Séries chronologiques

Arnaud Rousselle, Clément Rau

Outils maths pour la gestion (S2)

Définition

On appelle série chronologique ou série temporelle toute suite (finie) $\mathbf{y}$ *d'observations numériques d'une grandeur effectuées au cours de plusieurs années à intervalle de temps réguliers, appelés saisons.*

Notations

On note :

- y_t *la valeur de la* t^e *observation,*
- n *le nombre d'années,*
- p *le nombre de saisons (périodes) dans une année.*

Exemple 1

On a relevé les chiffres d'affaires trimestriels, exprimés en K€, d'une entreprise au cours des années 2012 à 2015. Ceux-ci sont présentés dans le tableau suivant.

Année \ Trimestre	1	2	3	4
2012	20	25	50	70
2013	35	30	65	105
2014	40	34	75	135
2015	50	37	80	170

Exemple 1 (en image)

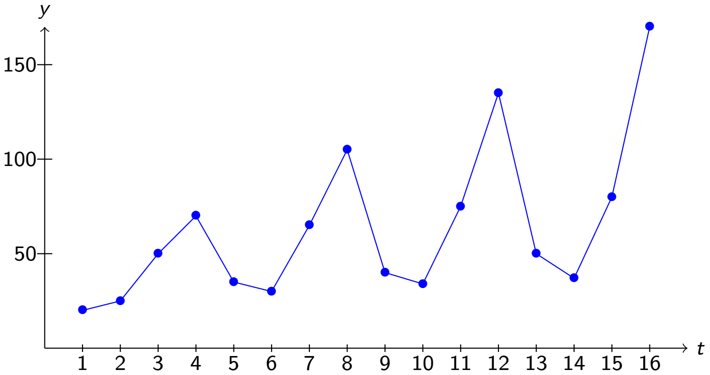

Figure: Chiffres d'affaires y_t (en K€) de la t^{e} période (Exemple 1).

Ici, le nombre d'années est $n = 4$, le nombre de périodes par année est $p = 4$ et on a $y_1 = 20$, $y_2 = 25$, $y_3 = 50$, $y_4 = 70$, $y_5 = 35$, ..., $y_{16} = 170$.

Exemple 2

On a relevé la consommation mensuelle en eau, exprimée en Hm^3, d'un producteur de melons au cours des années 2013 à 2015. Les relevés sont présentés dans le tableau suivant.

	Janv.	Fév.	Mars	Avr.	Mai	Juin	Juil.	Août	Sept.	Oct.	Nov.	Déc.
2013	1	1,5	3	5	10	20	45	50	30	2	1	0,5
2014	3,5	3	5,5	9	11	24	49	50	31	4	4	3,5
2015	7	6	8	9	15	25	52	55	37	7	5	6

Exemple 2 (en image)

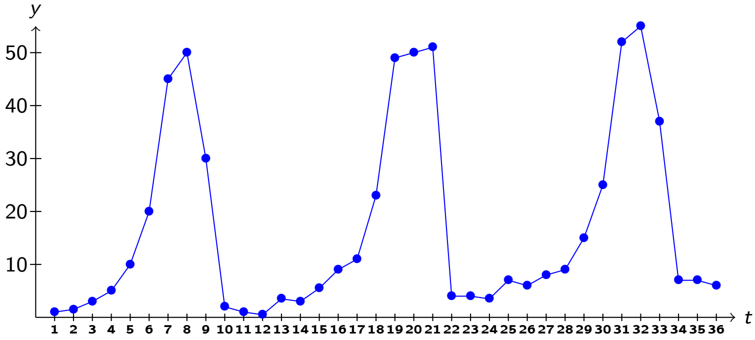

Figure: Consommation d'eau y_t (en Hm^3) au cours du t^e mois (Exemple 2).

Ici, le nombre d'années est $n = 3$, le nombre de périodes par année est $p = 12$ et on a $y_1 = 1$, $y_2 = 1,5$,..., $y_{12} = 0,5$, $y_{13} = 3,5$, ..., $y_{36} = 6$.

Motivations et ...

Pour toute série chronologique, trois questions essentielles se posent :

- Une *tendance générale* (« *trend* ») se dégage-t-elle de la série ? Quelle est l'évolution de la série en temps long ?
- Peut-on détecter des phénomènes saisonniers ? L'évolution de la série présente-t-elle des caractéristique se répétant d'année en année ?
- Peut-on prévoir des valeurs futures de la série ?

Composantes d'une série chronologique :

- la tendance générale (appelée « *trend* »),
- une composante *saisonnière*,
- une composante *aléatoire* (imprévisible).

objectifs

Pour répondre aux questions posées ci-dessus, nous fournirons des méthodes permettant :

- de décomposer une série chronologique en composantes *générale*, *saisonnière* et *aléatoire* afin de l'analyser (voir Section 2) ;
- d'estimer les composantes de la série chronologique (voir Section 3) ;
- prédire les valeurs futures de la série (voir Section 4).

Décomposition d'une série chronologique

On veut écrire toute valeur y_t d'une série chronologique sous la forme :

$$y_t = f(g_t, s_t, a_t),$$

où f est une certaine fonction et :

- g_t désigne la composante « tendance **g**énérale » de la série au temps t,
- s_t désigne la composante « **s**aisonnière » de la série au temps t,
- a_t désigne la composante « **a**léatoire » de la série au temps t.

Différentes fonctions f conduisent à des modèles différents. Dans la suite, on s'intéressera aux modèles *additif* et *multiplicatif*.

Exemple 1 (suite)

Graphiquement, on constate que les chiffres d'affaires ont tendance à augmenter d'année en année. Rigoureusement, on peut justifier ce fait en faisant une régression linéaire. On constate que la tendance générale à une faible courbure, ce qui rend légitimes les calculs que nous ferons dans la suite.

On voit que le chiffre d'affaire présente des « pics positifs » au 4$^{\text{e}}$ trimestre de chaque année et des « pics négatifs » lors des deux premiers trimestres de chaque année. On cherchera, dans la suite, à décrire l'impact d'une saison (ici un trimestre) sur la série chronologique. On supposera que la composante saisonnière est périodique de période p ($s_{k+p} = s_k$, pour tout k) et a une influence nulle sur une année.

La composante aléatoire est plus délicate à visualiser. Elle correspond aux irrégularités des cycles de la série. Nous supposerons par la suite que celle-ci est négligeable.

Modèle additif

Définition

On parle de modèle additif lorsque la série chronologique $\mathbf{y} = y_t$ *se décompose sous la forme :*

$$y_t = g_t + s_t + a_t,$$

où g_t *désigne la composante « tendance* **g***énérale »,* s_t *désigne la composante* **s***aisonnière et* a_t *désigne la composante* **a***léatoire de la série au temps* t*.*

Remarque

Rappelons que la composante saisonnière est supposée p-périodique (i.e. $s_{k+p} = s_k$*, pour tout* k*) et d'influence nulle sur une année (i.e.* $s_1 + s_2 + \cdots + s_p = 0$ *pour le modèle additif). La composante aléatoire est supposée négligeable (i.e.* $a_t \simeq 0$*, pour tout* t*, pour le modèle additif).*

Critère

Pour savoir si le modèle additif est adapté, on trace les lignes polygonales passant par les pics « positifs » d'une part et « négatifs » d'autre part. Si celles-ci sont proches de droites et si la largeur de la bande délimitée par celles-ci est essentiellement constante, on choisit d'utiliser le modèle additif.

Exemple 2 (suite)

Le modèle additif est adapté pour l'Exemple 2.

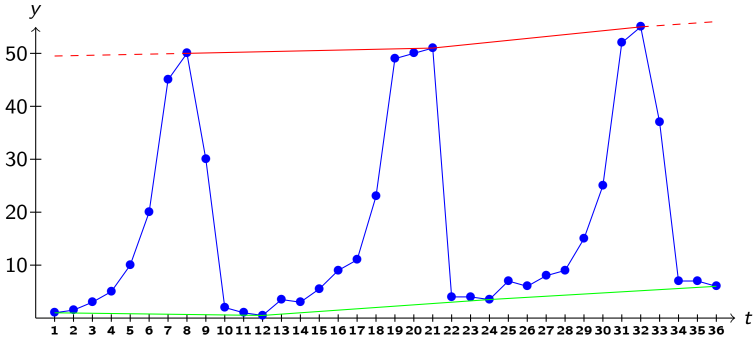

Figure: Consommation d'eau y_t (en Hm^3) au cours du t^e mois de l'Exemple 2 (bleu) et les lignes polygonales passant par les pics « positifs » (rouge) et « négatifs » (vert).

Modèle multiplicatif

Définition

On parle de modèle multiplicatif lorsque la série chronologique $\mathbf{y} = y_t$ *se décompose sous la forme :*

$$y_t = g_t \times s_t \times a_t,$$

où g_t *désigne la composante « tendance* **g***énérale »,* s_t *désigne la composante* **s***aisonnière et* a_t *désigne la composante* **a***léatoire de la série au temps* t*.*

Remarque

Rappelons que la composante saisonnière est supposée p-périodique (i.e. $s_{k+p} = s_k$*, pour tout* k*) et d'influence nulle sur une année (i.e.* $s_1 \times s_2 \times \cdots \times s_p = 1$ *pour le modèle multiplicatif). La composante aléatoire est supposée négligeable (i.e.* $a_t \simeq 1$*, pour tout* t*, pour le modèle multiplicatif).*

Critère

Pour savoir si le modèle multiplicatif est adapté, on trace les lignes polygonales passant par les pics « positifs » d'une part et « négatifs » d'autre part. Si celles-ci sont proches de droites et si la largeur de la bande délimitée par celles-ci est clairement croissante ou décroissante , on choisit d'utiliser le modèle multiplicatif.

Exemple 1 (suite)

Le modèle multiplicatif est adapté pour l'Exemple 1.

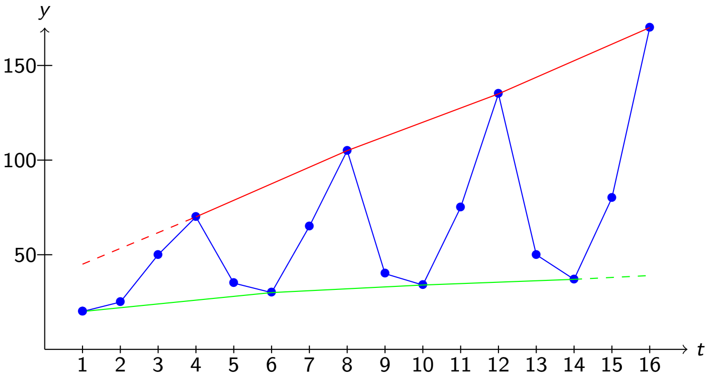

Figure: Chiffres d'affaires y_t (en K€) du $t^{\mathbf{e}}$ trimestre de l'Exemple 1 (bleu) et les lignes polygonales passant par les pics « positifs » (rouge) et « négatifs » (vert).

Estimation de la tendance générale

Objectif

Trouver une fonction du temps (simple) qui modélise au mieux la tendance de la série $\mathbf{y}$.

Estimation de la tendance générale

Objectif

Trouver une fonction du temps (simple) qui modélise au mieux la tendance de la série $\mathbf{y}$.

On distingue essentiellement deux méthodes.

- **Une première méthode paramétrique.** Pour estimer la tendance générale g_t, on effectue une régression de y_t en fonction de t. Celle-ci conduit à des calculs assez simples, mais n'est pas toujours adaptée à cause des effets saisonniers.
- **Une deuxième méthode non paramétrique**, utilisant le concept de *moyennes mobiles*, qui permet de "lisser" la série en atténuant les fluctuations saisonnières. Nous détaillerons cette approche.

Estimation de la tendance générale

Définition

Soit $\mathbf{y}$ *une série chronologique portant sur* n *années décomposées en* p *saisons et soit* $k \in \{2, \ldots, np\}$*. On appelle moyenne mobile (ou glissante) d'ordre* k *au temps* $t \in \{1 + \lfloor \frac{k}{2} \rfloor, \ldots, np - \lfloor \frac{k}{2} \rfloor\}$ *la quantité :*

$$\mathrm{M}_t^{(k)} = \begin{cases} \frac{1}{k}\left(y_{t-\frac{k-1}{2}} + \cdots + y_t + \cdots + y_{t+\frac{k-1}{2}}\right) & \textit{si } k \textit{ est impair} \\ \frac{1}{k}\left(\frac{y_{t-\frac{k}{2}}}{2} + y_{t-\frac{k}{2}+1} + \cdots + y_t + \cdots + y_{t+\frac{k}{2}-1} + \frac{y_{t+\frac{k}{2}}}{2}\right) & \textit{si } k \textit{ est pair} \end{cases}.$$

Moyennes mobiles (exemples)

Les moyennes mobiles d'ordre 2, 3, 4 et 5 sont données par :

$$\mathrm{M}_t^{(2)} = \frac{1}{2}\left(\frac{y_{t-1}}{2} + y_t + \frac{y_{t+1}}{2}\right),$$

$$\mathrm{M}_t^{(3)} = \frac{1}{3}\left(y_{t-1} + y_t + y_{t+1}\right),$$

$$\mathrm{M}_t^{(4)} = \frac{1}{4}\left(\frac{y_{t-2}}{2} + y_{t-1} + y_t + y_{t+1} + \frac{y_{t+2}}{2}\right)$$

et

$$\mathrm{M}_t^{(5)} = \frac{1}{5}\left(y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2}\right).$$

Remarque

1. *L'utilisation des moyennes mobiles permet de lisser la courbe et « supprimer » la composante aléatoire par moyennisation.*
2. *Lorsque k est un multiple de p, la composante saisonnière n'influe pas sur la moyenne mobile d'ordre k.*
3. *Il est nécessaire de choisir convenablement l'ordre k des moyennes mobiles utilisées. En général, on choisit $k = p$ (le nombre de période dans l'année).*

Estimation de la tendance générale

Proposition

Soit $\mathbf{y}$ *une série chronologique dont la tendance générale est notée* g_t*.*
Alors, un estimateur de g_t *est donné par :*

$$\hat{g}_t = \mathsf{M}_t^{(k)} .$$

Autrement dit, on peut considérer que :

$$g_t \simeq \hat{g}_t = \mathsf{M}_t^{(k)} .$$

Remarque

L'estimation de la tendance générale ne dépend pas du fait que l'on utilise un modèle additif ou multiplicatif.

Exemple 1 (suite)

L'estimation de la tendance des chiffres d'affaires trimestriels de l'Exemple 1 par des moyennes mobiles d'ordre 4 est donnée dans le tableau suivant et représentée dans la figure suivante.

Année \ Trimestre	1	2	3	4
2012			43,125	45,625
2013	48,125	54,375	59,375	60,5
2014	62,25	67,25	72,25	73,875
2015	74,875	79,875		

Exemple 1 (suite et image)

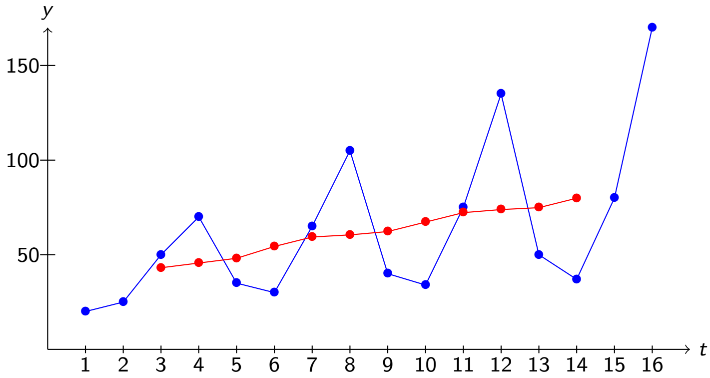

Figure: Chiffres d'affaires y_t (en K€) du $t^{\mathbf{e}}$ trimestre (bleu, Exemple 1) et estimation de la tendance g_t par $\hat{g}_t = \mathrm{M}_t^{(\mathbf{4})}$ (rouge).

Exemple 2 (suite)

Les valeurs approchées à une décimale de l'estimation de la tendance la consommation mensuelle d'eau de l'Exemple 2 par des moyennes mobiles d'ordre 7 sont données dans le tableau suivant et représentées dans la figure suivante.

	Janv.	Fév.	Mars	Avr.	Mai	Juin	Juil.	Août	Sept.	Oct.	Nov.	Déc.
2013				12,2	19,2	23,3	23,1	22,6	21,2	18,9	12,9	6,5
2014	3,5	4,8	8,1	15	21,6	25,6	25,4	24,7	23,6	21,2	15,1	9,1
2015	5,9	7,5	10,5	17,4	24,3	28,7	28,6	28	26,7			

Exemple 2 (suite en image)

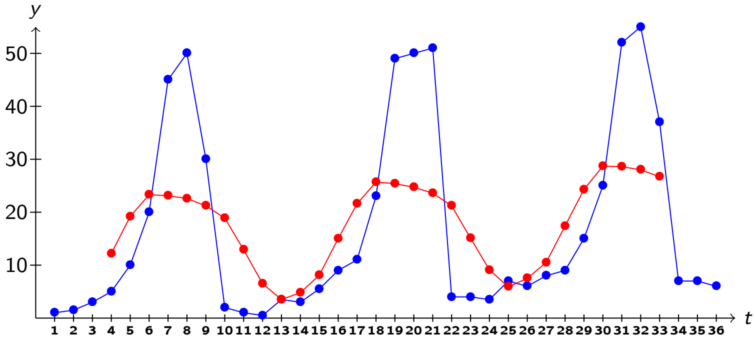

Figure: Consommation d'eau y_t (en $\mathrm{Hm}^{\mathbf{3}}$) au cours du $t^{\mathbf{e}}$ mois (bleu, Exemple 2) et estimation de la tendance g_t par $\hat{g}_t = \mathrm{M}_t^{(\mathbf{7})}$ (rouge).

Estimation de la composante saisonnière

L'estimation de la composante saisonnière s_t dépend du fait que l'on utilise un modèle additif ou multiplicatif.

Estimation de la composante saisonnière dans le modèle additif

On veut écrire les éléments de la série sous la forme :

$$y_t = g_t + s_t + a_t$$

où g_t est la tendance générale, s_t la composante saisonnière et a_t la composante aléatoire de la série au temps t.

$$y_t \simeq \hat{g}_t + \hat{s}_t.$$

On effectue les étapes suivantes :

1. pour $t \in \{1 + \lfloor \frac{k}{2} \rfloor, \ldots, np - \lfloor \frac{k}{2} \rfloor\}$, on calcule les écarts entre la valeur de la série et l'estimation de la tendance au temps t : $e_t = y_t - \hat{g}_t$;
2. pour $t \in \{1, \ldots, p\}$, on calcule la moyenne **arithmétique** $\bar{e}_t$ des écarts correspondant à la même période de l'année, *i.e.* des e_{t+kp} ;
3. on calcule la moyenne **arithmétique** $\bar{e}$ des $\bar{e}_t$: $\bar{e} = \frac{\bar{e}_{\mathbf{1}} + \cdots + \bar{e}_p}{p}$;
4. pour $t \in \{1, \ldots, p\}$, on calcule les *coefficients saisonniers normalisés* $\hat{s}_t = \bar{e}_t - \bar{e}$.

Proposition

Dans le cadre d'un modèle additif, une estimation de la composante saisonnière est donnée par $\hat{s}_t$ défini ci-dessus.

Exemple 2 (suite)

Le tableau suivant donne les valeurs approchées à une décimale près des écarts $e_t = y_t - \hat{g}_t$ obtenus pour chaque période $t \in \{4, \ldots, 33\}$, celles des moyennes $\bar{e}_t$ des écarts correspondant à chacun des mois d'une année :

$$\bar{e}_1 = \frac{e_{13} + e_{25}}{2}, \ldots, \bar{e}_3 = \frac{e_{15} + e_{27}}{2}, \bar{e}_4 = \frac{e_4 + e_{16} + e_{28}}{3}, \ldots, \bar{e}_9 = \frac{e_9 + e_{21} + e_{33}}{3},$$

$$\bar{e}_{10} = \frac{e_{10} + e_{22}}{2}, \ldots, \bar{e}_{12} = \frac{e_{12} + e_{24}}{2},$$

ainsi que l'estimation de la composante saisonnière $\hat{s}_t = \bar{e}_t - \bar{e}$ où $\bar{e} = \frac{\bar{e}_1 + \bar{e}_2 + \cdots + \bar{e}_{12}}{12} \simeq 0,05$.

	Janv.	Fév.	Mars	Avr.	Mai	Juin	Juil.	Août	Sept.	Oct.	Nov.	Déc.
2013				-7,2	-9,2	-3,3	21,9	27,4	8,8	-16,9	-11,9	-6
2014	0	-1,8	-2,6	-6	-10,6	-1,6	23,6	25,3	7,4	-17,2	-11,1	-5,6
2015	1,1	-1,5	-2,5	-8,4	-9,3	-3,7	23,4	27	10,3			
$\bar{e}_t$	0,55	-1,65	-2,5	-7,2	-9,7	-2,9	23	26,6	8,8	-17	-11,5	-5,8
$\hat{s}_t$	0,5	-1,7	-2,6	-7,2	-9,8	-2,9	22,9	26,5	8,8	-17,1	-11,5	-5,8

Estimation de la composante saisonnière dans le modèle multiplicatif

On veut écrire les éléments de la série sous la forme :

$$y_t = g_t \times s_t \times a_t$$

où g_t est la tendance générale, s_t la composante saisonnière et a_t la composante aléatoire de la série au temps t.

$$y_t \simeq \hat{g}_t \times \hat{s}_t.$$

Pour cela, on effectue les étapes suivantes :

1. pour $t \in \{1 + \lfloor \frac{k}{2} \rfloor, \ldots, np - \lfloor \frac{k}{2} \rfloor\}$, on calcule les ratios entre la valeur de la série et l'estimation de la tendance au temps t : $r_t = \frac{y_t}{\hat{g}_t}$;
2. pour $t \in \{1, \ldots, p\}$, on calcule la moyenne **géométrique** $\bar{r}_t$ des ratios correspondant à la même période de l'année, *i.e.* des r_{t+kp} ;
3. on calcule la moyenne **géométrique** $\bar{r}$ des $\bar{r}_t$: $\bar{r} = \sqrt[p]{\bar{r}_1 \times \cdots \times \bar{r}_p}$;
4. pour $t \in \{1, \ldots, p\}$, on calcule les *coefficients saisonniers normalisés* $\hat{s}_t = \frac{\bar{r}_t}{\bar{r}}$.

Proposition

Dans le cadre d'un modèle multiplicatif, une estimation de la composante saisonnière est donnée par $\hat{s}_t$ défini ci-dessus.

Exemple 1 (suite)

Le tableau suivant donne les valeurs approchées à deux décimales près des ratios $r_t = \frac{y_t}{\hat{g}_t}$ obtenus pour chaque période $t \in \{3, 4, \ldots, 14\}$, celles des moyennes $\bar{r}_t$ des ratios correspondant à chacun des trimestres d'une année :

$$\bar{r}_1 = \sqrt[3]{r_5 r_9 r_{13}}, \quad \bar{r}_2 = \sqrt[3]{r_6 r_{10} r_{14}}, \quad \bar{r}_3 = \sqrt[3]{r_3 r_7 r_{11}}, \quad \text{et} \quad \bar{r}_4 = \sqrt[3]{r_4 r_8 r_{12}},$$

ainsi que l'estimation de la composante saisonnière $\hat{s}_t = \frac{\bar{r}_t}{\bar{r}}$ où $\bar{r} = \sqrt[4]{\bar{r}_1 \bar{r}_2 \bar{r}_3 \bar{r}_4} \simeq 0,89$.

Année \ Trimestre	1	2	3	4
2012			1,16	1,53
2013	0,73	0,55	1,09	1,74
2014	0,64	0,51	1,04	1,83
2015	0,67	0,46		
$\bar{r}_t$	0,68	0,51	1,10	1,69
$\hat{s}_t$	0,76	0,57	1,23	1,90

Prédiction

Étant donnée une série chronologique $\mathbf{y}$ portant sur n années découpées en p saisons, on souhaite prédire les valeurs futures de la série. Pour cela, on effectue la démarche suivante.

1. On estime la tendance à l'aide de moyennes mobiles $\hat{g}_t = \mathrm{M}_t^{(k)}$, $t \in \{1, \ldots, np\}$.
2. On estime la composante saisonnière par $\hat{s}_t$.
3. On effectue une régression linéaire sur les $\hat{g}_t$ pour prédire leurs valeurs futures .
4. On en déduit des prédictions des valeurs futures de la série en utilisant que :
    - $y_t \simeq \hat{g}_t + \hat{s}_t$ si l'on utilise un modèle additif,
    - $y_t \simeq \hat{g}_t \times \hat{s}_t$ si l'on utilise un modèle multiplicatif.

Exemple 1 (suite)

En utilisant les résultats déjà obtenus dans les sections précédentes, on obtient que la droite de régression de $\hat{g}_t$ en t a pour équation :

$$y = \hat{g}_t = 3,37t + 33,15.$$

On en déduit les prédictions suivantes pour l'année 2016 :

$$y_{17} \simeq (3,37 \times 17 + 33,15)\hat{s}_{17} \simeq (3,37 \times 17 + 33,15) \times 0,76 \simeq 68,73,$$

$$y_{18} \simeq (3,37 \times 18 + 33,15)\hat{s}_{18} \simeq (3,37 \times 18 + 33,15) \times 0,57 \simeq 53,47,$$

$$y_{19} \simeq (3,37 \times 19 + 33,15)\hat{s}_{19} \simeq (3,37 \times 19 + 33,15) \times 1,22 \simeq 118,56,$$

$$y_{20} \simeq (3,37 \times 20 + 33,15)\hat{s}_{20} \simeq (3,37 \times 20 + 33,15) \times 1,89 \simeq 190,04.$$

Celles-ci sont représentées dans la figure suivante.

Exemple 1 (suite et fin en image)

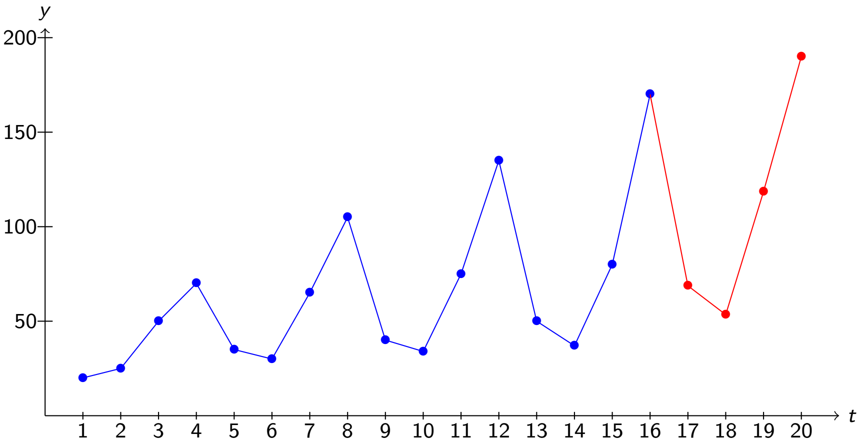

Figure: Chiffres d'affaires y_t (en K€) de la $t^{\mathbf{e}}$ période de l'Exemple 1 (bleu) et prédictions pour l'année 2016 (rouge).

Exemple 2 (suite)

En utilisant les résultats déjà obtenus dans les sections précédentes, on obtient que la droite de régression de $\hat{g}_t$ en t a pour équation :

$$y = \hat{g}_t = 0,04t + 17,1.$$

On en déduit les prédictions pour l'année 2016, en calculant, pour $t \in \{37, \dots, 48\}$,

$$y_t \simeq 0,04t + 17,1 + \hat{s}_t.$$

Les prédictions sont les suivantes et sont représentées dans la figure suivante :

$y_{37} \simeq 19,08,$ $y_{38} \simeq 16,92,$ $y_{39} \simeq 16,06,$ $y_{40} \simeq 11,4,$ $y_{41} \simeq 8,94,$ $y_{42} \simeq 15,88,$

$y_{43} \simeq 41,72,$ $y_{44} \simeq 45,36,$ $y_{45} \simeq 27,65,$ $y_{46} \simeq 1,84,$ $y_{47} \simeq 7,48,$ $y_{48} \simeq 13,19.$

Exemple 2 (suite et fin en image)

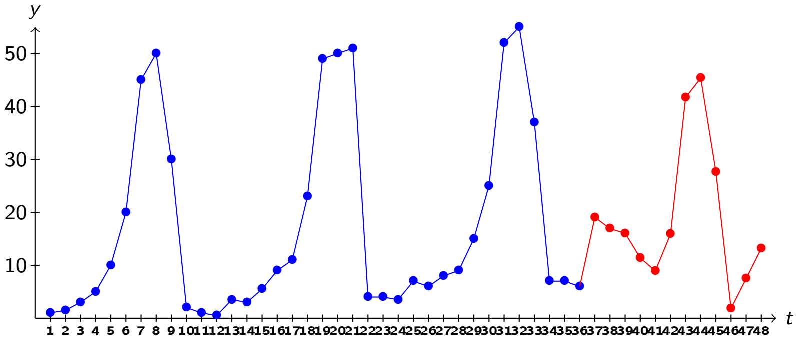

Figure: Consommation d'eau y_t (en Hm^3) au cours du t^{e} mois de l'Exemple 2 (bleu) et prévisions pour l'année 2016 (rouge).