bar{Y}_3 = \left(\frac{y_1}{2} + y_2 + y_3 + y_4 + \frac{y_5}{2}\right) / 4$

$\Rightarrow \bar{Y}_3 = \left(\frac{860}{2} + 794 + 1338 + 1148 + \frac{1096}{2}\right) / 4$

$\Rightarrow \bar{Y}_3 = (430 + 794 + 1338 + 1148 + 548) / 4$

$\Rightarrow \bar{Y}_3 = (4258) / 4 \Rightarrow \bar{Y}_3 = 1064{,}5$

$\bar{Y}_4 = \left(\frac{y_2}{2} + y_3 + y_4 + y_5 + \frac{y_6}{2}\right) / 4$

$\bar{Y}_4 = \left(\frac{794}{2} + 1338 + 1148 + 1096 + \frac{1021}{2}\right) / 4$

$\Rightarrow \bar{Y}_4 = (397 + 1338 + 1148 + 1096 + 510{,}5) / 4$

$\Rightarrow \bar{Y}_4 = (4489{,}5) / 4 \Rightarrow \bar{Y}_4 = 1122{,}375$

$\bar{Y}_5 = \left(\frac{y_3}{2} + y_4 + y_5 + y_6 + \frac{y_7}{2}\right) / 4 \Rightarrow \bar{Y}_5 = 1196{,}63$

$\bar{Y}_6 = \left(\frac{y_4}{2} + y_5 + y_6 + y_7 + \frac{y_8}{2}\right) / 4 \Rightarrow \bar{Y}_6 = 1287{,}13$

$\bar{Y}_7 = \left(\frac{y_5}{2} + y_6 + y_7 + y_8 + \frac{y_9}{2}\right) / 4 \Rightarrow \bar{Y}_7 = 1374{,}25$

$\bar{Y}_8 = \left(\frac{y_6}{2} + y_7 + y_8 + y_9 + \frac{y_{10}}{2}\right) / 4 \Rightarrow \bar{Y}_8 = 1459{,}50$

$\bar{Y}_9 = \left(\frac{y_7}{2} + y_8 + y_9 + y_{10} + \frac{y_{11}}{2}\right) / 4 \Rightarrow \bar{Y}_9 = 1579{,}00$

$\bar{Y}_{10} = \left(\frac{y_8}{2} + y_9 + y_{10} + y_{11} + \frac{y_{12}}{2}\right) / 4 \Rightarrow \bar{Y}_{10} = 1723{,}50$

$\overline{Y}_{11} = (\frac{y_9}{2} + y_{10} + y_{11} + y_{12} + \frac{y_{13}}{2}) / 4 \Rightarrow \overline{Y}_{11}$ = On ne peut pas le calculer car y_{13} n'existe pas.

$\overline{Y}_{12} = (\frac{y_{10}}{2} + y_{11} + y_{12} + y_{13} + \frac{y_{14}}{2}) / 4 \Rightarrow \overline{Y}_{12}$ = On ne peut pas le calculer car y_{13} et y_{14} n'existent pas.

Dans le cas des Séries Temporelles, l'ordre **(p)** désigne la période retenue pour les saisons :

- Semestrielles c'est **p = 2** car il y a deux semestres dans l'année ;
- Trimestrielles c'est **p = 4** car il y a quatre trimestres dans l'année ;
- Bimestrielles c'est **P = 6** car il y a six périodes de deux mois dans l'année ;
- Mensuelles c'est **p = 12** car il y a douze mois dans l'année.

Chapitre 5 :

Mesure des Composantes d'une Série Temporelle

Chapitre 5 : Mesure des Composantes d'une Série Temporelle

Le lissage des Séries Temporelles (ST) peut être intéressant en permettant une dessaisonalisation de celles-ci et en éliminant les effets non désirables dans leur évolution. Toutefois, le résultat dans certains cas n'est pas vraiment clair vu que certaines séries gardent tout de même après lissage certains de ces effets. Des fois aussi la tendance (T_t) trouvée n'est pas assez nette pour l'utiliser dans le processus prévisionnel ou celui de prise de décision. Ce qui explique, par conséquent, que la méthode utilisée de moyennes mobiles est assez limitée dans la réduction ou l'élimination des fluctuations aléatoires.

Nous avons donc besoin de mesures plus précises relatives à la tendance (T_t), aux effets cycliques (C_t) et aux effets saisonniers (S_t). Ce qui pourra éclaircir davantage la situation en déterminant les valeurs et le poids des différentes composantes que ce soient les composantes désirables comme le trend (T_t) ou bien non désirables comme les variations saisonnières (S_t), cycliques (C_t) et de bruit (ε_t).

Dans cette partie nous nous proposons d'analyser les modèles d'estimation permettant de mesurer avec plus de précision les composantes d'une Série Temporelle (ST). C'est le cas notamment de la composante de tendance (T_t) servant de prévisions ainsi que les composantes cycliques (C_t) et saisonnières (S_t).

Section 1 : Détermination du Trend (Tt)

Section 2 : Détermination de la Composante Saisonnière (St)

Section 3 : Détermination de la Composante Cyclique (Ct)

Section 1 : Détermination du Trend (T_t)

Un des principaux objectifs qu'on a pu développer dans la partie précédente concerne le fait que la série **lissée** doit garder le **trend (T_t) constant** une fois qu'elle est **corrigée** bien évidemment de tous les effets non normaux. Ce qui signifie que ce dernier doit être estimé, par conséquent, via la série **corrigée** laquelle représente les tendances normales et nettoyées de toute sorte de cause irrégulière, imprévisible et indésirable telles que les composantes cycliques et résiduelles ou accidentelles d'une ST.

C'est en ce moment qu'on peut avoir une bonne estimation de la série brute en procédant à des prévisions dans le future comme le chiffre d'affaires (CA) dans une entreprise à prévoir dans les années 2025, 2026, etc. en ayant ses résultats concernant les années précédentes ou bien encore les prévisions à faire sur les quantités à produire d'un produit à partir de la série brute de ses quantités mensuelles enregistrées avant, et ainsi de suite.

Pour ce faire, on peut procéder la première fois par lisser la Série Temporelle (ST) en utilisant la moyenne mobile d'ordre (p) notée **MMp** (t) tel que (p) représente la période du **mouvement saisonnier** dans la même série qui, elle, change selon la chronologie de la série initiale. Ce qui doit correspondre, nous l'avons déjà noté, à p = 2 pour les données semestrielles, p = 4 pour les données trimestrielles et p= 12 pour les données mensuelles.
Si dans cette logique le trend (T_t) peut être remplacé par le résultat du lissage obtenu à partir de la série brute, on peut donc écrire que :

$$\mathbf{T_t = MMp\ (t)}$$

Ce qui signifie que les prévisions des valeurs futures de la variable d'intérêt seront ainsi établies, par le trend (T_t) qui est à son tour évalué aux valeurs des moyennes mobiles correspondantes.

Si, par contre, les points du trend sont à peu près alignés (trend linéaire), on peut estimer sa valeur en utilisant la Droite de Régression Linéaire (DRL) telle que :

$$\mathbf{y_t = \alpha + \beta\ t + \varepsilon_t}$$

Les prévisions seront déterminées de la même manière qu'avant en utilisant la même droite et la méthode des Moindres Carrés Ordinaires (MCO - OLS) qui sera également utilisée. Ce qui explique aussi le fait de respecter les mêmes conditions de minimisation des écarts pour avoir finalement le trend (T_t) calculé comme suit :

Le principe des Moindres Carrés Ordinaires (MCO - OLS) stipule que la somme des carrés des écarts est minimum.

Si on note nos écarts $\boldsymbol{\varepsilon_t}$, on obtient : $\boldsymbol{\varepsilon_t} = (y_i - y_c)$

Avec ; $\Sigma \boldsymbol{\varepsilon_t}^2 \longrightarrow$ Min

$\mathbf{y_t\hat{} = \alpha\hat{} + \beta\hat{}\ t}$

Par ailleurs, lorsque le trend n'est pas **linéaire** de genre **exponentiel** ou **quadratique**, l'équation d'ajustement peut être remplacée par une autre expression plus adéquate à son type d'évolution, notamment :

S'il est exponentiel :

$\mathbf{y_t = \alpha * e^{\beta t} + \varepsilon_t}$

S'il est de genre Quadratique :

$\mathbf{y_t = \alpha + \beta_1 t + \beta_2 t^2 + \varepsilon_t}$

Dans les deux cas l'introduction du logarithme sera indéniable afin de faciliter le calcul de la valeur du trend. La démarche de calcul à suivre, quant à elle, est pratiquement la même que la précédente.

Exemple 1 : Cas linéaire

La série des indices trimestriels de vente de marchandises de l’entreprise ALPHA est fournie pour trois (3) années dans le tableau suivant :

Année \ Trimestre	I	II	III	IV
2020	118,2	129	138,9	157,1
2021	148,6	154,5	163	184
2022	163,3	175,3	189,1	217,9

L'utilisation de l'équation **linéaire** du **trend T**, $\mathbf{T = y_t = \alpha + \beta\ t + \varepsilon_t}$ permet d'avoir les résultats présentés dans le tableau suivant :

Année	**Trimestre (t)**	$\mathbf{y_t}$	$\mathbf{t_t - \bar{t}}$	$\mathbf{y_t - \bar{y}}$	$\mathbf{(t_t - \bar{t})(y_t - \bar{y})}$	$\mathbf{(t_t - \bar{t})^2}$	$\mathbf{y_t\hat{}}$
2020	1	118,2	-5,5	-43,375	238,5625	30,25	122,37
	2	129	-4,5	-32,575	146,5875	20,25	129,49
	3	138,9	-3,5	-22,675	79,3625	12,25	136,61
	4	157,1	-2,5	-4,475	11,1875	6,25	143,73
2021	5	148,6	-1,5	-12,975	19,4625	2,25	150,85
	6	154,5	-0,5	-7,075	3,5375	0,25	157,97
	7	163	0,5	1,425	0,7125	0,25	165,09
	8	184	1,5	22,425	33,6375	2,25	172,21
2022	9	163,3	2,5	1,725	4,3125	6,25	179,33
	10	175,3	3,5	13,725	48,0375	12,25	186,45
	11	189,1	4,5	27,525	123,8625	20,25	193,57
	12	217,9	5,5	56,325	309,7875	30,25	200,69
$\sum$	**78**	**1938,9**	**-**	**-**	**1019,05**	**143**	**-**

$y_t = \alpha + \beta\ t + \varepsilon_t$

$y_t\hat{} = \alpha\hat{} + \beta\hat{}\ t$

$$\bar{t} = \frac{\sum t_t}{N} = \frac{78}{12} = 6,5$$

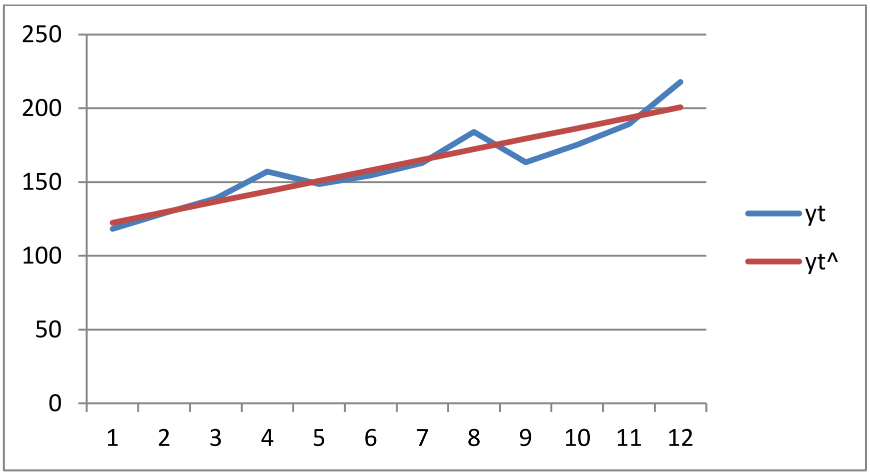
$$\bar{y} = \frac{\sum y_t}{N} = \frac{1938,9}{12} = 161,57$$

$\alpha = \bar{y} - \beta\ \bar{t} = 161,57 - (7,12) * (6,5) = 115,25$

$$\beta = \frac{(t_t - \bar{t})(y_t - \bar{y})}{(t_t - \bar{t})^2} = \frac{1019,05}{143} = 7,12$$

$y_t\hat{} = \alpha\hat{} + \beta\hat{}\ t \Rightarrow y_t\hat{} = 115,25 + 7,12\ t$

Ce résultat peut être encore mieux visualisé à travers la représentation graphique suivante qui montre une tendance croissante représentée par une droite (trend prévisionnel) :



Le graphe montre effectivement que la tendance est totalement linéaire. Ce qui signifie que l'ajustement est réussi et la mesure de la tendance sera effectuée par l'équation T, y_t^ = α^ + β^ t ⇒ y_t^ = 115,25 + 7,12 t. Dans toute sorte de prise de décisions ou de prévisions jugées nécessaires.

On peut donc prévoir l'indice de vente qui sera probablement réalisé pour le 13° trimestre, par exemple, ce qui correspond au 1° trimestre de l'année 2023 comme suit :

y_t^ = 115,25 + 7,12 t

y_{13}^ = 115,25 + 7,12 (13) ⇒ y_{13}^ = 207,81

On peut aussi prévoir l'indice de vente qui sera probablement réalisée pour le 14° trimestre, par exemple, ce qui correspond au 2° trimestre de l'année 2023 comme suit :

y_t^ = 115,25 + 7,12 t

y_{14}^ = 115,25 + 7,12 (14) ⇒ y_{14}^ = 214,93

Exemple 2 : Cas exponentiel

Soit la série X_t observée pendant 18 mois (premier trimestre 2023 au deuxième trimestre 2024.

	T_1 2023	**T_2 2023**	**T_3 2023**	**T_4 2023**	**T_1 2024**	**T_2 2024**
Y_t	255	330	435	570	740	960

A la lecture de cette statistique l'hypothèse d'une croissance exponentielle des ventes peut être retenue.

-Déterminer les paramètres de la fonction : $X = \beta\alpha^t$

-Déterminer la prévision pour le troisième et quatrième trimestre de 2024

$X_t = \beta\alpha^t \Rightarrow \ln X_t = \ln\beta * \alpha^t$

$\ln X_t = \ln\beta + ln \propto * t$

	T	Y_t	lnX_t	$tlnX_t$	t_2
2023	1	255	5,54	5,58	1
	2	330	5,79	11,5	4
	3	435	6,07	18,21	9
	4	570	6,34	25,36	16
2024	5	470	6,60	33	25
	6	960	6,86	41,16	36
Total	21	-	37,2	134,89	91

$$\ln \propto = \frac{\sum \mathrm{lny_t} * \mathrm{t} - \mathrm{nln}\bar{\mathrm{X}}_\mathrm{t} * \bar{\mathrm{t}}}{\sum \mathrm{t}^2 - \mathrm{n}\bar{\mathrm{t}}^2}$$

$$\ln \propto = \frac{134{,}89 - 6(6{,}2)(3{,}5)}{91 - 6(12{,}25)}$$

$$\ln \propto = \frac{4{,}69}{17} = 0{,}268$$

$$\propto = \mathrm{e}^{0{,}268} = 1{,}307$$

$$\mathrm{ln\beta} = \overline{\mathrm{lnX_t}} - \mathrm{ln\alpha}\bar{\mathrm{t}}$$

$$\mathrm{ln\beta} = 6{,}2 - 0{,}26(3{,}5)$$

$$\mathrm{ln\beta} = 5{,}29$$

$$\beta = \mathrm{e}^{5{,}29} = 198{,}343$$

$$\ln \mathrm{X_t} = 5{,}29 + 0{,}26\mathrm{t}$$

$$\ln \mathrm{X_{t+1}} = 5{,}29 + 0{,}26(7)$$

$$\ln \mathrm{X_{t+1}} = 7{,}11 \Rightarrow \mathrm{X_{t+1}} = \mathrm{e}^{7{,}11} = 1224{,}14$$

$$\ln \mathrm{X_{t+2}} = 5{,}29 + 0{,}26(8)$$

$$\ln \mathrm{X_{t+2}} = 7{,}37 \Rightarrow \mathrm{X_{t+2}} = \mathrm{e}^{7{,}37} = 1587{,}63$$

Section 2 : Détermination de la Composante Saisonnière (S_t)

La mesure de la composante saisonnière (S_t) doit se faire aussi dans le sens de purifier la série brute des effets saisonniers pour en faciliter l'utilisation au niveau des prises des décisions et les calculs prévisionnels.

La méthode à utiliser se base sur le calcul des **Indices de Saisons** appelés aussi les **Coefficients Saisonniers**, ou bien encore les **Coefficients de Variation Saisonniers (CVS)**, qui ont pour objet de **mesurer le degré de différence entre les différentes saisons** dans la période d'observation de la variable en question.

La démarche à suivre peut être récapitulée en quatre étapes différentes :

▪ **1° étape :** Désaisonnaliser la série observée toujours dans le sens de la nettoyer des effets saisonniers et aléatoires. Ce qui explique encore une fois l'utilisation des mêmes méthodes de lissage telle que la méthode des moyennes mobiles (MM), de préférence d'ordre (p) qui représente l'ordre du mouvement saisonnier, ou la méthode de régression linéaire (MCO – OLS) lorsqu'on sait que la série temporelle ne présente pas de variations cycliques.

Dans le cas de la méthode des moyennes mobiles (MM), le résultat sera le fait de calculer une série corrigée des effets saisonniers et résiduels telle que :

$$MM_p\,(t) = T_t * C_t$$

C'est le cas notamment d'un modèle multiplicatif et tel que (S_t), et (ε_t) seront éliminés.

▪ **2° étape :** La détermination des variations saisonnières et aléatoires en procédant par le rapport de la série brute à la série désaisonnalisée par la MM_p (t) et tel que :

$$\frac{yt}{MM_p\,(t)} = S_t * \varepsilon_t$$

▪ **3° étape :** La mesure des **variations saisonnières** en calculant (la moyenne des rapports) obtenus dans l'étape précédente et ce pour chaque type de saison dans la période d'observation tels que les semestres, les trimestres, les mois, etc. selon les cas çàd selon le degré de la saisonnalité.

Ce qui permettra d'éliminer la majorité des variations saisonnières.

▪ **4° étape :** Calculer, enfin, les **Coefficients de Variation Saisonniers (CVS), en ajustant les rapports moyens** obtenus dans **l'étape 3**, en divisant ces derniers par leur total et en **multipliant** par le nombre du mouvement saisonnier (p) correspondant (comme le nombre de trimestres par an) pour s'assurer que le coefficient saisonnier moyen est égal à l'unité.

On résume [1]:

▪ **Le modèle Additif :**

$S_j = X_t - \hat{X}_t$ Méthode de la différence à la tendance

Si : $\sum_{j=1}^{P} S_j = 0 \Rightarrow S_j = S_j^*$

Si : $\sum_{j=1}^{P} S_j \neq 0 \Rightarrow S_j^* = S_j - \bar{S}$

$\mathbf{S_j}$ **:** Coefficients saisonniers

$\mathbf{S_j^*}$: Coefficients saisonniers corrigés

▪ **Le modèle Multiplicatif :**

$S_j = \frac{X_i}{\hat{X}_t}$ Méthode des rapports à la tendance

Si $\sum_{j=1}^{P} S_j = P \Rightarrow S_j = S_j^*$

Si $\sum_{j=1}^{P} S_j \neq P \Rightarrow S_j^* = {S_j}/{\bar{S}}$ avec $\bar{S} = \sqrt{\prod_{j=1}^{P} S_j}$

▪ **Prévision pour un horizon de h périodes :**

Cas additif : $\hat{X}_{t+h}^{P} = \hat{X}_{t+h} + \text{CSMP}$

Cas multiplicatif : $\hat{X}_{t+h}^{P} = \hat{X}_{t+h} * \text{CSMP}$

[1] ABDERRAHMANI, **Cours des Séries Temporelles 1**, Master 1, Economie Quantitative, Département des Sciences Economiques, Faculté des Sciences Economique, Commerciale et des Sciences de Gestion, Université de Béjaia, 2020/2021, P : 5.

Exemples :

Nous essayons dans ce qui suit de montrer ces différentes étapes.

Exemple 1 :

Sur les mêmes observations concernant les indices de ventes enregistrés par l'entreprise ALPHA par trimestre entre 2020 et 2022.

Nous allons présenter, pour faciliter l'analyse, le résultat de chaque étape des deux premières successivement à savoir les moyennes mobiles pour la dessaisonalisation et le rapport des observations brutes par rapport à ces dernières dans le tableau ci-dessous.

Le résultat concernant la détermination des Coefficients de Variation Saisonniers (CVS), sera fait par la suite dans un deuxième tableau.

Année	Trimestre (t)	yt	$T = MM_4$	y_t / MM_4
2020	1	118,2	-	-
	2	129	-	-
	3	138,9	139,60	0,99
	4	157,1	146,58	1,07
2021	5	148,6	152,78	0,97
	6	154,5	159,16	0,97
	7	163	164,36	0,99
	8	184	168,80	1,09
2022	9	163,3	174,66	0,93
	10	175,3	182,16	0,96
	11	189,1	-	-
	12	217,9	-	-
Σ	**78**	**1938,9**	-	-

$$MM_p (t) = \overline{Y}_t = \left(\frac{y_{t-2}}{2} + y_{t-1} + y_t + y_{t+1} + \frac{y_{t+2}}{2}\right) / p$$

$$\overline{Y}_t = \left(\frac{y_{t-2}}{2} + y_{t-1} + y_t + y_{t+1} + \frac{y_{t+2}}{2}\right) / 4$$

$$\overline{Y}_1 = \left(\frac{y_{-1}}{2} + y_0 + y_1 + y_2 + \frac{y_3}{2}\right) / 4 \Rightarrow \overline{Y}_1 = \text{On ne peut pas le calculer car } y_{-1} \text{ et } y_0 \text{ n'existent pas.}$$

$\overline{Y}_2 = (\frac{y_0}{2} + y_1 + y_2 + y_3 + \frac{y_4}{2}) / 4 \Rightarrow \overline{Y}_2 =$ On ne peut pas le calculer car y_0 n'existe pas.

$$\overline{Y}_3 = (\frac{y_1}{2} + y_2 + y_3 + y_4 + \frac{y_5}{2}) / 4$$

$$\Rightarrow \overline{Y}_3 = (\frac{118{,}2}{2} + 129 + 138{,}9 + 157{,}1 + \frac{148{,}8}{2}) / 4$$

$$\Rightarrow \overline{Y}_3 = 139{,}60$$

Comme on peut le constater sur le tableau, la première étape consiste à déterminer les MM_4 pour un nombre des trimestres égal à 4 (p = 4) qui représente en principe l'ordre du mouvement saisonnier, qui permet, comme nous l'avons déjà mentionné auparavant, un meilleur lissage en nettoyant la série des effets saisonniers (S_t) et aléatoires (ε_t). Les moyennes mobiles centrées (MM_4) obtenues nous conduisent malheureusement à perdre les deux premières et les deux dernières observations.

Le résultat de la 2º étape est présenté par contre dans la colonne qui suit, en montrant des observations corrigées obtenues en rapportant les indices des ventes observés brutes par rapport aux moyennes mobiles centrées selon les dates de (t) concernées. De la même manière et pour les mêmes raisons aussi, les rapports des 1º et 2º trimestres de la première année 2020 ainsi que ceux des 3º et 4º trimestres de la dernière année 2022 seront indéfinis.

Quant aux résultats de la troisième étape récapitulés dans le tableau en bas, il faut regrouper tous les rapports obtenus mais par saison çad ; ici par trimestre (t = 1, 2, 3 et 4) pour calculer leur moyenne. Ce qui permet **d'éliminer la variation saisonnière**.

Année \ Trimestre	**I**	**II**	**III**	**IV**	**∑**
2020	-	-	0,99	1,07	
2021	0,97	0,97	0,99	1,09	
2022	0,93	0,96	-	-	
∑	1,9	1,93	1,98	2,16	
Moyenne	0,95	0,96	0,99	1,08	**3,98 ≈ 4**

0,97 + 0,93 = **1,9**

(0,97 + 0,93 = 1,9) / 2 = **0,95**

(0,95 + 0,96 + 0,99 + 1,08) / 4 = **3,98** ≈ 4

La première moyenne des rapports égale à 0,95 est calculée par exemple entre les 1º trimestres des années 2021 et 2022 (le rapport pour 2000 étant inexistant). La deuxième moyenne égale à 0,96 est calculée entre les 2º trimestres des années 2021 et 2022. Alors que les troisième et quatrième moyennes des rapports égales respectivement à 0,99 et 1,08 sont calculées entre les 3º et 4º trimestres respectivement des années 2020 et 2021.

Enfin, la **quatrième étape** qui consiste à ajuster ces différentes moyennes en divisant chacune d'elles par le total des moyennes égale à 3,98 et en multipliant par 4 comme cela est indiqué sur la dernière ligne du tableau ci-dessus.

A travers ces derniers résultats concernant les **Coefficients de Variation Saisonniers (CVS),** on peut effectivement montrer qu'en moyenne les 1º, 2º et 3º trimestres correspondent à des périodes de basses saisons étant donné leurs valeurs inférieures à la moyenne annuelle. Le 4º trimestre correspond, par contre, à une période de haute saison vu que leur valeur est supérieure à un (1).

Exemple 2 : Prévision dans le cas additif

Soit la série des ventes X observées pendant trois (3) ans. On demande de calculer la prévision des ventes pour les années 2024 et 2025.

	X_t	t	$X_t * t$	t^2	$\hat{X}$	S_j	S_j^*	X^{CVS}
2021	10	1	10	1	10,75	-0,75	-1,48	11,48
	15	2	30	4	12,78	2,22	1,48	13,51
2022	12	3	36	9	14,81	-2,81	-2,98	14,98
	20	4	80	16	16,84	3,16	2,98	17,01
2023	16	5	80	25	18,87	-2,87	-1,98	17,98
	22	6	132	36	20,9	1,1	1,98	20,01

$\hat{\beta}_1 = \frac{\sum X_i t - n\bar{t}\bar{X}}{\sum t^2 - nt^2}$, $\hat{\beta}_1 = \frac{368-6*3,5*15,83}{91-6*12,25} = 2,03$

$\hat{X}_i = 8,72 + 2,03t$

Prévision : Cas additif

$\hat{X}^{P}_{t+n} = \hat{X}_{t+n} + \mathrm{CSMP}$

$$\mathrm{CSMP}_{s1} = \frac{-1,48 + 2,98 - 1,98}{3} = -2,14$$

$$\mathrm{CSMP}_{s1} = \frac{1,48 + 2,98 + 1,98}{3} = +2,14$$

Année	S_j	t	N	X_{t+n}	CSMP	X^{P}_{t+n}
2024	S_1	7	1	22,93	2,14	20,79
	S_2	8	2	29,96	2,14	27,11
2025	S_1	9	3	26,99	-2,14	24,85
	S_2	10	4	29,02	2,14	31,16

Exemple 3 : **Prévision dans le cas multiplicatif**

Soit la série X_t observée pendant 4 ans.

	T_1	T_2	T_3	T_4
2020	2000	1000	700	2600
2021	2020	1100	600	3000
2022	2200	1120	740	3200
2023	2260	1130	800	3300

-Déterminer les composantes de cette série.

On suppose que la série X_t est générée par un modèle multiplicatif, calculer la prévision pour l'année 2024 sachant que l'équation de la tendance est donnée par : $c + 24,437t$ et les coefficients saisonniers trimestriels sont donnés dans le tableau suivant :

S_1	S_2	S_3	S_4
1,254	0,637	?	1,708

-Déterminer la série corrigée des variations saisonnières.

Solution :

La représentation graphique montre que la série X_t est fortement saisonnière, cette saisonnalité est liée aux troisième et quatrième trimestres de chaque année.

$\hat{\beta}_0 = \bar{X} - \hat{\beta}_1 \bar{t}$

$\hat{\beta}_0 = 1735{,}62 - (24{,}43)(8{,}5)$

$\hat{\beta}_0 = 1527{,}96$

$X_t = 1527{,}96 + 24{,}437t$

Les prévisions :

$X_{t+1} = 1943{,}38$

$X_{t+2} = 1967{,}82$

$X_{t+3} = 1992{,}26$

$X_{t+4} = 2016{,}7$

X_t Générée par un modèle multiplicatif :

$\sum_{j=1}^{P} S_j = P \Rightarrow \sum_{j=1}^{P} S_j = 4$

$CsMP_{t3} = 4 - (1{,}254 + 0{,}637 + 1{,}708) = 0{,}401$

$\chi^{CVS} = \frac{X_t}{S_j}$

$\bar{X}^{CVS} = \frac{\sum X^{CVS}}{16} = 1734{,}9$

La série corrigée des variations saisonnière :

	T_1	T_2	T_3	T_4
2020	1594,89	1569,85	1745,63	1522,24
2021	1610,84	1726,84	1496,25	1756,44
2022	1754,38	1758,24	1845,38	1873,53
2023	1802,23	1773,9	1995,01	1932,88

Section 3 : Détermination de la Composante Cyclique (C_t)

Dans le même objectif que la composante de la tendance (T_t), on suppose également de mesurer la composante cyclique (C_t) de la série temporelle. La méthode qu'on peut utiliser à cet égard est celle du **Rapport au Trend** qui s'articule autour de trois principales étapes récapitulées comme suit :

▪ **1° étape :** L'estimation de la tendance (T_t) à partir de la série brute en passant comme nous l'avons montré déjà par la méthode des moyennes mobiles (MM) ou la méthode de régression linéaire (MCO – OLS) $y_t = \alpha + \beta$ t.

▪ **2° étape :** La détermination des valeurs estimées de la série nettoyée (Trend T_t) Tt^ = y_t^ = α^ + β^ t, pour chaque période (t) et ce à partir du résultat du trend (T_t) estimé par l'équation d'ajustement si elle existe.

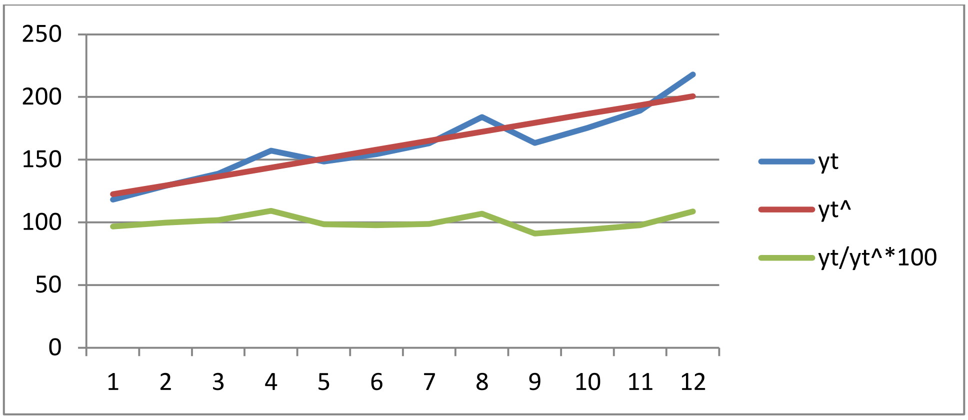
▪ **3° étape :** Le calcul des rapports au trend pour les périodes (t) concernées en passant par le rapport définit en pourcentage des valeurs brutes ou observées de la série (y_t) et les valeurs calculées du trend (y_t^ = T_t^) :

$$\frac{y_t}{y_t\text{^}} * 100 = \frac{y_t}{T_t\text{^}} * 100$$

Par rapport toujours à l'exemple précédant et si on suppose que le Trend (T) Linéaire est ajusté par la même équation linéaire de droite : y_t^ = 115,25 + 7,12 t, on peut montrer que le rapport au Trend (T) est déterminé comme suit :

Année	Trimestre (t)	y_t	y_t^	y_t / y_t^ * 100
2020	1	118,2	122,37	96,59
	2	129	129,49	99,62
	3	138,9	136,61	101,68
	4	157,1	143,73	109,30
2021	5	148,6	150,85	98,51
	6	154,5	157,97	97,80
	7	163	165,09	98,73
	8	184	172,21	106,85
2022	9	163,3	179,33	91,06
	10	175,3	186,45	94,02
	11	189,1	193,57	97,69
	12	217,9	200,69	108,58
Σ	**78**	**1938,9**	**1938,36**	**1200,4317**

Généralement, les rapports au trend (T) représentent les montants par lesquels les valeurs de la série observée dépassent ou se situent en-dessous de la droite du trend (T). Et c'est exactement ce qu'on peut visiblement remarquer sur le graphe suivant.



On aura donc tendance à parler de variations cycliques dans le cas où il existe des variations **assez régulières** qui se produisent au-dessus et en dessous de la droite du trend (T) comme c'est le cas dans ce graphe, vu que tous les rapports au trend (T) (la courbe en bas) sont calculés en dessous de la droite de tendance.

Si par contre, il existe des rapports au trend qui se situent au-dessus et en dessous de la droite du trend, alors les variations seront beaucoup plus **accidentelles** et pas **cycliques**.

Fréquemment, on a beaucoup de mal dans certains cas, à distinguer entre les deux. Raison pour laquelle on suppose disposer au moins d'une durée d'observation des variables suffisamment longue pour pouvoir comparer les deux sur plusieurs dates.

Chapitre 6 :

Les modèles linéaires ARIMA

Chapitre 6 : Les modèles linéaires ARIMA

Il existe deux catégories de modèles pour rendre compte d'une série temporelle. Les premiers considèrent que les données sont une fonction du temps (y = f(t)). Cette catégorie de modèle peut être ajustée par la méthode des moindres carrés, ou d'autres méthodes itératives. L'analyse des modèles par transformée de Fourier est une version sophistiquée de ce type de modèle.

Une seconde catégorie de modèles cherche à déterminer chaque valeur de la série en fonction des valeurs qui la précède (yt = f(yt-1, yt-2, …)). C'est le cas des modèles ARIMA (Auto - Regressive - Integrated - Moving Average).

On entend dire par un processus aléatoire une suite de variables aléatoires réelles indicées par le temps. Toutefois, l'étude des propriétés stochastiques (la moyenne, la variance et la covariance) d'une série temporelle permet de distinguer deux types de processus aléatoire à savoir le processus stationnaire et le processus non stationnaire.

Section 1 : Le processus stationnaire
Section 2 : Le processus non stationnaire

Section 1 : Le processus stationnaire

Un processus stochastique est dit stationnaire, si la distribution jointe de X_t et X_{t+h} ne dépend pas du **t**, mais seulement de **h**, autrement dit son espérance et sa variance se trouvent invariantes dans le temps[1]. D'une manière formalisée, la série temporelle X_t est stationnaire si :

$E(X_t) = E(X_{t+h}) = X_{t+h} = \mu$ la moyenne est constante et indépendant de temps ;

$Var(X_t) < \infty$ $\forall$ t, la variance est finie et indépendante du temps ;

Cov $(Xt;Xt+k) = E[(X_t - \mu)(X_{t+k} - \mu)] = \gamma_k$ la covariance est indépendante du temps.

1. Le processus Bruit Blanc :

Le bruit blanc est un cas particulier de processus stationnaire pour lequel la valeur prise par X_t à la date t s'écrit comme suit : $X_t = \varepsilon_t$ (où ε_t est une variable aléatoire dont son espérance et sa covariance sont nulles)[2]. Un tel processus n'a ni tendance ni mémoire. Autrement dit, la connaissance de la valeur du processus à une date donnée n'apporte aucune information pour la prédiction de sa valeur à une date ultérieure.

Donc la série $\{\varepsilon t\}$ dont $E(\varepsilon t) = 0$, $Var(\varepsilon t) = \sigma 2\ \varepsilon$, $Cov(\varepsilon t\ ,\ \varepsilon t+k) = 0$ est donc une série stationnaire. Elle est appelée aussi bruit blanc.

Une série stationnaire ne doit comporter ni tendance (Tt) et ni saisonnalité (St)[3].

Le processus bruit blanc peut être identifié en utilisant les statistiques de Box-Pierce et Liung-Box qui sont représentés dans le tableau suivant :

[1] Régis BOURBONNAIS, **Econométrie, Manuel et Exercices Corrigés**, Edition Dunod, Paris, France, 2005, P : 223.
[2] Idem, P : 225.
[3] Hélène HAMISULTANE, **Econométrie des Séries Temporelles**, Licence, France, 2002. ffcel01261174, HAL Id : cel-01261174, https://shs.hal.science/cel-01261174v1, Submitted on 24 Jan 2016, https://shs.hal.science/cel-01261174/document, P : 3.

Présentation du test de Box- Pierce et Liung-Box :

Test	Statistiques	Les hypothèses	La règle de décision
Box-Pierce	$Q = n\sum_{k=1}^{h} \boldsymbol{\rho}_k^2$, la statistique Q est distribuée comme un χ^2 (chi-deux) à h degré de liberté. **h** : le nombre de retards ; $\boldsymbol{\rho}_k$: autocorrélation d'ordre k ; **n** : nombre d'observations.	$\begin{cases} H_0: \rho_1 = \rho_2 \ldots = \rho_k = 0 \\ H_1: il\ existe\ au\ moins\ \ un\ \rho\ \ significativement\ différent\ de\ zéro \end{cases}$	Au seuil **α :** -Si Q > χ^2 (χ^2 lu dans la table au seuil (1- **α**) et h degré de liberté, dans ce cas, on va rejeter l'hypothèse H_0 donc la série n'est pas un bruit blanc. -Si Q < χ^2, on accepte l'hypothèse de bruit banc.
Liung-Box	$Q' = n(n+2)\sum_{k=1}^{h} \frac{\rho_k^2}{n-k}$	Identique à la précédente.	Identique à la précédente.

Source : ABDERRAHMANI, **Cours des Séries Temporelles 1**, Chapitre 2, Master 1, Economie Quantitative, Département des Sciences Economiques, Faculté des Sciences Economique, Commerciale et des Sciences de Gestion, Université de Bejaïa, 2020/2021, P : 2.

2. Modèle usuel (processus ARMA) :

Les modèles ARMA (p,q) représentent un processus généré par une combinaison des valeurs passées (processus Auto Régressif AR(p) et des erreurs passées (processus Moyenne Mobile MA(q))[1].

Ils sont définis par l'équation suivante :

$$X_t - \Phi_1 X_{t-1} - \Phi_2 X_{t-2} - \cdots - \Phi_p X_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

Un processus Autorégressif d'ordre *p,* de l'observation présente X_t est générée par une moyenne pondérée des observations passées jusqu'à la p-ième période sous la forme suivante :

$$AR_{(P)}: X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p} + \varepsilon_t$$

Le processus $AR_{(P)}$ est stationnaire si $|\Phi| < 1$.

Un processus moyenne mobile *(*Moving Average) d'ordre q est un processus stationnaire où chaque observation X_t est générée par une moyenne pondérée des aléas passés jusqu'à la q-ième période sous la forme suivante :

[1] Rainer VON SACHS, Sébastien VAN BELLEGEM, **Séries Chronologiques**, Institut de statistique, Université catholique de Louvain, Belgium, 2005, P : 42.

$MA_{(q)}: X_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q}$

Où $\Phi_1 \ldots \Phi_p$ et $\boldsymbol{\theta}_1 \ldots \boldsymbol{\theta}_q$: sont des paramètres pouvant être positifs ou négatifs et ε_t est un bruit blanc.

Section 2 : Le processus non stationnaire

Dans le cas d'une série aléatoire dans le temps c'est-à-dire sa variance et son espérance modifiée, la série est dite alors non stationnaire. Pour analyser la non stationnarité, deux types de processus à distinguer :

1. Le processus Trend Stationary (TS) :

S'écrit comme suite $X_t = f_t + \varepsilon_t$ où f_t est une fonction polynômiale de temps et ε_t est un bruit blanc. Le processus TS, le plus simple, est représenté par la fonction polynômiale de degré 1 c'est-à-dire $X_t = \alpha + \beta_t + \varepsilon_t$.

Le processus TS représente une non-stationnarité de type déterministe car :
$E(X_t) = \alpha + \beta_t$ dépend du temps. Ce dernier peut être stationnariser en retranchant à X_t la valeur estimée $\hat{\alpha} + \hat{\beta}t$ par la méthode des MCO/OLS, la série stationnaire est :

$X_{ts} = X_t - \widehat{X_t}$.[1]

2. Le processus Differency Stationary (DS) :

Provient non pas de la présence d'une composante déterministe tendancielle, mais d'une source stochastique car la tendance ainsi que la variance sont variables dans le temps.

En effet, la meilleure méthode pour stationnariser un tel processus est l'utilisation de la différenciation, c'est à dire qu'un processus aléatoire est stationnaire en différence (DS) ou encore intégré d'ordre « d »si l'on peut l'écrire sous la forme :
$(1 - D)^d X_t = \beta + \varepsilon_t$.

---

[1] Régis BOURBONNAIS, Op.cit., P : 229.

Où ε_t est un processus stationnaire, β une constante réelle encore appelée dérive, D l'opérateur décalage et d l'ordre d'intégration[1]. L'introduction de la constante β dans le processus permet de distinguer deux processus différents :

- Un processus DS sans dérive où β=0 s'écrit : $X_t = X_{t-1} + \varepsilon_t$, comme $\varepsilon_t \sim$ BB (0, σ_ε^2). Ce processus porte le nom d'une marche au hasard ;
- Un processus DS avec dérive qui est une marche au hasard avec dérive, c'est le cas le plus fréquemment rencontré lors de l'étude de séries d'observation où d = 1. On parle de processus intégré d'ordre 1.

3. Tests de stationnarité :

Il existe plusieurs tests de racine unitaire tels que les tests de Dickey-Fuller simple et Dickey-Fuller Augmenté, test de Phillips et Perron, test de Kwiatkowski et Phillips ... qui permettent non seulement de détecter l'existence d'une non-stationnarité mais aussi de déterminer de quel type de non-stationnarité (processus TS ou DS). C'est donc la bonne méthode pour stationnariser la série. Nous n'étudierons ici que les tests de Dickey-Fuller et le test de Phillips et Perron.

a) Test Dickey-Fuller simple (1979) :

Les tests de Dickey et Fuller sont construits à partir de trois modèles de base :

- $X_t = \rho_1 X_{r-1} + \varepsilon_t$…… [1] qui est un modèle autorégressif d'ordre 1, on le note AR(1) ;
- $X_t = \rho_1 X_{r-1} + c + \varepsilon_t$………………. [2] qui est un modèle AR(1) avec constante ;
- $X_t = \rho_1 X_{r-1} + \beta_t + c + \varepsilon_t$……………. [3] qui est un modèle AR(1) avec tendance.

L'hypothèse nulle du test est la présence de racine unitaire, soit la non-stationnarité de type stochastique. Le test consiste à tester :

$$\begin{cases} H0 : \rho = 1 \\ H1 : \rho < 1 \end{cases}$$

Le principe des tests est simple :

[1] Idem, P : 229.

Dans ces trois modèles :

- Si $|\rho 1| = 1$, l'hypothèse nulle H_0 est retenue et la chronique X_t est donc non stationnarité quel que soit le modèle retenu.
- Si $|\rho 1| < 1$ la série Xt est stationnaire[1].

Toute d'abord, on commence à étudier le modèle général [3]. On regarde si β est significativement différent de zéro ou non. Si β est significativement égal à zéro on passe à l'étude de modèle [2] et on cherche à savoir si c'est significativement différent de zéro ou pas. Si C'est significativement non différent de zéro, on étudie le modèle [1]. Enfin on applique le test de Dickey-Fuller pour savoir si la série est stationnaire ou non.

On estime par la méthode des MCO/OLS, le paramètre ρ noté $\hat{\rho}$ pour les trois modèles. L'estimation des coefficients et des écart-types des modèles par cette méthode fournit $t_{\widehat{\rho_t}}$ (statistique ADF) qui est calculée comme suite : $t_{\widehat{\rho_t}} = \frac{\hat{\rho}}{\delta\hat{\rho}}$

Les règles de décision sont les suivantes :

- Si $t_{\widehat{\rho_t}} < t_{ADF}$ où t_{DF} désigne la valeur critique donnée par la table de DF ; donc on accepte l'hypothèse H_1 c'est-à-dire celle de l'absence de racine unitaire. La série X_t dans ce cas est non-stationnaire ;
- Si $t_{\widehat{\rho_t}} \geq t_{ADF}$, on accepte l'hypothèse H_0, la série est stationnaire.

b) Test Dickey-Fuller augmenté (1981) :

Dickey-Fuller étendent ensuite cette procédure de test à des processus purement autorégressifs d'ordre *p* connu. Il s'agit alors des tests ADF pour «Augmented Dickey-Fuller». Cette procédure de test est fondée sur l'estimation par les MCO sous l'hypothèse alternative de trois modèles autorégressifs d'ordre *p* obtenus en

[1] Cem ERTUR, **Méthodologies de Test de la Racine Unitaire**, LATEC, Université de Bourgogne, 1998, P : 6.

soustrayant X_{t-1} aux deux membres des modèles [4], [5] et [6]et en ajoutant p-1 retards en différences premières [1]:

- Modèle [4] $\Delta X_t = \rho X_{t-1} - \sum_{j=2}^{p} \Phi_j \, \Delta X_{t-j+1} + \varepsilon_t$
- Modèle [5] $\Delta X_t = \rho X_{t-1} - \sum_{j=2}^{p} \Phi_j \, \Delta X_{t-j+1} + c + \varepsilon_t$
- Modèle [6] $\Delta X_t = \rho X_{t-1} - \sum_{j=2}^{p} \Phi_j \, \Delta X_{t-j+1} + c + \beta_t \varepsilon_t$

Avec $\varepsilon_t \rightarrow i.i.d.$

Le test se déroule de la même manière que le test DF simple, sauf que les tables statistiques se diffèrent. La valeur de p peut être déterminée selon les critères de Akaike[2] ou Schwarz[3] ou encore, en partant d'une valeur suffisamment importante de p. On estime un modèle à p-1 retards, puis p-2 retards, jusqu'à ce que le coefficient du $p^{ième}$ retards soit significatif.

[1] Idem, P : 8.
[2] La statistique de Akaike est donnée par : Aic = $\log \widehat{\delta_{\varepsilon}^{2}} + \frac{2(p+q)}{T}$.
[3] La statistique de Schwarz est la suivante : Sc = $\log \widehat{\delta_{\varepsilon}^{2}} + (p+q) \frac{\log_T}{T}$.

Chapitre 7 :

La méthodologie

de Box- Jenkins

Chapitre 7 : La méthodologie de Box- Jenkins

La méthode de Box et Jenkins est un outil systématique qui permet :

- De déterminer le meilleur modèle de type ARMA décrivant le processus stochastique d'une série observée ou d'une transformation stationnaire de celle-ci ;
- D'estimer ce modèle ;
- De l'utiliser pour extrapoler les valeurs de la série.

Section 1 : Les étapes de la méthodologie de Box et Jenkins

Section 2 : Stationarisation de la série

Section 3 : Identification

Section 1 : Les étapes de la méthodologie de Box et Jenkins

La méthodologie de Box et Jenkins comporte essentiellement cinq étapes :

1. Étape 1 : Transformation des données afin de stabiliser la variance (log) et différenciation des données pour les stationariser ;
2. Étape 2 : Visualiser les ACF et les PACF empiriques pour identifier les paramètres p et q appropriés ;
3. Étape 3 : Estimation des paramètres du (des) modèle(s) sélectionné(s) ;
4. Étape 4 : Diagnostique et tests adéquation du modèle ;
5. Étape 5 : Prévision : La dernière étape consiste la prévision des valeurs futures à travers le modèle retenu.

Section 2 : Stationarisation de la série

À travers la représentation graphique de la série, on peut décider si la série en question est stationnaire ou non. Dans la plus part des cas, une première différenciation rend la série stationnaire. Il arrive parfois d'appliquer une transformation (log) à la série pour stabiliser la variance.

Section 3 : Identification

Cette première étape a pour objet de trouver les valeurs des paramètres « p et q » des processus ARMA. A cette fin on se base sur l'étude des fonctions d'auto-corrélations et d'auto-corrélations partielles. Nous pouvons essayer de donner quelques règles simples facilitant la recherche des paramètres « p et q ».

1. Dans le cas d'une série affectée d'un mouvement saisonniers, il convient de le retirer préalablement à tout traitement statistique ;
2. La recherche de la stationnarité en terme de tendance (TEST DE DICKEY-FULLER)

Après stationnarisation, nous pouvons identifier les valeurs des paramètres « p et q » du modèle ARMA.

Chapitre 8 :

L'autocorrélation

et l'autocorrélation partielle

Chapitre 8 : L'autocorrélation et l'autocorrélation partielle

L'autocorrélation et l'autocorrélation partielle sont des mesures de l'association entre des valeurs de séries actuelles et passées ; elles indiquent les valeurs de séries passées les plus utiles à la prévision de valeurs futures.

Section 1 : La fonction d'auto-corrélation (FAC)

Section 2 : La fonction d'auto-corrélation partielle (FACP)

Section 1 : La fonction d'auto-corrélation (FAC)

La fonction d'autocorrélation (The Auto Correlation Function (ACF)) est une mesure de la corrélation entre des observations d'une série temporelle séparées par k unités de temps (y_t et y_{t-k}).

En statistiques, l'autocorrélation d'un processus aléatoire est la corrélation de Pearson entre les valeurs du processus à différents moments, comme une fonction de deux temps ou d'un écart de temps.

Pour l'autocorrélation, ce coefficient est calculé entre une série temporelle et la même série temporelle avec un écart d'un nombre de périodes spécifié. Par exemple, pour un écart de temps de 1 période, le coefficient de corrélation est calculé entre les premières valeurs *N-1*, soit $x_1, \dots, x_{N-1}$ et les prochaines valeurs *N-1* (valeurs déplacée de un), soit $x_2, \dots, x_N$.

Autocovariance de *X* pour un décalage de *k* :

$$\gamma_k = E((X_t - \mu)(X_{t+k} - \mu))$$

Autocorrélation de *X* pour un décalage de *k* :

$$\rho_k = \frac{\gamma_k}{\sigma^2}$$

1. Interprétation :

Utilisez les fonctions d'autocorrélation partielle et d'autocorrélation conjointement pour déterminer des modèles ARIMA. Etudiez les pics au niveau de chaque décalage pour déterminer s'ils sont significatifs. Un pic significatif dépasse les limites de signification, ce qui indique que la corrélation correspondant à ce décalage n'est pas égale à zéro.

Les schémas suivants peuvent vous aider à préciser les termes autorégressifs et de moyenne mobile dans un modèle ARIMA.

▪ Fonction d'autocorrélation (ACF). Au décalage *k*, il s'agit de la corrélation entre les valeurs de séries séparées par *k* intervalles.

▪ Fonction d'autocorrélation partielle (PACF). Au décalage *k*, il s'agit de la corrélation entre les valeurs de séries séparées par *k* intervalles, compte tenu des valeurs des intervalles intermédiaires.

Par exemple, une pointe au décalage 1 dans un tracé ACF indique une forte corrélation entre chaque valeur de série et la valeur précédente ; une pointe au décalage 2 indique une forte corrélation entre chaque valeur et la valeur apparaissant deux points auparavant, etc.

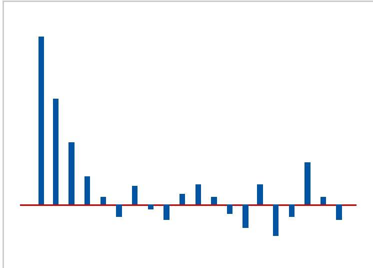
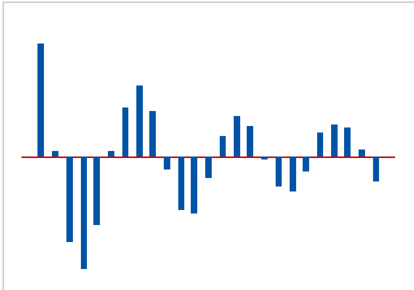
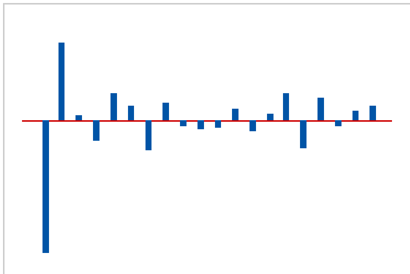
▪ Une corrélation positive indique que des valeurs actuelles élevées correspondent à des valeurs élevées au niveau du décalage spécifié ; une corrélation négative indique que des valeurs actuelles élevées correspondent à des valeurs faibles au niveau du décalage spécifié.

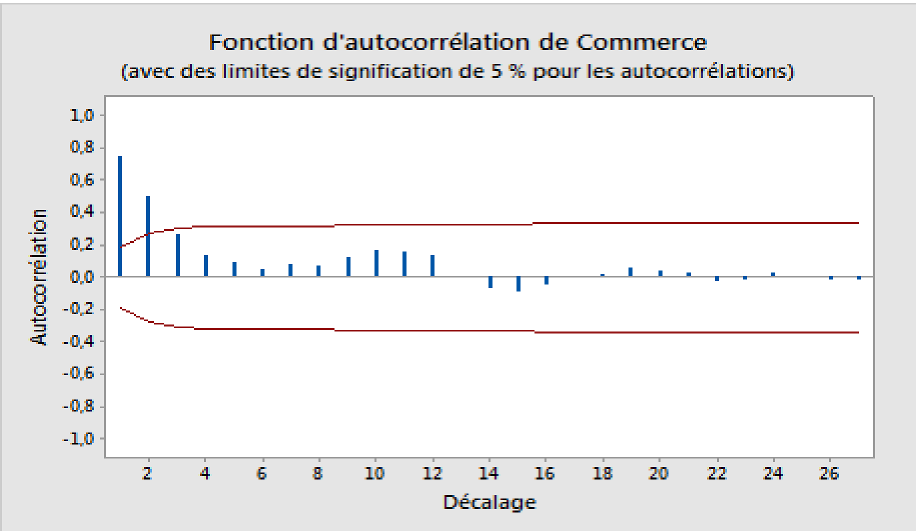
▪ La valeur absolue d'une corrélation est une mesure de la force de l'association, des valeurs absolues élevées indiquant des relations plus fortes.

▪ L'autocorrélation est particulièrement forte pour les décalages 3, 6, et 9. (Et 0, mais ça ne signifie rien : l'autocorrélation de la série avec elle-même -sans décalage- est forcément de 1!).

2. Remarque :

Les données doivent être stationnaires pour que vous puissiez interpréter le diagramme d'autocorrélation. Une série temporelle stationnaire possède une moyenne, une variance et une fonction d'autocorrélation plus ou moins constantes dans le temps.

Schéma	Interprétation du schéma	Exemple
Pic important au niveau du décalage 1 qui diminue après quelques décalages.	Terme autorégressif dans les données. Utilisez la fonction d'autocorrélation partielle pour déterminer l'ordre du terme autorégressif.	
Pic important au niveau du décalage 1, suivi par un motif en vague décroissant qui alterne entre corrélations positive et négative.	Terme autorégressif d'ordre supérieur dans les données. Utilisez la fonction d'autocorrélation partielle pour déterminer l'ordre du terme autorégressif.	
Corrélations significatives au niveau du premier ou du deuxième décalage, suivies de corrélations non significatives.	Terme de moyenne mobile dans les données. Le nombre de corrélations significatives indique l'ordre du terme de moyenne mobile.	



Sur ce diagramme, il existe une corrélation significative au niveau du décalage 1 qui diminue après quelques décalages. Ce schéma indique la présence d'un terme autorégressif. Vous devez utiliser la fonction d'autocorrélation partielle pour déterminer l'ordre de ce terme.

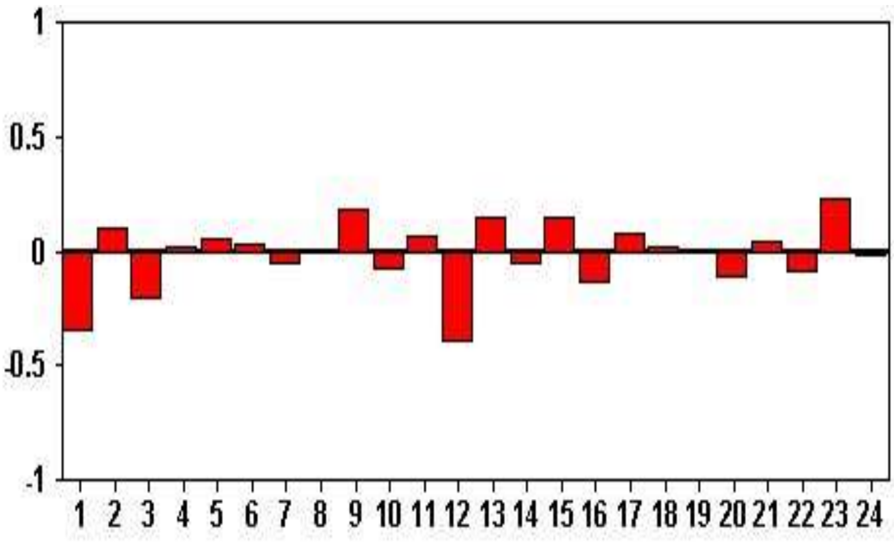


Figure : Tracé ACF d'une série temporelle par R.

Section 2 : La fonction d’auto-corrélation partielle (FACP)

1. Définition :

Rappelons que l’ordre q d’un modèle MA(q) peut être détecté grâce à la fonction d’auto-corrélation car (h) = 0 pour tout h > q alors que (q) 6 = 0.
La fonction d’auto-correlation partielle joue un rôle similaire dans la détection du paramètre p des modèles AR(p). En effet, comme on l’a vu précédemment, (h) = 0 pour tout h > p alors que (p) = 'p 6 = 0.

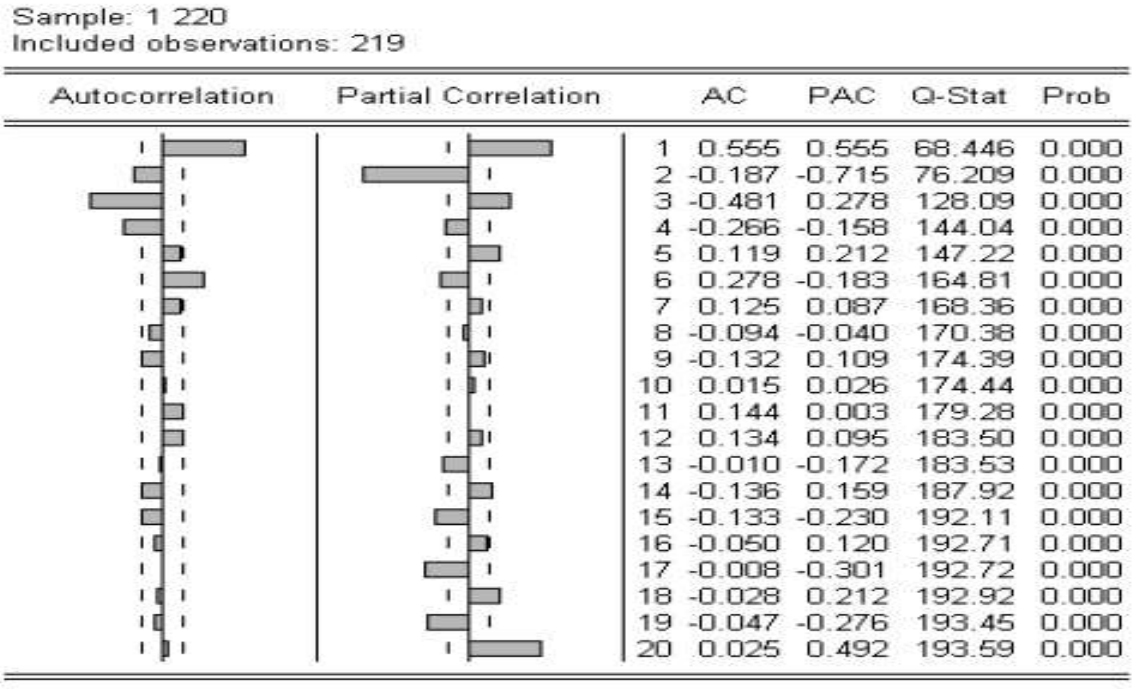
La fonction d’autocorrelation partielle empirique est disponible sous R via la commande pacf. On peut montrer que, si (Xt) est solution stationnaire d’un modèle AR causal telle que ("t) soit un BB (0) fort.
Pour tout h > p, ce qui permet de délimiter un intervalle de confiance à 5% comme on l’avait fait pour l’acf [1].

L’autocorrélation partielle de retard k comme le coefficient de corrélation partielle entre Yt et Yt−k , c’est-à-dire comme étant la corrélation entre Yt et Yt−k l’influence des autres variables décalées de k périodes (Yt−1 , Yt−2 , . . . , Yt−k+1) ayant été retirée.

[1] Adrien Hardy, **Séries Temporelles**, 30 mars 2020, P : 20.

La représentation graphique de la fonction d’auto-corrélation et d’autocorrélation partielle est appelée le correlogramme.

Exemple d’un correlogramme



Sample: 1 220
Included observations: 219

	AC	PAC	Q-Stat	Prob
1	0.555	0.555	68.446	0.000
2	-0.187	-0.715	76.209	0.000
3	-0.481	0.278	128.09	0.000
4	-0.266	-0.158	144.04	0.000
5	0.119	0.212	147.22	0.000
6	0.278	-0.183	164.81	0.000
7	0.125	0.087	168.36	0.000
8	-0.094	-0.040	170.38	0.000
9	-0.132	0.109	174.39	0.000
10	0.015	0.026	174.44	0.000
11	0.144	0.003	179.28	0.000
12	0.134	0.095	183.50	0.000
13	-0.010	-0.172	183.53	0.000
14	-0.136	0.159	187.92	0.000
15	-0.133	-0.230	192.11	0.000
16	-0.050	0.120	192.71	0.000
17	-0.008	-0.301	192.72	0.000
18	-0.028	0.212	192.92	0.000
19	-0.047	-0.276	193.45	0.000
20	0.025	0.492	193.59	0.000

▪ Si le corrélogramme simple n’a que ses « q » premiers termes différents de zéro, (max =3) et que les termes d’auto-corrélations partielles diminuent lentement nous pouvons pronostiquer un MA(q).

▪ Si corrélogramme partiel n’a que ses « p » premiers termes différents de zéro, (max P=3) et que les auto-corrélations simples diminuent lentement, cela caractérise un processus AR(p).

2. Estimation des processus ARMA :

Après avoir identifié les valeurs « P et q » d’un ou plusieurs processus ARMA, l’étape suivante consiste à estimer les coefficients associés aux termes autorégressifs et moyens mobiles. Dans certains cas, notamment dans le cas d’un AR(P), il est possible d’appliquer la méthode MCO. De façon générale, on utilise la méthode de maximum de vraisemblance. Cette méthode repose sur l'hypothèse de normalité des résidus.

Validation des processus ARMA :

Au début de cette étape, on dispose de plusieurs processus ARMA
Souvent, il n'est pas facile de déterminer un modèle unique qui représente le processus générateur de données, et il n'est pas rare pour estimer plusieurs modèles à l'étape initiale. Le modèle qui est finalement choisi est celui considéré comme le meilleur basé sur un ensemble de critères de contrôle et de diagnostic. Ces critères comprennent:

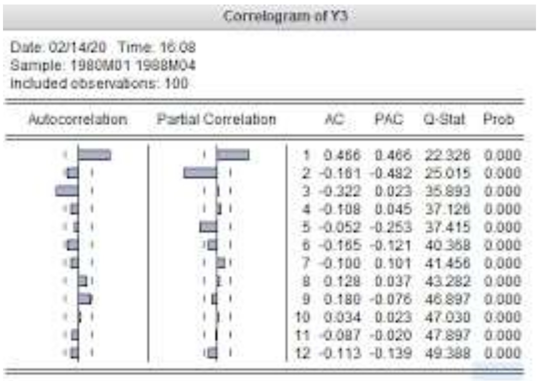
- Les t-tests de significativité des paramètres estimés.
- Analyse des résidus (normalité, absence d'auto-corrélation, homoscédasticité)
- Les critères d'informations.

Les résidus issus d'une estimation doivent vérifier quelques propriétés statistiques :

- La normalité.
- Absence d'auto-corrélation.
- Homoscédasticité.

3. Présentation graphique :

a) This is from EViews :

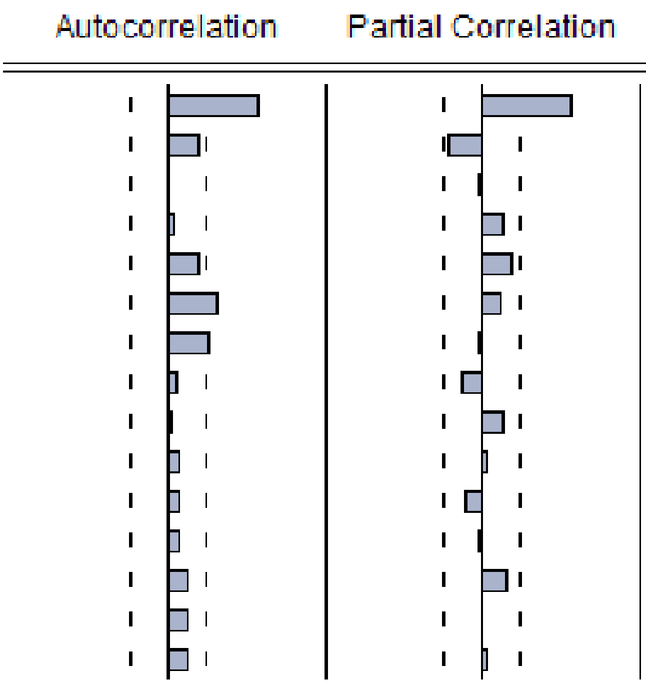

Correlogram of Y3

Date: 02/14/20 Time: 16:08
Sample: 1980M01 1988M04
Included observations: 100

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.466	0.466	22.326	0.000
		2	-0.161	-0.482	25.015	0.000
		3	-0.322	0.023	35.893	0.000
		4	-0.108	0.045	37.126	0.000
		5	-0.052	-0.253	37.415	0.000
		6	-0.165	-0.121	40.368	0.000
		7	-0.100	0.101	41.456	0.000
		8	0.128	0.037	43.282	0.000
		9	0.180	-0.076	46.897	0.000
		10	0.034	0.023	47.030	0.000
		11	-0.087	-0.020	47.897	0.000
		12	-0.113	-0.139	49.388	0.000

Correlogram of DLGDP

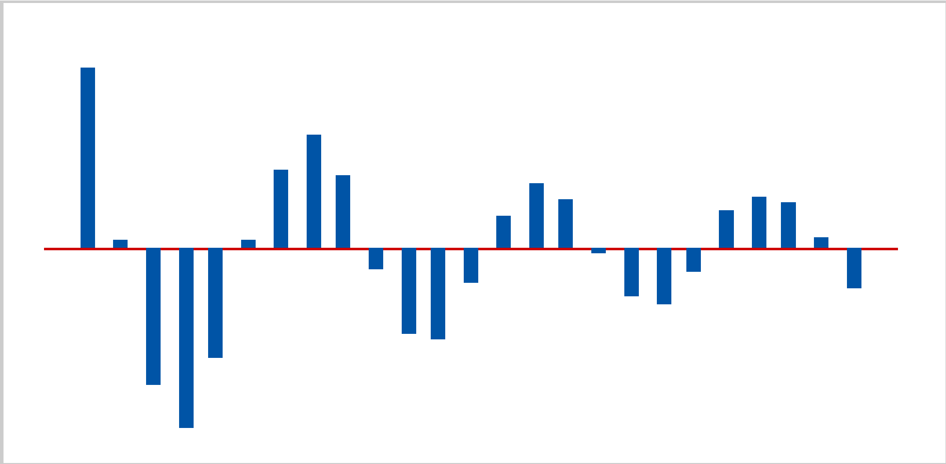
Date: 06/05/18 Time: 15:06
Sample: 1952 2015
Included observations: 63



	AC	PAC	Q-Stat	Prob
1	0.587	0.587	22.732	0.000
2	0.202	-0.217	25.460	0.000
3	0.006	-0.017	25.463	0.000
4	0.036	0.133	25.554	0.000
5	0.203	0.199	28.458	0.000
6	0.317	0.116	35.669	0.000
7	0.263	-0.010	40.719	0.000
8	0.064	-0.117	41.025	0.000
9	0.022	0.143	41.064	0.000
10	0.072	0.033	41.467	0.000
11	0.079	-0.109	41.962	0.000
12	0.075	-0.011	42.414	0.000
13	0.130	0.168	43.801	0.000
14	0.123	0.001	45.074	0.000
15	0.126	0.040	46.429	0.000

Autocorrelation and partial autocorrelation function graphs of the DLGDP series**.**

b) This is from R



Ministry of Higher Education & Scientific Research
Mouloud MAMMERI University of Tizi Ouzou – Algeria
Faculty of Economics, Business & Management
Department of Economics
1st Year Master - Business Economics

Academic Year : 2024 – 2025
Module : Time Series Analysis
Worksheet 1st

Lecturer : G. BELBACHIR
Email : gouraya.belbachir@ummto.dz

Exercise 1 :

Une agence de voyage a réalisé une étude pour savoir s'il existe une relation entre le prix moyen mensuel de billet (X en 1000 DA) et le nombre de clients qui ont réservé un voyage (Y). Les résultats sont donnés ci-après :

X	7	9	9.5	11	12.5	13	15	17	19	21
Y	90	110	130	150	170	190	210	230	250	270

1. Tracer le diagramme de dispersion du nombre de clients qui ont réservé un voyage en fonction du prix moyen de billet. Conclure quant à la nature de la relation.

2. Calculer un paramètre qui permet de mesurer l'intensité de la relation linéaire entre les deux variables étudiées. Interpréter votre résultat.

3. Déterminer l'équation de la droite donnant le nombre de client qui ont réservé un voyage en fonction du prix moyen de billet.

4. Déterminer le nombre de clients qui ont réservé pour des prix moyens de billet qui sont respectivement de 8000 DA et 10000 DA.

Exercise 2 :

Le nombre des ventes trimestrielles réalisées par l'entreprise ALPHA est récapitulé dans le tableau suivant :

Année \ Trimestre	**1**	**2**	**3**	**4**
2020	860	794	1338	1148
2021	1096	1021	1705	1505
2022	1436	1363	2319	2047

1. Représenter graphiquement cette série temporelle.

2. Que peut-on déduire.

3. Désaisonnaliser la série en utilisant une $MM_{(3)}$, $MM_{(4)}$ et $MM_{(5)}$.

4. Représenter graphiquement les trois sur le même graphe que la série. Votre conclusion.

5. Désaisonnaliser la série en utilisant un modèle linéaire. Que pensez-vous ?

Exercise 1 :

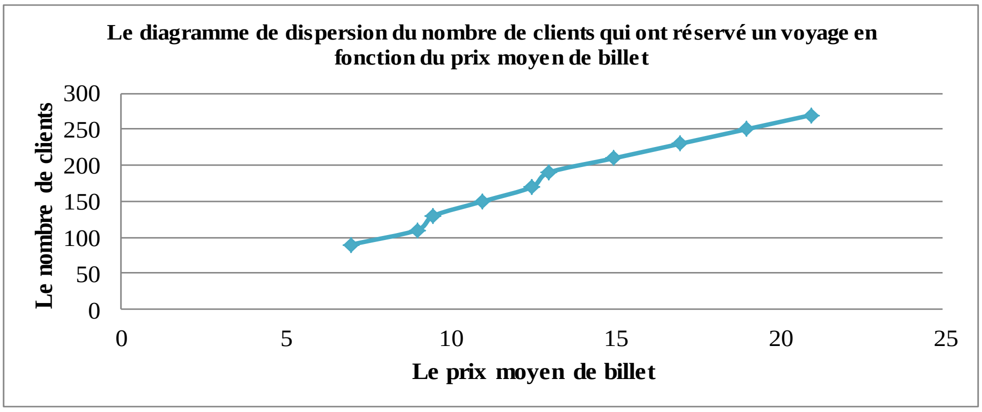
La population est représentée par dix (10) observations portant sur dix (10) clients d'une agence de voyage.

Alors que les variables utilisées sont définies par :

▪ **X** : « le prix moyen mensuel de billet en 1000 DA » comme variable exogène.

▪ **Y** : « le nombre de clients qui ont réservé un voyage suite à un prix affiché par l'agence » comme variable endogène.

5. Le diagramme de dispersion du nombre de clients qui ont réservé un voyage en fonction du prix moyen de billet est présenté comme suit :



On constate que le nombre de clients ayant réservé un voyage évolue positivement et linéairement au prix moyen de billet.

6. Le paramètre qui permet de **mesurer l'intensité de la relation** linéaire entre les deux variables étudiées est représenté par le **coefficient de corrélation** (r) définit comme suit :

$$r = \frac{Cov(x,y)}{\delta(x) * \delta(y)}$$

D'où le tableau et les résultats suivants :

	$\bar{X}$	$\bar{Y}$	$\bar{X}^2$	$\bar{Y}^2$	$\bar{X}\ \bar{Y}$
	X	**Y**	**X^2**	**Y^2**	**X.Y**
	7	90	49	8100	630
	9	110	81	12100	990
	9.5	130	90.25	16900	1235
	11	150	121	22500	1650
	12.5	170	156.25	28900	2125
	13	190	169	36100	2470
	15	210	225	44100	3150
	17	230	289	52900	3910
	19	250	361	62500	4750
	21	270	441	72900	5670
Σ	**134**	**1800**	**1982.5**	**357000**	**26580**

Avec :

$$r = \frac{Cov(x,y)}{\delta(x) * \delta(y)}$$

$$r = \frac{XY - \bar{X} * \bar{Y}}{\sqrt{\bar{X}^2 - (\bar{X})^2} * \sqrt{\bar{Y}^2 - (\bar{Y})^2}}$$

- La moyenne arithmétique de X ($\bar{X}$) :

$\bar{X} = \frac{\sum X_i}{N}$ / Sachant que N = 10 (Nombre de situations)

$\bar{X} = \frac{134}{10}$

$\bar{X} = 13,4$

Le prix moyen des prix mensuels fixé pour les 10 clients (ou 10 derniers mois) pris dans cet échantillon est de 13400 DA.

- La moyenne arithmétique au carrée des X ($\bar{X})^2$:

$(\bar{X})^2 = (13,4)^2$

$(\bar{X})^2 = 179,56$

- La moyenne arithmétique des X au carrée ($\overline{X^2}$) :

$$\overline{X^2} = \frac{\sum X_i^2}{N}$$

$$\overline{X^2} = \frac{1982{,}5}{10}$$

$$\overline{X^2} = 198{,}25$$

- La moyenne arithmétique de Y ($\overline{Y}$) :

$$\overline{Y} = \frac{\sum Y_i}{N}$$

$$\overline{Y} = \frac{1800}{10}$$

$$\overline{Y} = 180$$

Le nombre moyen des clients ayant réservé un voyage sur les dix derniers mois est de 180 clients.

- La moyenne arithmétique au carrée des Y ($\overline{Y})^2$:

$$(\overline{Y})^2 = (180)^2$$

$$(\overline{Y})^2 = 32400$$

- La moyenne arithmétique des Y au carrée ($\overline{Y^2}$) :

$$\overline{Y^2} = \frac{\sum Y_i^2}{N}$$

$$\overline{Y^2} = \frac{357000}{10}$$

$$\overline{Y^2} = 35700$$

- La moyenne arithmétique de $\overline{X} * \overline{Y}$:

$$\overline{X} * \overline{Y} = (13{,}4) * (180)$$

$$\overline{X} * \overline{Y} = 2412$$

- La moyenne arithmétique de XY :

$$XY = \frac{\sum X_i Y_i}{N}$$

$$XY = \frac{26580}{10}$$

$XY = 2658$

- Cov(x,y) :

$Cov(x,y) = XY - \bar{X} * \bar{Y}$

$Cov(x,y) = 2658 – 2412$

$Cov(x,y) = 246$

- L’écart type δ(x) :

$\delta(x) = \sqrt{\bar{X}^2 – (\bar{X})^2}$

$\delta(x) = \sqrt{198{,}25 – 179{,}56}$

$\delta(x) = \sqrt{18{,}66}$

$\delta(x) = 4{,}32$

En moyenne, le prix de billet s'écarte du prix moyen calculé sur les 10 derniers mois d'environ 4320 DA.

- L’écart type δ(y) :

$\delta(y) = \sqrt{\bar{Y}^2 – (\bar{Y})^2}$

$\delta(y) = \sqrt{35700 – 32400}$

$\delta(y) = \sqrt{3300}$

$\delta(y) = 57{,}44 \approx 58$

En moyenne, le nombre des clients ayant réservé un voyage s'éloigne du nombre moyen des clients sur les dix derniers mois d'environ 58 clients.

- $r = \dfrac{XY - \bar{X} * \bar{Y}}{\sqrt{\bar{X}^2 – (\bar{X})^2} * \sqrt{\bar{Y}^2 – (\bar{Y})^2}}$

$$r = \frac{2658 – 2412}{4{,}32 * 57{,}44}$$

$$r = \frac{246}{248{,}14}$$

r = 0,99

Le prix de billet et le nombre de clients sont dépendants positivement.

Enfin, on trouve :

Les deux variables observées dans cette population sont donc ***fortement*** et ***positivement corrélées*** puisque jusqu'à ***99%*** de la variabilité du nombre de clients est expliquée par la variabilité du prix moyen fixé par mois. Ainsi, il y a seulement ***1%*** des changements susceptibles sur le nombre de clients de cette agence qui est expliqué par ***les autres variables non observées***.

7. L'équation de la droite donnant le nombre de client qui ont réservé un voyage en fonction du prix moyen de billet est représentée par :

Y = aX + b $\Rightarrow$ Dont les valeurs des deux paramètres ***a*** et ***b*** sont estimées par :

- $a = \dfrac{Cov(x,y)}{V(X)}$

$$a = \frac{\overline{XY} - \overline{X} * \overline{Y}}{\overline{X^2} - (\overline{X})^2}$$

$$a = \frac{2658 - 2412}{198{,}25 - 179{,}56}$$

$$a = \frac{246}{18{,}69}$$

a = 13,16

Ce qui signifie que lorsque le prix moyen par mois ***augmente*** de ***1%***, le nombre de clients ***augmente aussi*** d'environ ***13%***.

- $b = \overline{Y} - a\,\overline{X}$

b = 180 – (13,16) * (13,4)

b = 180 – 176,34

b = 3,65

Ce qui représente le nombre constant de clients qui peut réserver un voyage ***indépendamment*** du prix moyen par mois du billet.

Ce qui donne, enfin, l'équation de la droite d'ajustement donnée par :

$Y = aX + b$

$Y = 13{,}16X + 3{,}65$

8. Sur la base de la droite précédente on peut donner les prévisions suivantes :

▪ Si le prix moyen de billet par mois est fixé à 8000 DA alors le nombre de clients qui peut réserver sera prévu à :

$Y = (13{,}16) * (8) + 3{,}65$

$Y = 105{,}28 + 3{,}65$

$Y \approx 109$

Soit donc un nombre prévisionnel d'environ 109 clients.

▪ Si, par contre, le prix moyen de billet par mois est fixé à 10000 DA alors le nombre de clients qui peut réserver sera prévu à :

$Y = (13{,}16) * (10) + 3{,}65$

$Y = 131{,}6 + 3{,}65$

$Y = 135{,}25$

$Y \approx 135$

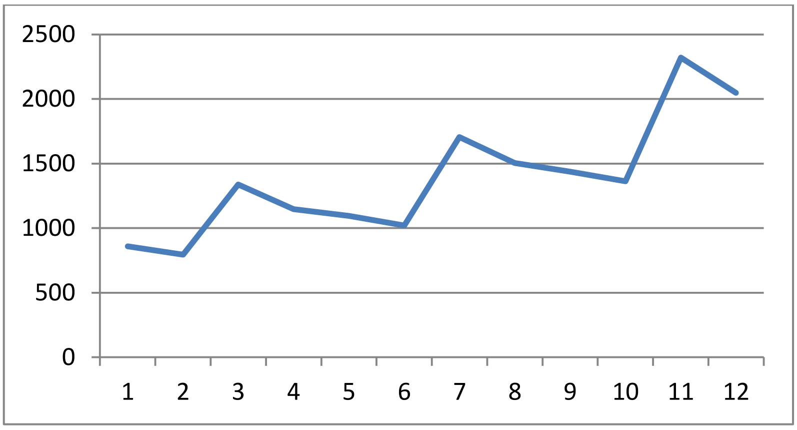
Soit donc un nombre prévisionnel d'environ 135 clients.

Exercise 2 :

Il s'agit dans cette application d'une série temporelle composée par des ventes trimestrielles réalisées par l'entreprise ALPHA pendant la période allant de 2020 à 2022. Ce qui représente 12 observations.

1. La représentation graphique de la série temporelle brute doit se faire normalement en représentant les dates « t » sur l'axe des abscisses et les valeurs des ventes sur l'axe des ordonnées. Elle est donnée comme suit :

Les observations y_t



Les dates d'observations t = 12 trimestres (t)

Le diagramme de la série brute (y_t)

2. On peut en déduire effectivement certaines caractéristiques qui sont essentielles pour l'évolution de la série brute. On constate ainsi que :

▪ La série des ventes observées est touchée par un effet d'évolution tantôt à la hausse tantôt à la baisse ;

▪ Dans cette évolution la série marque une hausse nette malgré les hauts et les bas. C'est ce qu'on appelle la *tendance* ou le *trend* qui peut être représenté par une courbe ou une ligne croissante dans ce cas ;

▪ La série brute est infectée aussi par des variations *saisonnières* puisque ce sont les évolutions qui se répètent pratiquement de la même façon mais sur des courtes périodes (les années prises séparément) ;

▪ On peut même détecter des variations accidentelles qui font que les variations négatives soient atténuées dans leur rythme entre les années. En effet, on peut imaginer une sorte de *faits accidentels* qui vont diminuer l'effet négatif des ventes, comme par exemple une diminution des prix des produits, une augmentation du pouvoir d'achat, un afflux de demande, etc. ;

▪ Le modèle de décomposition parait visiblement de nature *additif* vu que les extremums (les piques sur le graphe) positifs et négatifs gardent la même amplitude.

3. La dessaisonalisation de la série brute en utilisant la méthode des Moyennes Mobiles (MM) doit se faire en appliquant deux équations différentes selon le degré k, en effet on a :

- Si k est impaire tel que $k = 2m + 1$: (voir le cours)

Exercise 3

Soit la série temporelle suivante :

t	1	2	3	4	5	6	7
$\mathbf{y_t}$	1	1	2	5	11	20	32

1. Représenter graphiquement cette série temporelle.

2. Que peut-on déduire sur la nature de Trend (T) : Linéaire, exponentielle ou polynomiale ?

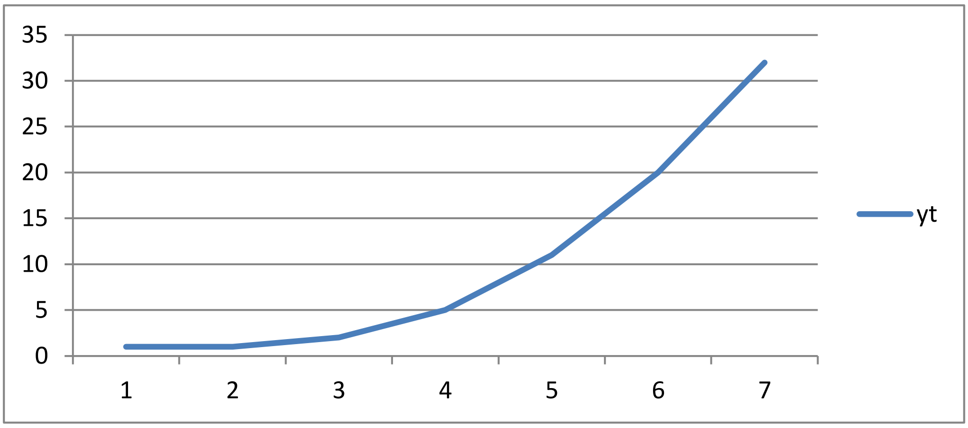
3. Déterminer le Trend (T) mathématiquement : $y_t = \alpha * e^{\beta t}$.

4. Comment procéder à l'élimination de la composante tendancielle dite Trend (T) ?

5. Représenter graphiquement $y\wedge_t$ et ε_t.

Solution :

1. La représentation graphique de cette série temporelle :



2. La nature de Trend (T) est exponentielle de la forme : $y_t = \alpha * e^{\beta t}$.

3. La détermination mathématique de Trend (T) : $y_t = \alpha * e^{\beta t}$.

t	y_t	$Ln(y_t) = Z_t$	$Z_t - \bar{Z}$	$t_t - \bar{t}$	$(t_t - \bar{t})(Z_t - \bar{Z})$	$(t_t - \bar{t})^2$	$Z\hat{}_t$	$e^{z\hat{}t} = y\hat{}_t$	$\varepsilon_t = y_t - y\hat{}_t$
1	1	Ln (1) = 0	-1,59	-3	4,77	9	-0,34	$e^{-0,34} = 0,71$	0,29
2	1	Ln (1) = 0	-1,59	-2	3,18	4	0,31	$e^{0,31} = 1,36$	-0,36
3	2	Ln (2) = 0,69	-0,90	-1	0,90	1	0,96	$e^{0,96} = 2,61$	-0,61
4	5	Ln (5) = 1,61	0,02	0	0,00	0	1,61	$e^{1,61} = 5,00$	0,00
5	11	Ln (11) = 2,4	0,81	1	0,81	1	2,26	$e^{2,26} = 9,58$	1,42
6	20	Ln (20) = 3	1,41	2	2,81	4	2,91	$e^{2,91} = 18,36$	1,64
7	32	Ln (32) = 3,47	1,88	3	5,63	9	3,56	$e^{3,56} = 35,16$	-3,16
28	**72**	**11,16**	**-**	**-**	**18,09**	**28**	**-**	**-**	**-**

$y_t = \alpha * e^{\beta t}$

$Ln(y_t) = Ln(\alpha * e^{\beta t})$

$Ln(y_t) = Ln(\alpha) + Ln(e^{\beta t})$ $\quad / \; Ln(y_t) = Z_t$

$Ln(\alpha) = A$

$Ln(e^{\beta t}) = \beta t$

$Ln(y_t) = Ln(\alpha) + Ln(e^{\beta t})$

$Z_t = A + \beta t$

$A\hat{} = \bar{Z} - \beta\hat{} * \bar{t}$

$$\beta\hat{} = \frac{\sum(t_t - \bar{t})(y_t - \bar{y})}{\sum(t_t - \bar{t})^2}$$

$$\bar{t} = \frac{\sum t_t}{N} = \frac{28}{7} = 4$$

$$\bar{Z} = \frac{\sum Z_t}{N} = \frac{11,16}{7} = 1,59$$

$A\hat{} = \bar{Z} - \beta\hat{} * \bar{t} = 1,59 - (0,65) * (4) = -0,99$

$$\beta\hat{} = \frac{(t_t - \bar{t})(Z_t - \bar{Z})}{(t_t - \bar{t})^2} = \frac{18,09}{28} = 0,65$$

L'équation du Trend (T) donnée par le modèle exponentielle sera présentée ainsi :

$T = Z_t = A + \beta t$

$\Rightarrow Z_t = -0,99 + 0,65\, t$

$Z\hat{}_1 = -0,99 + (0,65) * (1) = -0,34$

$Z\hat{}_2 = -0,99 + (0,65) * (2) = 0,31$

$Z\hat{}_3 = -0,99 + (0,65) * (3) = 0,96$

.
.
.

$Z\wedge_7 = -0{,}99 + (0{,}65) * (7) = 3{,}56$

4. L'élimination de la composante tendancielle dite Trend (T) :

On a :

$e^{z\wedge t} = y\wedge_t$

$y\wedge_t = T$

Et sachant que :

$y_t = T + \varepsilon_t$

$\varepsilon_t = y_t - T$

$\varepsilon_t = y_t - y\wedge_t \qquad / \; y\wedge_t = T$

$\varepsilon_t = y_t - y\wedge_t$

$\varepsilon_1 = y_1 - y\wedge_1$

$\varepsilon_1 = 1 – 0{,}71 = 0{,}29$

$\varepsilon_2 = y_2 - y\wedge_2$

$\varepsilon_2 = 1 – 1{,}36 = 0{,}36$

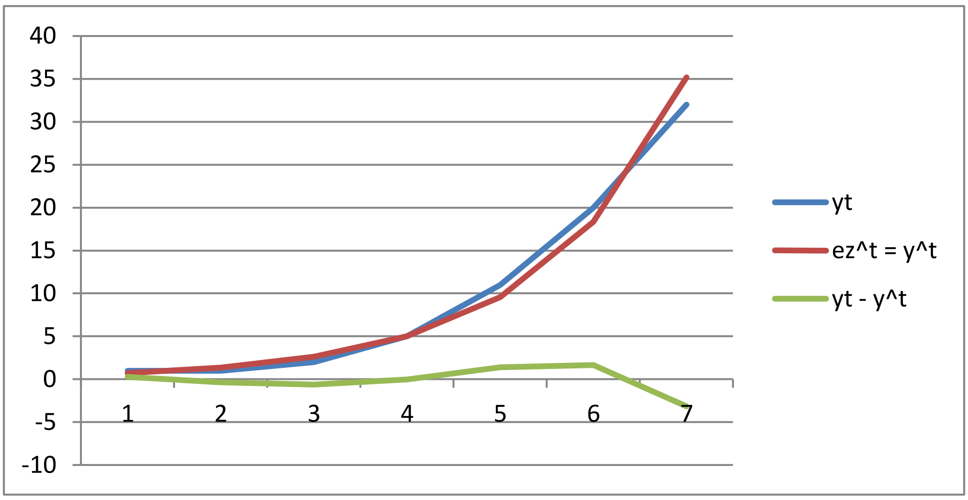
$\varepsilon_3 = y_3 - y\wedge_3$

$\varepsilon_3 = 2 - 2{,}61 = -0{,}61$

.
.
.

$\varepsilon_7 = 32 – 35{,}16 = -3{,}16$

5. La représentation graphique de $y\wedge_t$ et ε_t :



Conclusion générale

Conclusion générale :

A travers ces cours nous avons initié l'étudiant à quelques unes des techniques les plus importantes dans le domaine économique, notamment en finance et gestion, à savoir ; les techniques de prévision, un aspect ô combien important pour l'économiste.

Les séries temporelles comme il est démontré, constituent des séries statistiques à deux variables dont l'une est le temps. C'est cette dernière qui confère à la série sa dimension temporelle et, par là même, la possibilité de déterminer des situations futures à partir de situations présentes ou passées, en tenant comptes des différentes influences qui peuvent affecter le phénomène étudié à différents moments au cours de la période globale considérée. Les influences saisonnières sont les plus fréquentes et celles qui conditionnent fortement le travail du gestionnaire.

Nous avons également vu que les techniques d'analyse des séries temporelles sont diverses et plus ou moins complexes selon les circonstances et le contexte qui caractérisent le phénomène étudié. Les différentes méthodes d'ajustement exposées en sont l'expression.

Bibliographie

Bibliographie :

1. ABDERRAHMANI, **Cours des Séries Temporelles 1**, Master 1, Economie Quantitative, Département des Sciences Economiques, Faculté des Sciences Economique, Commerciale et des Sciences de Gestion, Université de Béjaia, 2020/2021.

2. ANDERSON D. R., SWEENEY D. J., CAMM J. D., WILLIAMS T. A., COCHRAN J. J, **Statistiques pour L'Economie et la Gestion**, De Boeck, Bruxelles, 2015.

3. BALTAGI H. Badi (2011), **Econometrics**, 5ème édition, Springer-Verlag, Heidelberg, Berlin.

4. BOURBONNAIS Regis (2018), **Econométrie**, 10ème édition, Dunod, Paris, France.

5. BOURBONNAIS Régis, **Econométrie, Manuel et Exercices Corrigés**, Edition Dunod, Paris, France, 2005.

6. DOANE P. David, SEWARD E. Lori, **Applied Statistics in Business and Economics**, Fifth Edition, McGraw-Hill Education, 2016.

7. ERTUR Cem, **Méthodologies de Test de la Racine Unitaire**, LATEC, Université de Bourgogne, 1998.

8. FLOYD John E., **Statistics for Economists : A Beginning**, University of Toronto, July 2, 2010.

9. FLUX Jamie, **Business Statistics All in One Skills Practice Workbook: With Full Step by Step Solutions**, Broché, ISBN-13 979-8340780041, Septembre 2024.

10. GREENE William (2005), Traduction de la 5ème édition par AZOMAHOU Théophile & COUDEC Nicolas, **Econométrie**, édition française dirigée par SCHLACTHER Didier, IEP Paris, Université Paris I, France.

11. HADJEM Madjid, **Polycopié Cours de Statistique I**, Semestre 1, 1ere Année Licence, 1, Département des Sciences Commerciales, Faculté des Sciences Economique, Commerciale et des Sciences de Gestion, Université Mouloud MAMMERI de Tizi-Ouzou, 2023/2024, https://dspace.ummto.dz/server/api/core/bitstreams/01256031-7855-41ea-ad3c-dbb7ab165058/content

12. HAMISULTANE Hélène, **Econométrie des Séries Temporelles**, Licence, France, 2002. ffcel01261174, HAL Id : cel-01261174, https://shs.hal.science/cel-01261174v1, Submitted on 24 Jan 2016, https://shs.hal.science/cel-01261174/document.

13. HARDY Adrien, **Séries Temporelles**, 30 mars 2020.

14. PIROTTE Alain et BRESSON Georges (1995), **Econométrie des Séries Temporelles : Théorie et Applications**, 1ère édition, Presse Universitaire, Paris, France.

15. VON SACHS Rainer, VAN BELLEGEM Sébastien, **Séries Chronologiques**, Institut de statistique, Université catholique de Louvain, Belgium, 2005.