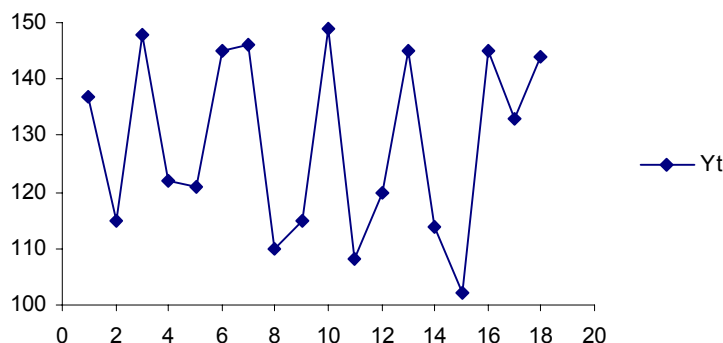


Resolução do problema 3

Começamos por traçar o gráfico da série cronológica



- a) Caracterizar uma série é dizer se ela apresenta **tendência** (se há um desenvolvimento global, ao longo do tempo, de crescimento ou decrescimento dos valores observados) e/ou **sazonalidade** (se existe uma flutuação periódica dos valores que sobem ou descem relativamente ao desenvolvimento global da série, mas com periodicidade *fixa*).

Da observação do gráfico, pode verificar-se que, globalmente, a série não cresce nem decresce, pelo que pode dizer-se que, aparentemente, a série não tem tendência.

Quanto à sazonalidade, verifica-se que, por exemplo, os valores máximos não se repetem em intervalos de tempo fixos (a partir da observação no instante três – Y_3 – encontra-se novo máximo ao fim de 4 intervalos de tempo – Y_7 –, depois são necessários 3 para Y_{10} e sucessivamente outros 3 para Y_{13} , 3 para Y_{16} e 2 para Y_{18} . Verificação semelhante se poderia fazer para os valores mínimos, em que a periodicidade destes valores também não existe, o que permite concluir que a série não tem sazonalidade.

Trata-se, portanto, de uma série que não apresenta tendência nem sazonalidade.

- b) Uma série cronológica deste tipo é modelável pela expressão $Y_t = n_t + \varepsilon_t$ - em que n_t é o nível da série e ε_t uma variável aleatória incontável.

Assim, o modelo de previsão a adoptar será $\hat{Y}_{t+k} = n_t + e_t$ em que k é o número de passos à frente que se pretende fazer a previsão, n_t representa a estimativa que se vai fazer do nível da série e e_t o erro de previsão.

Para a determinação da estimativa do nível da série foram leccionados dois tipos de modelos com diferentes abordagens sobre o peso a atribuir aos valores conhecidos da série cronológica – *amortecimento exponencial simples e médias móveis aritméticas*.

Modelo de amortecimento exponencial simples

Neste modelo, o nível da série no instante t é estimado pela seguinte fórmula recursiva

$$n_t = \alpha Y_t + (1 - \alpha) n_{t-1}$$

que permite actualizar a estimativa do nível a partir da observação feita no instante t – Y_t – incorporando a estimativa feita até ao instante $t-1$ – n_{t-1} através da constante de amortecimento α .

Neste tipo de modelos, é necessário proceder sempre à sua inicialização, isto é, atribuir um valor ao nível inicial n_1 . Um procedimento possível, e frequentemente utilizado, é admitir $n_1 = Y_1$, o que concretizando no nosso caso será

$$n_1 = 137$$

Pode, agora, utilizar-se a fórmula recursiva para calcular os sucessivos valores de n_t até n_{18} (último instante para o qual se conhece o valor da observação) utilizando, como proposto no enunciado $\alpha = 0.10$.

$$n_2 = 0.10 \cdot 115 + (1 - 0.10) \cdot 137 = 134.80$$

$$n_3 = 0.10 \cdot 148 + (1 - 0.10) \cdot 134.80 = 136.12$$

$$n_4 = 0.10 \cdot 122 + (1 - 0.10) \cdot 136.12 = 134.71$$

$$n_5 = 0.10 \cdot 121 + (1 - 0.10) \cdot 134.71 = 133.34$$

$$n_6 = 0.10 \cdot 145 + (1 - 0.10) \cdot 133.34 = 134.50$$

$$n_7 = 0.10 \cdot 146 + (1 - 0.10) \cdot 134.50 = 135.65$$

$$n_8 = 0.10 \cdot 110 + (1 - 0.10) \cdot 135.65 = 133.09$$

$$n_9 = 0.10 \cdot 115 + (1 - 0.10) \cdot 133.09 = 131.28$$

$$n_{10} = 0.10 \cdot 149 + (1 - 0.10) \cdot 131.28 = 133.05$$

$$n_{11} = 0.10 \cdot 108 + (1 - 0.10) \cdot 133.05 = 130.55$$

$$n_{12} = 0.10 \cdot 120 + (1 - 0.10) \cdot 130.55 = 129.49$$

$$n_{13} = 0.10 \cdot 145 + (1 - 0.10) \cdot 129.49 = 131.04$$

$$n_{14} = 0.10 \cdot 114 + (1 - 0.10) \cdot 131.04 = 129.34$$

$$n_{15} = 0.10 \cdot 102 + (1 - 0.10) \cdot 129.34 = 126.60$$

$$n_{16} = 0.10 \cdot 145 + (1 - 0.10) \cdot 126.60 = 128.44$$

$$n_{17} = 0.10 \cdot 133 + (1 - 0.10) \cdot 128.44 = 128.90$$

$$n_{18} = 0.10 \cdot 144 + (1 - 0.10) \cdot 128.90 = 130.41$$

No enunciado do problema é pedida a previsão da “procura do próximo mês”, o mesmo é dizer 1 passo à frente ou seja a $\hat{Y}_{19}$, pelo que a resposta será

$$\hat{Y}_{19} = 130$$

Nota importante – as previsões devem ter sempre o mesmo número de casas decimais que as observações, razão porque a previsão é 130 e *não* 130.41 !

É pedido, também, no enunciado que seja determinado o intervalo de confiança a 99%, que (admitindo que o erro de previsão - e_t - segue uma distribuição normal com média 0 e uma variância igual ao erro quadrático médio – EQM) é definido como

$$\hat{Y}_t \pm z_{\alpha/2} \cdot \sqrt{EQM}$$

em que $z_{\alpha/2}$, como habitualmente, é o valor da variável normal padrão tal que $P(Z > z_{\alpha/2}) = \alpha/2$ (sendo $1 - \alpha$ o nível de confiança definido, no nosso caso 99%, pelo que $\alpha = 1\%$) e o EQM dado pela expressão $\Sigma (Y_t - \hat{Y}_t)^2 / N$.

Consultando as tabelas da normal padrão pode verificar-se que o valor de $z_{\alpha/2}$ para o qual $P(Z > z_{\alpha/2}) = \alpha/2 = 0.005$ é de 2.58.

A determinação do EQM tem de ter em atenção o número de passos à frente da previsão que se está a fazer. No nosso caso, como estamos a prever 1 passo à frente, tal significa que o erro de previsão $e_t = Y_t - \hat{Y}_t$ será dado para cada instante t pela expressão $Y_t - n_{t-1}$.

Por outro lado, estando a utilizar o modelo de amortecimento exponencial simples, os primeiros níveis n_t calculados poderão estar algo distorcidos dado que dependem fortemente do erro, eventualmente cometido, aquando da inicialização. Por isso, é aconselhável só começar a determinar os erros de previsão alguns instantes após a inicialização, mas não exagerando nesta prudência, de modo a que haja, ainda, um número significativo de parcelas para o cálculo do EQM.

Admita-se, então, para o nosso caso que o primeiro nível que já não é afectado pelo erro de inicialização é n_7 , pelo que os erros de previsão que vamos considerar serão:

$$\begin{aligned} e_8 &= Y_8 - n_7 = 110 - 135.65 = -25.65 \\ e_9 &= Y_9 - n_8 = 115 - 133.09 = -18.09 \\ e_{10} &= Y_{10} - n_9 = 149 - 131.28 = 17.72 \\ &\vdots \\ e_{18} &= Y_{18} - n_{17} = 144 - 128.90 = 15.10 \end{aligned}$$

pelo que

$$EQM = \frac{(-25.65)^2 + (-18.09)^2 + (17.72)^2 + \dots + (15.10)^2}{11} = 354.87$$

Assim o intervalo de confiança da previsão será

$$130.41 \pm 2.58 \cdot \sqrt{354.87} = 130.41 \pm 48.60 \Rightarrow 81.81 \text{ e } 179.01$$

No entanto, tendo em atenção a nota importante da página anterior o intervalo de confiança a 99% desta previsão é

[82, 179]

Apresenta-se a seguir um quadro resumo da resolução apresentada

t	Y_t	n_t	$\hat{Y}_t$	e_t
1	137	137		
2	115	134.80	137	
3	148	136.12	134.80	
4	122	134.71	136.12	
5	121	133.34	134.71	
6	145	134.50	133.34	
7	146	135.65	134.50	
8	110	133.09	135.65	-25.65
9	115	131.28	133.09	-18.09

t	Y_t	n_t	$\hat{Y}_t$	e_t
10	149	133.05	131.28	17.72
11	108	130.55	133.05	-25.05
12	120	129.49	130.55	-10.55
13	145	131.04	129.49	15.51
14	114	129.34	131.04	-17.04
15	102	126.60	129.34	-27.34
16	145	128.44	126.60	18.40
17	133	128.90	128.44	4.56
18	144	130.41	128.90	15.10
19			130	

Modelo de médias móveis aritméticas

Neste modelo, o nível da série no instante t é estimado pela seguinte expressão:

$$n_t = \frac{Y_{t-N+1} + Y_{t-N+2} + \dots + Y_{t-1} + Y_t}{N}$$

em que N é o comprimento da média móvel.

Assim, para $N = 5$, como indicado no enunciado, será:

$$n_5 = \frac{Y_1 + Y_2 + Y_3 + Y_4 + Y_5}{5} = \frac{137 + 115 + 148 + 122 + 121}{5} = 128.6$$

$$n_6 = \frac{115 + 148 + 122 + 121 + 145}{5} = 130.2$$

$$n_7 = \frac{148 + 122 + 121 + 145 + 146}{5} = 136.4$$

:

$$n_{18} = \frac{114 + 102 + 145 + 133 + 144}{5} = 127.6$$

Donde, conforme já se explicou, para o modelo anterior, a previsão para a procura “no próximo mês” será:

$$\hat{Y}_{19} = 128$$

O cálculo do EQM para este modelo seguirá um processo semelhante ao descrito anteriormente, sem ter que haver, no entanto, a preocupação então havida por causa do eventual erro provocado pela inicialização, que aqui não existe.

Então

$$e_6 = Y_6 - n_7 = 145 - 128.6 = 16.4$$

$$e_7 = Y_7 - n_8 = 146 - 130.2 = 15.8$$

$$e_8 = Y_8 - n_9 = 110 - 136.4 = -26.2$$

:

$$e_{18} = Y_{18} - n_{17} = 144 - 127.8 = 16.2$$

pelo que

$$EQM = \frac{(16.4)^2 + (15.8)^2 + (-26.4)^2 + \dots + (16.2)^2}{13} = 385.52$$

Assim o intervalo de confiança da previsão será

$$127.6 \pm 2.58 * \sqrt{385.52} = 127.6 \pm 50.6 \Rightarrow 77.0 \text{ e } 178.2$$

pelo que, tendo em atenção a nota importante já referida, o intervalo de confiança a 99% desta previsão é

$$[77, 178]$$

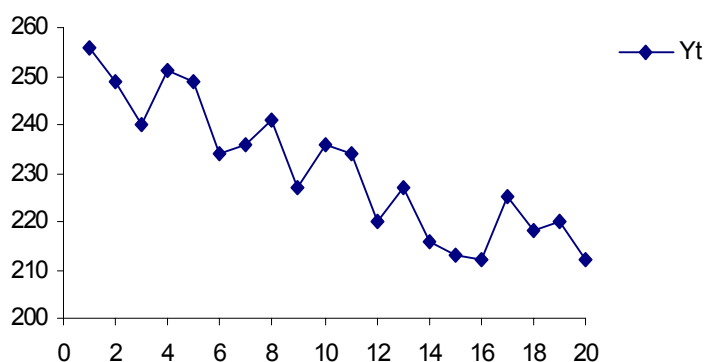
Apresenta-se a seguir um quadro resumo da resolução apresentada

t	Y_t	n_t	$\hat{Y}_t$	e_t
1	137			
2	115			
3	148			
4	122			
5	121	128.6		
6	145	130.2	128.6	16.4
7	146	136.4	130.2	15.8
8	110	128.8	136.4	-26.4
9	115	127.4	128.8	-13.8

t	Y_t	n_t	$\hat{Y}_t$	e_t
10	149	133.0	127.4	21.6
11	108	125.6	133.0	-25.0
12	120	120.4	125.6	-5.6
13	145	127.4	120.4	24.6
14	114	127.2	127.4	-13.4
15	102	117.8	127.2	-25.2
16	145	125.2	117.8	27.2
17	133	127.8	125.2	7.8
18	144	127.6	127.8	16.2
19			128	

Resolução do problema 6

Começamos por traçar o gráfico da série cronológica



a) Da observação do gráfico, pode verificar-se que, globalmente, a série decresce, pelo que pode dizer-se que, a série apresenta tendência linear decrescente.

Quanto à sazonalidade, verifica-se que não existe nenhum padrão fixo de flutuações em torno da tendência global de decrescimento da série, o que permite concluir que a série não tem sazonalidade.

Trata-se, portanto, de uma série que apresenta tendência linear decrescente e não tem sazonalidade.

b) Uma série cronológica deste tipo é modelável pela expressão $Y_t = T_t + \varepsilon_t$ - em que T_t representa a tendência e ε_t uma variável aleatória incontrolável.

Assim, o modelo de previsão a adoptar será $\hat{Y}_t = \hat{T}_t + e_t$ em que $\hat{T}_t$ representa a estimativa que se vai fazer da tendência da série e e_t o erro de previsão.

Para estimar a previsão de uma série só com tendência foram leccionados dois tipos de modelos- *Holt* e *regressão linear simples*.

Modelo de Holt

Neste modelo, a previsão para o instante t será estimada pela expressão

$$\hat{Y}_{t+k} = n_t + k \cdot b_t$$

Em que k representa o número de passos à frente para o qual se vai fazer a previsão, n_t é o nível e b_t a tendência, valores que se obtêm a partir das seguintes fórmulas recursivas

$$n_t = \alpha Y_t + (1 - \alpha) (n_{t-1} + b_{t-1})$$

$$b_t = \beta (n_t - n_{t-1}) + (1 - \beta) b_{t-1}$$

que permitem actualizar as estimativas do nível e da tendência a partir da observação feita no instante $t - Y_t$ - incorporando as estimativas feitas até ao instante $t-1$ - n_{t-1} e b_{t-1} através das constantes de amortecimento α e β .

Neste tipo de modelos, é necessário proceder sempre à sua inicialização, isto é, atribuir valores iniciais a n_1 , n_2 e b_2 . Um procedimento possível, e frequentemente utilizado, é admitir $n_1 = Y_1$, $n_2 = Y_2$ e $b_2 = Y_2 - Y_1$, o que concretizando no nosso caso dará

$$n_1 = 256 \quad n_2 = 249 \quad b_2 = 256 - 249 = -7$$

Podem, agora, utilizarem-se as fórmulas recursivas para calcular os sucessivos valores de n_3 e de b_3 até n_{20} e de b_{20} (último instante para o qual se conhece o valor da observação) utilizando, como proposto no enunciado $\alpha = 0.4$ e $\beta = 0.3$

$n_3 = 0.40 \cdot 240 + (1 - 0.40) \cdot [(249 + (-7))] = 241.2$	$b_3 = 0.3 \cdot (241.2 - 249) + (1 - 0.3) \cdot (-7) = -7.2$
$n_4 = 0.40 \cdot 251 + (1 - 0.40) \cdot [(241.2 + (-7.2))] = 240.8$	$b_4 = 0.3 \cdot (240.8 - 241.2) + (1 - 0.3) \cdot (-7.2) = -5.2$
$\vdots$	$\vdots$
$\vdots$	$\vdots$
$n_{19} = 0.40 \cdot 220 + (1 - 0.40) \cdot [(215.5 + (-0.9))] = 216.8$	$b_{19} = 0.3 \cdot (216.8 - 215.5) + (1 - 0.3) \cdot (-0.9) = -0.3$
$n_{20} = 0.40 \cdot 212 + (1 - 0.40) \cdot [(216.8 + (-0.3))] = 214.7$	$b_{20} = 0.3 \cdot (214.7 - 216.8) + (1 - 0.3) \cdot (-0.3) = -0.8$

Deste modo, o volume de vendas para a próxima semana cuja previsão é pedida no enunciado (1 passo à frente, portanto) será

$$\hat{Y}_{21} = n_{20} + 1 \cdot b_{20} = 214.7 + 1 \cdot (-0.8) = 213.9 \Rightarrow \mathbf{214}$$

Apresenta-se a seguir um quadro com todos os cálculos já efectuados, bem como os que irão ser necessários para a determinação do intervalo de confiança a 90%, ou seja o cálculo do EQM.

t	Y_t	n_t	b_t	$\hat{Y}_t$	e_t
1	256	256			
2	249	249	-7		
3	240	241.2	-7.2		
4	251	240.8	-5.2		
5	249	240.9	-3.6		
6	234	236.0	-4.0		
7	236	233.6	-3.5	232.0	4.0
8	241	234.5	-2.2	230.1	10.9
9	227	230.2	-2.8	232.3	-5.3
10	236	230.8	-1.8	227.3	8.7

t	Y_t	n_t	b_t	$\hat{Y}_t$	e_t
11	234	231.0	-1.2	229.0	5.0
12	220	225.9	-2.4	229.8	-9.8
13	227	224.9	-2.0	223.5	3.5
14	216	220.2	-2.8	223.0	-7.0
15	213	215.6	-3.3	217.4	-4.4
16	212	212.2	-3.4	212.3	-0.3
17	225	215.3	-1.4	208.8	16.2
18	218	215.5	-0.9	213.9	4.1
19	220	216.8	-0.3	214.6	5.4
20	212	214.7	-0.8	216.5	-4.5
21				214	

Face ao erro eventualmente cometido com o processo de inicialização (ver o comentário feito aquando da resolução do problema anterior – pag 13), admitamos que para o cálculo do EQM só se vão considerar os erros de previsão a partir do instante 7.

Assim,

$$e_7 = Y_7 - \hat{Y}_7 = Y_7 - (n_6 + b_6) = 234 - [236.0 + (-4.0)] = 234 - 232.0 = 4.0$$

$$e_8 = Y_8 - \hat{Y}_8 = Y_8 - (n_7 + b_7) = 241 - [233.6 + (-3.5)] = 241 - 23.1 = 10.9$$

:

$$e_{20} = Y_{20} - \hat{Y}_{20} = Y_{20} - (n_{19} + b_{19}) = 212 - [216.8 + (-0.3)] = 212 - 216.5 = -4.5$$

pelo que

$$EQM = \frac{(4.0)^2 + (10.9)^2 + \dots + (-4.5)^2}{14} = 54.69$$

Assim o intervalo de confiança da previsão, recordando que para um nível de confiança de 90% o valor de $z_{\alpha/2}$ é 1.645, será

$$213.9 \pm 1.645 * \sqrt{54.69} = 213.9 \pm 12.2 \Rightarrow 201.7 \text{ e } 226.1$$

Como a série cronológica não tem nenhuma casa decimal, a resposta deverá ser:

[202, 226]

Regressão linear simples

Neste modelo, a previsão para o instante t será estimada pela expressão

$$\hat{Y}_t = \hat{a} + \hat{b}t$$

em que $\hat{a}$ e $\hat{b}$ são, respectivamente, a ordenada na origem e o declive da recta que melhor se ajusta a um conjunto de pontos (t, Y_t) aplicando o critério da minimização da soma do quadrado dos erros.(ou seja $\text{Min } \sum e_t^2$)

As expressões para a determinação destes dois parâmetros são:

$$\hat{b} = \frac{\sum t * Y_t - n * \bar{t} * \bar{Y}_t}{\sum t^2 - n * \bar{t}^2}$$

$$\hat{a} = \bar{Y}_t - \hat{b} * \bar{t}$$

Para aplicar estas expressões há que, naturalmente, efectuar os cálculos dos somatórios e médias que delas constam:

$n = 20$ (pontos)

$\sum t = 210$ ($\frac{1+20}{2} * 20$ - soma do primeiro mais o último a dividir por 2, vezes o nº de parcelas)

$\bar{t} = 10.5$ ($210/20$)

$\sum t^2 = 2870$ calculável, expeditamente, pela fórmula $\frac{20}{3} * (20 + 0.5) * (20 + 1)$

$\sum Y_t = 256 + 249 + \dots + 220 + 212 = 4616$

$\bar{Y}_t = 230.8$ ($4616/20$)

$\sum t Y_t = 1 * 256 + 2 * 249 + \dots + 19 * 220 + 20 * 212 = 47055$

Assim,

$$\hat{b} = \frac{47055 - 20 * 10.5 * 230.8}{2870 - 20 * 10.5^2} = -2.125$$

$$\hat{a} = 230.8 - (-2.12) * 10.5 = 253.11$$

pelo que o modelo de previsão que se vai adoptar, será

$$\hat{Y}_t = 253.11 - 2.125 * t$$

Logo, teremos para a previsão para o instante 21

$$\hat{Y}_{21} = 253.11 - 2.125 * 21 = 208.49 \Rightarrow \mathbf{208}$$

Dado que a estimativa feita a partir da regressão linear segue uma distribuição t de Student com $n - 2$ graus de liberdade, a expressão para o cálculo do intervalo de confiança da previsão é

$$\hat{Y}_t \pm t_{\alpha/2, n-2} \cdot \hat{\sigma}_\varepsilon \sqrt{\frac{1}{n} + \frac{(t - \bar{t})^2}{\sum t^2 - n * \bar{t}^2}}$$

onde

$t_{\alpha/2, n-2}$ é o valor que resulta da consulta da tabela da distribuição t de Student consoante o nível de confiança $(1 - \alpha)$ e o número de graus de liberdade $(n-2)$

$\hat{\sigma}_\varepsilon$ é calculado pela expressão $\sqrt{\frac{\sum e_t^2}{n-2}}$

No caso concreto deste problema em que o nível de confiança é de 90%, virá

$$t_{0.95, 18} = 1.734$$

e os erros de previsão – $e_t = Y_t - \hat{Y}_t$ – serão

$$e_1 = 256 - (253.11 - 2.125 \cdot 1) = 256 - 251.0 = 5.0$$

$$e_2 = 249 - (253.11 - 2.125 \cdot 2) = 249 - 248.9 = 0.1$$

:

$$e_{19} = 220 - (253.11 - 2.125 \cdot 19) = 220 - 212.7 = 7.3$$

$$e_{20} = 212 - (253.11 - 2.125 \cdot 20) = 212 - 210.6 = 1.4$$

pelo que

$$\sum e_t^2 = (5.0)^2 + (0.1)^2 + \dots + (7.3)^2 + (1.4)^2 = 668.84 \quad \text{e, portanto}$$

$$\hat{\sigma}_\varepsilon = \sqrt{\frac{668.84}{20-2}} = 6.10$$

Podendo obter-se, finalmente, o intervalo de confiança a 90% da previsão feita

$$208.49 \pm 1.734 \cdot 6.10 \cdot \sqrt{\frac{1}{20} + \frac{(21-10.5)^2}{2870-20 \cdot 10.5^2}} = 208.49 \pm 4.91 \Rightarrow 203.6 \text{ e } 213.4$$

ou seja **[204, 213]**

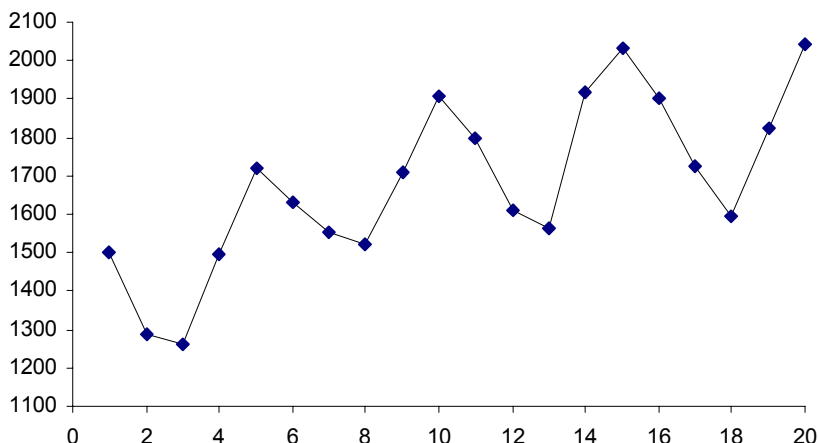
Apresenta-se a seguir um quadro com todos os cálculos efectuados

t	Y_t	$\hat{Y}_t$	e_t	t	Y_t	$\hat{Y}_t$	e_t
1	256	251.0	5.0	11	234	229.7	4.3
2	249	248.9	0.1	12	220	227.6	-7.6
3	240	246.7	-6.7	13	227	225.5	1.5
4	251	244.6	6.4	14	216	223.4	-7.4
5	249	242.5	6.5	15	213	221.2	-8.2
6	234	240.4	-6.4	16	212	219.1	-7.1
7	236	238.2	-2.2	17	225	217.0	8.0
8	241	236.1	4.9	18	218	214.9	3.1
9	227	234.0	-7.0	19	220	212.7	7.3
10	236	231.9	4.1	20	212	210.6	1.4
				21		208	

Resolução do problema 13

a) Face ao gráfico representativo da série,

naturalmente se conclui tratar-se de uma série com tendência crescente (que vamos assumir



como linear) e com sazonalidade de período 5.

Com efeito, a série apresenta, globalmente, um desenvolvimento crescente e a sazonalidade é facilmente reconhecível dado que de 5 em 5 instantes repete-se o mesmo tipo de flutuação em torno do crescimento global da série (máximos nos instantes 5, 10 e 15, mínimos nos instantes 3, 8, 13 e 18, etc) ou seja, há um padrão *fixo* de comportamento.

b) Para modelar séries com sazonalidade que tenham (ou não) tendência, foram leccionados dois tipos de métodos – *Holt-Winters* e *Decomposição Clássica* – qualquer deles com 2 modelos – *aditivo* e *multiplicativo*.

Representando por S_t a componente sazonal, de um modo geral, pode dizer-se que a uma série com sazonalidade é-lhe aplicável o modelo

- *aditivo* se for do tipo $Y_t = T_t + S_t + \varepsilon_t$ (se tiver tendência) ou $Y_t = n_t + S_t + \varepsilon_t$ (se não tiver tendência) e corresponde às séries que graficamente se desenvolvem ao longo de duas linhas paralelas,

- *multiplicativo* se for do tipo $Y_t = T_t * S_t + \varepsilon_t$ (se tiver tendência) e cuja representação gráfica se desenvolve ao longo de duas linhas divergentes (se a tendência for crescente) ou convergentes (se a tendência for decrescente)

$$Y_t = n_t * S_t + \varepsilon_t \text{ (se não tiver tendência).}$$

Se nalgumas séries é claro qual o modelo a aplicar (ver problema 11 onde é evidente a aplicabilidade do modelo multiplicativo), outras há em que a modelação mais “correcta” só pode ser feita utilizando outros processos (por exemplo, o modelo que tiver menor EQM).

No caso vertente do problema 13 de que estamos tratando, vamos admitir que o modelo que se aplica é o aditivo, pelo que o vamos abordar utilizando as duas metodologias.

Holt-Winters (modelo aditivo)

Neste modelo, a previsão para o instante t será estimada pela expressão

$$\hat{Y}_{t+k} = n_t + k \cdot b_t + f_{t+k-s}$$

Em que s representa o período sazonal, k o número de passos à frente para o qual se vai fazer a previsão, n_t é o nível, b_t a tendência e f_t o índice sazonal, valores que se obtêm a partir das seguintes fórmulas recursivas

$$n_t = \alpha (Y_t - f_{t-s}) + (1 - \alpha) (n_{t-1} + b_{t-1})$$

$$b_t = \beta (n_t - n_{t-1}) + (1 - \beta) b_{t-1}$$

$$f_t = \gamma (Y_t - n_t) + (1 - \gamma) f_{t-s}$$

que permitem actualizar as estimativas do nível, da tendência e dos índices sazonais a partir da observação feita no instante $t - Y_t$ - incorporando as estimativas feitas até ao instante $t-1 - n_{t-1}$ e b_{t-1} e do índice sazonal feita um período antes - f_{t-s} - através das constantes de amortecimento α , β e γ .

Neste tipo de modelos, é necessário proceder sempre à sua inicialização, isto é, atribuir valores iniciais de n_1 a n_s , b_2 , e de f_1 a f_s .

Um procedimento possível, e frequentemente utilizado, é fazer a regressão dos s primeiros pontos da série (obtendo-se um valor para a ordenada na origem (c) e um declive (d) e atribuir a

$$n_s = c + dt \quad \text{com } t \text{ de } 1 \text{ a } s,$$

$$b_s = d$$

$$f_t = Y_t - n_t \quad \text{com } t \text{ de } 1 \text{ a } s$$

No nosso caso, após fazer a regressão dos 5 primeiros pontos, obter-se-ia $c = 1260.5$ e $d = 64.5$, donde

$$n_1 = 1260.5 + 64.5 = 1325.0 \quad n_2 = 1260.5 + 2 \cdot 64.5 = 1389.5 \quad n_3 = 1454.0$$

$$n_4 = 1518.5 \quad n_5 = 1583.0$$

$$b_2 = 64.5$$

$$f_1 = 1503 - 1325.0 = 178.0 \quad f_2 = 1287 - 1389.5 = -102.5 \quad f_3 = 1262 - 1454.0 = -192.0$$

$$f_4 = 1498 - 1518.5 = -20.5 \quad f_5 = 1720 - 1583.0 = 137.0$$

Podem, agora, utilizarem-se as fórmulas recursivas para calcular os sucessivos valores de n_t , de b_t e f_t até $t = 20$ (último instante para o qual se conhece o valor da observação) utilizando, como proposto no enunciado $\alpha = 0.79$, $\beta = 0.10$ e $\gamma = 0.99$.

$$n_6 = 0.79 \cdot (1630 - 178.0) + (1 - 0.79) \cdot (1583.0 + 64.5) = 1493.1$$

$$b_6 = 0.10 \cdot (1493.1 - 1583.0) + (1 - 0.10) \cdot 64.5 = 49.06$$

$$f_6 = 0.99 \cdot (1630 - 1493.1) + (1 - 0.99) \cdot 178.0 = 137.36$$

$$n_7 = 0.79 \cdot [1551 - (-102.5)] + (1 - 0.79) \cdot (1493.1 + 49.06) = 1630.1$$

$$b_7 = 0.10 \cdot (1630.1 - 1493.1) + (1 - 0.10) \cdot 49.06 = 57.86$$

$$f_7 = 0.99 \cdot (1551 - 1630.1) + (1 - 0.99) \cdot (-102.5) = -79.34$$

e sucessivamente até

$$n_{20} = 0.79 \cdot (2042 - 118.31) + (1 - 0.79) \cdot (1818.8 + 23.54) = 1906.6$$

$$b_{20} = 0.10 \cdot (1630.1 - 1493.1) + (1 - 0.10) \cdot 23.54 = 29.97$$

$$f_{20} = 0.99 \cdot (2042 - 1906.6) + (1 - 0.99) \cdot 118.31 = 135.21$$

Apresenta-se no quadro abaixo os cálculos já feitos, bem como os que irão ser efectuados a seguir quer para as previsões quer para a determinação do intervalo de confiança a 90%. Com o intuito de facilitar a identificação dos instantes que pertencem à mesma família, introduziu-se uma coluna com essa indicação de I a V.

t	Y_t	n_t	b_t	f_t		$\hat{Y}_t$	e_t
1	1503	1325.0		178.0	I		
2	1287	1389.5		-102.5	II		
3	1262	1454.0		-192.0	III		
4	1498	1518.5		-20.5	IV		
5	1720	1583.0	64.50	137.0	V		
6	1630	1493.1	49.06	137.36	I		
7	1551	1630.1	57.86	-79.34	II		
8	1523	1709.3	59.99	-186.38	III		
9	1710	1738.7	56.92	-28.57	IV		
10	1907	1775.4	54.90	131.68	V		
11	1798	1696.3	41.50	102.09	I	1967.63	-169.63
12	1611	1700.3	37.76	-89.20	II	1658.43	-47.43
13	1565	1748.6	38.81	-183.61	III	1551.68	13.32
14	1917	1912.4	51.31	4.32	IV	1758.82	158.18
15	2031	1912.8	46.22	118.31	V	2095.34	-64.34
16	1902	1833.3	33.65	69.00	I	2061.14	-159.14
17	1723	1823.7	29.32	-100.59	II	1777.78	-54.78
18	1597	1795.8	23.60	-198.67	III	1669.42	-72.42
19	1823	1818.8	23.54	4.16	IV	1823.74	-0.74
20	2042	1906.6	29.97	135.21	V	1960.69	81.31

Deste modo, a previsão da produção para os próximos 5 anos será:

$$\hat{Y}_{21} = n_{20} + 1 \cdot b_{20} + f_{16} = 1906.6 + 1 \cdot 29.97 + 69.00 = 2005.59 \quad \Rightarrow \mathbf{2006}$$

$$\hat{Y}_{22} = n_{20} + 2 \cdot b_{20} + f_{17} = 1906.6 + 2 \cdot 29.97 + (-100.59) = 1865.96 \quad \Rightarrow \mathbf{1866}$$

$$\hat{Y}_{23} = n_{20} + 3 \cdot b_{20} + f_{18} = 1906.6 + 3 \cdot 29.97 + (-198.67) = 1797.85 \quad \Rightarrow \mathbf{1798}$$

$$\hat{Y}_{24} = n_{20} + 4 \cdot b_{20} + f_{19} = 1906.6 + 4 \cdot 29.97 + 4.16 = 2030.65 \quad \Rightarrow \mathbf{2031}$$

$$\hat{Y}_{25} = n_{20} + 5 \cdot b_{20} + f_{20} = 1906.6 + 5 \cdot 29.97 + 135.21 = 2191.67 \quad \Rightarrow \mathbf{2192}$$

Para a definição dos intervalos de confiança, é necessário calcular o EQM, o que implica determinar os erros de previsão dos instantes passados.

Face ao erro eventualmente cometido com o processo de inicialização (ver o comentário feito aquando da resolução do problema 3 – pag 13), admitamos que para o cálculo do EQM só se vão considerar os erros de previsão a partir do instante 11. De notar que para este tipo de modelo, o resguardo a dar relativamente à inicialização deve ser maior, pois os índices de sazonalidade são actualizados de s em s instantes (no nosso caso $s = 5$).

Os erros de previsão seriam (ver quadro):

$$e_{11} = Y_{11} - \hat{Y}_{11} = 1798 - (1775.4 + 54.90 + 137.36) = 1798 - 1967.63 = -169.63$$

$$e_{12} = Y_{12} - \hat{Y}_{12} = 1611 - (1696.3 + 41.50 - 79.34) = 1611 - 1658.43 = -47.43$$

$$\vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots$$

$$e_{20} = Y_{20} - \hat{Y}_{20} = 2042 - (1818.8 + 23.54 + 118.31) = 2042 - 1960.69 = 81.31$$

pelo que o erro quadrático médio seria:

$$EQM = \frac{(-169.63)^2 + (-47.43)^2 + \dots + (81.31)^2}{10} = 10054.4$$

Assim, os intervalos de confiança das previsões, recordando que para um nível de confiança de 90% o valor de $z_{\alpha/2}$ é 1.645, será

$$2005.59 \pm 1.645 \cdot \sqrt{10054.4} = 2005.59 \pm 164.95 \Rightarrow 1840.6 \text{ e } 2170.5 \Rightarrow \mathbf{[1841, 2171]}$$

$$1865.96 \pm 1.645 \cdot \sqrt{10054.4} = 1865.96 \pm 164.95 \Rightarrow 1701.0 \text{ e } 2030.9 \Rightarrow \mathbf{[1701, 2031]}$$

$$1797.85 \pm 1.645 \cdot \sqrt{10054.4} = 1797.85 \pm 164.95 \Rightarrow 1632.9 \text{ e } 1962.8 \Rightarrow \mathbf{[1633, 1963]}$$

$$2030.65 \pm 1.645 \cdot \sqrt{10054.4} = 2030.65 \pm 164.95 \Rightarrow 1865.7 \text{ e } 2195.6 \Rightarrow \mathbf{[1866, 2196]}$$

$$2191.67 \pm 1.645 \cdot \sqrt{10054.4} = 2191.67 \pm 164.95 \Rightarrow 2026.7 \text{ e } 2356.6 \Rightarrow \mathbf{[2027, 2357]}$$

Decomposição clássica (modelo aditivo)

Neste modelo a previsão para o instante t é dada pela expressão

$$\hat{Y}_t = \hat{T}_t + \hat{S}_t + e_t \qquad \text{no caso de a série ter tendência}$$

$$\hat{Y}_t = n_t + \hat{S}_t + e_t \qquad \text{no caso de a série não ter tendência}$$

Em que $\hat{T}_t$ ou n_t são os estimadores da tendência ou do nível, consoante o caso e $\hat{S}_t$ representa os estimadores dos s índices de sazonalidade.

No método da decomposição clássica, começa-se por tentar eliminar a sazonalidade, para o que se utilizam *médias móveis centradas* com um comprimento igual (ou múltiplo) ao período de sazonalidade s .

Como no nosso caso o período de sazonalidade é 5, as médias móveis a aplicar serão

$$M_t = \frac{Y_{t-2} + Y_{t-1} + Y_t + Y_{t+1} + Y_{t+2}}{5}$$

(De notar, que a expressão para as médias móveis de comprimento par têm uma fórmula ligeiramente diferente).

Assim as médias móveis que temos de calcular serão:

$$M_3 = \frac{Y_1 + Y_2 + Y_3 + Y_4 + Y_5}{5} = \frac{1503 + 1287 + 1262 + 1498 + 1720}{5} = 1454.0$$

$$M_4 = \frac{Y_2 + Y_3 + Y_4 + Y_5 + Y_6}{5} = \frac{1287 + 1262 + 1498 + 1720 + 1630}{5} = 1479.4$$

: : : :

$$M_{18} = \frac{Y_{16} + Y_{17} + Y_{18} + Y_{19} + Y_{20}}{5} = \frac{1902 + 1723 + 1597 + 1823 + 2042}{5} = 1817.4$$

É sobre a série cronológica das médias móveis centradas que, teoricamente, isolaram a tendência, (isto é, $M_t \cong T_t$) que vamos aplicar a regressão linear simples para encontrar os estimadores $\hat{a}$ e $\hat{b}$ dessa tendência (no caso da série Y_t não ter tendência, será o nível n_t que tem de ser estimado).

O quadro abaixo sistematiza os cálculos até aqui efectuados, como todos os que sequencialmente se irão efectuar.

t	Y_t	M_t	$\hat{T}_t$	S_t		S'_t	$\hat{Y}_t$	e_t
1	1503		1451.7	51.3	I	64.9	1524.5	-21.5
2	1287		1477.3	-190.3	II	-125.8	1366.6	-79.6
3	1262	1454.0	1502.8	-240.8	III	-207.6	1320.3	-58.3
4	1498	1479.4	1528.4	-30.4	IV	17.1	1547.5	-49.5
5	1720	1532.2	1553.9	166.1	V	179.5	1755.1	-35.1
6	1630	1584.4	1579.4	50.6	I	-71.9	1652.2	-22.2
7	1551	1626.8	1605.0	-54.0	II	595.0	1494.4	56.6
8	1523	1664.2	1630.5	-107.5	III		1448.0	75.0
9	1710	1697.8	1656.1	53.9	IV		1675.2	34.8
10	1907	1709.8	1681.6	225.4	V		1882.8	24.2
11	1798	1718.2	1707.2	90.8	I	72.8	1780.0	18.0
12	1611	1759.6	1732.7	-121.7	II	-110.6	1622.1	-11.1
13	1565	1784.4	1758.2	-193.2	III	-182.5	1575.7	-10.7
14	1917	1805.2	1783.8	133.2	IV	19.1	1802.9	114.1
15	2031	1827.6	1809.3	221.7	V	201.2	2010.6	20.4
16	1902	1834.0	1834.9	67.1	I	0.00	1907.7	-5.7
17	1723	1815.2	1860.4	-137.4	II		1749.8	-26.8
18	1597	1817.4	1886.0	-289.0	III		1703.4	-106.4
19	1823		1911.5	-88.5	IV		1930.6	-107.6
20	2042		1937.0	105.0	V		2138.3	-96.3

Assim, correlacionando, pela regressão linear simples, os 16 pontos (t , M_t) desde o ponto (3, 1454.0) até (18, 1817.4), calculando, portanto, os valores de $\Sigma t = 168$, $\bar{t} = 10.5$, $\Sigma t^2 = 2104$, $\Sigma M_t = 27110.2$, $\bar{M}_t = 1694.4$ e de $\Sigma t \cdot M_t = 293341.4$, obter-se á:

$$\hat{a} = 1426.2 \quad \text{e} \quad \hat{b} = 25.54$$

o que permite estimar os valores da tendência a partir da expressão $\hat{T}_t = 1426.2 + 25.54 \cdot t$, valores que no quadro estão indicados na quarta coluna.

Estimada a componente tendência, podem agora calcular-se os valores dos índices de sazonalidade através da diferença $Y_t - \hat{T}_t$, que são apresentados na quinta coluna do quadro anterior. Exemplificando:

$$S_1 = 1503 - 1451.7 = 51.3, \quad S_2 = 1287 - 1477.3 = -190.3, \quad \text{até} \quad S_{20} = 2042 - 1937.0 = 105.0$$

(Se o modelo adoptado fosse o multiplicativo, em vez de ser $Y_t - \hat{T}_t$, seria $Y_t / \hat{T}_t$)

Sendo o período de sazonalidade igual a 5, pertencerão à mesma família os índices separados de 5 instantes, isto é, S_1 (51.3), S_6 (50.6), S_{11} (90.8) e S_{16} (67.1) são da família que se denominou de I (vide quadro), do mesmo modo S_2 (-190.3), S_7 (-54.0), S_{12} (-121.7), e S_{17} (-137.4) pertencem à família II, etc.

A genese do método da decomposição clássica implica a procura do padrão médio dos índices de sazonalidade, pelo que há que determinar o índice médio de cada uma das famílias.

$$\bar{S}_I = \frac{51.3 + 50.6 + 90.8 + 67.1}{4} = 64.9$$

$$\bar{S}_{II} = \frac{-190.3 - 54.0 - 121.7 - 137.4}{4} = -125.8$$

$$\bar{S}_{III} = \frac{-240.8 - 107.5 - 193.2 - 289.0}{4} = -207.6$$

$$\bar{S}_{IV} = \frac{-30.4 + 53.9 + 133.2 - 88.5}{4} = 17.1$$

$$\bar{S}_V = \frac{166.1 + 225.4 + 221.7 + 105.0}{4} = 179.5$$

Por definição, no modelo aditivo do método da decomposição clássica, a soma dos índices de sazonalidade deve ser igual a zero, pelo que se deve utilizar a seguinte fórmula correctiva (para que a soma seja efectivamente igual a zero)

$$S'_k = S_k - |S_k| * \frac{\sum S_j}{\sum |S_j|}$$

(No modelo multiplicativo, a soma dos índices de sazonalidade tem de ser igual ao período de sazonalidade – s – e a fórmula correctiva é $S'_k = S_k * \frac{s}{\sum S_j}$).

Aplicando ao caso do problema de vimos tratando, dado que

$$\sum S_j = 64.9 - 125.8 - 207.6 + 17.1 + 179.5 = -71.9$$

$$\sum |S_j| = 64.9 + 125.8 + 207.6 + 17.1 + 179.5 = 595.0$$

virá,

$$S'_I = 64.9 - 64.9 * \frac{(-71.9)}{595.0} = 72.8$$

$$S'_{II} = -125.8 - 125.8 * \frac{(-71.9)}{595.0} = -110.6$$

$$S'_{III} = -207.6 - 207.6 * \frac{(-71.9)}{595.0} = -182.5$$

$$S'_{IV} = 17.1 - 17.1 * \frac{(-71.9)}{595.0} = 19.1$$

$$S'_{V} = 179.5 - 179.5 * \frac{(-71.9)}{595.0} = 201.2$$

Estamos, finalmente, aptos a definir o modelo de previsão a utilizar

$$\hat{Y}_t = 1426.20 + 25.54 * t + \begin{cases} +72.8 & \text{se o instante } t \text{ a prever pertencer à família I} \\ -110.6 & \text{se o instante } t \text{ a prever pertencer à família II} \\ -182.5 & \text{se o instante } t \text{ a prever pertencer à família III} \\ +19.1 & \text{se o instante } t \text{ a prever pertencer à família IV} \\ +201.2 & \text{se o instante } t \text{ a prever pertencer à família V} \end{cases}$$

Assim, as previsões serão:

$$\hat{Y}_{21} = 1426.20 + 25.54 * 21 + 72.8 = 2035.4 \Rightarrow \mathbf{2035}$$

$$\hat{Y}_{22} = 1426.20 + 25.54 * 22 - 110.6 = 1877.4 \Rightarrow \mathbf{1877}$$

$$\hat{Y}_{23} = 1426.20 + 25.54 * 23 - 182.5 = 1831.1 \Rightarrow \mathbf{1831}$$

$$\hat{Y}_{24} = 1426.20 + 25.54 * 24 + 19.1 = 2058.3 \Rightarrow \mathbf{2058}$$

$$\hat{Y}_{25} = 1426.20 + 25.54 * 25 + 201.2 = 2266.0 \Rightarrow \mathbf{2266}$$

Para a definição dos intervalos de confiança destas previsões, o procedimento é igual ao que já foi descrito:

- Determinação das estimativas para os instantes de 1 a 20, utilizando o modelo de previsão atrás indicado e cujos valores estão indicados na penúltima coluna do quadro da página 24

$$\text{Exemplo } \hat{Y}_1 = 1426.20 + 25.54 * 1 + 72.8 = 1524.5$$

- Cálculo dos erros de previsão ocorrido em cada instante (vide última coluna do mesmo quadro)

$$\text{Exemplo } e_1 = 1503 - 1524.5 = -21.5$$

- Cálculo do somatório dos quadrados dos erros de previsão

$$\sum e_i^2 = (-21.5)^2 + (-79.6)^2 + \dots + (-96.3)^2 = 71933$$

- Determinação do EQM = $71933/20 = 3596.7$

Relembrando que para um nível de confiança de 90%, $z_{\alpha/2}$ é 1.645, os intervalos de confiança pretendidos serão:

$$2035.4 \pm 1.645 * \sqrt{3596.7} = 2035.4 \pm 98.7 \Rightarrow 1936.7 \text{ e } 2134.0 \Rightarrow \mathbf{[1937, 2134]}$$

$$1877.4 \pm 1.645 * \sqrt{3596.7} = 1877.4 \pm 98.7 \Rightarrow 1778.8 \text{ e } 1976.1 \Rightarrow \mathbf{[1779, 1976]}$$

$$1831.1 \pm 1.645 * \sqrt{3596.7} = 1831.1 \pm 98.7 \Rightarrow 1732.4 \text{ e } 1929.8 \Rightarrow \mathbf{[1732, 1930]}$$

$$2058.3 \pm 1.645 * \sqrt{3596.7} = 2058.3 \pm 98.7 \Rightarrow 1959.7 \text{ e } 2157.0 \Rightarrow \mathbf{[1960, 2157]}$$

$$2266.0 \pm 1.645 * \sqrt{3596.7} = 2266.0 \pm 98.7 \Rightarrow 2167.3 \text{ e } 2364.6 \Rightarrow \mathbf{[2167, 2365]}$$