# TD01- AJUSTEMENT LINÉAIRE, METHODE DES MOINDRES CARRES (MCO), COEFFICIENT DE DETERMINATION

Rappel de cours :

Soient deux jeux de de données $\{y_1, y_2,..., y_N\}$, $\{x_1, x_2,..., x_N\}$, réalisation de deux variables statistiques X et Y.

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{i=N} x_i \qquad \sigma_X = \sqrt{\frac{1}{N}\sum_{i=1}^{i=N}(x_i - \bar{x})^2 = \frac{1}{N}(\sum_{i=1}^{i=N} {x_i}^2) - \bar{x}^2}$$

$$\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{i=N}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{N}(\sum_{i=1}^{i=N} x_i y_i) - \bar{x}\bar{y} \qquad \rho_{x,y} \coloneqq r_{x,y} \coloneqq \frac{Cov(X,Y)}{\sigma_X \times \sigma_Y} = \frac{\sigma_{xy}}{\sigma_X \times \sigma_y}$$

Soit $D_{Y/X}$ : y=$\beta_0$+$\beta_1 x$ la droite de régression obtenue par la méthode des moindres carrés, alors

$$\beta_1 = \frac{Cov(X,Y)}{Var(X)} \quad et\ \beta_0 = \bar{y} - \beta_1\, \bar{x}$$

Le coefficient de détermination $R^2 = \frac{Variation\ Expliquée}{Variation\ Totale} = \frac{SCE}{SCT} = \rho_{x,y}^2$ avec

SCE: $= \sum_{k=1}^{k=N}(y_k - \bar{y})^2$ et SCR: $= \sum_{k=1}^{k=N}(y_k - \widehat{y_k})^2$ où $\widehat{y_k} = \beta_0 + \beta_1 x_k$

## 1. Exercice 1.1 (QCM)

1. Lors d'une régression simple, si le $R^2$ vaut 1, les points sont-ils alignés ?

A. Non ; B. Oui ; C. Pas obligatoirement.

2. La droite des MCO d'une régression simple passe-t-elle par le point $(\bar{x}, \bar{y})$ ?

A. Toujours ; B. Jamais ; C. Parfois

## 2. Exercice 1.2 ($R^2$ et coefficient de corrélation)

Rappeler la formule définissant le coefficient de détermination $R^2$ et la développer pour montrer qu'il est égal au carré du coefficient de corrélation empirique entre x et y, noté $\rho_{x,y}$, c'est-à-dire qu'on a :

$$R^2 = \rho_{x,y}^2 = \left(\frac{\sum_{i=1}^{i=N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{i=N}(x_i - \bar{x})^2 \times \sum_{i=1}^{i=N}(y_i - \bar{y})^2}\right)^2$$

## 3. Exercice 1.3 (Poids des pères et des fils)

L'étude statistique ci-dessous porte sur les poids respectifs des pères et de leur fil aîné.

Père : 65 63 67 64 68 62 70 66 68 67 69 71

Fils : 68 66 68 65 69 66 68 65 71 67 68 70

Voici les résultats numériques que nous avons obtenus :

$$\sum_{i=1}^{i=12} p_i = 800 \qquad \sum_{i=1}^{i=12} {p_i}^2 = 53418 \qquad \sum_{i=1}^{i=12} f_i p_i = 54107 \qquad \sum_{i=1}^{i=12} f_i = 811 \qquad \sum_{i=1}^{i=12} {f_i}^2 = 54849$$

1. Calculez la droite des moindres carrés du poids des fils en fonction du poids des pères.

2. Calculez la droite des moindres carrés du poids des pères en fonction du poids des fils.

3. Montrer que le produit des pentes des deux droites est égal au carré du coefficient de corrélation empirique entre les $p_i$ et les $f_i$ (ou encore au coefficient de détermination).

## 4. Exercice 1.4 (Hauteur d'un arbre)

Nous souhaitons exprimer la hauteur y (en pieds) d'un arbre d'une essence donnée en fonction de son diamètre x (en pouces) à 1m30 du sol. Pour ce faire, nous avons mesuré 20 couples (diamètre, hauteur) et effectué les calculs suivants : $\bar{x}$ = 4.53, $\bar{y}$ = 8.65 et

$$\frac{1}{20}\sum_{i=1}^{i=N}(x_i-\bar{x})^2 = 10.97 \qquad \frac{1}{20}\sum_{i=1}^{i=N}(y_i-\bar{y})^2 = 2.24 \qquad \sum_{i=1}^{i=N}(x_i-\bar{x})(y_i-\bar{y}) = 3.77$$

1. On note $y = \hat{\beta}_0 + \hat{\beta}_1 x$ la droite de régression. Calculer $\hat{\beta}_0$ et $\hat{\beta}_1$.

2. Donner et commenter une mesure de la qualité de l'ajustement des données au modèle. Exprimer cette mesure en fonction des statistiques élémentaires. Commenter le résultat.

## 5. Exercice 1.5 (Secteur hôtelier)

Un organisme bancaire du secteur hôtelier a besoin de prévoir en fonction du montant des prêts accordés aux professionnels quelles sommes il doit lui-même emprunter sur les marchés financiers. Pour cela il réalise une étude sur les 12 trimestres écoulés. Le montant global des prêts accordés chaque trimestre est donné en millions d'euros. Tous les calculs sont faits à 0,1 près.

| Rang du trimestre $x_i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Montant $y_i$ | 40 | 42 | 44 | 45 | 48 | 50 | 52 | 55 | 58 | 63 | 68 | 70 |

0. Dans un repère orthogonal, représenter le nuage des 12 points. Unités graphiques: 1 cm pour 1 rang sur l'axe des abscisses et 1 cm pour 5 millions d'euros sur l'axe des ordonnées en commençant à la graduation 35.

1. On envisage de résumer quantitativement les valeurs observées à l'aide d'un ajustement affine des 12 points (xi ; yi).

a. Calculer $\bar{x}, \bar{y}$, $\sigma_x$, $\sigma_y$ $\sigma_{XY}$ et le coefficient de corrélation $\rho_{x,y}$

b. Donner une équation de la droite Δ1 : $y = \beta_0 + \beta_1 x$ d'ajustement par la méthode des moindres carrés.

c. Déterminer une estimation des prêts accordés au quatrième trimestre 2003 sachant que le point (1; 40) est celui du premier trimestre 2000.

2. Cette extrapolation des données repose sur les 12 trimestres précédents l'étude, mais pour tenir compte de l'évolution récente, on se limite aux quatre dernières observations. Reprendre alors la méthode des moindres carrés pour déterminer une équation du nouvel ajustement affine Δ2 et donner la prévision plus réaliste pour le quatrième trimestre 2003.

3. Dans un repère orthogonal, représenter le nuage des 12 points et les deux droites Δ1 et Δ2. Unités graphiques: 1 cm pour 1 rang sur l'axe des abscisses et 1 cm pour 5 millions d'euros sur l'axe des ordonnées en commençant à la graduation 35.

## 6. Exercice 1.6 (Mortalité infantile)

Remarque : Le but de cette étude est de constater que des résultats statistiques portant sur une certaine période et possédant des propriétés remarquables (comme ici un ajustement linéaire tout à fait justifié) ne peuvent être extrapolés sans précaution.

L'Annuaire statistique de la France donne le tableau suivant indiquant le taux pour mille (‰) de mortalité infantile par période quinquennales de 1886 à 1940 :

| période | ‰ | période | ‰ |
|---|---|---|---|
| 1886-1890 | 168 | 1916-1920 | 119 |
| 1891-1895 | 170 | 1921-1925 | 95 |
| 1896-1900 | 161 | 1926-1930 | 89 |
| 1901-1905 | 142 | 1931-1935 | 73 |
| 1906-1910 | 129 | 1936-1940 | 70 |
| 1911-1915 | 126 | | |

1. Représenter graphiquement les données précédentes par un nuage (ensemble de points non reliés) de 11 points. On prendra comme abscisse 1,2, ...11 : 1 sera l'année 1888, 2 l'année 1893, ... , 11 l'année 1938.

2. Vous constatez que les points sont susceptibles d'être ajustés linéairement. Calculer le coefficient de corrélation et procédez à l'ajustement par la méthode des moindres carrés.

Tracer la droite.

3. On admet que l'abscisse n d'une année An > 1887 est donnée par : An - 1888 = 5(n - 1).

Quelle est l'abscisse de l'année 1988?

Quel serait le taux en admettant que la droite d'ajustement trouvée est encore valable ?

## 7. Exercice 1.7 (Consommation de graisse par an et par personne en Norvège)

Le tableau suivant donne la consommation de graisse par an et par personne en Norvège, ainsi que le taux de mortalité par athérosclérose pour 100 000 habitants, pendant une période qui couvre la seconde guerre mondiale.

1. Construire le nuage de points.
2. Calculer les caractéristiques numériques des variables X et Y (moyenne, variance et écart-type).
3. Calculer la covariance entre X et Y.
4. Calculer le coefficient de corrélation linéaire.
5. Utiliser les questions 1. et 4. pour conclure.
taux en utilisant la droite de régression obtenue à la question précédente.

| *Année* | *Consommation de graisse en kg par an et par personne, X* | *Taux de mortalité par athérosclérose pour 100 000 habitants, Y* |
|---|---|---|
| 1938 | 14,4 | 29,1 |
| 1939 | 16 | 29,7 |
| 1940 | 11,6 | 29,2 |
| 1941 | 11 | 26 |
| 1942 | 10 | 24 |
| 1943 | 9,6 | 23,1 |
| 1944 | 9,2 | 23 |
| 1945 | 10,4 | 23,1 |
| 1946 | 11,4 | 25,2 |
| 1947 | 12,5 | 26,1 |

6. Calculer les coefficients de régression linéaire de Y sur X. Tracer sur le graphique de la question 1. la droite de régression.

7. En 1948, la consommation de graisse était de 13 kg par personne. Le taux de mortalité par athérosclérose n'a pas été relevé cette année-là. Donner une estimation de ce taux en utilisant la droite de régression obtenue à la question précédente.

## 8. Exercice 1.8

Le tableau suivant indique les différentes valeurs obtenues pour deux variables quantitatives X et Y dans un échantillon de taille 8.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| X | 1,3 | 1,5 | 2,5 | 2,5 | 2,7 | 3 | 4 | 5 |
| Y | 2,3 | 3,5 | 3,5 | 4,5 | 4,7 | 3 | 2 | 1 |

1. Construire le nuage de points.

2. Calculer le coefficient de corrélation linéaire.

3. Peut-on conclure dans ce cas à l'existence d'un lien linéaire entre X et Y ? Que suggère plutôt le nuage de points ?

4. À l'aide du nuage de points, extraire de l'échantillon initial deux sous-échantillons de taille 4 chacun et calculer le coefficient de corrélation linéaire pour ces deux échantillons. Que constate-t-on ?

# TD02- MOYENNES MOBILES

## 1. Exercice (Vrai-Faux & QCM)

1. QCM

| Ennoncé | VRAI | FAUX |
|---|---|---|
| 1 - Le graphique de la bande permet de connaître de façon simple si le schéma de décomposition est de type additif ou multiplicatif | | |
| 2 - Le filtre des moyennes mobiles utilisé dans le cours est centré. | | |
| 3 - Appliquer une moyenne mobile à une série composée d'une tendance et d'une saison, élimine la composante saisonnière | | |
| 4 - Une moyenne mobile centrée réalisée sur une série, fait perdre autant de points en fin de série que la longueur de la moyenne mobile | | |
| 5 - La moyenne mobile centrée permet d'approcher la composante de long terme | | |

2. Pour une chronique à 12 termes :

a) on peut calculer 8 moyennes mobiles centrées de longueur 4

b) on peut calculer une moyenne mobile centrée de longueur 12

c) on peut calculer 10 médianes mobiles centrées de longueur 3

d) on peut calculer 2 moyennes mobiles centrées de longueur 11

3. Si une chronique X a une composante saisonnière de période p, alors :

a) les moyennes mobiles centrées de longueur 2 p éliminent la saisonnalité

b) on peut approximer la tendance par la suite des moyennes mobiles centrées de longueur p

c) la somme de p termes successifs de X donne une approximation de la moyenne de la tendance

d) on peut toujours calculer (T-p) moyennes mobiles centrées de longueur p si elle a T termes

## 2. Exercice (air ATMO)

On dispose aussi de la répartition mensuelle du niveau de l'indice de la qualité de l'air ATMO dans l'agglomération parisienne selon trois classes de niveau pour les six années agrégées.
Légende : Niveau 1 à 4 : très bon à bon.
Légende : Niveau 5 à 7 : moyen à médiocre.
Légende : Niveau 8 à 10 : mauvais à très mauvais.
On s'intéresse à la classe de niveau « 5 à 7 ».

*Tableau 1. Fréquences mensuelles d'apparition des indices de 1998 à 2003*

| Niveau | 1 à 4 | 5 à 7 | 8 à 10 | Nombre total de jours |
|---|---|---|---|---|
| Janvier | 164 | 22 | 0 | 186 |
| Février | 136 | 29 | 4 | 169 |
| Mars | 151 | 35 | 0 | 186 |
| Avril | 152 | 28 | 0 | 180 |
| Mai | 132 | 54 | 0 | 186 |
| Juin | 115 | 65 | 0 | 180 |
| Juillet | 123 | 59 | 4 | 186 |
| Août | 93 | 83 | 10 | 186 |
| Septembre | 155 | 25 | 0 | 180 |
| Octobre | 155 | 31 | 0 | 186 |
| Novembre | 172 | 8 | 0 | 180 |
| Décembre | 177 | 9 | 0 | 186 |
| Nombre total de jours | 1725 | 448 | 18 | 2191 |

1. Représentez graphiquement son évolution au cours des 12 mois.

2. Calculez la suite des moyennes mobiles de longueur 3 et représentez-la sur le même graphique. Quelle propriété de la moyenne mobile venez-vous d'illustrer ?

## 3. Exercice (Moyenne mobile)

| t | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $x_t$ | 3 | -1 | 5 | 1 | 3 | -1 | 5 | 1 | 3 | -1 | 5 | 1 |

1. Calculez les suites des moyennes mobiles de longueurs 2, 3, 4 et 5.

Quelles sont les propriétés de la moyenne mobile qui sont illustrées par cet exemple ?

2. Soit la chronique $x1_t = 10 - 2t + x_t$ , calculez la suite des moyennes mobiles de longueur 4 de la nouvelle série $z_t$

## 4. Exercice (Moyennes mobiles)

On considère la série chronologique représentant les ventes d'huitres en tonnes par trimestres
1. Calculez la moyenne mobile centrée d'ordre 4 pour t = 3 63,125 ou 65,525
2. Calculez la moyenne mobile centrée d'ordre 4 pour t = 4 66.775 ou 65,875
3. Calculez la moyenne mobile centrée d'ordre 4 pour t = 5 69,125 ou 68,197
4.Calculez l'ensemble des moyennes mobiles centrées d'ordre 4 et les représenter dans un tableau, par année et par trimestre

| | Série initiale | | | | Moyenne mobile d'ordre 4 | | | |
|---|---|---|---|---|---|---|---|---|
| | Tri-1 | Tri-2 | Tri-3 | Tri-4 | Tri-1 | Tri-2 | Tri-3 | Tri-4 |
| 2010 | 52 | 36 | | | | | | |
| 2011 | 65 | | | | | | | |
| 2012 | | | | | | | | |
| 2013 | | | | | | | | |

*Tableau 2. Ventes d'huitres en tonnes par trimestres*

Ventes d'huitres

| Période | Tonnes |
|---|---|
| 1er tri 10 | 52 |
| 2e tri 10 | 36 |
| 3e tri 10 | 69 |
| 4e tri 10 | 89 |
| 1er tri 11 | 65 |
| 2e tri 11 | 45 |
| 3e tri 11 | 86 |
| 4e tri 11 | 111 |
| 1er tri 12 | 81 |
| 2e tri 12 | 56 |
| 3e tri 12 | 108 |
| 4e tri 12 | 139 |
| 1er tri 13 | 102 |
| 2e tri 13 | 70 |
| 3e tri 13 | 135 |
| 4e tri 13 | 174 |

# TD03- MODELES MULTIPLICATIF ET ADDITIF

## 1. Exercice (nombre de mariages, CVS et modèle additif)

On a relevé le nombre de mariages dans une ville du sud-ouest de la France chaque trimestre pendant 3 ans :
On notera Y la variable dont on étudie l'évolution.
1. Représenter graphiquement cette série chronologique (avec périodes superposées puis avec périodes successives). Commenter.
2. Calculer la série des moyennes mobiles, lisser la courbe.

| Trimestre \ Année | 1998 | 1999 | 2000 |
|---|---|---|---|
| 1 | 10 | 11 | 12 |
| 2 | 12 | 14 | 15 |
| 3 | 13 | 15 | 17 |
| 4 | 11 | 12 | 12 |

3. Calculer l'équation de la droite de tendance et tracer cette droite sur le graphique précédent.

4. Calculer les quatre coefficients saisonniers (pour le modèle additif).

5. Utiliser le modèle construit pour prévoir le nombre de mariages dans cette ville en 2002.

## 2. Exercice (nombre de mariages, CVS et modèle additif)

*Tableau 3. Source : INSEE- Division indicateurs conjoncturels d'activité.*

Le tableau ci-contre donne les indices trimestriels de la production industrielle (base 100 en 1980) en ce qui concerne les produits de la parachimie et de la pharmacie :
1. Faire une étude complète de cette série chronologique (reprendre les 4 premières questions de l'exercice 1.1).

| Trimestre \ Année | 1982 | 1983 | 1984 |
|---|---|---|---|
| 1 | 106 | 106,1 | 109,6 |
| 2 | 108,8 | 107,1 | 108,5 |
| 3 | 97,9 | 98,5 | 103 |
| 4 | 108,5 | 111,2 | 115,5 |

2. On note toujours Y la variable dont on étudie l'évolution et afin de réduire les calculs, voici un tableau (à compléter) qui fournit quelques résultats

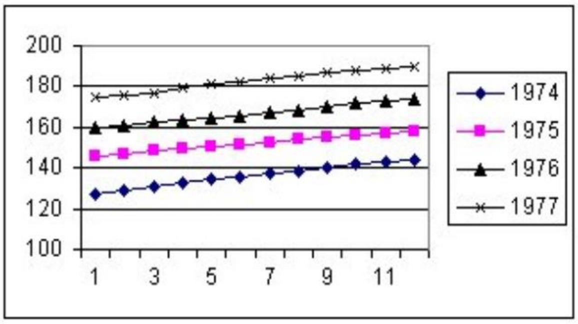
Quel est l'indice prévisible pour le 3-ième trimestre 1985 ?

| $t$ | $y(t)$ | $z(t)$ | $t*z(t)$ | $x(t)=0,49t+103,04$ | $y(t)-x(t)$ | $s'(t)$ | $s(t)$ |
|---|---|---|---|---|---|---|---|
| 1 | 106 | | | 103,53 | 2,47 | 1,743 | |
| 2 | 108,8 | | | 104,02 | 4,78 | 2,153 | |
| 3 | 97,9 | 105,31 | 315,94 | 104,51 | -6,61 | | |
| 4 | 108,5 | 105,11 | 420,45 | 105 | 3,5 | | |
| 5 | 106,1 | 104,98 | 524,88 | 105,49 | 0,61 | | |
| 6 | 107,1 | 105,39 | 632,33 | 105,98 | 1,12 | | |
| 7 | 98,5 | 106,16 | 743,14 | 106,47 | -7,97 | | |
| 8 | 111,2 | 106,78 | 854,20 | 106,96 | 4,24 | | |
| 9 | 109,6 | | | | | | |
| 10 | 108,5 | | | | | | |
| 11 | 103 | | | | | | |
| 12 | 115,5 | | | | | | |

## 3. Exercice (Moyenne mobile et CVS)

|  | 1974 | 1975 | 1976 | 1977 |
|---|---|---|---|---|
| Janvier | 127,4 | 145,9 | 159,9 | 174,3 |
| Février | 129,1 | 147 | 161 | 175,5 |
| Mars | 130,6 | 148,2 | 162,4 | 177,1 |
| Avril | 132,7 | 149,5 | 163,8 | 179,4 |
| Mai | 134,3 | 150,6 | 164,9 | 181,1 |
| Juin | 135,8 | 151,7 | 165,6 | 182,5 |
| Juillet | 137,5 | 152,8 | 167,2 | 184,1 |
| Août | 138,6 | 153,8 | 168,4 | 185,1 |
| Septembre | 140,1 | 155,1 | 170,2 | 186,7 |
| Octobre | 141,8 | 156,3 | 171,8 | 188,2 |
| Novembre | 143,1 | 157,3 | 173,2 | 188,9 |
| Décembre | 144,3 | 158,2 | 173,8 | 189,4 |

On se propose d'étudier l'évolution de l'indice mensuel des prix à la consommation de 1974 à 1977. Les observations mensuelles sont consignées dans le tableau ci-contre.
1. Calculer la première et la dernière moyenne mobile.




2. Voici la représentation graphique de cette série chronologique avec périodes superposées :
Est-il intéressant ici de calculer la série des moyennes mobiles ?
3. Qu'apporterait le calcul des variations saisonnières ?
4. Quelle méthode vous paraît la mieux adaptée pour faire des prévisions ?

## 4. Exercice (Disneyland-Paris & Moyenne mobile et CVS)

*Tableau 4. Source : Disneyland-Paris, Marne la Vallée, 1998.*

On étudie la fréquentation des hôtels américains de Disneyland-Paris entre 1995 et 1997.
Le nombre de clients (en milliers) est donné dans le tableau ci-contre :

| Année | Trimestre 1 | Trimestre 2 | Trimestre 3 | Trimestre 4 |
|---|---|---|---|---|
| 1995 | 8,2 | 12,3 | 32,7 | 8,3 |
| 1996 | 10 | 14,5 | 37,3 | 11,5 |
| 1997 | 11 | 17,3 | 41,3 | 13,3 |

1. Compléter la série des moyennes mobiles dont les 6 dernières valeurs sont : 16,95 ; 17,925 ; 18,45 ; 18,925 ; 19,775 ; 20,5.

Représenter sur un même graphique la série des données et la série des moyennes mobiles. Commenter.

2. Déterminer l'équation de la droite de tendance et la tracer sur le même graphique.

3. On donne la somme des composantes saisonnières : S0 = 0,567. Calculer les coefficients saisonniers du 1$^{er}$ et du 3$^{ième}$ trimestre.

4. En utilisant le modèle que l'on a construit, quelle fréquentation peut-on prévoir pour le troisième trimestre 2000 ?

## 5. Exercice 1.5 (billets vendus & moyennes mobiles et modèle additif)

On étudie l'évolution du nombre de billets vendus (en milliers) dans un complexe cinématographique lors des trois premières années :

| | Janv-Fév | Mars-Avril | Mai-Juin | Juillet-Août | Sep-Oct | Nov-Déc |
|---|---|---|---|---|---|---|
| 1997 | 100 | 82 | 70 | 40 | 62 | 91 |
| 1998 | 105 | 94 | 73 | 43 | 72 | 106 |
| 1999 | 111 | 99 | 84 | 52 | 77 | 118 |

1. Représenter graphiquement ces données dans un repère cartésien.

2. Calculer la série des moyennes mobiles.

3. Montrer que l'équation de la droite de tendance est : x(t) = 1,34t + 68,93. Tracer cette droite sur le graphique précédent.

4. On donne la somme des composantes saisonnières : S0 = 3,04.

Calculer le coefficient saisonnier pour Juillet-Août. Utiliser ce coefficient pour faire des prévisions pour Juillet-Août 2000.

## 6. Exercice 1.6 (magasin d'insecticides, moyennes mobiles et modèle additif)

Voici les résultats du chiffre d'affaire d'un magasin d'insecticides d'Egletons (département de la Corrèze) au cours des douze derniers trimestres (en milliers de francs) :

| | 1998 | 1999 | 2000 |
|---|---|---|---|
| 1er trimestre | 27 | 29 | 30 |
| 2ème trimestre | 10 | 12 | 13 |
| 3ème trimestre | 15 | 16 | 17 |
| 4ème trimestre | 26 | 27 | 28 |

1. Sur le graphique de la page suivante, faire une représentation cartésienne avec périodes successives de la série chronologique précédente.
Quelles conjectures peut-on tirer de ce graphique ?

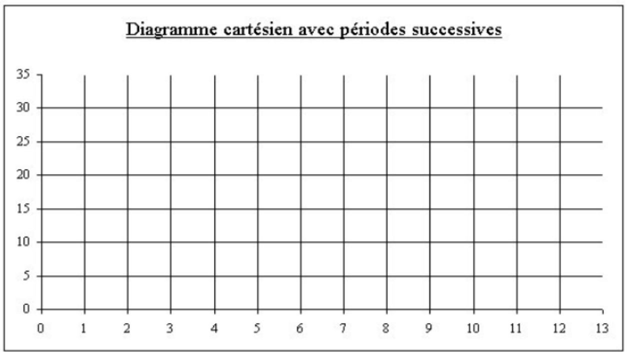



*Figure 1. Diagramme cartésien pour les insecticides*

2. Afin de dégager la tendance, le propriétaire du magasin décide d'utiliser la méthode des moyennes mobiles ; compléter le tableau avec les moyennes mobiles, z(t), non calculées.
3. Pour avancer les calculs, on donne $\bar{Z} = 20{,}938$. Montrer que l'équation de la droite de tendance est x(t) = 0,289t + 19,061.
Tracer sur le graphique précédent cette droite.
4. Compléter sur le tableau les différents calculs relatifs aux coefficients saisonniers(δ(t), s'(t) et s(t)).
5. Tracer sur le graphique les estimations (les valeurs de $\hat{y}(t)$) que donne le modèle pour l'année 2000. Au vu des résultats obtenus, le modèle vous paraît-il pertinent ?
6. Déduire des calculs précédents, la prévision du chiffre d'affaire du magasin pour le troisième trimestre 2001.

| $t$ | $z(t)$ | $\delta(t)$ | $s'(t)$ | $s(t)$ |
|---|---|---|---|---|
| 1 | | 7,650 | | |
| 2 | | -9,639 | | |
| 3 | 19,750 | | -5,084 | -4,978 |
| 4 | | | 5,627 | 5,733 |
| 5 | 20,625 | 8,494 | | |
| 6 | 20,875 | -8,795 | | |
| 7 | 21,125 | -5,084 | -5,084 | -4,978 |
| 8 | 21,375 | 5,627 | 5,627 | 5,733 |
| 9 | 21,625 | 8,338 | | |
| 10 | | -8,951 | | |
| 11 | | -5,240 | -5,084 | -4,978 |
| 12 | | 5,471 | 5,627 | 5,733 |

## 7. Exercice 1.7 (construction, moyennes mobiles et modèle additif)

Le tableau suivant donne l'évolution de l'indice du coût de la construction (base 100 au 4ème trimestre 1953). Cet indice est utilisé par exemple pour la révision annuelle des loyers. On notera Y la variable « indice du coût de la construction » et T le temps.

*Tableau 5. Source INSEE*

| Trimestres \ Années | 1998 | 1999 | 2000 |
|---|---|---|---|
| 1 | 1058 | 1071 | 1083 |
| 2 | 1058 | 1074 | 1089 |
| 3 | 1057 | 1080 | 1093 |
| 4 | 1074 | 1065 | 1127 |

On pourra utiliser le tableau et le graphique ci-dessous.
1. Représenter graphiquement cette série chronologique avec périodes successives.
Peut-on parler ici de variations saisonnières ?

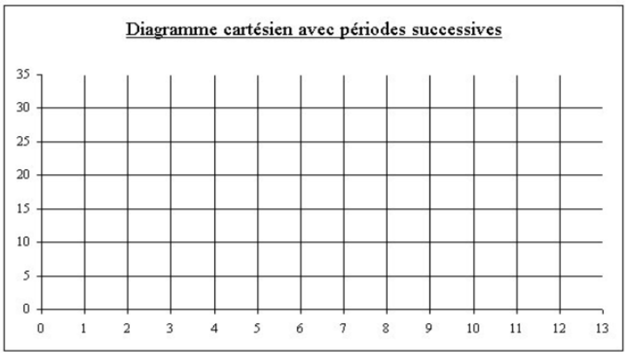



*Figure 2. Diagramme cartésien pour la construction*

2. Compléter la série des moyennes mobiles (expliciter le calcul des z(t)) et représenter cette série sur le graphique ci-dessus.
Quel est l'intérêt des moyennes mobiles ?
Peut-on utiliser le modèle additif pour décrire cette série chronologique ?

| $t$ | $y(t)$ | $z(t)$ |
|---|---|---|
| 1 | 1058 | |
| 2 | 1058 | |
| 3 | 1057 | 1063,375 |
| 4 | 1074 | 1067 |
| 5 | 1071 | 1071,875 |
| 6 | 1074 | 1073,625 |
| 7 | 1080 | 1074 |
| 8 | 1065 | |
| 9 | 1083 | |
| 10 | 1089 | |
| 11 | 1093 | |
| 12 | 1127 | |

## 8. Exercice (TrucNet, Modèle additif + moyenne mobile)

Le tableau ci-contre donne le chiffre d'affaires trimestriel de l'entreprise TrucNet sur la période 2017 à 2017.
1- Tracer la série chronologique X.
2- Quel modèle peut-on appliquer sur la série chronologique X ? Justifier votre choix.
3- Après avoir justifié la longueur de la moyenne mobile à employer, calculer la série Y des moyennes mobiles et représenter les deux courbes sur un même graphique.

| | Trimestre 1 | Trimestre 2 | Trimestre 3 | Trimestre 4 |
|---|---|---|---|---|
| 2014 | 120 | 181 | 71 | 119 |
| 2015 | 128 | 190 | 73 | 124 |
| 2016 | 140 | 196 | 84 | 133 |
| 2017 | 145 | 206 | 96 | 142 |

*Figure 3. Chiffre d'affaires trimestriel de TrucNet*

# TD04- LISSAGE EXPONENTIEL

## 1. Exercice (QCM)

Une prévision par lissage exponentiel simple :

a) tient d'autant plus compte des valeurs récentes de la série que la constante α est faible

b) peut s'envisager pour une chronique possédant une composante saisonnière

c) ne peut pas s'envisager pour une chronique possédant une tendance à la hausse

d) s'adapte d'autant plus rapidement à un changement de niveau de la chronique que α est élevée

## 2. Exercice (Nombre d'immatriculations de voitures avec un moteur diesel)

L'Institut National de la Statistique et des Etudes Economiques (INSEE) vous charge d'un projet comprenant deux objectifs : mener une étude exploratoire sur le nombre d'immatriculations de voitures particulières neuves avec un moteur diesel, puis réaliser une prévision de ce nombre d'immatriculations pour chacun des mois des années 2006 et 2007. Pour mener à bien ce projet, vous disposez des données suivantes exprimées en milliers (source INSEE - site internet)

| | Jan. | Feb. | Mar. | Apr. | May | Jun. | Jul. | Aug. | Sep. | Oct. | Nov. | Dec. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1982 | 17.9 | 17.8 | 22.3 | 18.8 | 15.4 | 17.6 | 16.4 | 14.0 | 17.3 | 20.2 | 21.0 | 22.6 |
| 1983 | 17.2 | 14.6 | 17.6 | 13.8 | 14.1 | 16.7 | 15.0 | 11.8 | 13.4 | 17.7 | 20.2 | 20.7 |
| 1984 | 17.1 | 15.7 | 18.6 | 17.4 | 20.9 | 18.9 | 19.0 | 16.9 | 15.9 | 27.3 | 27.4 | 25.0 |
| 1985 | 23.9 | 22.2 | 24.0 | 21.5 | 20.4 | 19.0 | 23.6 | 15.7 | 17.8 | 26.5 | 25.0 | 25.2 |
| 1986 | 24.9 | 19.3 | 21.6 | 23.6 | 20.5 | 19.7 | 26.0 | 22.5 | 21.0 | 31.2 | 30.7 | 38.5 |
| 1987 | 27.7 | 27.4 | 32.5 | 32.3 | 27.0 | 24.9 | 37.9 | 26.6 | 23.1 | 41.2 | 38.2 | 45.4 |
| 1988 | 37.6 | 32.4 | 43.1 | 40.9 | 40.8 | 31.0 | 53.9 | 36.9 | 32.2 | 45.7 | 53.1 | 74.9 |
| 1989 | 66.2 | 59.9 | 57.8 | 56.4 | 46.5 | 46.7 | 72.4 | 50.9 | 47.5 | 68.3 | 62.7 | 51.8 |
| 1990 | 82.0 | 59.7 | 65.8 | 58.5 | 57.2 | 44.5 | 75.0 | 53.5 | 44.4 | 77.9 | 74.9 | 68.7 |
| 1991 | 76.7 | 60.8 | 66.5 | 68.2 | 54.4 | 46.0 | 91.8 | 55.0 | 50.8 | 76.3 | 66.2 | 67.8 |
| 1992 | 74.3 | 60.7 | 68.8 | 68.0 | 55.9 | 53.5 | 93.0 | 57.2 | 58.2 | 81.2 | 74.6 | 75.2 |
| 1993 | 58.5 | 60.0 | 70.2 | 61.1 | 51.3 | 49.3 | 88.5 | 52.7 | 53.4 | 78.5 | 79.5 | 80.0 |
| 1994 | 64.6 | 64.0 | 86.0 | 80.8 | 68.8 | 53.8 | 102.8 | 68.9 | 67.3 | 87.6 | 101.0 | 94.4 |
| 1995 | 84.9 | 81.2 | 88.9 | 71.5 | 65.4 | 73.1 | 97.9 | 66.4 | 55.9 | 72.7 | 77.3 | 62.7 |
| 1996 | 81.9 | 74.6 | 76.8 | 74.7 | 57.7 | 47.8 | 97.1 | 61.8 | 71.7 | 70.3 | 58.5 | 65.0 |
| 1997 | 53.7 | 55.4 | 59.6 | 67.0 | 49.1 | 37.3 | 94.1 | 50.2 | 55.0 | 68.1 | 60.8 | 65.9 |
| 1998 | 60.1 | 59.2 | 70.5 | 67.4 | 55.6 | 40.8 | 101.7 | 54.7 | 63.4 | 66.3 | 71.1 | 70.1 |
| 1999 | 64.5 | 63.7 | 82.5 | 80.1 | 62.8 | 44.1 | 136.7 | 76.5 | 70.7 | 86.3 | 87.3 | 92.3 |
| 2000 | 82.8 | 87.4 | 99.6 | 86.8 | 92.5 | 89.5 | 94.1 | 75.4 | 76.4 | 91.8 | 91.9 | 78.3 |
| 2001 | 99.8 | 92.8 | 110.1 | 108.5 | 112.3 | 124.4 | 121.8 | 84.5 | 88.8 | 116.4 | 110.4 | 98.0 |
| 2002 | 117.9 | 106.7 | 123.4 | 128.7 | 114.7 | 126.6 | 127.7 | 75.7 | 95.9 | 122.7 | 109.5 | 105.4 |
| 2003 | 118.2 | 107.8 | 125.2 | 121.3 | 108.1 | 136.6 | 125.4 | 70.1 | 105.0 | 125.8 | 99.9 | 110.6 |
| 2004 | 108.8 | 104.8 | 133.5 | 118.3 | 117.4 | 148.5 | 116.6 | 73.5 | 108.0 | 119.3 | 124.3 | 119.9 |
| 2005 | 117.2 | 113.0 | 143.8 | 134.3 | 126.6 | 151.4 | 108.0 | 77.0 | 114.5 | 109.6 | 117.8 | 115.7 |

<u>A - Phase exploratoire</u>

1. Compléter les tableaux suivants :

| Année | Total | Moyenne |
|---|---|---|
| 1982 | · | 18.44 |
| 1983 | 192.8 | 16.07 |
| 1984 | 240.1 | 20.01 |
| 1985 | 264.8 | 22.07 |
| 1986 | 299.5 | 24.96 |
| 1987 | 384.2 | 32.02 |
| 1988 | 522.5 | 43.54 |
| 1989 | · | 57.26 |
| 1990 | 762.1 | 63.51 |
| 1991 | 780.5 | · |
| 1992 | 820.6 | 68.38 |
| 1993 | 783.0 | 65.25 |

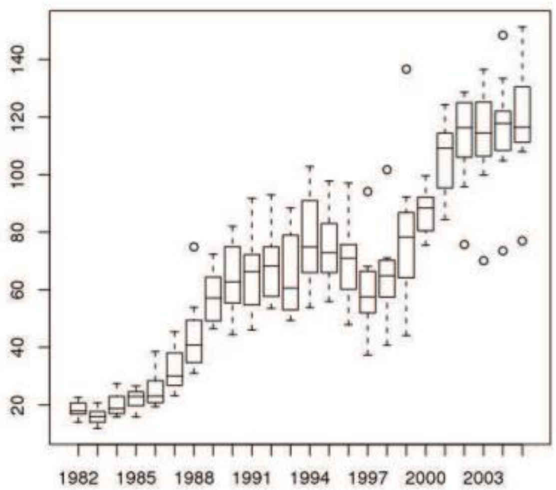
| Année | Total | Moyenne |
|---|---|---|
| 1994 | 940.0 | 78.33 |
| 1995 | 897.9 | 74.83 |
| 1996 | 837.9 | 69.83 |
| 1997 | 716.2 | 59.68 |
| 1998 | 780.9 | 65.08 |
| 1999 | 947.5 | 78.96 |
| 2000 | 1046.5 | 87.21 |
| 2001 | 1267.8 | · |
| 2002 | 1354.9 | 112.91 |
| 2003 | 1354.0 | 112.83 |
| 2004 | · | 116.08 |
| 2005 | 1428.9 | 119.08 |

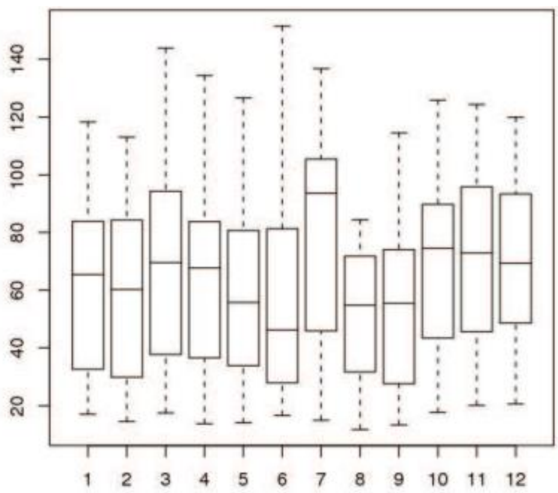
2. Représenter graphiquement la série correspondant à la moyenne mensuelle d'immatriculations de voitures particulières neuves avec un moteur diesel par an. Interpréter succinctement ce graphique en mettant en évidence les différentes phases.

3. Compléter le tableau suivant, puis commenter :

| Mois | Total | Moyenne | Ecart-type |
|---|---|---|---|
| Janvier | · | 65.77 | 33.24 |
| Février | 1461.1 | 60.88 | 31.29 |
| Mars | 1708.7 | 71.20 | 37.98 |
| Avril | 1619.9 | 67.50 | 36.20 |
| Mai | 1455.4 | 60.64 | 34.64 |
| Juin | 1461.7 | 60.90 | 43.98 |
| Juillet | 1936.4 | 80.68 | 38.68 |
| Août | 1248.4 | 52.02 | 23.01 |
| Septembre | 1367.6 | 56.98 | 30.74 |
| Octobre | 1728.9 | · | 33.09 |
| Novembre | 1683.3 | 70.14 | 31.69 |
| Décembre | 1674.1 | 69.75 | 29.72 |

4. Identifier chacun des graphiques suivants dans le contexte, puis les commenter (médianes, dispersions, et valeurs atypiques).



5. La représentation graphique de la série est la suivante. Commenter ce graphique :

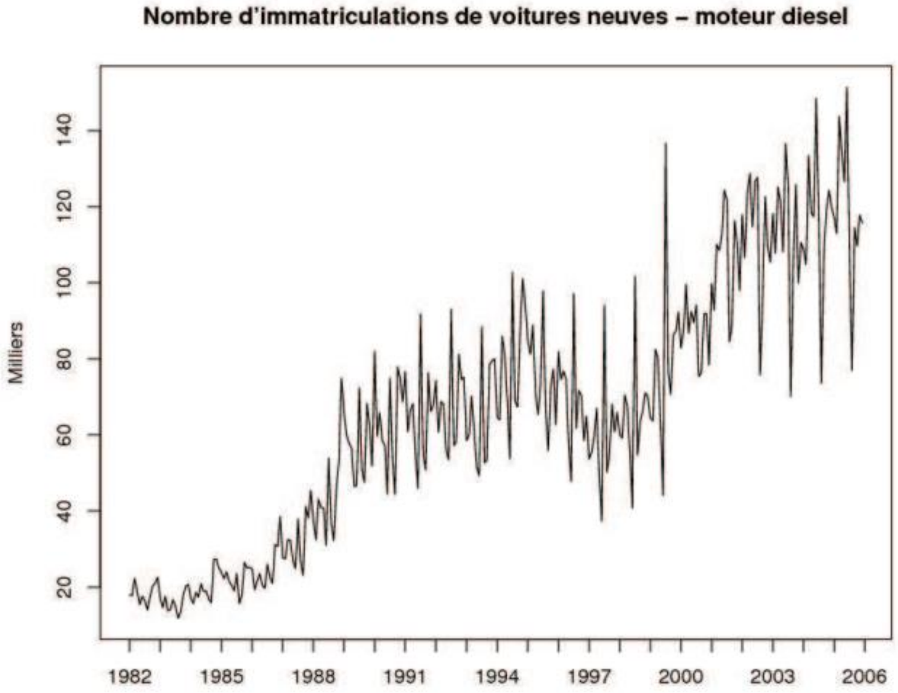
Le phénomène observé est-il croissant, décroissant, ou stationnaire ?

Visualise-ton une composante tendancielle déterministe ou aléatoire ?

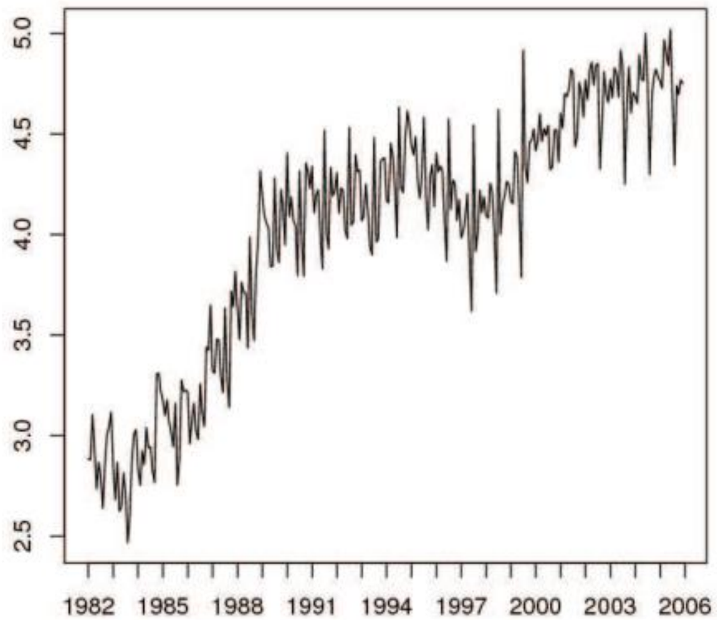
Observe-t-on des ruptures de tendance ?

Détecte-t-on un phénomène saisonnier ?

Existe-t-il des points atypiques ?




6. Déterminer la transformation que l'on devrait mener afin d'homogénéiser la volatilité de la série sur la période d'observation. Via cette transformation on obtient le graphique suivant. Cette transformation a-t-elle été efficace ?




B - Phase de modélisation

La première phase a permis de mettre en évidence l'existence d'une composante tendancielle aléatoire. Dans la phase de modélisation, on propose dans un premier temps de faire usage d'un modèle de type Holt, c'est-à-dire de réaliser un lissage exponentiel double. On rappelle à ce propos que cette méthode consiste à estimer la valeur de la variable à l'instant t + h, étant donné que l'on se situe à l'instant t, selon la relation :

$$\hat{y}_t(h) = a_t + h * b_t,$$

où les coefficients $a_t$ et $b_t$, représentant respectivement l'ordonnée à l'origine et la pente de la droite à l'instant $t$ (selon une notation anglosaxone), sont calculés par récurrence via le système suivant :

$$\begin{cases} a_t = \alpha y_t + (1-\alpha)\,(a_{t-1} + b_{t-1}) \\ b_t = \beta(a_t - a_{t-1}) + (1-\beta) b_{t-1} \end{cases}$$

où α et β sont deux paramètres appelées constantes de lissage. Cette double relation de récurrence nécessite une initialisation qui, par défaut, consiste en générale à choisir les quantités suivantes : $a_2 = y_2$ et $b_2 = y_2 - y_1$

1. Dans le cas présent, on va travailler avec le logarithme de la série dont les valeurs pour l'année 1982 sont respectivement :

2.885 2.879 3.105 2.934 2.734 2.868 2.797 2.639 2.851 3.006 3.045 3.118

Déduire de ces valeurs les quantités a2 et b2, puis la quantité $\hat{y}_2(1)$

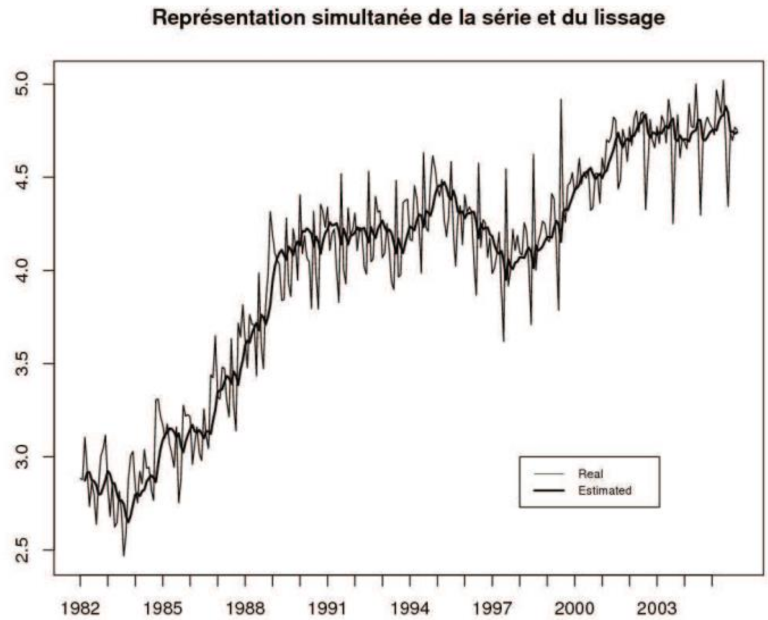
2. En choisissant α = 0.2 et β = 0.6, compléter le tableau suivant :

| | Jan. | Feb. | Mar. | Apr. | May | Jun. | Jul. | Aug. | Sep. | Oct. | Nov. | Dec. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1982 | - | - | . | . | . | . | . | . | . | . | . | . |
| 1983 | 2.97 | 2.99 | 2.93 | 2.91 | 2.82 | 2.72 | 2.69 | 2.65 | 2.55 | 2.49 | 2.55 | 2.68 |
| 1984 | 2.83 | 2.91 | 2.94 | 3.00 | 3.01 | 3.06 | 3.07 | 3.06 | 3.00 | 2.91 | 3.00 | 3.11 |
| 1985 | 3.19 | 3.24 | 3.25 | 3.27 | 3.23 | 3.17 | 3.08 | 3.06 | 2.92 | 2.84 | 2.90 | 2.98 |
| 1986 | 3.07 | 3.16 | 3.15 | 3.17 | 3.19 | 3.16 | 3.11 | 3.14 | 3.13 | 3.10 | 3.20 | 3.30 |
| 1987 | 3.47 | 3.52 | 3.53 | 3.57 | 3.59 | 3.53 | 3.43 | 3.46 | 3.39 | 3.28 | 3.36 | 3.44 |
| 1988 | 3.58 | 3.67 | 3.68 | 3.76 | 3.81 | 3.83 | 3.75 | 3.82 | 3.78 | 3.68 | 3.69 | 3.76 |
| 1989 | 3.95 | 4.11 | 4.21 | 4.27 | 4.28 | 4.20 | 4.09 | 4.12 | 4.04 | 3.95 | 3.98 | 4.00 |
| 1990 | 3.98 | 4.10 | 4.14 | 4.19 | 4.19 | 4.17 | 4.06 | 4.11 | 4.06 | 3.96 | 4.03 | 4.12 |
| 1991 | 4.18 | 4.28 | 4.28 | 4.30 | 4.30 | 4.23 | 4.09 | 4.16 | 4.10 | 4.02 | 4.07 | 4.10 |
| 1992 | 4.14 | 4.21 | 4.21 | 4.25 | 4.26 | 4.21 | 4.13 | 4.23 | 4.19 | 4.14 | 4.20 | 4.25 |
| 1993 | 4.29 | 4.25 | 4.21 | 4.21 | 4.17 | 4.07 | 3.97 | 4.06 | 4.03 | 3.99 | 4.08 | 4.20 |
| 1994 | 4.31 | 4.34 | 4.34 | 4.42 | 4.46 | 4.43 | 4.31 | 4.38 | 4.34 | 4.28 | 4.32 | 4.41 |
| 1995 | 4.48 | 4.52 | 4.52 | 4.54 | 4.48 | 4.37 | 4.30 | 4.34 | 4.27 | 4.16 | 4.13 | 4.15 |
| 1996 | 4.12 | 4.18 | 4.23 | 4.29 | 4.33 | 4.28 | 4.16 | 4.24 | 4.21 | 4.22 | 4.23 | 4.18 |
| 1997 | 4.16 | 4.09 | 4.02 | 4.00 | 4.02 | 3.97 | 3.82 | 3.98 | 3.97 | 3.99 | 4.07 | 4.12 |
| 1998 | 4.19 | 4.21 | 4.20 | 4.24 | 4.26 | 4.21 | 4.05 | 4.17 | 4.12 | 4.11 | 4.13 | 4.17 |
| 1999 | 4.21 | 4.22 | 4.22 | 4.29 | 4.36 | 4.33 | 4.18 | 4.37 | 4.40 | 4.39 | 4.44 | 4.48 |
| 2000 | 4.53 | 4.53 | 4.54 | 4.58 | 4.57 | 4.57 | 4.55 | 4.55 | 4.48 | 4.40 | 4.39 | 4.40 |
| 2001 | 4.37 | 4.43 | 4.47 | 4.56 | 4.65 | 4.73 | 4.83 | 4.91 | 4.83 | 4.74 | 4.73 | 4.70 |
| 2002 | 4.64 | 4.65 | 4.64 | 4.68 | 4.74 | 4.76 | 4.82 | 4.86 | 4.73 | 4.65 | 4.66 | 4.64 |
| 2003 | 4.63 | 4.65 | 4.66 | 4.72 | 4.76 | 4.77 | 4.84 | 4.88 | 4.71 | 4.66 | 4.67 | 4.63 |
| 2004 | 4.62 | 4.62 | 4.62 | 4.70 | 4.75 | 4.78 | 4.89 | 4.91 | 4.76 | 4.71 | 4.69 | 4.71 |
| 2005 | 4.72 | 4.73 | 4.73 | 4.81 | 4.87 | 4.90 | 4.98 | 4.93 | 4.76 | 4.70 | 4.64 | 4.63 |

3. Dans le cas présent, on a imposé les deux constantes de lissage.
Est-on sûr que ces constantes donnent un lissage optimal ?

Comment faudrait-il procéder pour déterminer les constantes donnant le meilleur ajustement ?

4. Selon une approche consistant à chercher la solution optimale, on obtient le graphique suivant.
Quel est l'intérêt d'obtenir une série lissée ?




5. On propose de recommencer la modélisation en intégrant cette fois-ci une composante saisonnière. Dans ce but, on va faire usage du lissage exponentiel de type Holt-Winters. Dans ce cadre, on estime la valeur de la variable à l'instant t + h par :

$$\hat{y}_t(h) = a_t + h * b_t + s_{t+h/\text{mod p}}$$

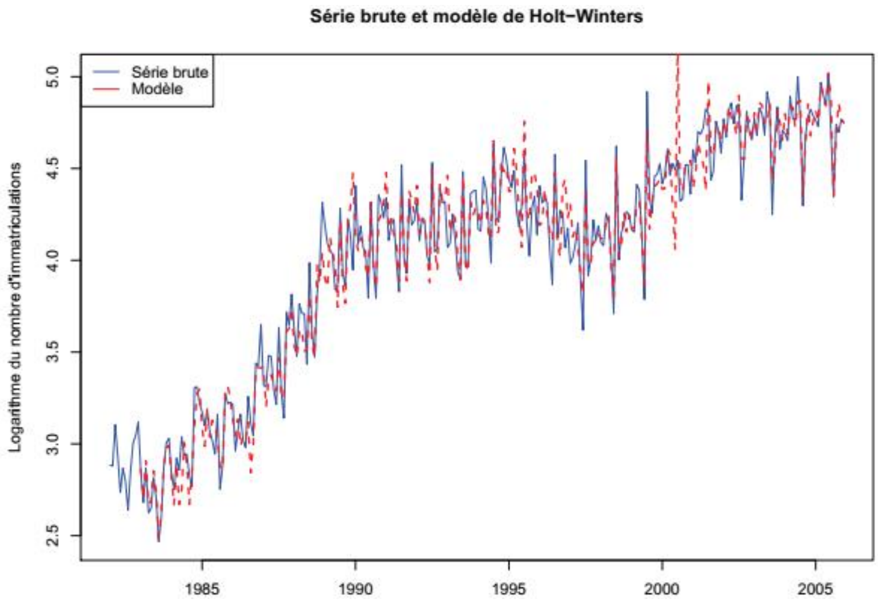
où les coefficients $a_t$, $b_t$ et $s_t$ sont calculés via le système suivant :

$$\begin{cases} a_t = \alpha(y_t - s_{t-p}) + (1-\alpha)\,(a_{t-1} + b_{t-1}) \\ b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1} \\ s_t = \gamma(y_t - a_t) + (1-\gamma)s_{t-p} \end{cases}$$

où α, β et γ sont trois paramètres (les constantes de lissage).

En réalisant cette nouvelle stratégie, on obtient le graphique suivant.

Pensez-vous que cette modélisation est meilleure ? Justifier votre réponse.



| | |
|---|---|
| a | 4.6952 |
| b | 0.0049 |
| s1 | 0.0598 |
| s2 | 0.0197 |
| s3 | 0.2395 |
| s4 | 0.1599 |
| s5 | 0.1005 |
| s6 | 0.2724 |
| s7 | 0.0232 |
| s8 | -0.3602 |
| s9 | 0.0104 |
| s10 | 0.0517 |
| s11 | 0.0781 |
| s12 | 0.0553 |

C - Phase de prévision

Une modélisation satisfaisante étant obtenue, il convient maintenant de passer à l'étape de prévision. Dans le cadre du lissage de Holt-Winters, les prévisions sont réalisées via la relation

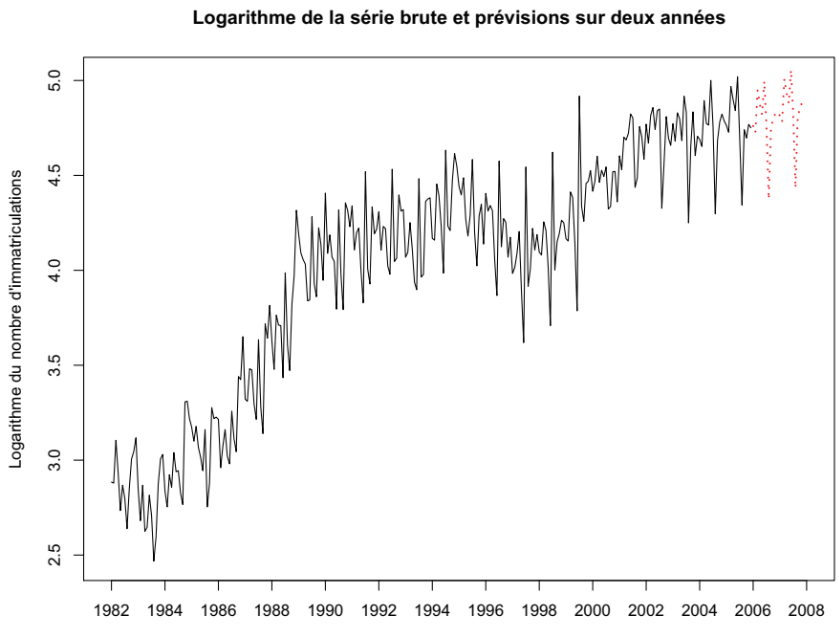
$$\hat{y}_t(h) = a_t + h \times b_t + s_{t+h/\text{mod p}}, \qquad \forall\ h \geq 1$$

A l'issue de la modélisation précédente, on a obtenu les coefficients suivants.

1. Compléter alors le tableau suivant :

| $\log \hat{y}_t$ | Janv. | Fév. | Mars | Avril | Mai | Juin | Juil. | Août | Sep. | Oct. | Nov. | Déc. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2006 | . | . | . | . | . | . | . | . | . | . | . | . |
| 2007 | . | . | . | . | . | . | . | . | . | . | . | . |

2. Selon cette approche, on obtient la représentation graphique suivante :




Vos résultats sont-ils en accord avec ce graphique ?

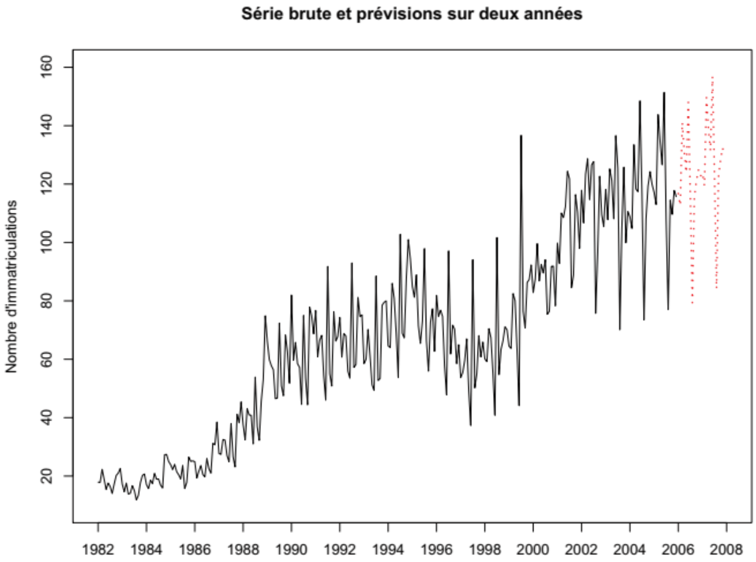
3. Quelle transformation doit-on effectuer sur ces prévisions afin qu'elles soient cohérentes avec la série initiale ? Justifier votre réponse.

4. Compléter alors le tableau suivant :

| $\hat{y}_t$ | Janv. | Fév. | Mars | Avril | Mai | Juin | Juil. | Août | Sep. | Oct. | Nov. | Déc. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2006 | . | . | . | . | . | . | . | . | . | . | . | . |
| 2007 | . | . | . | . | . | . | . | . | . | . | . | . |

5. On obtient finalement la représentation graphique suivant.

Ce graphique correspond-il à vos calculs ? Commenter succinctement ces prévisions.



# TD01- AJUSTEMENT LINÉAIRE, METHODE DES MOINDRES CARRES (MCO), COEFFICIENT DE DETERMINATION-CORRECTION

## 1. Exercice

B-A

## 2. Exercice

$$\begin{aligned}
\mathrm{R}^2 &= \frac{\sum_{i=1}^n \left(\hat\beta_1 + \hat\beta_2 x_i - \bar y\right)^2}{\sum_{i=1}^n (y_i - \bar y)^2} = \frac{\sum_{i=1}^n \left(\bar y - \hat\beta_2 \bar x + \hat\beta_2 x_i - \bar y\right)^2}{\sum_{i=1}^n (y_i - \bar y)^2} \\
&= \frac{\hat\beta_2^2 \sum_{i=1}^n (x_i - \bar x)^2}{\sum_{i=1}^n (y_i - \bar y)^2} = \frac{\left[\sum_{i=1}^n (x_i - \bar x)(y_i - \bar y)\right]^2 \sum_{i=1}^n (x_i - \bar x)^2}{\left[\sum_{i=1}^n (x_i - \bar x)^2\right]^2 \sum_{i=1}^n (y_i - \bar y)^2} \\
&= \frac{\left[\sum_{i=1}^n (x_i - \bar x)(y_i - \bar y)\right]^2}{\sum_{i=1}^n (x_i - \bar x)^2 \sum_{i=1}^n (y_i - \bar y)^2} = \rho^2(X, Y).
\end{aligned}$$

## 3. Exercice

**Exercice 1.3 (Poids des pères et des fils)**

1. La droite des moindres carrés du poids des fils en fonction du poids des pères s'écrit (cf. figure 1.10 à gauche) : $f = \hat\alpha_1 + \hat\alpha_2 p = 35.8 + 0.48p$.
2. La droite des moindres carrés du poids des fils en fonction du poids des pères s'écrit (cf. figure 1.10 à droite) : $p = \hat\beta_1 + \hat\beta_2 f = -3.38 + 1.03f$.
3. Le produit des pentes des deux droites est

$$\hat\alpha_2 \hat\beta_2 = \frac{\left(\sum (f_i - \bar f)(p_i - \bar p)\right)^2}{\left(\sum (f_i - \bar f)^2\right)\left(\sum_{i=1}^n (p_i - \bar p)^2\right)} = R^2,$$

où $R^2$ est le coefficient de détermination, carré du coefficient de corrélation linéaire.

## 4. Exercice 1.4

1. Les estimateurs de la droite des moindres carrés $y = \hat\beta_0 + \hat\beta_1 x$ sont respectivement :

$$\hat\beta_1 = \frac{\sum (x_i - \bar x)(y_i - \bar y)}{\sum (x_i - \bar x)^2} \approx 0.344$$

et

$$\hat\beta_0 = \bar y - \hat\beta_1 \bar x \approx 7.09$$

2. Une mesure de la qualité de l'ajustement des données au modèle est donnée par le coefficient de détermination $R^2$, dont on a vu qu'il correspond au carré du coefficient de corrélation linéaire empirique :

$$R^2 = \left( \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \right)^2 \approx 0.58.$$

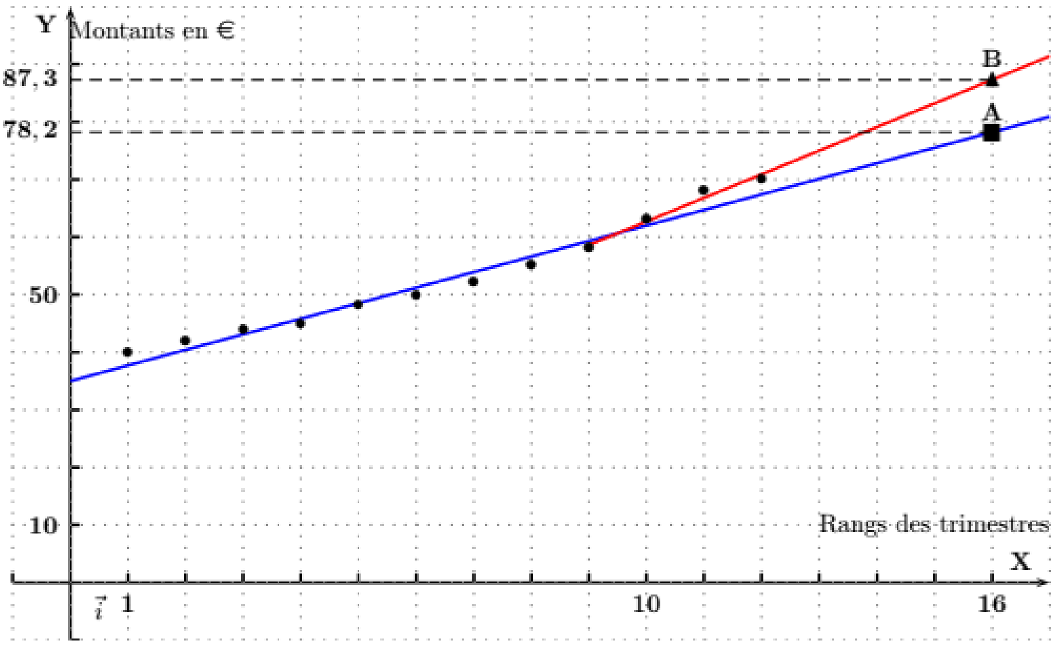
Le modèle de régression linéaire simple explique donc un peu plus de la moitié de la variance présente dans les données.

## 5. Exercice 1.5 (secteur hôtelier)

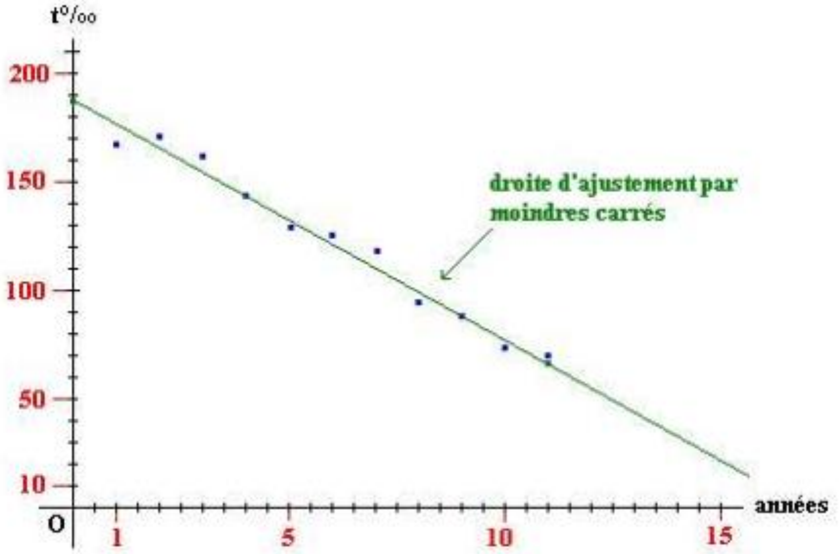
**1.** On obtient sur TI 82: $y = 2,7x + 35$ et $r = 0,98$. L'ajustement est justifié.

**2.** $x = 16$ donne la valeur estimée: $\hat{y} = 2,7 \times 16 + 35 = 78,2$ millions d'€. Point $A$.

**3.** On obtient sur TI 82 en gardant la même numérotation des trimestres: $y = 4,1x + 21,7$ et $r = 0,98$. L'ajustement est de bonne qualité. La prévision est alors: $\hat{y} = 4,1 \times 16 + 21,7 = 87,3$ millions d'€ . Point $B$.



## 6. **Exercice 1.6 (mortalité infantile)**




2°/ Le coefficient de corrélation linéaire est donné par la formule :

$$r = \frac{\text{cov}(X,Y)}{\sqrt{V(X) \times V(Y)}} \text{ avec } \text{cov}(X,Y) = \frac{1}{n}\sum x_i y_i - \overline{x} \times \overline{y}$$

Le tableau ci-dessous résume les différents paramètres statistiques (V = variance, σ = écart-type) des séries observées.
| r | = 0,989 est proche de 1 : forte présomption d'alignement. La droite de régression de y en x (ajustement linéaire) par la méthode des moindres carrés a sensiblement pour équation : y = -11x + 187.

| | | | |
|---|---|---|---|
| Point moyen : G(6,122) | $\Sigma xi = 66$ | $\Sigma yi = 1342$ | $\Sigma xi^2 = 506$ |
| $\Sigma yi^2 = 177062$ | $\Sigma xiyi = 6854$ | V(xi) = 10 | σ(xi) = 3,1623 |
| V(yi) = 1212,546 | σ(yi) = 34,8216 | cov(xi,yi) = -108,9091 | r = - 0,989 |

→ Rappelons que si (d) : y = ax + b est l'équation de la droite de régression de y en x, le coefficient directeur de cette droite est donné par la formule : a = cov(X,Y)/V(X); l'ordonnée à l'origine se calcule facilement sachant que (d) passe par le point moyen G(Σxi,Σyi).

3°/ Les classes ont une amplitude de 5 ans; si 1888, centre de la classe, correspond à l'abscisse n = 1, les années sont en progression arithmétique ($A_n$) de raison 5, de 1er terme 1888; on a donc : $A_n = 1888 + (n - 1) \times 5$, ou encore :

$$A_n - 1888 = 5(n - 1)$$

! On constate qu'une prédiction à long terme (x = 21 : 100 ans entre 1888 et 1988) n'a pas de sens : le taux serait négatif ! Toute extrapolation de résultats statistiques doit bien sûr tenir compte de changements significatifs dans une population donnée.

➔ Dans la première moitié du 20e siècle les progrès de la médecine, dus pour beaucoup à Pasteur (1822-1895), ont permis d'éradiquer un grand nombre de maladies du nourrisson.

De plus, en Europe en particulier, les naissances ont lieu désormais dans des maternités avec un suivi de l'enfant, ce qui n'était pas le cas jusque dans les années 1950. La mortalité infantile a beaucoup baissé depuis cette époque mais cette évolution est plus lente. En France, le taux était de 7,8‰ en 1988; 4,2 en 2002; 3,4 en 2008.

## 7. Exercice 1.7 (Consommation de graisse par an et par personne en Norvège)

1. Voir figure ci-dessous.

*2. En notant $X$ la variable consommation de graisse et $Y$ la variable taux de mortalité, on a :*

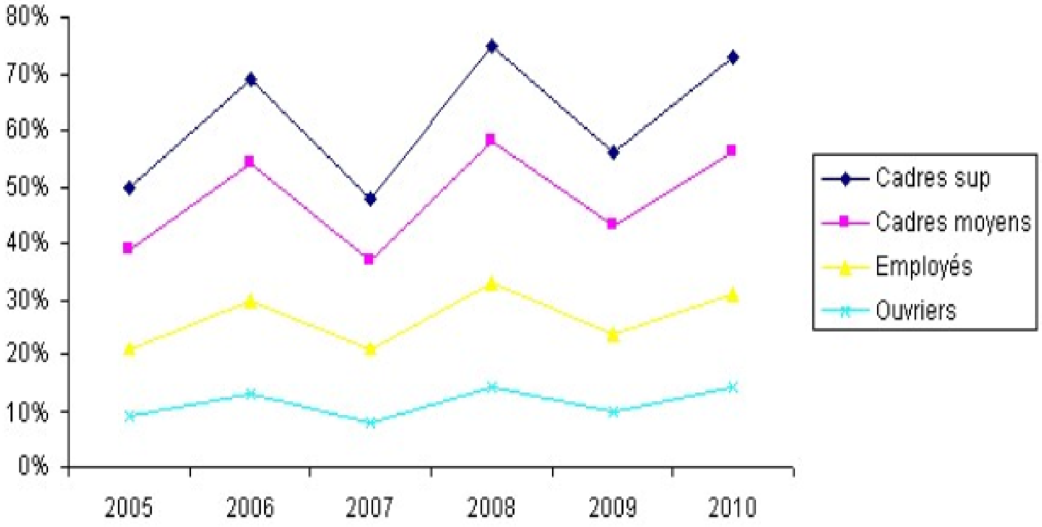



FIGURE 8.2 – *Graphique arithmétique des départs en vacances d'hiver.*

| $x_i$ | $y_i$ | $x_i^2$ | $y_i^2$ | $x_i \times y_i$ |
|---|---|---|---|---|
| 14,4 | 29,1 | 207,36 | 846,81 | 419,04 |
| 16 | 29,7 | 256 | 882,09 | 475,2 |
| 11,6 | 29,2 | 134,56 | 852,64 | 338,72 |
| 11 | 26 | 121 | 676 | 286 |
| 10 | 24 | 100 | 576 | 240 |
| 9,6 | 23,1 | 92,16 | 533,61 | 221,76 |
| 9,2 | 23 | 84,64 | 529 | 211,6 |
| 10,4 | 23,1 | 108,16 | 533,61 | 240,24 |
| 11,4 | 25,2 | 129,96 | 635,04 | 287,28 |
| 12,5 | 26,1 | 156,25 | 681,21 | 326,25 |
| 116,1 | 258,5 | 1390,09 | 6746,01 | 3046,09 |

*On en déduit* $\overline{X} = 11,61$, $\sigma_X^2 = 4,22$, $\sigma_X = 2,05$, $\overline{Y} = 25,85$, $\sigma_Y^2 = 6,38$ *et* $\sigma_Y = 2,53$.

*3. Calcul de la covariance :*

$$\mathrm{Cov}(X,Y) = \frac{3046,09}{10} - (11,61 \times 25,85) = 4,4905.$$

*4. Calcul du coefficient de corrélation linéaire :*

$$r(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{4,4905}{2,05 \times 2,53} = 0,87.$$

*5. La forme du nuage de points "étirée" (les points sont proches d'une droite) et la valeur du coefficient de corrélation linéaire proche de 1 rendent plausible l'existence d'une liaison linéaire forte entre* $X$ *et* $Y$.

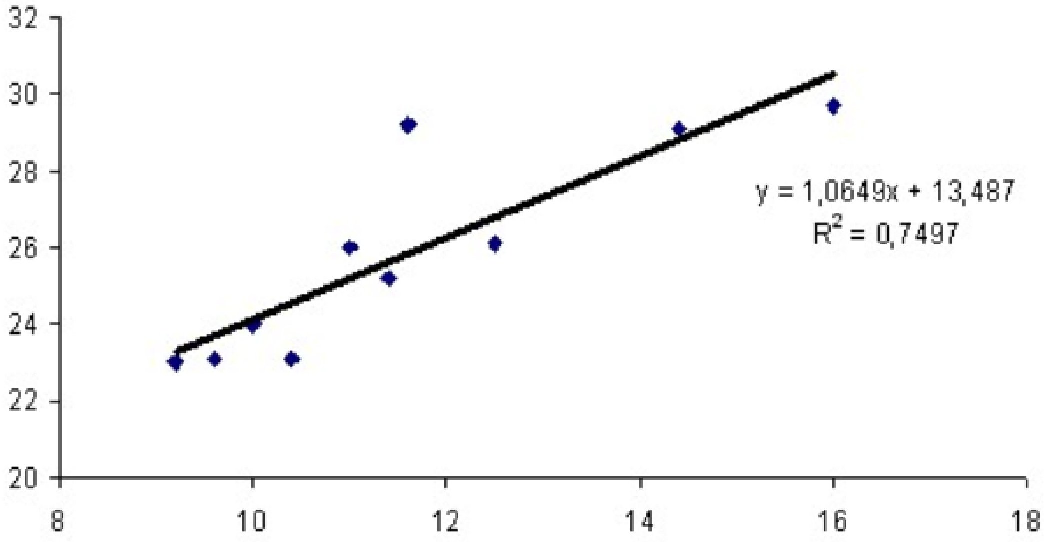



FIGURE 8.3 – *Nuage de points et droite de régression pour les variables graisse/taux de mortalité.*

*6. La pente a de la droite de régression de $Y$ sur $X$ est donnée par :*

$$a = \frac{\mathrm{Cov}(X,Y)}{\sigma_X^2} = 1,06,$$

*et la valeur à l'origine $b$ est donnée par :*

$$b = \overline{Y} - a\overline{X} = 13,49.$$

*L'équation de la droite de régression de $Y$ sur $X$ est donc :*

$$y = 1,06x + 13,49.$$

*7. Une estimation du taux de mortalité par athérosclérose en 1948 est :*

$$\widehat{y}(1948) = 1,06 \times 13 + 13,49 = 27,27.$$

## 8. **Exercice (moyenne mobile, modele additif)**

L'entreprise X communique le montant de son chiffre d'affaires (en milliers d'euro) pour les années de 1989 à 1992

1. Représenter la série chronologique. On considère que janvier 1989 corresponds à 1 et que $y_1$=1230
2. Calculer le coefficient de corrélation et pour cette série et l'équation de la droite obtenu par la méthode des moindres carres
3. Choisir un modèle pour l'influence saisonnière (additif ou multiplicatif). Justifier
4. Calculer les coefficients saisonniers sur une période de 1 an.
5. Calculer la moyenne des coefficients saisonniers
6. Calculer la série corrigée des variations saisonnières

| mois/année | 1989 | 1990 | 1991 | 1992 |
|---|---|---|---|---|
| janvier | 1230 | 1590 | 1750 | 1840 |
| février | 1280 | 1640 | 1650 | 1790 |
| mars | 1400 | 1800 | 1800 | 2370 |
| avril | 1600 | 1990 | 1900 | 2360 |
| mai | 1450 | 1870 | 1950 | 2280 |
| juin | 1390 | 1910 | 1910 | 2510 |
| juillet | 1280 | 1670 | 1980 | 2320 |
| aout | 930 | 1260 | 1410 | 1870 |
| septembre | 1080 | 1430 | 1520 | 1820 |
| octobre | 1400 | 1780 | 1920 | 2210 |
| novembre | 1500 | 1750 | 1900 | 2440 |
| décembre | 1550 | 1670 | 1780 | 2270 |

7. résumer les résultats précédents dans le tableau suivant et prédire le chiffre d'affaire en 1993.

| m=a | 1989 | 1990 | 1991 | 1992 | αj | CVS | 1989(c) | 1990(c) | 1991(c) | 1992(c) | 1993(l) | 1993 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| jan. | 1230 | 1590 | 1750 | 1840 | 1602,5 | -147,5 | 1377,5 | 1737,5 | 1897,5 | 1987,5 | 2250,5 | 2103 |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| Dec. | 1550 | 1670 | 1780 | 2270 | | | | | | | | |

8. Representer · la Série · corrigee · des variations saisonnieres ainsi que les previsions · effectuees · sur un autre graphique graphique

# 9. **Exercice 1.9**

*2. Après calculs, on obtient :* $r(X,Y) = -0,55$.

*3. La valeur (moyenne) du coefficient de corrélation linéaire et surtout la forme du nuage de points (en forme de "T") suggère qu'il n'y a pas de liaison linéaire entre* $X$ *et* $Y$ *mais qu'il y a éventuellement deux populations pour lesquelles la liaison linéaire entre* $X$ *et* $Y$ *est forte.*

*4. Considérons un premier sous-échantillon constitué par les points 1, 2, 4 et 5. On obtient le tableau :*

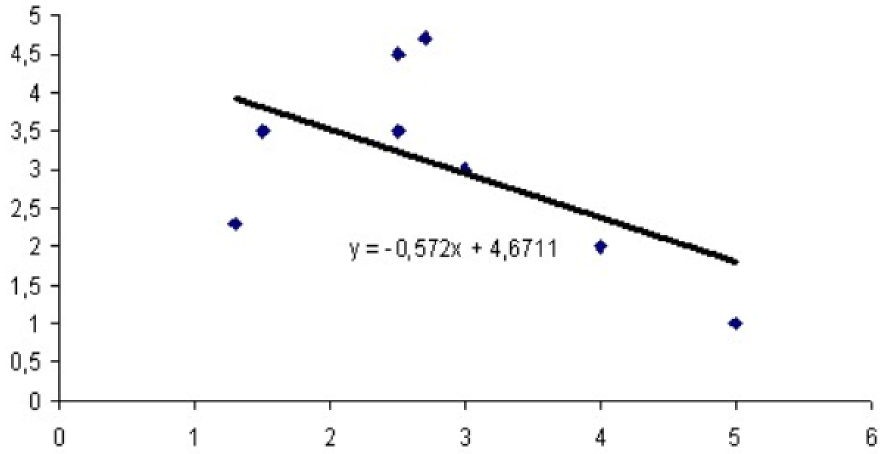



FIGURE 8.4 – *Nuage de points et droite de régression pour les variables $X$ et $Y$.*

| $X$ | 1,3 | 1,5 | 2,5 | 2,7 |
|---|---|---|---|---|
| $Y$ | 2,3 | 3,5 | 4,5 | 4,7 |

*Le coefficient de corrélation linéaire pour ce sous-échantillon est :*

$$r_1(X,Y) \ = \ 0,94.$$

*Considérons un second sous-échantillon formé par les points 3, 6, 7 et 8. On obtient le tableau :*

| $X$ | 2,5 | 3 | 4 | 5 |
|---|---|---|---|---|
| $Y$ | 3,5 | 3 | 2 | 1 |

*Le coefficient de corrélation linéaire pour ce sous-échantillon est :*

$$r_2(X,Y) \ = \ -1.$$

*On constate donc que les deux coefficients de corrélation linéaire sur chaque sous-échantillon sont proches de 1 en valeur absolue (égal à -1 pour le second), ce qui confirme l'existence d'une liaison linéaire forte sur chaque sous-échantillon.*

# TD02- MOYENNES MOBILES-CORRECTION

## 1. Exercice (Vrai-Faux & QCM)

1. 1 - VRAI, 2 - VRAI, 3- VRAI, 4 - FAUX, 5-VRAI.

2. a) c) et d)

Il y a (T-p + 1) moyennes mobiles centrées de longueur impaire p et (T-p) moyennes mobiles centrées de longueur paire p

3. a) et b)

La moyenne mobile centrée de longueur p rend constante les séries périodiques de période p et de période sous-multiple de p.

La moyenne mobile centrée de longueur 2p éliminent la composante saisonnière de période p puisque la somme des coefficients saisonniers sur une période est nulle.

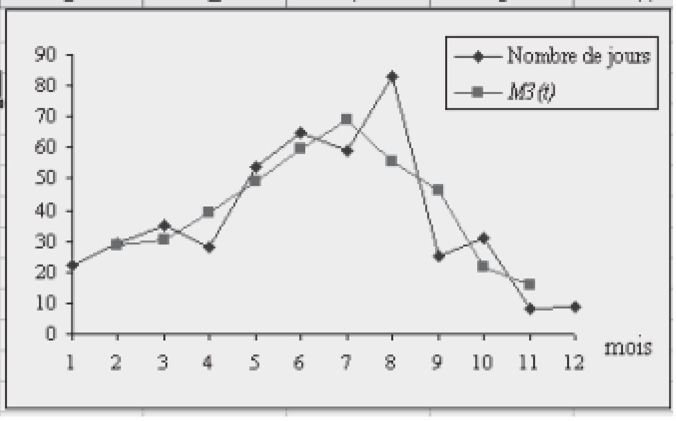
La somme de p termes successifs divisée par p donne une évaluation de la tendance pour la date correspondant à celle du terme du milieu des p termes.

On peut calculer (T-p) moyennes mobiles centrées de longueur p si p est pair, et (T-p + 1) moyennes mobiles centrées de longueur p si p est impair, on a donc toujours au moins (T-p) moyennes mobiles centrées.

## 2. Exercice (Air ATMO)

C3 $f_x$ =(B2+B3+B4)/3

| | A | B | C |
|---|---|---|---|
| 1 | Mois | Nombre de jours | $M_3(t)$ |
| 2 | 1 | 22 | |
| 3 | 2 | 29 | 28,67 |
| 4 | 3 | 35 | 30,67 |
| 5 | 4 | 28 | 39,00 |
| 6 | 5 | 54 | 49,00 |
| 7 | 6 | 65 | 59,33 |
| 8 | 7 | 59 | 69,00 |
| 9 | 8 | 83 | 55,67 |
| 10 | 9 | 25 | 46,33 |
| 11 | 10 | 31 | 21,33 |
| 12 | 11 | 8 | 16,00 |
| 13 | 12 | 9 | |




La moyenne mobile lisse la série chronologique, et permet d'évaluer la tendance.

## 3. **Exercice**

G5 $f_x$ =10-2*A5+E5

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | $t$ | $x_t$ | $M_2(t)$ | $M_3(t)$ | $M_4(t)$ | $M_5(t)$ | $M_4(z_t)$ | |
| 2 | 1 | 3 | | | | | | |
| 3 | 2 | -1 | 1,5 | 2,33 | | | | |
| 4 | 3 | 5 | 2,5 | 1,67 | 2 | 2,2 | 6 | |
| 5 | 4 | 1 | 2,5 | 3,00 | 2 | 1,4 | 4 | |
| 6 | 5 | 3 | 1,5 | 1,00 | 2 | 2,6 | 2 | |
| 7 | 6 | -1 | 1,5 | 2,33 | 2 | 1,8 | 0 | |
| 8 | 7 | 5 | 2,5 | 1,67 | 2 | 2,2 | -2 | |
| 9 | 8 | 1 | 2,5 | 3,00 | 2 | 1,4 | -4 | |
| 10 | 9 | 3 | 1,5 | 1,00 | 2 | 2,6 | -6 | |
| 11 | 10 | -1 | 1,5 | 2,33 | 2 | 1,8 | -8 | |
| 12 | 11 | 5 | 2,5 | 1,67 | | | | |
| 13 | 12 | 1 | | | | | | |
| 14 | | | | | | | | |
| 15 | | | | | | | | |
| 16 | | | | | | | | |

=SOMME(B3:B7)/5
=(B9+B10+B11)/3
=(B5/2+B6+B7/2)/2
=(B2/2+SOMME(B3:B5)+B6/2)/4

La série $x_t$ est périodique de période 4.
Toutes les suites de moyennes mobiles sont aussi périodiques de période 4.
La suite des moyennes mobiles de longueur 4 est constituée de termes constants égaux à la moyenne des termes sur une période.

**2.** La moyenne mobile tranforme une série alignée en elle-même, donc la série $y_t = 10 - 2t$ est transformée en elle-même, et la suite des moyennes mobiles de longueur 4 de la série $z_t$ est égale à : $10 - 2t + 2 = 12 - 2t$ ($t = 3$ à $10$).

## 4. Exercice (Moyennes mobiles)

Question 1 :
Calculez la moyenne mobile centrée d'ordre 4 pour t = 3

- 65,525
- 63,125

$$mm_{4,3} = 1/4\ (52/2 + 36 + 69 + 89 + 65/2) = 63{,}125$$

Question 2
Calculez la moyenne mobile centrée d'ordre 4 pour t = 4

- 65,875
- 66,775

$$mm_{4,4} = 1/4\ (36/2 + 69 + 89 + 65 + 45/2) = 65{,}875$$

Question 3
Calculez la moyenne mobile centrée d'ordre 4 pour t = 5

- 68,1975
- 69,125

$$mm_{4,5} = 1/4\ (69/2 + 89 + 65 + 45 + 86/2) = 69{,}125$$

Question 4
Calculez l'ensemble des moyennes mobiles centrées d'ordre 4
Représentation dans un tableau, par année et par trimestre
Les réponses se trouvent dans le tableau ci-dessous

| | Série initiale | | | | Moyenne mobile d'ordre 4 | | | |
|---|---|---|---|---|---|---|---|---|
| | tr1 | tr2 | tr3 | tr4 | tr1 | tr2 | tr3 | tr4 |
| année 2010 | 52 | 36 | 69 | 89 | | | 63,125 | 65,875 |
| année 2011 | 65 | 45 | 86 | 111 | 69,125 | 74 | 78,75 | 82,125 |
| année 2012 | 81 | 56 | 108 | 139 | 86,25 | 92,5 | 98,625 | 103 |
| année 2013 | 102 | 70 | 135 | 174 | 108,125 | 115,875 | | |

# TD03- MODELES MULTIPLICATIF ET ADDITIF-CORRECTION

## 1. Exercice (nombre de mariages, CVS et modèle additif)

1. Les représentations graphiques font apparaître des variations saisonnières.

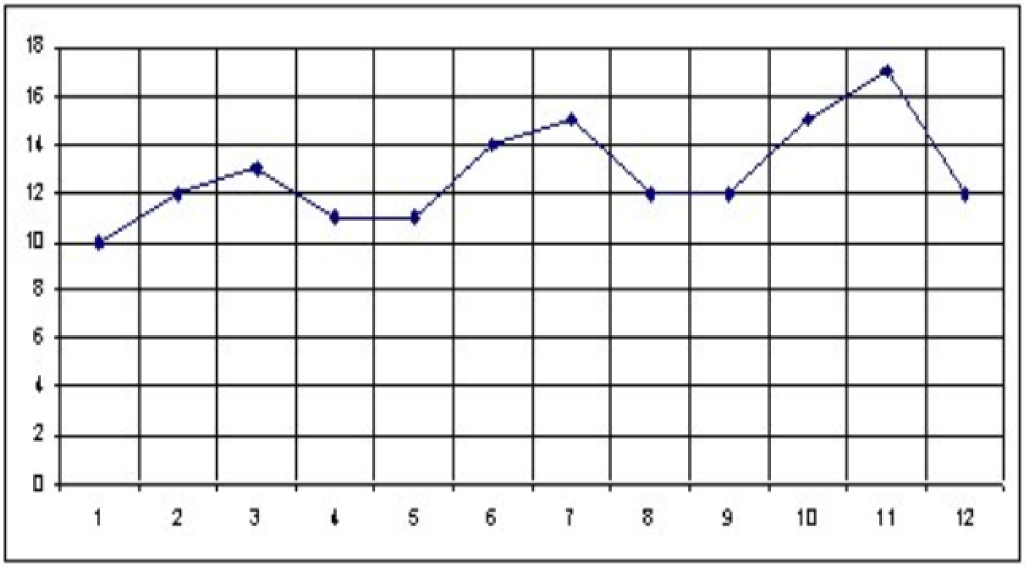



FIGURE 8.19 – Représentation graphique des mariages avec périodes successives

*2.3.4. Les relevés étant trimestriels, on considère les moyennes mobiles d'ordre 4 :*

$$z(t) = \frac{y(t-2)/2 + y(t-1) + y(t) + y(t+1) + y(t+2)}{4}.$$

*Ensuite on réalise la régression de z sur* $t_z$*. On a d'après le cours*

$$\bar{t}_z = \frac{N+1}{2} = 6,5 \qquad \text{Var}(t_z) = \frac{N_z^2 - 1}{12} = 5,25$$

*où* $N = 12$ *et* $N_z = 8$*. On a aussi* $\bar{z} = \dfrac{103,25}{8} \approx 12,91$ *puis*

$$\text{Cov}(z, t_z) = \frac{685,25}{8} - \frac{103,25}{8} 6,5 \approx 1,77$$

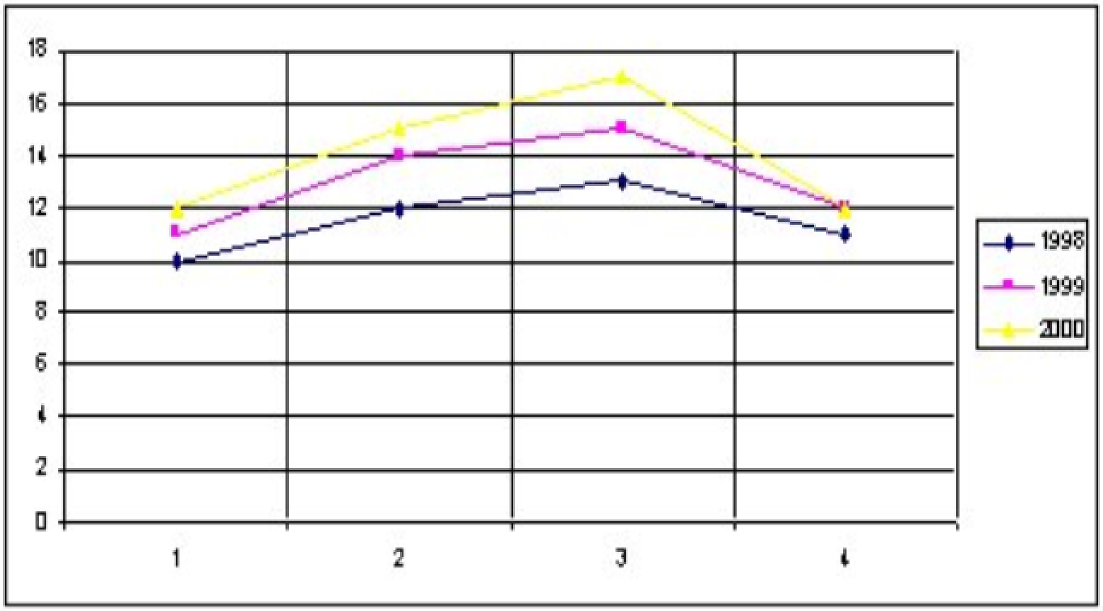



FIGURE 8.20 – Représentation graphique des mariages avec périodes superposées

*et* $a = \dfrac{\mathrm{Cov}(z, t_z)}{\mathrm{Var}(t_z)} \approx 0,34$ *et* $b = \overline{z} - a\overline{t}_z \approx 12,72.$

*On peut présenter l'ensemble des calculs dans le tableau suivant :*

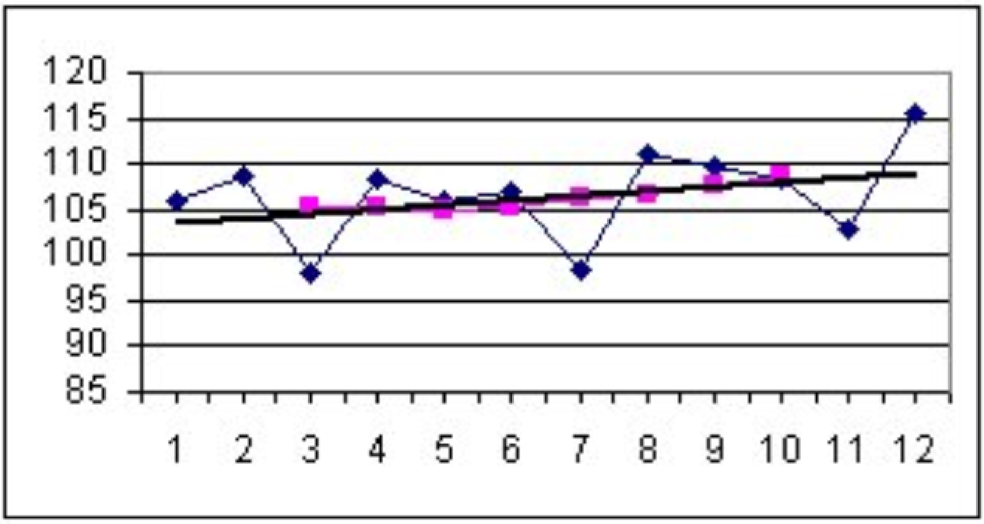
| $t$ | $y(t)$ | $z(t)$ | $t.z(t)$ | $x(t) = 0,336t + 10,724$ | $y(t) - x(t)$ | $s'(t)$ | $s(t)$ |
|---|---|---|---|---|---|---|---|
| 1 | 10 | | | 11,06 | -1,06 | -1,404 | -1,329 |
| 2 | 12 | | | 11,396 | 0,604 | 0,927 | 1,002 |
| 3 | 13 | 11,63 | 34,88 | 11,732 | 1,268 | 1,924 | 1,999 |
| 4 | 11 | 12,00 | 48,00 | 12,068 | -1,068 | -1,745 | -1,670 |
| 5 | 11 | 12,50 | 62,50 | 12,404 | -1,404 | | |
| 6 | 14 | 12,88 | 77,25 | 12,74 | 1,26 | | |
| 7 | 15 | 13,13 | 91,88 | 13,076 | 1,924 | | |
| 8 | 12 | 13,38 | 107,00 | 13,412 | -1,412 | | |
| 9 | 12 | 13,75 | 123,75 | 13,748 | -1,748 | | |
| 10 | 15 | 14,00 | 140,00 | 14,084 | 0,916 | | |
| 11 | 17 | | | 14,42 | 2,58 | | |
| 12 | 12 | | | 14,756 | -2,756 | | |
| | | 103,25 | 685,25 | - | - | $S0 = -0,3$ | |

5.

| $t$ | $x(t) = 0,336t + 10,724$ | $s(t)$ | $x(t) + s(t)$ |
|---|---|---|---|
| 17 | 16,436 | -1,329 | 15,107 |
| 18 | 16,772 | 1,002 | 17,774 |
| 19 | 17,108 | 1,999 | 19,107 |
| 20 | 17,444 | -1,670 | 15,774 |

## 2. Exercice 1.2

| $t$ | $y(t)$ | $z(t)$ | $t \times z(t)$ | $x(t) = 0,49t + 103,04$ | $y(t) - x(t)$ | $s'(t)$ | $s(t)$ |
|---|---|---|---|---|---|---|---|
| 1 | 106 | | | 103,53 | 2,47 | 1,743 | 1,243 |
| 2 | 108,8 | | | 104,02 | 4,78 | 2,153 | 1,653 |
| 3 | 97,9 | 105,31 | 315,94 | 104,51 | -6,61 | -6,670 | -7,170 |
| 4 | 108,5 | 105,11 | 420,45 | 105 | 3,5 | 4,773 | 4,273 |
| 5 | 106,1 | 104,98 | 524,88 | 105,49 | 0,61 | | |
| 6 | 107,1 | 105,39 | 632,33 | 105,98 | 1,12 | | |
| 7 | 98,5 | 106,16 | 743,14 | 106,47 | -7,97 | | |
| 8 | 111,2 | 106,78 | 854,20 | 106,96 | 4,24 | | |
| 9 | 109,6 | 107,51 | 967,61 | 107,45 | 2,15 | | |
| 10 | 108,5 | 108,61 | 1086,13 | 107,94 | 0,56 | | |
| 11 | 103 | | | 108,43 | -5,43 | | |
| 12 | 115,5 | | | 108,92 | 6,58 | | |
| | | 849,85 | 5544,66 | - | - | $S_0 = 2$ | |




*3ième trimestre 1985 signifie* $t = 15$ *d'où* $\hat{y}(15) = 0,49 \times 15 - 7,170 = 103,22$.

## 3. Exercice 1.3 (moyennes mobiles)

*1. On veut calculer des moyennes mobiles d'ordre 12. La première que l'on peut calculer correspond au mois de juillet 1974* ($t = 7$) *:*

$$z(7) = \frac{\frac{127,4}{2} + 129,1 + \ldots + 144,3 + \frac{145,9}{2}}{12} = 137,046.$$

*La dernière que l'on peut calculer correspond au mois de juin 1977* ($t = 42$) *:*

$$z(42) = \frac{\frac{173,8}{2} + 174,3 + \ldots + 188,9 + \frac{189,4}{2}}{12} = 182,042.$$

*2. La courbe est déjà lisse, il est inutile de calculer la série des moyennes mobiles : les valeurs seraient très proches des valeurs observées.*

*3. Il n'y a pas de variations saisonnières.*

*4. Pour faire des prévisions, on pourrait étudier simplement la corrélation linéaire.*

## 4. Exercice 1.5 ((Disneyland-Paris & Moyenne mobile et CVS)

| $t$ | $y(t)$ | $z(t)$ | $x(t) = 0,26t + 1,15$ | $y(t) - x(t)$ | $s'(t)$ | $s(t) = s'(t) + 0,6$ |
|---|---|---|---|---|---|---|
| 1 | 1 | | 1,41 | -0,41 | -0,36 | 0,24 |
| 2 | 0 | | 1,67 | -1,67 | -0,37 | 0,23 |
| 3 | 5 | 1,6 | 1,93 | 3,07 | 4,37 | 4,97 |
| 4 | 2 | 1,8 | 2,19 | -0,19 | -1,39 | -0,79 |
| 5 | 0 | 2,4 | 2,45 | -2,45 | -2,65 | -2,05 |
| 6 | 2 | 2,8 | 2,71 | -0,71 | | |
| 7 | 3 | 3,2 | 2,97 | 0,03 | | |
| 8 | 7 | 3,4 | 3,23 | 3,77 | | |
| 9 | 4 | 3,8 | 3,49 | 0,51 | | |
| 10 | 1 | 4 | 3,75 | -2,75 | | |
| 11 | 4 | 4,6 | 4,01 | -0,01 | | |
| 12 | 4 | 4,2 | 4,27 | -0,27 | | |
| 13 | 10 | 4,4 | 4,53 | 5,47 | | |
| 14 | 2 | 4,6 | 4,79 | -2,79 | | |
| 15 | 2 | 5 | 5,05 | -3,05 | | |
| 16 | 5 | 5,2 | 5,31 | -0,31 | | |
| 17 | 6 | 5,4 | 5,57 | 0,43 | | |
| 18 | 11 | 5,8 | 5,83 | 5,17 | | |
| 19 | 3 | | 6,09 | -3,09 | | |
| 20 | 4 | | 6,35 | -2,35 | | |
| | | | | | $S_0 = -3$ | |

*1. On calcule ici des moyennes mobiles d'ordre 4.*

*2. On est obligé de calculer tous les coefficients $s'(t)$ pour faire la somme puis calculer les coefficients saisonniers $s(t)$.*

*3.*

| $t$ | $x(t) = 0,26t + 1,15$ | $s(t) = s'(t) + 0,6$ | $x(t) + s(t)$ |
|---|---|---|---|
| 21 | 6,61 | 0,24 | 6,85 |
| 22 | 6,87 | 0,23 | 7,1 |
| 23 | 7,13 | 4,97 | 12,1 |

| $t$ | $y(t)$ | $z(t)$ | $t.z(t)$ | $x(t) = 0,70t + 13,45$ | $y(t) - x(t)$ | $s'(t)$ | $s(t)$ |
|---|---|---|---|---|---|---|---|
| 1 | 8,2 | | | 14,15 | -5,95 | -7,217 | -7,359 |
| 2 | 12,3 | | | 14,85 | -2,55 | -2,950 | -3,092 |
| 3 | 32,7 | 15,6 | 46,8 | 15,55 | 17,15 | 18,750 | 18,608 |
| 4 | 8,3 | 16,1 | 64,4 | 16,25 | -7,95 | -8,017 | -8,159 |
| 5 | 10 | 16,95 | 84,75 | 16,95 | -6,95 | | |
| 6 | 14,5 | 17,925 | 107,55 | 17,65 | -3,15 | | |
| 7 | 37,3 | 18,45 | 129,15 | 18,35 | 18,95 | | |
| 8 | 11,5 | 18,925 | 151,4 | 19,05 | -7,55 | | |
| 9 | 11 | 19,775 | 177,975 | 19,75 | -8,75 | | |
| 10 | 17,3 | 20,5 | 205 | 20,45 | -3,15 | | |
| 11 | 41,3 | | | 21,15 | 20,15 | | |
| 12 | 13,3 | | | 21,85 | -8,55 | | |
| | | 144,23 | 6967,03 | - | - | $S0 = 0,57$ | |

*4. Prévisions pour le 3ième trimestre 2000 :*

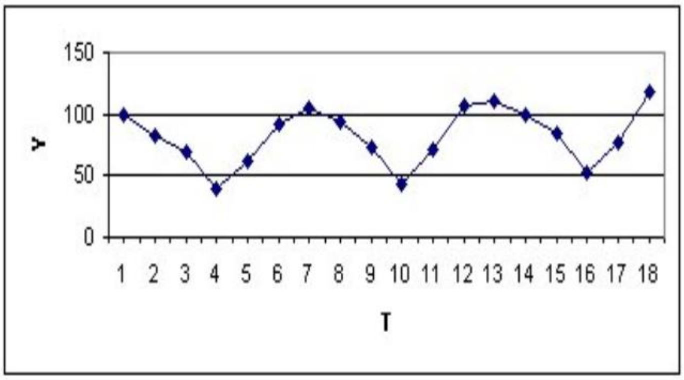
$$\hat{y}(23) = x(23) + s(3) = 0,7 \times 23 + 13,45 + 18,608 = 48,158.$$

*On prévoit donc avec ce modèle* $\boxed{\textit{48158 visiteurs}}$ *pour l'été 2000.*

## 5. Exercice 1.6 (billets vendus & moyennes mobiles et modèle additif)

*1. Notons $T$ la variable temps numérotée dans l'ordre chronologique.*

*Notons $Y$ la variable "nombre de billets vendus (en milliers)".*




*2. Il faut calculer des moyennes mobiles d'ordre 6.*

| $t$ | $y(t)$ | $z(t)$ | $t.z(t)$ | $x(t) = 1,34t + 68,93$ | $y(t) - x(t)$ | $s'(t)$ | $s(t) = s'(t) - 0,51$ |
|---|---|---|---|---|---|---|---|
| 1 | 100 | | | 70,27 | 29,73 | 27,02 | 26,51 |
| 2 | 82 | | | 71,61 | 10,39 | 12,02 | 11,51 |
| 3 | 70 | | | 72,95 | -2,95 | -5,32 | -5,83 |
| 4 | 40 | 74,58 | 298,33 | 74,29 | -34,29 | -37,33 | -37,84 |
| 5 | 62 | 76,00 | 380,00 | 75,63 | -13,63 | -13,34 | -13,85 |
| 6 | 91 | 77,25 | 463,50 | 76,97 | 14,03 | 19,99 | 19,48 |
| 7 | 105 | 77,75 | 544,25 | 78,31 | 26,69 | | |
| 8 | 94 | 78,83 | 630,67 | 79,65 | 14,35 | | |
| 9 | 73 | 80,92 | 728,25 | 80,99 | -7,99 | | |
| 10 | 43 | 82,67 | 826,67 | 82,33 | -39,33 | | |
| 11 | 72 | 83,58 | 919,42 | 83,67 | -11,67 | | |
| 12 | 106 | 84,92 | 1019,00 | 85,01 | 20,99 | | |
| 13 | 111 | 86,58 | 1125,58 | 86,35 | 24,65 | | |
| 14 | 99 | 87,75 | 1228,50 | 87,69 | 11,31 | | |
| 15 | 84 | 89,17 | 1337,50 | 89,03 | -5,03 | | |
| 16 | 52 | | | 90,37 | -38,37 | | |
| 17 | 77 | | | 91,71 | -14,71 | | |
| 18 | 118 | | | 93,05 | 24,95 | | |
| | | 980 | 9501,67 | - | - | $S0 = 3,04$ | |

*3. Pour tracer la droite de tendance, on peut placer le point $B(0;68,93)$ et le point moyen :*

$$\overline{T} = \quad = 9,5 \quad \overline{Z} = \frac{980}{12} = 81,67$$

*i.e. le point $M(9,5;81,67)$ appartient à la droite.*

*4. Prévisions pour Juillet-Août 2000 :*

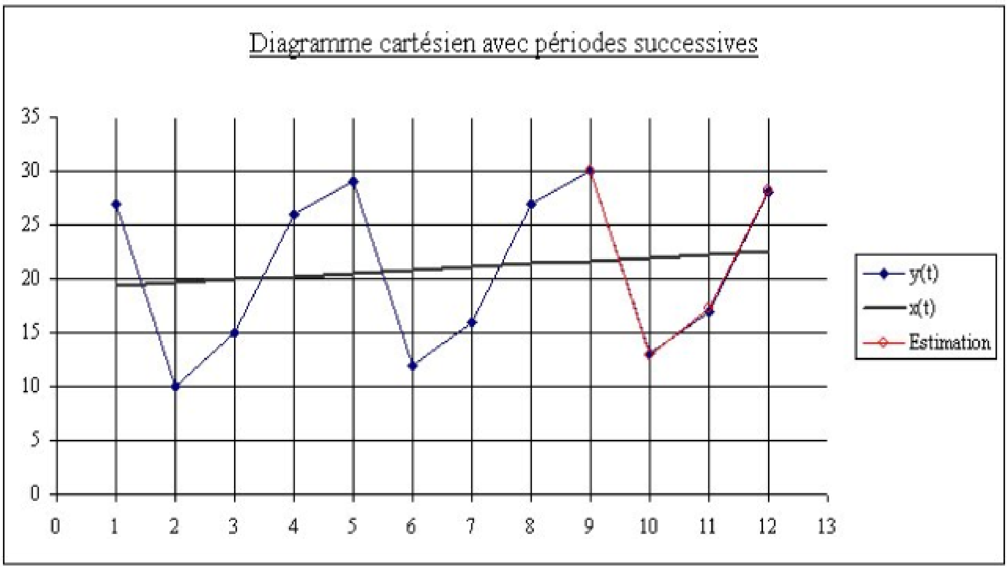
$$\hat{y}(22) = x(22) + s(4) = 1,34 \times 22 + 68,93 - 37,84 = 60,57,$$

*soit* $\boxed{\textit{60570 billets vendus}}$.

6. **Exercice (magasin d'insecticides, moyennes mobiles et modèle additif)**

*1. On commence par faire le diagramme cartésien :*

*2. 3. et 4. On fait ensuite le tableau :*




| $t$ | $y(t)$ | $z(t)$ | $tz(t)$ | $x(t)$ | $\delta(t)$ | $s'(t)$ | $s(t)$ | $estimation$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 27 | | | 19,35 | 7,65 | 8,16066667 | 8,26683333 | 27,61683 |
| 2 | 10 | | | 19,639 | -9,639 | -9,12833333 | -9,02216667 | 10,61683 |
| 3 | 15 | 19,75 | 59,25 | 19,928 | -4,928 | -5,084 | -4,97783333 | 14,95017 |
| 4 | 26 | 20,25 | 81 | 20,217 | 5,783 | 5,627 | 5,73316667 | 25,95017 |
| 5 | 29 | 20,625 | 103,125 | 20,506 | 8,494 | | 8,26683333 | 28,77283 |
| 6 | 12 | 20,875 | 125,25 | 20,795 | -8,795 | | -9,02216667 | 11,77283 |
| 7 | 16 | 21,125 | 147,875 | 21,084 | -5,084 | | -4,97783333 | 16,10617 |
| 8 | 27 | 21,375 | 171 | 21,373 | 5,627 | | 5,73316667 | 27,10617 |
| 9 | 30 | 21,625 | 194,625 | 21,662 | 8,338 | | 8,26683333 | 29,92883 |
| 10 | 13 | 21,875 | 218,75 | 21,951 | -8,951 | | -9,02216667 | 12,92883 |
| 11 | 17 | | | 22,24 | -5,24 | | -4,97783333 | 17,26217 |
| 12 | 28 | | | 22,529 | 5,471 | | 5,73316667 | 28,26217 |

*D'où l'on déduit :*

$$\overline{T'} = 6,5 \ ; \qquad \mathrm{Var}(T') = 0,25 \ ; \qquad \overline{Z} = 20,9375$$

$$\mathrm{Cov}(T', Z) = 1,515625 \ ; \qquad S_0/4 = -0,10616667$$

$$a = 0,28869048 \ ; \qquad b = 19,0610119.$$

*6. Prévision pour le 3ième trimestre 2001 :* $\boxed{18,418}$ *milliers de francs (soit 18418 francs).*

## 7. **Exercice (construction, moyennes mobiles et modèle additif)**

*1. Les variations observées ne sont pas des variations saisonnières.*

*2.* $z(8) = 1077,375$*,* $z(9) = 1080,875$ *et* $z(10) = 1090,25$*. Les moyennes mobiles lissent la courbe et font apparaître une tendance linéaire croissante. Le modèle additif est inadapté.*

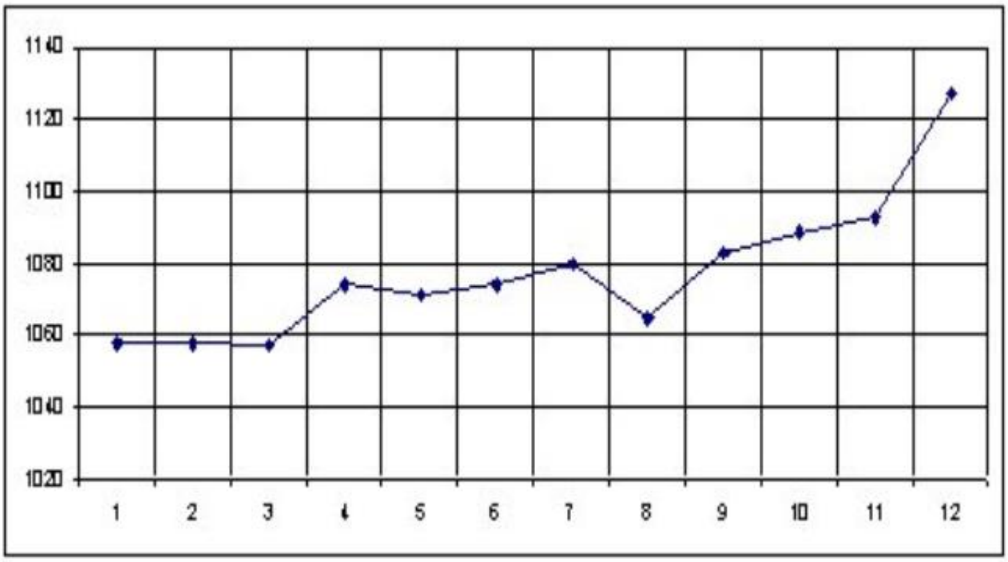



FIGURE 8.21 – Représentation graphique de la construction avec périodes successives

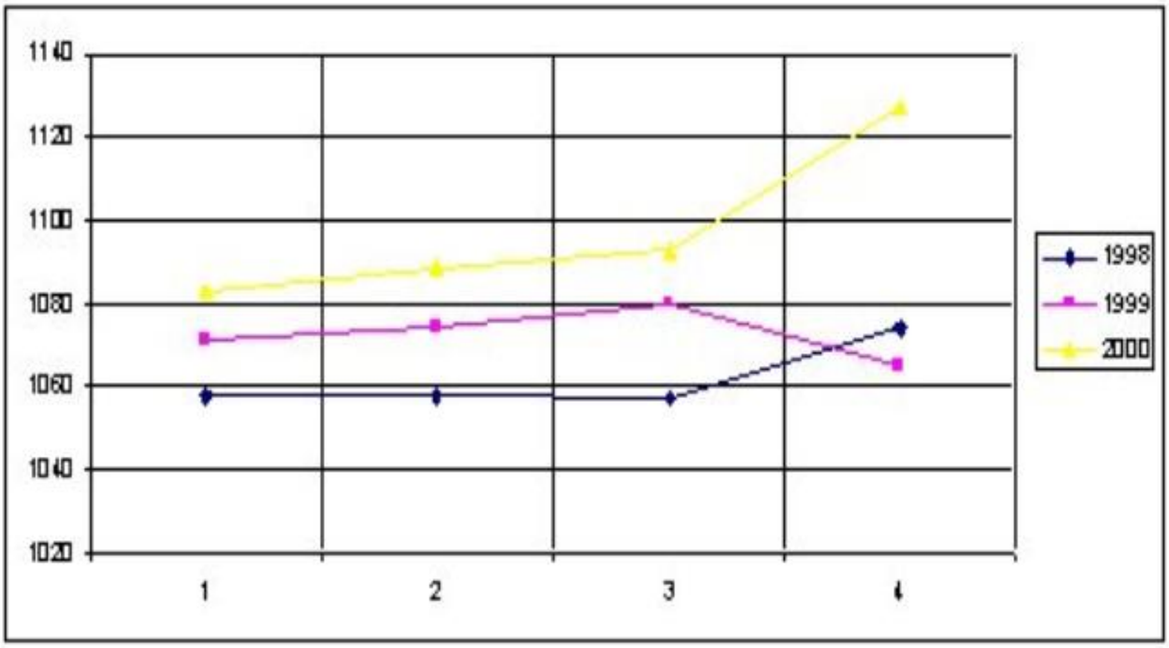



FIGURE 8.22 – Représentation graphique de la construction avec périodes superposées

## 8. Exercice 8

a) Représentation Soit $M$ le caractère donnant le nombre de mois écoulés depuis décembre 1988 et $X$ donnant le chiffre d'affaires pour le mois correspondant. Ayant des données sur 4 ans, on a à disposition une série $(m_i, x_i)_{i=1,\dots,N}$, où $N = 48, m_1 = 1, m_2 = 2, \cdots, m_{48} = 48$ et $x_1 = 1230, x_2 = 1280, \dots, x_{12} = 1550, x_{13} = 1590, x_{14} = 1640, \dots, x_{48} = 2270$. On représente cette série sur le graphique suivant :

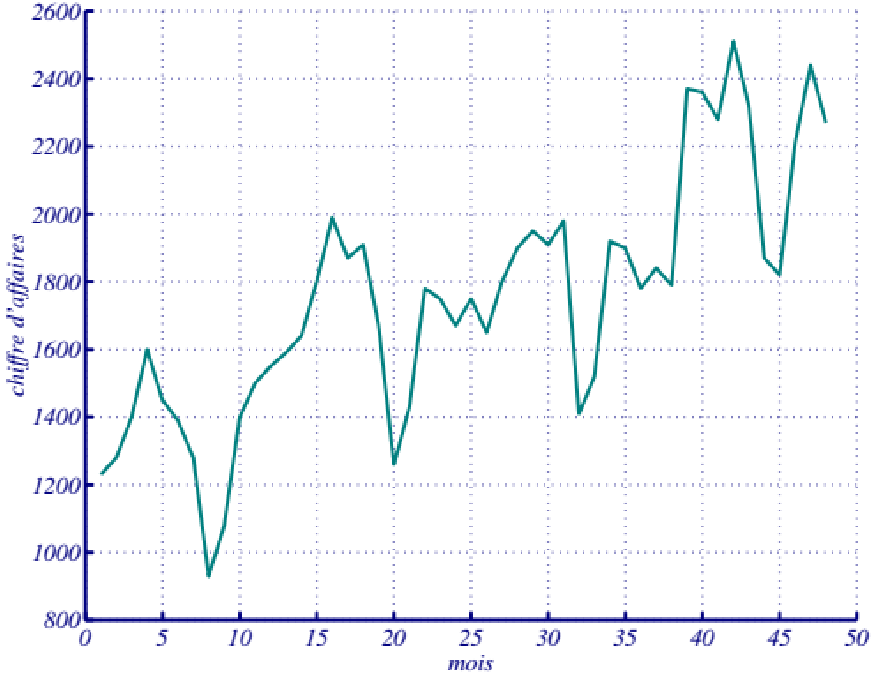



Figure 13. *Évolution du chiffre d'affaires de Janvier* 1989 *à décembre* 1992.

b) Ajustement par la méthode des moindres carrés Le tableau de calcul n'est pas représenté ici. On se contente de donner les résultats, notamment, on indique les différentes sommes suivantes à partir desquelles on calculer toutes les quantités voulues pour l'ajustement : $\sum_{i=1}^{N} m_i = \frac{N(N+1)}{2} = \frac{48\times 49}{2} = 1176$, $\sum x_i = 84000$, $\sum m_i^2 = 38024$, $\sum x_i^2 = 153218400$,

$\sum m_i x_i = 2246180$. Commençons par préciser $\overline{M} = 24.5$, $\overline{X} = 1750$, $\mathrm{V}(M) = 191.92$, $\mathrm{V}(X) = 129550$ et $\mathrm{Cov}(M,X) = 3920.4$. L'ajustement affine de cette série par la méthode des moindres carrés est donné par la droite $D$ d'équation $x = am + b$, où

$$a = \frac{\mathrm{Cov}(M,X)}{\mathrm{V}(M)} \text{ et } b = \overline{X} - a\overline{M}\,,$$

ce qui donne numériquement, l'équation $x = 2.4251m + 1690.6$ puisque

$$a = \frac{3920.4}{191.92} = 20.4277 \text{ et } b = 1750 - (20.4277)24.5 = 1249.5\,.$$

c) Modèle pour l'influence saisonnière On choisit un modéle additif, car l'amplitude des creux (mois d'août) ne varie pas avec les années.

d) Calcul des coefficients saisonniers On va ici calculer les coefficients saisonniers sur une période de 1 an. Autrement dit on doit déterminer 12 coefficient $\alpha_j$. Plus précisemment on écrit, pour $j = 1, \dots, 12$,

$$\alpha_j = \frac{1}{4}(X_j + X_{12+j} + X_{24+j} + X_{36+j})$$

On obtient les résultats suivants : $\alpha_1 = (1230 + 1590 + 1750 + 1840)/4 = 1602.5$, $\alpha_2 = (1280+1640+1650+1790)/4 = 1590$, $\dots$, $\alpha_{12} = (1550+1670+1780+2270)/4 = 1817.5$ (voir tableau).

e) Moyenne des coefficients saisonniers on a simplement

$$\overline{\alpha} = \frac{1}{12}\sum_{j=1}^{12} \alpha_j = \frac{1}{12}(1602.5 + 1590 + \cdots + 1817.5) = \frac{21000}{12} = 1750\,.$$

On en déduit les CVS, grâce à la formule $\alpha'_j = \alpha_j - \overline{\alpha}$ (modèle additif). Par exemple

$$\alpha'_1 = \alpha_1 - \overline{\alpha} = 1602.5 - 1750 = -147.5\,,$$

$$\alpha'_2 = \alpha_2 - \overline{\alpha} = 1590 - 1750 = -160.0\,, \dots,$$

$$\alpha'_{12} = \alpha_{12} - \overline{\alpha} = 1817.5 - 1750 = 67.5\,.$$

f) Série corrigées des variations saisonnières La série corrigée des variations saisonnières est obtenue simplement en écrivant que

$$Y_{12(i-1)+j} = X_{12(i-1)+j} - \alpha'_j$$

où, on a noté $(Y_n)$ la série corrigée des variations saisonnières (voir le tableau). On a par exemple $Y_1 = X_1 - \alpha'_1$, $Y_2 = X_2 - \alpha'_2$, $\dots$, $Y_{12} = X_{12} - \alpha'_{12}$, $Y_{13} = X_{13} - \alpha'_1$, $Y_{14} = X_{14} - \alpha'_2$, $\dots$, $Y_{25} = X_{25} - \alpha'_1$, *etc.*, soit numériquement $Y_1 = 1230 - (-147.5) = 1377.5$, $Y_2 = 1280 - (-160) = 1440$, $\dots$, $Y_{13} = 1590 - (-147.5) = 1737.5$, *etc.*.

g) Chiffre d'affaires en 1993 On commence par résumé les résultats précédents dans le tableau suivant

| m/a | 1989 | 1990 | 1991 | 1992 | $\alpha_j$ | CVS | $1989(c)$ | $1990(c)$ | $1991(c)$ | $1992(c)$ | $1993(l)$ | 1993 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| jan. | 1230 | 1590 | 1750 | 1840 | 1602.5 | −147.5 | 1377.5 | 1737.5 | 1897.5 | 1987.5 | 2250.5 | 2103 |
| fév. | 1280 | 1640 | 1650 | 1790 | 1590 | −160 | 1440 | 1800 | 1810 | 1950 | 2270.9 | 2110.9 |
| mars | 1400 | 1800 | 1800 | 2370 | 1842.5 | 92.5 | 1307.5 | 1707.5 | 1707.5 | 2277.5 | 2291.3 | 2383.8 |
| avr. | 1600 | 1990 | 1900 | 2360 | 1962.5 | 212.5 | 1387.5 | 1777.5 | 1687.5 | 2147.5 | 2311.8 | 2524.3 |
| mai | 1450 | 1870 | 1950 | 2280 | 1887.5 | 137.5 | 1312.5 | 1732.5 | 1812.5 | 2142.5 | 2332.2 | 2469.7 |
| juin | 1390 | 1910 | 1910 | 2510 | 1930 | 180 | 1210 | 1730 | 1730 | 2330 | 2352.6 | 2532.6 |
| jui. | 1280 | 1670 | 1980 | 2320 | 1812.5 | 62.5 | 1217.5 | 1607.5 | 1917.5 | 2257.5 | 2373 | 2435.5 |
| aout | 930 | 1260 | 1410 | 1870 | 1367.5 | −382.5 | 1312.5 | 1642.5 | 1792.5 | 2252.5 | 2393.5 | 2011 |
| sep. | 1080 | 1430 | 1520 | 1820 | 1462.5 | −287.5 | 1367.5 | 1717.5 | 1807.5 | 2107.5 | 2413.9 | 2126.4 |
| oct. | 1400 | 1780 | 1920 | 2210 | 1827.5 | 77.5 | 1322.5 | 1702.5 | 1842.5 | 2132.5 | 2434.3 | 2511.8 |
| nov. | 1500 | 1750 | 1900 | 2440 | 1897.5 | 147.5 | 1352.5 | 1602.5 | 1752.5 | 2292.5 | 2454.8 | 2602.3 |
| déc. | 1550 | 1670 | 1780 | 2270 | 1817.5 | 67.5 | 1482.5 | 1602.5 | 1712.5 | 2202.5 | 2475.2 | 2542.7 |

Les deux dernières colonnes représentent les valeurs prédites. Plus précisemment l'avant dernière colonne représente les prévisions de la série corrigée des variations saisonnières $Y_n$ données par

$$Y_n = a \times n + b\,, \quad n = 49, \dots, 60.$$

Ces valeurs de $n$ correspondent aux 12 mois de l'année 1993. La dernière colonne est la prévision de la série $X_n$ donnée par

$$X_n = Y_n + \alpha'_{n-48}\,.$$

En réalité, on devrait écrire $X_{12(i-1)+j} = Y_{12(i-1)+j)} + \alpha'_j$ pour $i = 5$ (correspond à l'année 1993) et pour tout $j = 1, 2, \dots, 12$.

On représente la série corrigée des variations saisonnières ainsi que les prévisions effectuées sur le graphique suivant

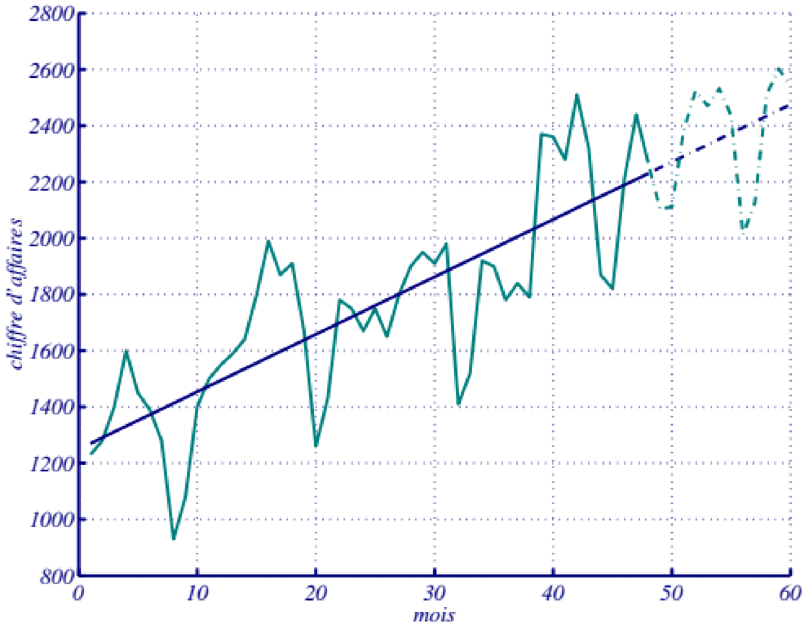



Figure 14. *Prévisions pour l'année* 1993. *La courbe grise représente la série corrigée des variations saisonnières. Les courbes en pointillé représentent les valeurs prédites.*

# 9. Exercice (Truc Net, Modèle additif+ moyenne mobile)

***1.*** La série chronologique de l'entreprise TrucNet sur la période 2014 à 2017 est représentée ci-contre.

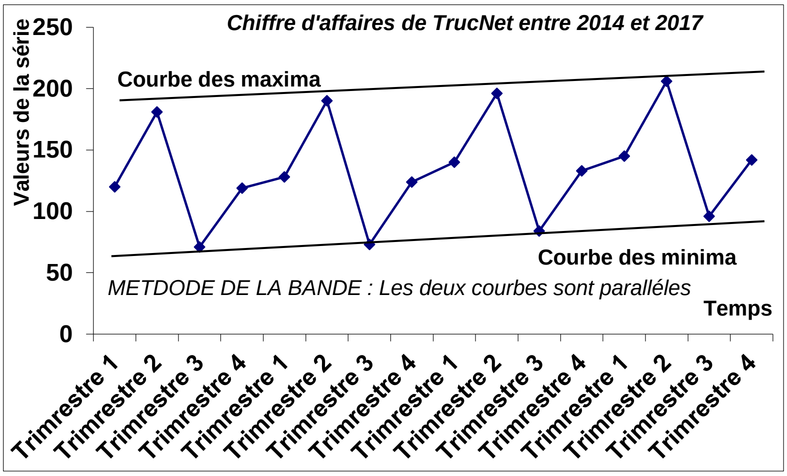



*Tableau 6. Tableau des moyennes mobiles du chiffres d'affaires*

| | Série X | $mm_4(X)$ |
|---|---|---|
| Trimrestre 1 | 120 | X |
| Trimrestre 2 | 181 | X |
| Trimrestre 3 | 71 | 123.75 |
| Trimrestre 4 | 119 | 125.88 |
| Trimrestre 1 | 128 | 127.25 |
| Trimrestre 2 | 190 | 128.13 |
| Trimrestre 3 | 73 | 130.25 |
| Trimrestre 4 | 124 | 132.50 |
| Trimrestre 1 | 140 | 134.63 |
| Trimrestre 2 | 196 | 137.13 |
| Trimrestre 3 | 84 | 138.88 |
| Trimrestre 4 | 133 | 140.75 |
| Trimrestre 1 | 145 | 143.50 |
| Trimrestre 2 | 206 | 146.13 |
| Trimrestre 3 | 96 | X |
| Trimrestre 4 | 142 | X |

***2.*** Le modèle de la série des chiffres d'affaires de l'entreprise TrucNet est additif. En effet, si nous employons la méthode de la bande, qui consiste à tracer la droite qui passe par les minima et celle qui passe par les maxima, ces deux droites sont parallèles.

***3.*** La série du chiffre d'affaires de l'entreprise TrucNet est trimestrielle, la période est donc de quatre. Nous devons, pour éliminer la saison, appliquer une moyenne mobile de longueur 4. Nous obtenons les calculs suivants :

Traçons la série des chiffres d'affaires et sa moyenne mobile centrées :

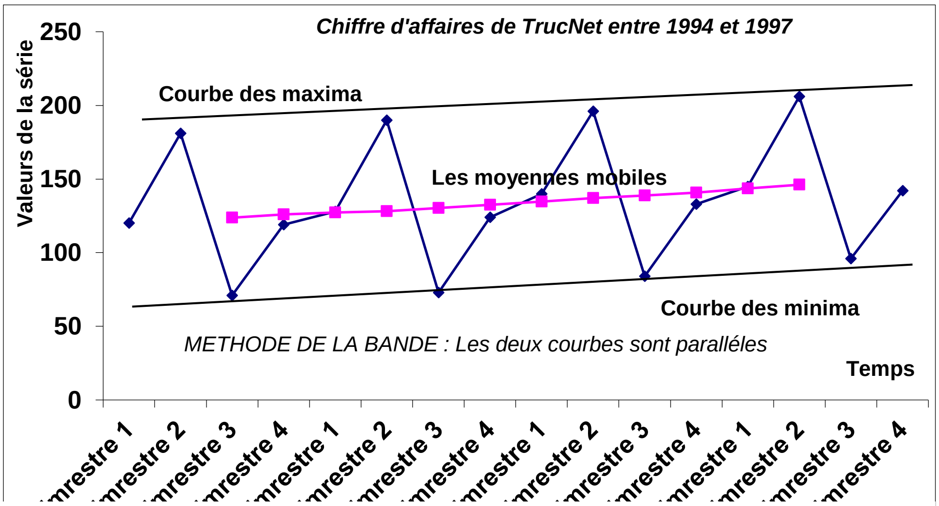



*Figure 4. La série du chiffre d'affaires et moyenne mobile centrée.*

Commentaires : Le graphique ci-dessus nous permet de voir très nettement que le chiffre d'affaires de l'entreprise TrucNet est en progression constante sur la période étudiée. Nous remarquons également que nous avons perdu deux points au départ et deux en fin de série des moyennes mobiles.ds

# TD04- LISSAGE EXPONENTIEL-CORRECTION

## 1. Exercice (QCM)

c) et d)

Le lissage exponentiel simple ne peut s'envisager que pour une chronique sans saisonnalité et sans évolution tendancielle. La prévision tient d'autant plus compte des valeurs récentes de la série que la constante de lissage α est élevée.

## 2. Exercice (Nombre d'immatriculations de voitures avec un moteur diesel)

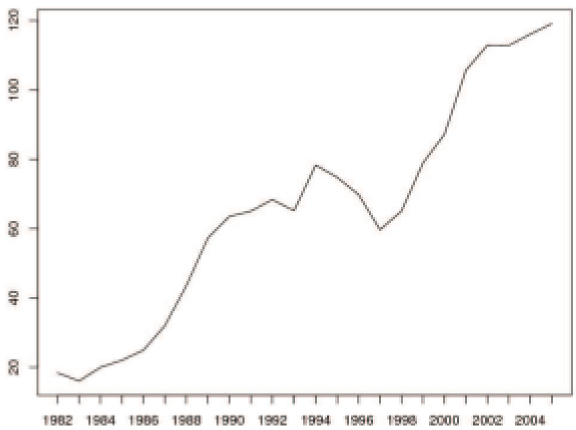
A - Phase exploratoire

1. On obtient :

| Année | Total | Moyenne |
|---|---|---|
| 1982 | 221.3 | 18.44 |
| 1983 | 192.8 | 16.07 |
| 1984 | 240.1 | 20.01 |
| 1985 | 264.8 | 22.07 |
| 1986 | 299.5 | 24.96 |
| 1987 | 384.2 | 32.02 |
| 1988 | 522.5 | 43.54 |
| 1989 | 687.1 | 57.26 |
| 1990 | 762.1 | 63.51 |
| 1991 | 780.5 | 65.04 |
| 1992 | 820.6 | 68.38 |
| 1993 | 783.0 | 65.25 |

| Année | Total | Moyenne |
|---|---|---|
| 1994 | 940.0 | 78.33 |
| 1995 | 897.9 | 74.83 |
| 1996 | 837.9 | 69.83 |
| 1997 | 716.2 | 59.68 |
| 1998 | 780.9 | 65.08 |
| 1999 | 947.5 | 78.96 |
| 2000 | 1046.5 | 87.21 |
| 2001 | 1267.8 | 105.65 |
| 2002 | 1354.9 | 112.91 |
| 2003 | 1354.0 | 112.83 |
| 2004 | 1392.9 | 116.08 |
| 2005 | 1428.9 | 119.08 |

2. On obtient le graphique suivant :




Sur ce graphique, on peut observer globalement trois périodes : la première de 1982 à 1994, où le nombre d'immatriculation de voitures neuves avec moteur diesel a connu une très forte croissance passant de 20M à 80M, soit une augmentation de 300% ; une seconde période, de 1995 à 1997, où ce nombre d'immatriculations a chuté, passant de 80M à 60M, soit une baisse de 25%. Cette baisse n'est pas facile à expliquer, bien que l'on puisse penser que le marché est peut-être arrivé à saturation. Le mieux serait de poser la question aux constructeurs. Enfin, on note une troisième période, de 1998 à 2005, où ce nombre d'immatriculations connaît de nouveau une très forte progression. Cette progression s'explique vraisemblablement par une problématique de coût d'essence et par le fait que les moteurs diesel sont maintenant comparables aux moteurs essence au niveau des caractéristiques : souples, nerveux, fiables, etc.

3. On obtient :

| Mois | Total | Moyenne | Ecart-type |
|---|---|---|---|
| Janvier | **1578.4** | 65.77 | 33.24 |
| Février | 1461.1 | 60.88 | 31.29 |
| Mars | 1708.7 | 71.20 | 37.98 |
| Avril | 1619.9 | 67.50 | 36.20 |
| Mai | 1455.4 | 60.64 | 34.64 |
| Juin | 1461.7 | 60.90 | 43.98 |
| Juillet | 1936.4 | 80.68 | 38.68 |
| Août | 1248.4 | 52.02 | 23.01 |
| Septembre | 1367.6 | 56.98 | 30.74 |
| Octobre | 1728.9 | **72.04** | 33.09 |
| Novembre | 1683.3 | 70.14 | 31.69 |
| Décembre | 1674.1 | 69.75 | 29.72 |

Tout d'abord, on note que les écarts-types sont assez importants, traduisants ainsi une grande diversité du nombre d'immatriculations pour un mois fixé d'une année à l'autre. Néanmoins, on constate qu'en moyenne, c'est le mois de juillet qui connat le plus grand nombre d'immatriculations. A contrario, c'est le mois de d'aot qui connat le nombre moyen d'immatriculations le plus faible. On peut imaginer qu'un grand nombre de personnes achète un véhicule pour partir en vacances, ce qui expliquerait ce pic puis ce creux d'immatriculations. L'analyse du nombre total d'immatriculations confirme évidemment cette observation, avec 80M d'immatriculations en juillet contre 52M en aot.

3. On obtient :

| Mois | Total | Moyenne | Ecart-type |
|---|---|---|---|
| Janvier | **1578.4** | 65.77 | 33.24 |
| Février | 1461.1 | 60.88 | 31.29 |
| Mars | 1708.7 | 71.20 | 37.98 |
| Avril | 1619.9 | 67.50 | 36.20 |
| Mai | 1455.4 | 60.64 | 34.64 |
| Juin | 1461.7 | 60.90 | 43.98 |
| Juillet | 1936.4 | 80.68 | 38.68 |
| Août | 1248.4 | 52.02 | 23.01 |
| Septembre | 1367.6 | 56.98 | 30.74 |
| Octobre | 1728.9 | **72.04** | 33.09 |
| Novembre | 1683.3 | 70.14 | 31.69 |
| Décembre | 1674.1 | 69.75 | 29.72 |

Tout d'abord, on note que les écarts-types sont assez importants, traduisants ainsi une grande diversité du nombre d'immatriculations pour un mois fixé d'une année à l'autre. Néanmoins, on constate qu'en moyenne, c'est le mois de juillet qui connat le plus grand nombre d'immatriculations. A contrario, c'est le mois de d'aot qui connat le nombre moyen d'immatriculations le plus faible. On peut imaginer qu'un grand nombre de personnes achète un véhicule pour partir en vacances, ce qui expliquerait ce pic puis ce creux d'immatriculations. L'analyse du nombre total d'immatriculations confirme évidemment cette observation, avec 80M d'immatriculations en juillet contre 52M en aot.

4. Le graphique de gauche représente sous forme de boxplots la distribution du volume mensuel des immatriculations par an sur la période d'étude. Sur ce graphique, on retrouve les trois périodes détectées précédemment via une analyse de l'évolution de la médiane. La nouveauté est que l'on observe une volatilité qui augmente avec le temps. Ainsi, sur la première période, on note une dispersion relativement faible, traduisant un nombre d'immatriculations d'un mois sur l'autre, pour une année fixée, relativement homogène. Par contre, sur les deux autres périodes, on observe une grande dispersion du nombre d'immatriculations, avec la présence de points atypiques. Cela traduit le fait que sur ces deux périodes, le nombre mensuel d'immatriculations peut être très différent d'un mois sur l'autre pour une même année. On note d'ailleurs la présence d'un mois très "creux" pour les années 2002, 2003, 2004, et 2005, avec à peine 80M d'immatriculations pour une médiane aux environs de 120M ! Sans surprise, ce mois correspond au mois d'août. Le graphique de droite représente sous forme de boxplots la distribution du volume mensuel des immatriculations par mois sur la période d'étude. On note que sur la période d'étude les dispersions sont très importantes et toutes assez semblables, sauf peut-être pour le mois d'août. Cette double observation s'explique en partie par le fait que le nombre d'immatriculations a connu sur la période 1982-2005 une forte progression. Finalement, on ne peut tirer beaucoup d'informations de ce graphique, si ce n'est que le mois de juillet se démarque des autres mois avec une médiane importante, et donc un nombre élevé d'immatriculations.
5. Le graphique de la série brute permet tout d'abord de confirmer l'existence des trois périodes, même si elles apparaissent moins disctinctement du fait du "bruit". Néanmoins, le phénomène est globalement croissant avec une augmentation maximale de 600%. La composane tendancielle reste toutefois aléatoire sur l'ensemble de la période, et cela du fait du "trou" observé sur la période 1995-1997. Cette situation fait alors apparaître des ruptures de tendances : en 1995 et en 1997. Sur ce graphique, il est très diffcile d'observer un quelconque phénomène saisonnier. Enfin, la variabilité est clairement hétéroscédastique puisqu'elle augmente au cours du temps.
6. Dans le cas présent, il importe de transformer la série en prenant le logarithme de la série initiale. Le graphique montre que cette transformation a permis de stabiliser la variabilité de la série.

B - Phase de modélisation

1. Clairement, on obtient : $a_2 = 2.879$ et $b_2 = -0.006$. Ainsi, on obtient : $\hat{y}_2(1) = 2.873$.
2. En prenant $\alpha = 0.2$ et $\beta = 0.6$, on obtient :

| | **Jan.** | **Feb.** | **Mar.** | **Apr.** | **May** | **Jun.** | **Jul.** | **Aug.** | **Sep.** | **Oct.** | **Nov.** | **Dec.** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1982 | **-** | **-** | **2.87** | **2.94** | **2.96** | **2.91** | **2.89** | **2.85** | **2.76** | **2.74** | **2.79** | **2.87** |
| 1983 | 2.97 | 2.99 | 2.93 | 2.91 | 2.82 | 2.72 | 2.69 | 2.65 | 2.55 | 2.49 | 2.55 | 2.68 |
| 1984 | 2.83 | 2.91 | 2.94 | 3.00 | 3.01 | 3.06 | 3.07 | 3.06 | 3.00 | 2.91 | 3.00 | 3.11 |
| 1985 | 3.19 | 3.24 | 3.25 | 3.27 | 3.23 | 3.17 | 3.08 | 3.06 | 2.92 | 2.84 | 2.90 | 2.98 |
| 1986 | 3.07 | 3.16 | 3.15 | 3.17 | 3.19 | 3.16 | 3.11 | 3.14 | 3.13 | 3.10 | 3.20 | 3.30 |
| 1987 | 3.47 | 3.52 | 3.53 | 3.57 | 3.59 | 3.53 | 3.43 | 3.46 | 3.39 | 3.28 | 3.36 | 3.44 |
| 1988 | 3.58 | 3.67 | 3.68 | 3.76 | 3.81 | 3.83 | 3.75 | 3.82 | 3.78 | 3.68 | 3.69 | 3.76 |
| 1989 | 3.95 | 4.11 | 4.21 | 4.27 | 4.28 | 4.20 | 4.09 | 4.12 | 4.04 | 3.95 | 3.98 | 4.00 |
| 1990 | 3.98 | 4.10 | 4.14 | 4.19 | 4.19 | 4.17 | 4.06 | 4.11 | 4.06 | 3.96 | 4.03 | 4.12 |
| 1991 | 4.18 | 4.28 | 4.28 | 4.30 | 4.30 | 4.23 | 4.09 | 4.16 | 4.10 | 4.02 | 4.07 | 4.10 |
| 1992 | 4.14 | 4.21 | 4.21 | 4.25 | 4.26 | 4.21 | 4.13 | 4.23 | 4.19 | 4.14 | 4.20 | 4.25 |
| 1993 | 4.29 | 4.25 | 4.21 | 4.21 | 4.17 | 4.07 | 3.97 | 4.06 | 4.03 | 3.99 | 4.08 | 4.20 |
| 1994 | 4.31 | 4.34 | 4.34 | 4.42 | 4.46 | 4.43 | 4.31 | 4.38 | 4.34 | 4.28 | 4.32 | 4.41 |
| 1995 | 4.48 | 4.52 | 4.52 | 4.54 | 4.48 | 4.37 | 4.30 | 4.34 | 4.27 | 4.16 | 4.13 | 4.15 |
| 1996 | 4.12 | 4.18 | 4.23 | 4.29 | 4.33 | 4.28 | 4.16 | 4.24 | 4.21 | 4.22 | 4.23 | 4.18 |
| 1997 | 4.16 | 4.09 | 4.02 | 4.00 | 4.02 | 3.97 | 3.82 | 3.98 | 3.97 | 3.99 | 4.07 | 4.12 |
| 1998 | 4.19 | 4.21 | 4.20 | 4.24 | 4.26 | 4.21 | 4.05 | 4.17 | 4.12 | 4.11 | 4.13 | 4.17 |
| 1999 | 4.21 | 4.22 | 4.22 | 4.29 | 4.36 | 4.33 | 4.18 | 4.37 | 4.40 | 4.39 | 4.44 | 4.48 |
| 2000 | 4.53 | 4.53 | 4.54 | 4.58 | 4.57 | 4.57 | 4.55 | 4.55 | 4.48 | 4.40 | 4.39 | 4.40 |
| 2001 | 4.37 | 4.43 | 4.47 | 4.56 | 4.65 | 4.73 | 4.83 | 4.91 | 4.83 | 4.74 | 4.73 | 4.70 |
| 2002 | 4.64 | 4.65 | 4.64 | 4.68 | 4.74 | 4.76 | 4.82 | 4.86 | 4.73 | 4.65 | 4.66 | 4.64 |
| 2003 | 4.63 | 4.65 | 4.66 | 4.72 | 4.76 | 4.77 | 4.84 | 4.88 | 4.71 | 4.66 | 4.67 | 4.63 |
| 2004 | 4.62 | 4.62 | 4.62 | 4.70 | 4.75 | 4.78 | 4.89 | 4.91 | 4.76 | 4.71 | 4.69 | 4.71 |
| 2005 | 4.72 | 4.73 | 4.73 | 4.81 | 4.87 | 4.90 | 4.98 | 4.93 | 4.76 | 4.70 | 4.64 | 4.63 |

3. On ne peut être sûr que ces deux constantes amènent un lissage optimal. Pour obtenir le meilleur ajustement, il faut dans un premier temps déterminer un critère qui permette de quantifier l'ajustement. Souvent, on fait usage de la somme des écarts carrés SSE (Sum of Square Error). Ensuite, il convient de faire tourner un algorithme lpermettant de tester un ensemble de valeurs possibles pour le couple $(\alpha, \beta)$.
4. L'intérêt d'obtenir une série lissée est que l'on visualise plus facilement la tendance, c'est-à-dire en faisant abstraction des irrégularités. Il est alors plus facile d'avoir une approche subjective pour réaliser des prévisions à court terme.

5.
Clairement, cette modélisation semble meilleure car elle s'adapte beaucoup mieux à la série, et donc diminue l'erreur d'ajustement.

C - Phase de prévision

1. On obtient :

| $\log \hat{y}_t$ | **Janv.** | **Fév.** | **Mars** | **Avril** | **Mai** | **Juin** | **Juil.** | **Aot** | **Sep.** | **Oct.** | **Nov.** | **Déc.** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2006 | 4.7599 | 4.7247 | 4.9494 | 4.8747 | 4.8202 | 4.9969 | 4.7526 | 4.3741 | 4.7496 | 4.7958 | 4.8271 | 4.8091 |
| 2007 | 4.8185 | 4.7833 | 5.0080 | 4.9333 | 4.8788 | 5.0555 | 4.8112 | 4.4327 | 4.8082 | 4.8544 | 4.8857 | 4.8678 |

2. Oui, les résultats numériques sont en accord avec le graphique.
3. On doit réaliser la transformation inverse de celle effectuée précédemment. Ainsi, on doit prendre l'exponentielle de la série.
4. On obtient :

| $\hat{y}_t$ | **Janv.** | **Fév.** | **Mars** | **Avril** | **Mai** | **Juin** | **Juil.** | **Aot** | **Sep.** | **Oct.** | **Nov.** | **Déc.** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2006 | 116.73 | 112.70 | 141.09 | 130.93 | 123.98 | 147.86 | 115.88 | 79.37 | 115.51 | 121.03 | 124.84 | 122.61 |
| 2007 | 123.78 | 119.52 | 149.61 | 138.84 | 131.46 | 156.79 | 122.88 | 84.16 | 122.48 | 128.34 | 132.38 | 130.01 |

5. Evident.