Exercício – 1 [Alisamento Exponencial Simples]

Vamos considerar a série de produção de veículos [automóveis, comerciais leves, caminhões e ônibus] no Brasil para o período de 2003 a 2020. Os dados podem ser observados na Figura–1 [https://g1.globo.com/economia/noticia/2021/01/08/producao-de-veiculos-no-brasil-cai-316percent-em-2020-diz-anfavea.ghtml].

Figura-1 Produção de veículos novos no Brasil.

... ***a indústria automobilística possui um impacto considerável*** *na economia nacional, representando cerca de 12,3% do PIB industrial e* ***gerando, para cada emprego direto, de 35 a 40 empregos indiretos*** (Ferran, 1998 *apud* Dias *et al.*, 1999).

O vetor com os dados da Figura-1 são apresentados a seguir.

```
veiculos <- c(1684715, 2124177, 2357172, 2403680, 2825224, 3050629,
              3076000,3382135,3417782, 3404384, 3713813, 3151831,
              2428089, 2176784,2737256, 2881018, 2944988, 2014055)
```

(a) Construa a série temporal, utilizando o comando `ts`, para os dados do vetor de produção de veículos no Brasil de 2003 a 2020;

(b) Utilize o comando `autoplot` para plotar a série do item (a);

(c) A partir da série criada no item (a), construa uma outra série temporal com dados para o intervalo de anos de 2003 até 2019;

(d) Utilize o comando `ses` [Simple Exponential Soomthing] para realizar as previsões para os anos de 2020 até 2025, para isso use os dados do item (c). Nesse caso, considere o resultado ótimo do `ses`;

(e) Utilize os comandos `autoplot` e `autolayer` para plotar a série no nível, a série ajustada e as previsões do item (d);

(f) Utilize o comando `ses` [**s**imple **e**xponential **s**moothing] para realizar as previsões para os anos de 2021 até 2025, para isso use os dados do item (a). Nesse caso, considere o resultado ótimo do `ses`;

(g) Utilize os comandos `autoplot` e `autolayer` para plotar a série no nível, a série ajustada e as previsões do item (f);

(h) Compare as previsões dos itens (d) e (f), comente sobre os resultados utilizando como apoio o artigo:
https://g1.globo.com/economia/noticia/2021/01/11/ford-fecha-fabricas-e-encerra-producao-no-brasil-em-2021.ghtml

Referências

FERRAN, L. Fornecedores do Complexo Industrial Ford de Guaíba. Porto Alegre, 16 set. 1998 (Apresentado à Federação das Indústrias do Rio Grande do Sul – FIERGS)

DIAS, A. V. C.; GALINA, S. V. R.; SILVA, F. D. Análise Contemporânea da Cadeia Produtiva do Setor Automobilístico: Aspectos relativos à Capacitação Tecnológica, ENEGEP-1999.

Exercício – 2 [Método de Holt]

Na Figura–2 temos o gráfico do Preço de Fechamento da ação da empresa RaiaDrogasil [RADL3] que é uma empresa brasileira do setor varejista farmacêutico criada em 2011 após a fusão da Drogasil S.A. com a Raia S.A.

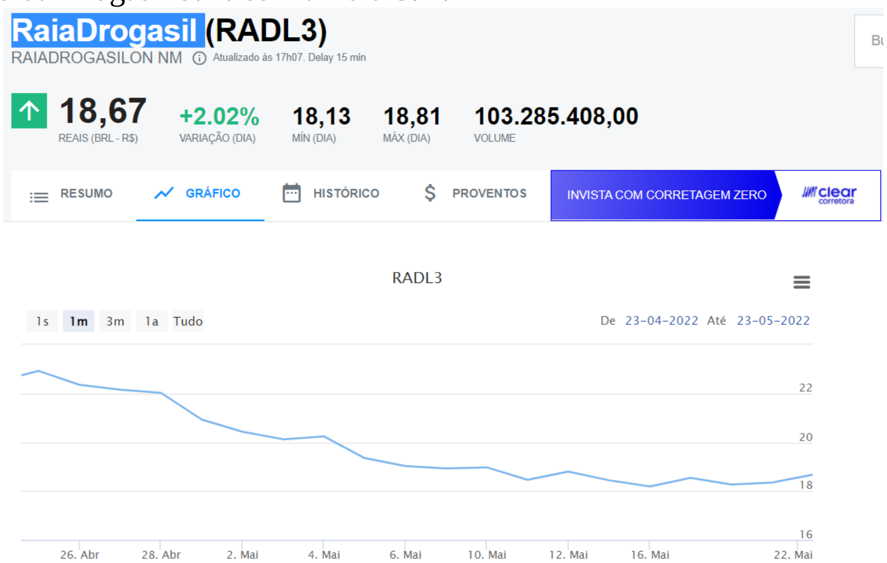


Figura-1 Preço de Fechamento da RADL3.

O Método de Holt é uma técnica bastante aplicada para séries temporais que apresentam poucos dados e exibem tendência, como mostrado na Figura–2. Na Tabela–1 temos os dados para a ação da RADL3 no período de 19/04/2022 até 20/05/2022.

DATA	ABERTURA	FECHAMENTO	VARIAÇÃO	MÍNIMO	MÁXIMO	VOLUME
20/05/2022	18,46	18,30	-0,27	18,13	18,50	89,29M
19/05/2022	18,28	18,35	0,44	18,17	18,54	134,03M
18/05/2022	18,53	18,27	-1,46	17,83	18,81	275,90M
17/05/2022	18,28	18,54	1,92	18,28	18,80	125,60M
16/05/2022	18,46	18,19	-1,36	18,07	18,58	146,71M
13/05/2022	18,79	18,44	-1,91	18,44	19,20	184,71M
12/05/2022	18,31	18,80	1,84	18,21	18,89	182,78M
11/05/2022	18,96	18,46	-2,69	18,38	19,01	170,70M
10/05/2022	19,00	18,97	0,21	18,59	19,22	116,64M
09/05/2022	18,84	18,93	-0,47	18,68	19,27	188,15M
06/05/2022	19,39	19,02	-1,71	18,81	19,39	144,85M
05/05/2022	20,01	19,35	-4,40	18,75	20,01	209,68M
04/05/2022	19,40	20,24	0,60	19,16	20,30	287,26M
03/05/2022	20,43	20,12	-1,52	19,81	20,53	170,53M
02/05/2022	21,00	20,43	-2,39	20,22	21,28	97,88M
29/04/2022	22,38	20,93	-4,95	20,65	22,39	300,49M
28/04/2022	22,19	22,02	-0,59	21,91	22,23	90,23M
27/04/2022	22,27	22,15	-0,89	22,01	22,60	104,49M
26/04/2022	22,81	22,35	-2,49	22,27	22,92	224,05M
25/04/2022	22,33	22,92	1,87	22,24	23,13	172,11M
22/04/2022	22,18	22,50	0,22	22,11	22,73	247,17M
20/04/2022	22,45	22,45	-0,22	22,26	23,03	337,97M
19/04/2022	21,97	22,50	1,90	21,94	22,59	207,76M

Tabela-1 Dados sobre a negociação da RADL3 no período de 19/04/20 a 20/05/2022.

A partir do vetor de dados do Preço de Fechamento mostrado a seguir

```
radl3 <- c(22.50, 22.45, 22.50, 22.92, 22.35, 22.15, 22.02, 20.93, 20.43,
           20.12, 20.24, 19.35, 19.02, 18.93, 18.97, 18.46, 18.80, 18.44,
           18.19, 18.54, 18.27, 18.35, 18.30)
```

(a) Crie a série temporal usando o comando `ts`;

(b) Plote a série temporal usando o comando `autoplot`;

(c) Utilizando o Método de Holt, obtenha as previsões com intervalos de confiança de 80% e 95% para os próximos 5 dias;

(d) Plote num mesmo gráfico, a série original, a série ajustada e as previsões com ICs obtidas no item (c).

Exercício – 3 [Método de Holt-Winters]

O Centro de Estudos Avançados em Economia Aplicada (Cepea) é parte do Departamento de Economia, Administração e Sociologia da Escola Superior de Agricultura “Luiz de Queiroz” (Esalq), unidade da Universidade de São Paulo (USP), localizada em Piracicaba – 150 km da capital. É um grupo de pesquisas registrado no CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico).

g1 AGRO

fique por **dentro** **Eleições** **Covid em alta** **Troca na Petrobras**

Preço da cenoura dispara e gera memes; entenda o que está por trás da alta

Valor da hortaliça subiu depois que fortes chuvas atingiram grandes cidades produtoras nos primeiros meses do ano. Com colheita menor, agricultores cobram o maior preço real da série histórica do Cepea, iniciada em 2008.

Por Paola Patriarca, g1

11/03/2022 06h00 · Atualizado há um mês

A seguir são apresentados os dados dos preços obtidos do CEPEA para:
- Cenoura AAA [Saca de 20Kg], para os meses de Janeiro/2019 até Abril/2022;
- Mandioca [Reais por Tonelada], para os meses de Janeiro/2018 até Maio/2022.

Os dados estão respectivamente nos vetores:

```
cenoura <- c(48.51,50.00,49.50,51.45,66.20,59.15,48.05,31.94,26.65,24.11,23.32,23.57,
             37.93,51.44,68.50,60.59,44.56,31.29,26.33,36.56,39.08,31.40,42.35,41.39,
             34.33,43.58,29.04,30.53,26.50,23.75,42.93,39.33,39.33,41.04,39.00,37.13,
             67.08,126.88,144.38,100.00)

mandioca <- c(584.495,603.725,531.92,514.19,458.095,378.254,430.455,437.5,431.515,
              448.5775,397.82,342.8775,347.9633,373.325,343.932,291.395,302.186,
              286.025,308.0425,307.114,288.9875,300.035,366.7,460.7375,424.276,
              355.615,369.06,342.44,306.0825,314.65,328.544,322.4175,340.9425,
              534.842,472.385,462.595,449.7675,448.345,421.765,441.544,470.0475,
              460.1825,434.452,495.03,545.75,581.012,676.13,719.7,732.645,746.255,
              774.4075,802.212,826.255)
```

Para os dados de Preços da Cenoura e da Mandioca, responda os seguintes itens:

(a) Construa as séries temporais, utilizando o comando `ts`;

(b) Utilize o comando `autoplot` para plotar as séries do item (a);

(c) Para cada uma das séries, ajuste um Modelo de Holt-Winters Aditivo e um Multiplicativo com horizonte de previsão `h = 5`;

(d) Obtenha os valores de AIC [Critério de Informação de Akaike] para os Modelos Aditivo e Multiplicativo do item (c). Selecione o modelo que gerar o menor AIC;

(e) Para cada um dos produtos agrícolas, plote num mesmo gráfico, a série original, a série ajustada e as previsões com ICs obtidas no item (c).

Exercício – 4 [Componentes da Série Temporal]

Impacto de 11 de Setembro nas viagens aéreas nos Estados Unidos: O Bureau of Transportation Statistics (BTS) da Administração de Pesquisa e Tecnologia Inovadora conduziu um estudo para avaliar o impacto do ataque terrorista de 11 de Setembro de 2001 no transporte dos EUA. O relatório do estudo e os dados podem ser encontrados em www.bts.gov/publications/estimated_impacts_of_9_11_on_us_travel. O objetivo do estudo foi declarado da seguinte forma:
"*O objetivo deste estudo é proporcionar uma melhor compreensão dos padrões de comportamento de viagens de passageiros de pessoas que fazem viagens de longa distância antes e depois de 11 de Setembro.*"
O relatório analisa os dados mensais de movimentação de passageiros entre Janeiro de 1990 e Abril de 2004. Os dados de três séries temporais mensais são fornecidos no arquivo Sept11Travel.xls para este período: (1) receita real das companhias aéreas por passageiros-milhas (Air), (2) passageiros-milhas ferroviárias (Rail) e (3) milhas percorridas por automóveis (VMT). Para avaliar o impacto de 11 de setembro, o BTS adotou a seguinte abordagem: usando dados anteriores a 11 de setembro/2001, eles previram dados futuros (presumindo que não haveria ataque terrorista). Em seguida, eles compararam as séries previstas com os dados reais para avaliar o impacto do evento. Nosso primeiro passo, portanto, é dividir cada uma das séries temporais em duas partes: antes e depois de 11 de setembro. Agora nos concentramos apenas nas séries temporais anteriores a 11/Setembro/2001.

(a) Plote cada uma das três séries temporais pré-evento (aéreo, ferroviário, automóvel).
(b) Quais componentes da série temporal aparecem no gráfico **(a)**?
(c) Que tipo de tendência aparece? Altere a escala da série, adicione linhas de tendência e suprima a sazonalidade para visualizar melhor o padrão de tendência.

Exercício – 5 [Alisamento Exponencial Simples – SES]

Nesse exercício vamos replicar os resultados do livro do Montgomery [Introduction to Time Series Analysis and Forecasting, 2a. Edição, Wiley, 2015].
Estamos interessados na velocidade média em um trecho específico de uma rodovia fora do horário de pico. Para o último ano e meio (78 semanas), temos disponíveis médias semanais da velocidade média em milhas / hora entre 10h00 e 15h00. Os dados são fornecidos na Tabela-1. A Figura-1 mostra que os dados da série temporal seguem um processo constante. Para suavizar a variação excessiva, entretanto, a suavização exponencial de primeira ordem pode ser usada. A "melhor" constante de suavização pode ser determinada encontrando o valor da constante de suavização que minimiza a soma dos erros de predição de um passo à frente ao quadrado.

A soma dos erros de predição de um passo à frente ao quadrado para vários valores de λ é dada na Tabela-2. Além disso, a Figura-2 mostra que o SSE mínimo é obtido para λ = 0,4.
Vamos supor que também somos solicitados a fazer previsões para as próximas 12 semanas depois da semana 78. A Figura-3 mostra os valores suavizados para as primeiras 78 semanas junto com as previsões para as semanas 79-90 com intervalos de previsão. Também mostra a velocidade semanal real durante esse período. Observe que, como utilizamos o alisamento exponencial simples, as previsões para as próximas 12 semanas são as mesmas. Da mesma forma, os intervalos de previsão são constantes para esse período.

Tabela-1 A velocidade média semanal durante o horário fora do rush.

Week	Speed	Week	Speed	Week	Speed	Week	Speed
1	47.12	26	46.74	51	45.71	76	45.69
2	45.01	27	46.62	52	43.84	77	44.59
3	44.69	28	45.31	53	45.09	78	43.45
4	45.41	29	44.69	54	44.16	79	44.75
5	45.45	30	46.39	55	46.21	80	45.46
6	44.77	31	43.79	56	45.11	81	43.73
7	45.24	32	44.28	57	46.16	82	44.15
8	45.27	33	46.04	58	46.50	83	44.05
9	46.93	34	46.45	59	44.88	84	44.83
10	47.97	35	46.31	60	45.68	85	43.93
11	45.27	36	45.65	61	44.40	86	44.40
12	45.10	37	46.28	62	44.17	87	45.25
13	43.31	38	44.11	63	45.18	88	44.80
14	44.97	39	46.00	64	43.73	89	44.75
15	45.31	40	46.70	65	45.14	90	44.50
16	45.23	41	47.84	66	47.98	91	45.12
17	42.92	42	48.24	67	46.52	92	45.28
18	44.99	43	45.59	68	46.89	93	45.15
19	45.12	44	46.56	69	46.01	94	46.24
20	46.67	45	45.02	70	44.98	95	46.15
21	44.62	46	43.67	71	45.76	96	46.57
22	45.11	47	44.53	72	45.38	97	45.51
23	45.18	48	44.37	73	45.33	98	46.98
24	45.91	49	44.62	74	44.07	99	46.64
25	48.39	50	46.71	75	44.02	100	44.31

Figura-1 A velocidade média semanal durante o horário fora do rush.

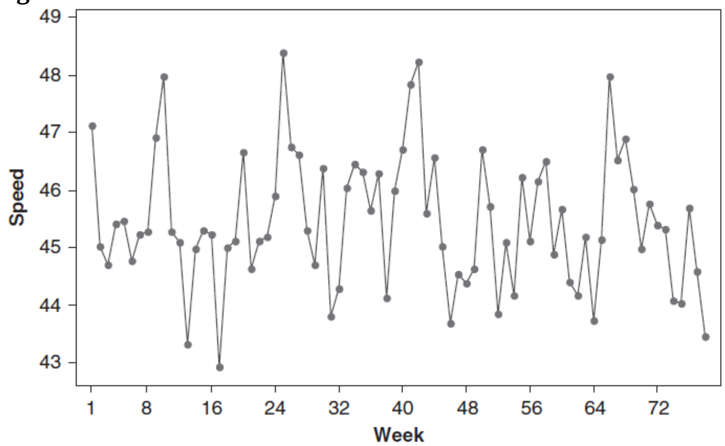


Tabela-2 SSE para diferentes valores $\boldsymbol{\lambda}$ para os dados de velocidade média.

λ		0.1		0.2		0.3		0.4		0.5		0.9	
Week	Speed	Forecast	$e(t)$	Forecast	$e(t)$	Forecast	$e(t)$	Forecast	$e(t)$	Forecast	$e(t)$	Forecast	$e(t)$
1	47.12	47.12	0.00	47.12	0.00	47.12	0.00	47.12	0.00	47.12	0.00	47.12	0.00
2	45.01	47.12	−2.11	47.12	−2.11	47.12	−2.11	47.12	−2.11	47.12	−2.U	47.12	−2.11
3	44.69	46.91	−2.22	46.70	−2.01	46.49	−1.80	46.28	−1.59	46.07	−1.38	45.22	−0.53
4	45.41	46.69	−1.28	46.30	−0.89	45.95	−0.54	45.64	−0.23	45.38	0.03	44.74	0.67
5	45.45	46.56	−1.11	46.12	−0.67	45.79	−0.34	45.55	−0.10	45.39	0.06	45.34	0.11
6	44.77	46.45	−1.68	45.99	−1.22	45.69	−0.92	45.51	−0.74	45.42	−0.65	45.44	−0.67
7	45.24	46.28	−1.04	45.74	−0.50	45.41	−0.17	45.21	0.03	45.10	0.14	44.84	0.40
8	45.27	46.18	−0.91	45.64	−0.37	45.36	−0.09	45.22	0.05	45.17	0.10	45.20	0.07
9	46.93	46.09	0.84	45.57	1.36	45.33	1.60	45.24	1.69	45.22	1.71	45.26	1.67
10	47.97	46.17	1.80	45.84	2.13	45.81	2.16	45.92	2.05	46.07	1.90	46.76	1.21
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
75	44.02	45.42	−1.40	45.30	−1.28	45.12	−1.10	44.93	−0.91	44.75	−0.73	44.20	−0.18
76	45.69	45.28	0.41	45.05	0.64	44.79	0.90	44.56	1.13	44.39	1.30	44.04	1.65
77	44.59	45.32	−0.73	45.18	−0.59	45.06	−0.47	45.01	−0.42	45.04	−0.45	45.52	−0.93
78	43.45	45.25	−1.80	45.06	−1.61	44.92	−1.47	44.84	−1.39	44.81	−1.36	44.68	−1.23
SS_E			124.14		118.88		117.27		116.69		116.95		128.98

Figura-2 Gráfico de SSE para vários valores de λ para dados de velocidade média.

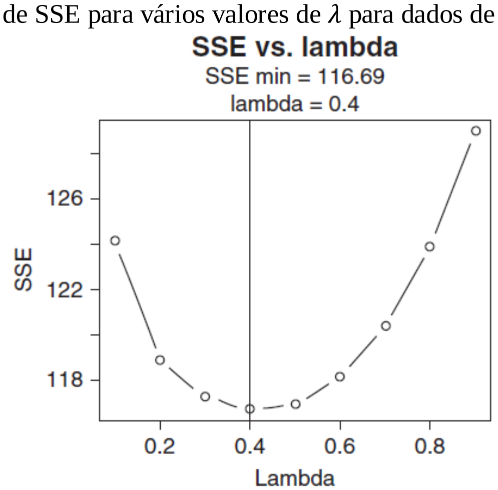
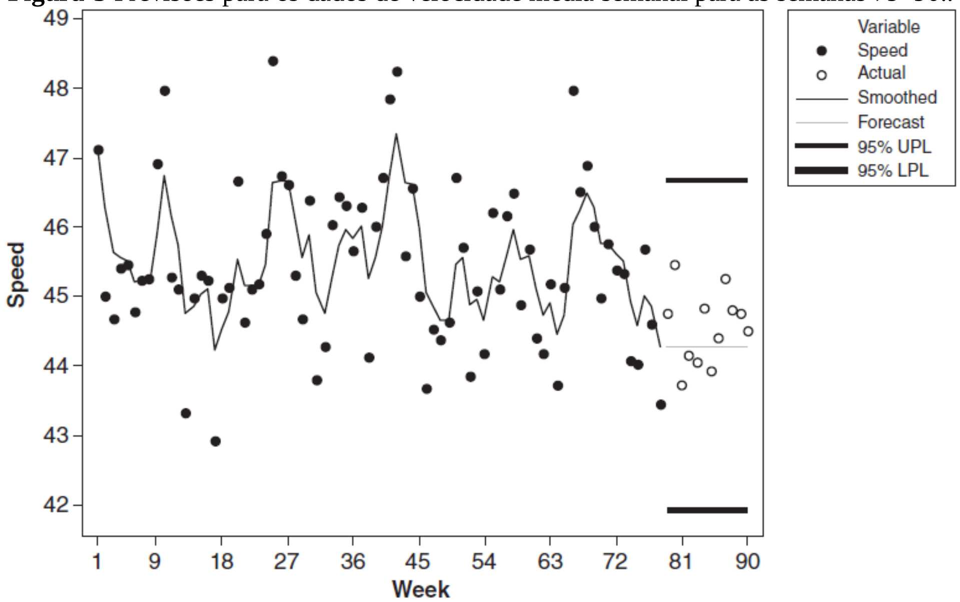


Figura-3 Previsões para os dados de velocidade média semanal para as semanas 79–90..



Nesse livro [Montgomery, 2015] não é usado o R. Vamos reconstruir os seguintes resultados:

(a) Plotar a série conforme mostrado na Figura-1;
(b) Calcular as predições usando, ajustar um modelo de alisamento exponencial, com λ assumindo os valores 0,1; 0,2; 0,3; 0,4; 0,5 e 0,9;
(c) Utilize a função `ses` (**s**imple **e**xponential **s**moothing) do R para determinar o melhor λ. Compare com o resultado obtido pelo Montgomery [$\lambda = 0{,}4$];
(d) Reconstrua a Figura-3, ou seja, determine as previsões dos valores para os instantes de 79 a 90. Além disso, plotar junto com os dados reais nos instantes de 1 a 78; a curva obtida da suavização; previsões e intervalo de 95%. [A figura não precisa ficar exatamente a mesma].

```
velocidade <- c(47.12,45.01,44.69,45.41,45.45,44.77,45.24,45.27,46.93,47.97,
                45.27,45.10,43.31,44.97,45.31,45.23,42.92,44.99,45.12,46.67,
                44.62,45.11,45.18,45.91,48.39,46.74,46.62,45.31,44.69,46.39,
                43.79,44.28,46.04,46.45,46.31,45.65,46.28,44.11,46.00,46.70,
                47.84,48.24,45.59,46.56,45.02,43.67,44.53,44.37,44.62,46.71,
                45.71,43.84,45.09,44.16,46.21,45.11,46.16,46.50,44.88,45.68,
                44.40,44.17,45.18,43.73,45.14,47.98,46.52,46.89,46.01,44.98,
                45.76,45.38,45.33,44.07,44.02,45.69,44.59,43.45,44.75,45.46,
                43.73,44.15,44.05,44.83,43.93,44.40,45.25,44.80,44.75,44.50,
                45.12,45.28,45.15,46.24,46.15,46.57,45.51,46.98,46.64,44.31)
```

Exercício – 6 [Alisamento Exponencial Simples – SES]

Nesse exercício vamos replicar os resultados do livro do Doane [Applied Statistics in Business and Economics, 3a. Edição, McGraw-Hill/Irwin, 2010].

O modelo de suavização exponencial é um tipo especial de média móvel. É usado para a previsão contínua de um período à frente para dados que apresentam movimentos para cima e para baixo, mas nenhuma tendência consistente. Por exemplo, uma empresa pode fazer pedidos de milhares de unidades diferentes para seu estoque a cada semana, de modo a manter seu estoque de cada item no nível desejado (para evitar chamadas de emergência para armazéns ou fornecedores). Para tais previsões, muitas empresas escolhem a suavização exponencial, um modelo de previsão simples com apenas duas entradas e uma constante. A fórmula de atualização para as previsões é

$$\hat{y}_t = \alpha y_t + (1-\alpha)\hat{y}_{t-1}.$$

A partir da fórmula acima, vemos que $\hat{y}_t$ depende de $\hat{y}_{t-1}$, que por sua vez depende de $\hat{y}_{t-2}$, e assim por diante, até y_1. Mas onde obtemos y_1 (a predição inicial)? Existem muitas maneiras de inicializar o processo de predição. Por exemplo, o **Excel** simplesmente define a previsão inicial igual ao primeiro valor real dos dados:

Método A

Defina $\hat{y}_1 = y_1$ (use o primeiro valor de dados).

Esse método tem a vantagem da simplicidade, mas se y_1 for incomum, pode levar algumas iterações para que as previsões se estabilizem. Outra abordagem é definir a previsão inicial igual à média dos primeiros vários valores de dados observados. Por exemplo, o **MINITAB** usa os primeiros seis valores de dados:

Método B

Defina $\hat{y}_1 = \dfrac{y_1 + y_2 + y_3 + y_4 + y_5 + y_6}{6}$ (média dos 6 primeiros valores de dados).

Este método tende a eliminar os efeitos de valores incomuns de y, mas consome mais dados e ainda é vulnerável a valores de y incomuns (outliers).
A Tabela-1 mostra as vendas semanais de um Material Selante (um produto de pintura vendido em galões) em um grande varejista do tipo "faça você mesmo". Para previsões de suavização exponencial, a empresa usa α = 0,10. Sua escolha de α é baseada na experiência. Como α é bastante pequeno, ele fornecerá um alisamento forte. As duas últimas colunas comparam os dois métodos de inicialização das previsões. Vendas excepcionalmente altas na semana 5 têm um forte efeito no ponto de partida do método B. No início, a diferença nas previsões é perceptível, mas com o tempo os métodos convergem.
Apesar de pontos diferentes de partida, as predições para o período 19 não diferem muito.
Arredondando para o próximo número inteiro mais alto, para a semana 19, a empresa pediria 130 galões (usando o método A) ou 134 galões (usando o método B). As Figuras-1 e 2 mostram a semelhança nos padrões das predições, embora o nível das predições seja sempre maior no método B devido ao seu maior valor inicial. Isso demonstra que a escolha dos valores iniciais afeta as previsões.

Onde aparecer F, considere F como sendo $\hat{y}$.

Tabela-1 Suavização exponencial (n = 18 semanas).

Week	*Sales in Gallons*	*Method A: $F_1 = y_1$*	*Method B: F_1 = Average (1st 6)*
1	106	106.000	127.833
2	110	106.000	125.650
3	108	106.400	124.085
4	97	106.560	122.477
5	210	105.604	119.929
6	136	116.044	128.936
7	128	118.039	129.642
8	134	119.035	129.478
9	107	120.532	129.930
10	123	119.179	127.637
11	139	119.561	127.174
12	140	121.505	128.356
13	144	123.354	129.521
14	94	125.419	130.969
15	108	122.277	127.272
16	168	120.849	125.344
17	179	125.564	129.610
18	120	130.908	134.549

Figura-1 Usando o primeiro valor de y.

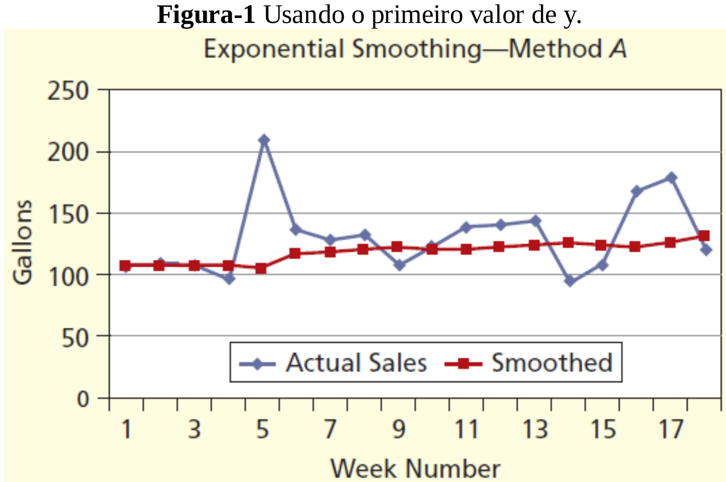
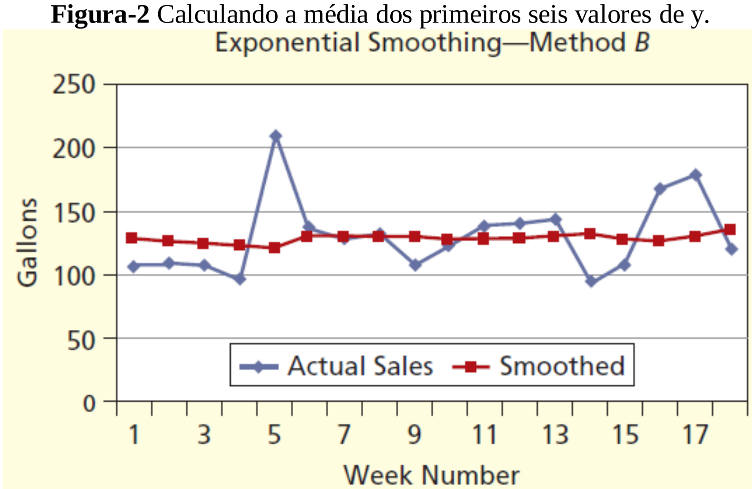


Figura-2 Calculando a média dos primeiros seis valores de y.



Determine:

(a) A previsão para o instante 19;
(b) Refaça as Figuras-1 e 2 usando o R;
(c) O ajuste usando $\alpha = 0{,}10$ e o α ótimo usando a função `ses` (**s**imple **e**xponential **s**moothing) do R. Compare os resultados da Tabela-1.

Exercício – 7 [Alisamento Exponencial Simples – SES e Método de Holt]

Vamos aqui replicar os resultados de um exercício de um livro que escrevi com uma professora da USP e um especialista em Logística [Gestão estratégica de estoques e demanda, IESDE, 2009].
Um dos objetivos desse exercício é realizar o ajuste [predição] do modelo e a previsão de um valor no futuro para uma série que tem exatamente 12 valores mensais. Essa tarefa é relativamente comum, temos os valores dos últimos 12 meses e queremos prever para o próximo mês. Essa é uma situação em que o método de Holt se adequa muito bem.

A Tabela–1 abaixo apresenta uma série de doze valores de vendas.

Tabela-1 Dados mensais da variável vendas.

Mês	T	Vendas	$\alpha = 0{,}1$	$\alpha = 0{,}5$	$\alpha = 0{,}9$
Jan	1	6,05			
Fev	2	5,99			
Mar	3	5,62			
Abr	4	5,59			
Mai	5	5,80			
Jun	6	6,45			
Jul	7	5,72			
Ago	8	5,92			
Set	9	6,35			
Out	10	5,94			
Nov	11	6,20			
Dez	12	6,29			
Jan					

Usando o R:

(a) Determine os valores de ponderação exponencial simples considerando $\alpha = 0{,}1$, $\alpha = 0{,}5$ e $\alpha = 0{,}9$.

(b) Construa um gráfico contendo a série original e as séries para cada um dos coeficientes. Para cada coeficiente de ponderação calcule a previsão para janeiro do próximo ano.

(c) Aplique o método de Holt, com coeficiente de ponderação exponencial (α) igual a 0,2 e coeficiente de amortecimento para a tendência (β) igual a 0,5, determine o valor das vendas para o mês de janeiro.

(d) Calcule o erro quadrático médio (EQM, ou MSE = mean squared error) e o erro absoluto médio (EAM, ou MAE = mean absolute error) para as predições obtidas.

(e) Determine o valor previsto para janeiro do próximo ano usando o método de Holt.

(f) Construa o gráfico das vendas originais e predições utilizando o resultado do método de Holt do item **(c)**. Também nesse gráfico plotar o valor previsto no item **(e)**.

(g) Utilize a função `holt` do R para determinar os valores ótimos de α e β.

(h) Um aluno de MBA resolveu esse exercício usando o **Excel**. Os resultados dele estão no Apêndice-A. Utilize esse apêndice como referência e corrija-o se for necessário.

Exercício – 8 [Treinamento e Validação]

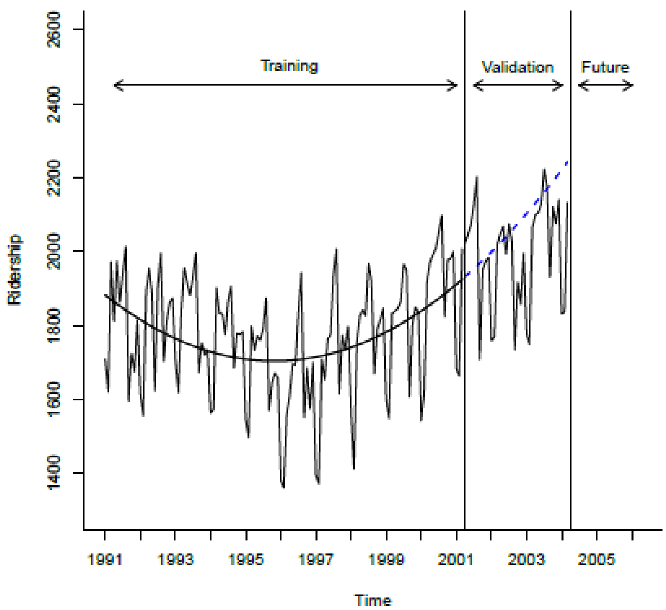
Nesse exercício vamos replicar os resultados do livro da Galit Shmueli [Practical Time Series Forecasting with R: A Hands-On Guide, 2a. Edição, Axelrod Schnall Publishers, 2016].

Antes de tentar prever os valores futuros da série, podemos considerar dois períodos: (i) de treinamento e (ii) validação. Algumas vantagens dessa estratégia são:

1. O período de validação, que é o período mais recente, geralmente contém as informações mais valiosas em termos de ser o mais próximo no tempo do período previsto.
2. Com mais dados (a série temporal completa é comparada apenas com o período de treinamento), alguns modelos podem ser estimados com mais precisão.
3. Se apenas o período de treinamento for usado para gerar previsões, será necessário fazer previsões no futuro. Por exemplo, se o conjunto de validação contém quatro pontos de tempo, a previsão da próxima observação exigirá uma previsão de cinco passos à frente do conjunto de treinamento ($\hat{y}_{T+5}$).

A Figura-1 mostra as previsões no período de validação de um modelo de tendência quadrática aplicado aos dados de treinamento de passageiros da Amtrak.

Figura-1 Previsões no período de validação a partir de um modelo de tendência quadrática estimado a partir do período de treinamento (dados de viagens da Amtrak).



Escolha do período de validação

A escolha da duração do período de validação depende do objetivo da previsão, da frequência dos dados e do horizonte de previsão. O princípio fundamental é escolher um período de validação que simule o horizonte de previsão, para permitir a avaliação do desempenho preditivo real. Por exemplo, no exemplo de viagens da Amtrak, se o modelo de previsão for usado para produzir previsões para o próximo ano, o período de validação deve incluir pelo menos um ano. A escolha de um período de validação mais curto não permitirá a avaliação do desempenho das previsões de longo prazo. Escolher um período de validação de mais de um ano significa que nosso período de treinamento contém informações menos recentes e, portanto, o modelo baseado no período de treinamento perderá as informações mais recentes disponíveis no momento da previsão.

Os comandos mostrados abaixo, permitem construir uma figura parecida com a Figura-1.

```
library(forecast)

Amtrak_data <- read.csv("Amtrak_Data.csv")
passageiros_ts <- ts(Amtrak_data$Passageiros, start = c(1991,1), end = c(2004, 3), freq = 12)
plot(passageiros_ts, xlab = "Tempo", ylab = "Passageiros", ylim = c(1300, 2300), bty = "l")

nValid <- 36
nTrain <- length(passageiros_ts) - nValid
train.ts <- window(passageiros_ts, start = c(1991, 1), end = c(1991, nTrain))
valid.ts <- window(passageiros_ts, start = c(1991, nTrain + 1), end = c(1991, nTrain + nValid))
passageiros.lm <- tslm(train.ts ~ trend + I(trend^2))
passageiros.lm.pred <- forecast(passageiros.lm, h = nValid, level = 0)
plot(passageiros.lm.pred, ylim = c(1300, 2600), ylab = "Passageiros", xlab = "Tempo", bty = "l",
     xaxt = "n", xlim = c(1991,2006.25), main = "", flty = 2)
axis(1, at = seq(1991, 2006, 1), labels = format(seq(1991, 2006, 1)))
lines(passageiros.lm$fitted, lwd = 2)
lines(valid.ts)
```

Obtenha:

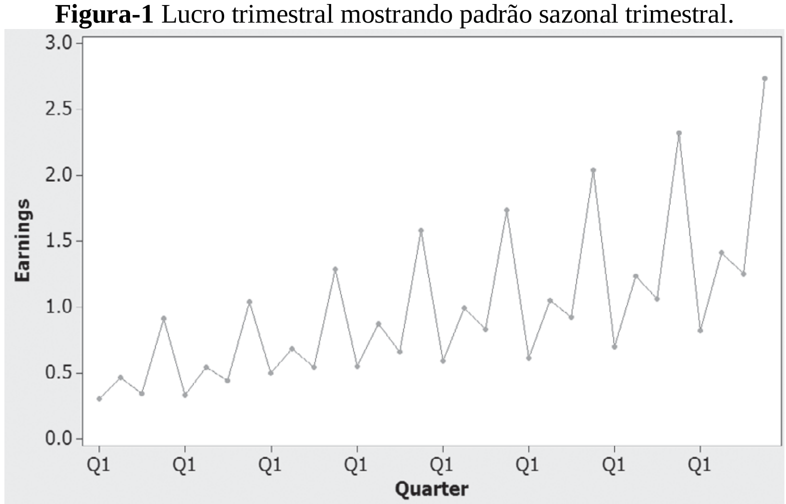
(a) `passageiros.lm.pred$residuals`, interprete o que significa;
(b) `valid.ts - passageiros.lm.pred$mean`, interprete o que significa;
(c) crie um vetor de `erro` formado pelos vetores dos itens (a) e (b). Determine sua média e desvio-padrão;
(d) plote o vetor de `erro` do item (c);
(e) faça um teste de normalidade para o vetor de `erro` do item (c). O que você conclui sobre o teste de normalidade?

Exercício – 9

Nesse exercício vamos replicar os resultados do livro do Newbold [Statistics for Business and Economics, Edição Global, Pearson, 2013].

Outro componente importante é o padrão sazonal. A Figura-1 mostra o lucro trimestral por ação de uma empresa. Os ganhos do quarto trimestre são substancialmente maiores, e os ganhos do segundo trimestre são um pouco maiores em comparação com os outros períodos. Observe como esse padrão continua a se repetir ao longo do ciclo de quatro trimestres representando cada ano. Além do componente de sazonalidade, também há uma tendência de alta perceptível no lucro por ação. Nosso tratamento da sazonalidade depende de nossos objetivos. Por exemplo, se for importante prever cada trimestre com a maior precisão possível, incluímos um componente de sazonalidade em nosso modelo. Se anteciparmos que o padrão de sazonalidade continuará, a estimativa do componente de sazonalidade deve ser incluída em nosso modelo de previsão.
Para alguns propósitos, a sazonalidade pode ser um incômodo. Em muitas aplicações, o analista requer uma avaliação dos movimentos gerais em uma série de tempo, não contaminada pela influência de fatores sazonais.
Os padrões sazonais em uma série temporal constituem uma forma de comportamento oscilatório regular. Além disso, muitas séries de negócios e econômicas exibem padrões oscilatórios ou cíclicos não relacionados ao comportamento sazonal. Por exemplo, muitas séries econômicas seguem os padrões do ciclo de negócios de altas e baixas.

Figura-1 Lucro trimestral mostrando padrão sazonal trimestral.

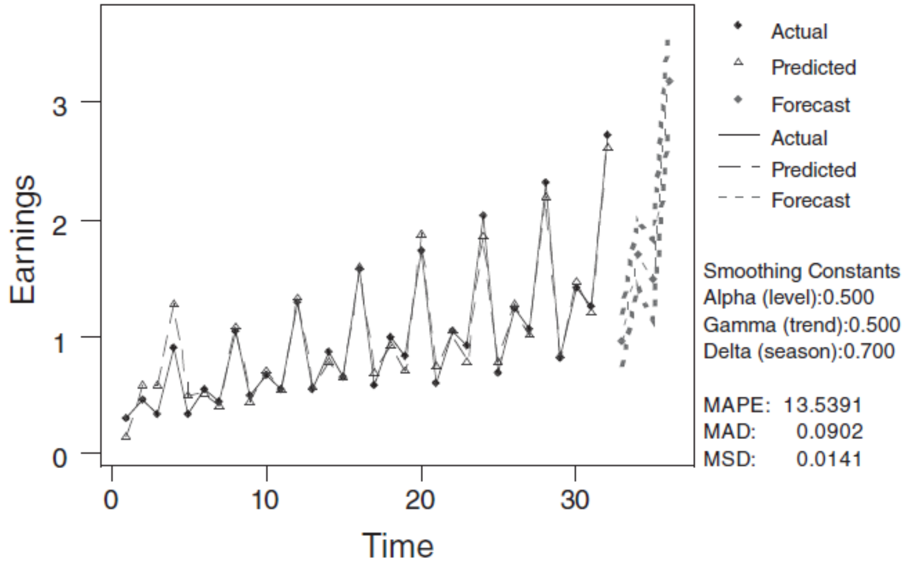


Muitos procedimentos de previsão utilizados em negócios são baseados em extensões da suavização exponencial simples. O procedimento de suavização exponencial de Holt-Winters permite a tendência, e também a sazonalidade, em uma série temporal.
O vetor de dados usados pelo Newbold para construir a figura acima é dado por:

```
ganhos <- c(0.3, 0.46, 0.345, 0.91, 0.33, 0.545, 0.44, 1.04, 0.495, 0.68,
0.545, 1.285, 0.55, 0.87, 0.66, 1.58, 0.59, 0.99, 0.83, 1.73, 0.61, 1.05,
0.92, 2.04, 0.7, 1.23, 1.06, 2.32, 0.82, 1.41, 1.25, 2.73)
```

Usando o **MINITAB**, o Newbold obtem o resultado mostrado na Figura-2.

Figura-2 Modelo multiplicativo de Holt-Winters para ganhos.



Usando o R, reproduza:

(a) a Figura-2 com a série no nível e a ajustada pelo Holt-Winters. Além disso, também plotar na mesma figura a previsão para os próximos 4 trimestres com o intervalo de confiança de 95%;

(b) usando os parâmetros do modelo Holt-Winters mostrados na figura, obtenha os valores de MAPE [mean absolute percentage error, erro médio absoluto percentual], MAD [mean absolute deviation, desvio absoluto médio] e MSD [mean squared deviation, desvio quadrado médio];

$$MAPE = \frac{\sum \left| \frac{y_t - \hat{y}_t}{y_t} \right|}{n} \times 100, \quad (y_t \neq 0) \qquad MAD = \frac{\sum |y_t - \hat{y}_t|}{n} \qquad MSD = \frac{\sum |y_t - \hat{y}_t|^2}{n}.$$

(c) usando a função `hw`, com os parâmetros especificados pelo Newbold, `alpha = 0.5`, `beta = 0.5` e `gamma = 0.7`, obtenha os valores ajustados;

(d) aplique o comando `round(accuracy(seu_modelo_aqui),2)` para o item **(c)**;

(e) determine os valores dos parâmetros ótimos usando a função `hw` com: `seasonal = "additive"` e `seasonal = "multiplicative"`;

(f) aplique o comando `round(accuracy(seu modelo aqui),2)` para o caso `seasonal = "multiplicative"` do item **(e)**;

(g) aplique o comando `seu_modelo_aditivo_aqui$model` e `seu_modelo_multiplicativo_aqui$model` para o item **(e)**;

(h) aplique o comando `seu_modelo_aditivo_aqui$fitted` e `seu_modelo_multiplicativo_aqui$fitted` para o item **(e)**;

(i) a Figura-2 com a série no nível e a ajustada pelo Holt-Winters ótimo **multiplicativo**. Além disso, também plotar na mesma figura a previsão para os próximos 4 trimestres com o intervalo de confiança de 95% obtido pelo Holt-Winters ótimo **multiplicativo**.

Exercício – 10 [Método de Holt-Winters]

Nesse exercício vamos replicar os resultados do livro do Barlow [Excel Models for Business and Operations Management, 2a. Edição, Wiley, 2005].

Ainda é muito comum o uso do **Excel** como ferramenta de apoio para modelar séries temporais e realizar previsões. As Figura-1 e 2 referem-se a um exemplo disso.

Figura-1 Modelo de Holt-Winters - fórmulas de planilha e parâmetros do Solver.

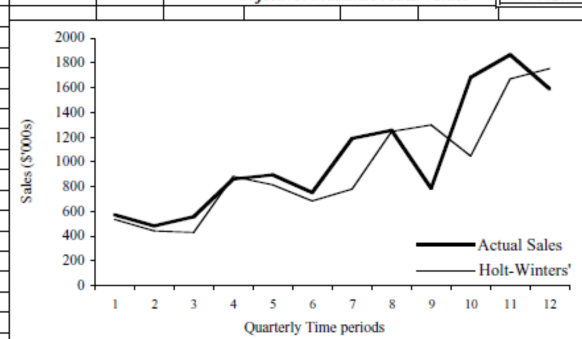
Cell	Formula	Copied to
E11	D11/AVERAGE(D$11:D$14)	E12:E14
F14	(AVERAGE(D15:D18) – AVERAGE(D11:D14))/4	
G14	D14/E14	
G15	D$5*D15/E11 + (1−D$5)*(G14 + F14)	G16:G26
E15	D$7*D15/G15 + (1−D$7)*E11	E16:E26
F15	D$6*(G15-G14) + (1−D$6)*F14	F16:F26
H15	(G14 + F14)*E11	H16:H26
H28	SUMXMY2(D15:D26,H15:H26)/COUNT(H15:H26)	

Solver parameters

Set target cell:	H28	
Equal to:	Min	
By changing cells:	D5:D7	
Subject to constraints:	D5:D7 ≥ 0	= smoothing constants must be ≥ 0
	D5:D7 ≤ 1	= smoothing constants must be ≤ 1

Figura-2 Um modelo para o método de suavização exponencial de Holt-Winters..

	A	B	C	D	E	F	G	H	I
1	**Example 6.4 A model for the Holt-Winters' exponential smoothing method**								
2									
3			*User input cells are shaded*						
4									
5			$\alpha =$	0.05					
6			$\beta =$	1.00					
7			$\gamma =$	1.00					
8									
9		***Year/***	***Period***	***Sales***	***Seasonality***	***Trend***	***Level***	***Forecast***	
10		***Quarter***		Y_t	S_t	T_t	L_t	F_{t+n}	
11		200W - Q1	−3	490	1.089	*Values within this dashed*			
12		Q2	−2	370	0.822	*rectangle are initial estimates*			
13		Q3	−1	330	0.733	---	---	---	
14		Q4	0	610	1.356	41.9	450.0	---	
15		200X - Q1	1	571	1.157	43.6	493.6	535.7	
16		Q2	2	482	0.893	46.2	539.8	441.7	
17		Q3	3	557	0.936	55.2	595.0	429.7	
18		Q4	4	861	1.326	54.4	649.5	881.4	
19		200Y - Q1	5	895	1.265	58.1	707.5	814.2	
20		Q2	6	753	0.978	62.1	769.6	683.6	
21		Q3	7	1190	1.392	85.0	854.7	778.6	
22		Q4	8	1255	1.335	85.4	940.0	1245.7	
23		200Z - Q1	9	786	0.783	64.3	1004.4	1297.1	
24		Q2	10	1684	1.527	98.3	1102.7	1045.6	
25		Q3	11	1868	1.546	105.6	1208.3	1672.2	
26		Q4	12	1589	1.215	99.2	1307.5	1754.2	
27									
28				***Objective: Minimise MSE values =***				77882.7	



Obtenha:

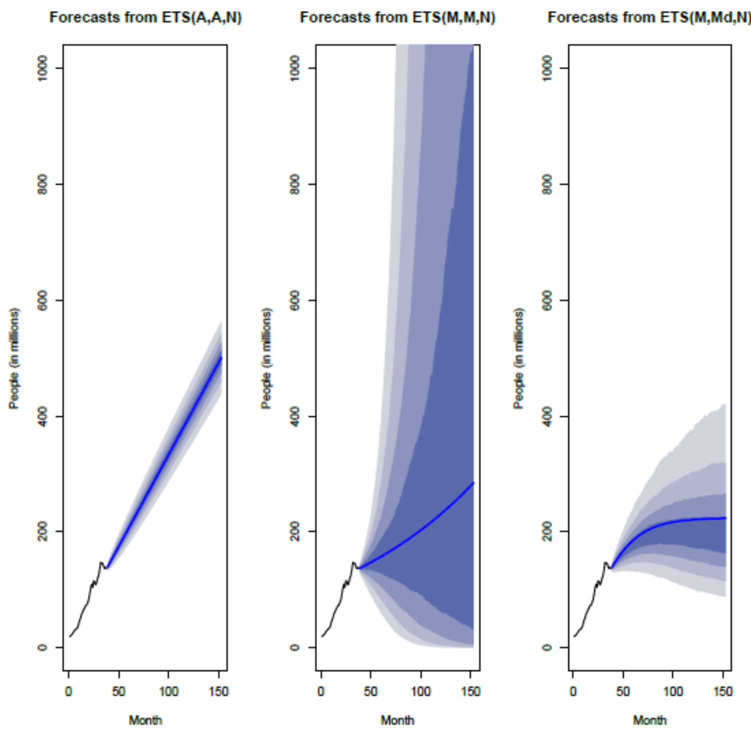
(a) o ajuste da série temporal usando os mesmos parâmetros do Barlow. Compare;
(b) a figura com os valores reais e do modelo de Holt-Winters, conforme Figura-2, mas com os valores ajustados que você obteve em **(a)**;
(c) aplique o comando `round(accuracy(seu_modelo_aqui),2)` para o item **(a).** Compare com o MSE obtido pelo Barlow;
(d) determine os valores dos parâmetros ótimos usando `seasonal = "multiplicative"`;
(e) aplique o comando `round(accuracy(seu_modelo_aqui),2)` para o item **(d).** Compare com o MSE obtido pelo Barlow;
(f) aplique o comando `seu_modelo_multiplicativo_aqui$model` para o item **(d)**;
(h) aplique o comando `seu_modelo_multiplicativo_aqui$fitted` para o item **(d)**.

Exercício – 11 [ETS – ExponenTial Smoothing – ou Erro, Tendência, Sazonalidade - ou Modelo de Espaço de Estados]

Nesse exercício vamos replicar os resultados do livro da Galit Shmueli [Practical Time Series Forecasting with R: A Hands-On Guide, 2a. Edição, Axelrod Schnall Publishers, 2016].

É interessante e importante notar que diferentes modelos podem gerar diferentes cones de previsão.

Figura-1 Cones de previsão de três modelos de suavização exponencial ajustados para os dados da Tumblr.



Cones de Previsão

Em R, pode ser útil comparar conjuntos de intervalos de predição de diferentes modelos. Alguns modelos de previsão têm diferentes níveis de certeza de previsão para diferentes períodos futuros, normalmente com certeza decrescente à medida que fazemos previsões no futuro. Quando os intervalos de previsão de tal modelo em um nível específico de certeza são contíguos no tempo, eles formam um cone de predição.

Acima, consideramos os cones de previsão de três modelos diferentes. A classe de modelos de suavização exponencial no pacote de previsão de R é abreviada por ETS, que significa **E**rro, **T**endência e **S**azonalidade.

Os componentes de erro, tendência e sazonalidade nesta classe de modelos podem assumir uma variedade de configurações diferentes, como aditiva (A) ou multiplicativa (M).

Para cada modelo na Figura-1, a linha grossa no meio dos cones de previsão representa as previsões de pontos do modelo em cada etapa à frente. Para esses modelos específicos ajustados aos dados da Tumblr, vemos que entre os modelos os cones de previsão diferem

ainda mais do que suas previsões pontuais. Essas diferenças podem ser muito importantes quando se considera a incerteza em torno de qualquer previsão pontual.
A seguir são apresentados os comandos do R para gerar a Figura-1.

```
dev.off(dev.list()["RStudioGD"])
rm(list=ls())
cat("\f")

library(forecast)
library(tidyverse)
library(ggplot2)
library(fpp2)
library(smooth)

tumblr.data <- read.csv("Tumblr.csv")
#View(tumblr.data)

pessoas.ts <- ts(tumblr.data$People.Worldwide) / 1000000 # Crie uma série temporal a
                                                         # partir dos dados

pessoas.ets.AAN <- ets(pessoas.ts, model = "AAN")                    # Fit Modelo 1
pessoas.ets.MMN <- ets(pessoas.ts, model = "MMN", damped = FALSE)    # Fit Modelo 2
pessoas.ets.MMdN <- ets(pessoas.ts, model = "MMN", damped = TRUE)    # Fit Modelo 3

pessoas.ets.AAN.pred <- forecast(pessoas.ets.AAN, h = 115, level = c(0.2, 0.4, 0.6, 0.8))
pessoas.ets.MMN.pred <- forecast(pessoas.ets.MMN, h = 115, level = c(0.2, 0.4, 0.6, 0.8))
pessoas.ets.MMdN.pred <- forecast(pessoas.ets.MMdN, h = 115, level = c(0.2, 0.4, 0.6, 0.8))

par(mfrow = c(1, 3))

plot(pessoas.ets.AAN.pred, xlab = "Meses", ylab = "Pessoas (em milhões)", ylim = c(0, 1000))
plot(pessoas.ets.MMN.pred, xlab = "Meses", ylab="Pessoas (em milhões)", ylim = c(0, 1000))
plot(pessoas.ets.MMdN.pred, xlab = "Meses", ylab="Pessoas (em milhões)", ylim = c(0, 1000))
```

Responda aos itens:

(a) o que significa `model = "AAN"` e `model = "MMN"`?;
(b) refaça a Figura-1 usando `h = 5` e depois `h = 10`;
(c) No `Modelo 3` ajustado a opção `damped` é setada como `TRUE`, escreva esse modelo matematicamente explicando o que significa.

APÊNDICE – A

Para α = 0,1

Nesse caso tem-se que:

$\hat{y}_t = \alpha \hat{y}_{t-1} + (1-\alpha) y_t \quad , \quad 0 < \alpha \leq 1$

sendo que:

$\hat{y}_1 = y_1$

O valor previsto para janeiro é igual ao verdadeiro valor de janeiro. A previsão para fevereiro é dada por:

$\hat{y}_2 = (0,1).\hat{y}_{\mathbf{1}} + (0,9).y_2 = (0,1).(6,05) + (0,9).(5,99) = 6,00$

Para o mês de março tem-se que:

$\hat{y}_3 = (0,1).\hat{y}_2 + (0,9).y_3 = (0,1).(6,00) + (0,9).(5,62) = 5,66$

Dessa forma, procede-se sucessivamente até obtenção de todos os valores previstos utilizando α = 0,1.

Para α = 0,5

Nesse caso tem-se que:

$\hat{y}_t = \alpha \hat{y}_{t-1} + (1-\alpha) y_t \quad , \quad 0 < \alpha \leq 1$

sendo que:

$\hat{y}_1 = y_1$

O valor previsto para janeiro é igual ao verdadeiro valor de janeiro. A previsão para fevereiro é dada por:

$\hat{y}_2 = (0,5).\hat{y}_{\mathbf{1}} + (0,5).y_2 = (0,5).(6,05) + (0,5).(5,99) = 6,02$

Para o mês de março tem-se que:

$\hat{y}_3 = (0,5).\hat{y}_2 + (0,5).y_3 = (0,5).(6,02) + (0,5).(5,62) = 5,82$

Dessa forma, procede-se sucessivamente até a obtenção de todos os valores previstos utilizando α = 0,5.

Para α = 0,9

Nesse caso tem-se que:

$\hat{y}_t = \alpha \hat{y}_{t-1} + (1-\alpha) y_t \quad , \quad 0 < \alpha \leq 1$

sendo que:

$\hat{y}_1 = y_1$

O valor previsto para janeiro é igual ao verdadeiro valor de janeiro. A previsão para fevereiro é dada por:

$\hat{y}_2 = (0,9).\hat{y}_{\mathbf{1}} + (0,1).y_2 = (0,9).(6,05) + (0,1).(5,99) = 6,04$

Para o mês de março tem-se que:

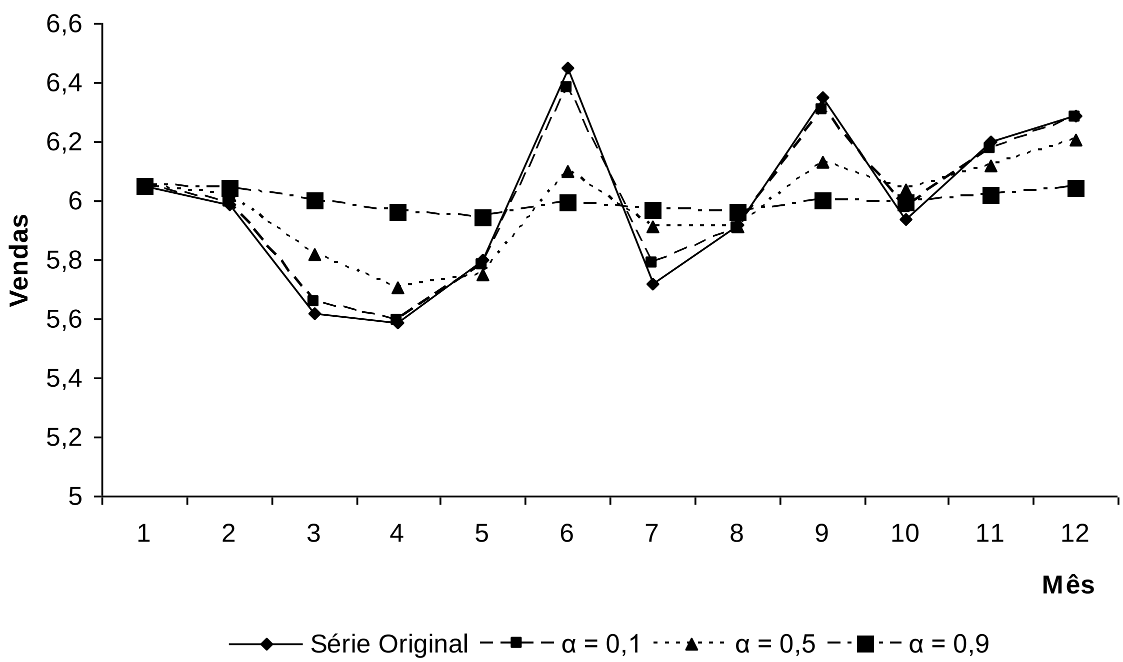
$\hat{y}_3 = (0,9).\hat{y}_2 + (0,1).y_3 = (0,9).(6,04) + (0,1).(5,62) = 6,00$

Dessa forma, procede-se sucessivamente até a obtenção de todos os valores previstos utilizando α = 0,9.

Nesse caso temos que:

Mês	T	Vendas	α = 0,1	α = 0,5	α = 0,9
Jan	1	6,05	6,05	6,05	6,05
Fev	2	5,99	6,00	6,02	6,04
Mar	3	5,62	5,66	5,82	6,00
Abr	4	5,59	5,60	5,71	5,96
Mai	5	5,80	5,78	5,75	5,94
Jun	6	6,45	6,38	6,10	5,99
Jul	7	5,72	5,79	5,91	5,97
Ago	8	5,92	5,91	5,92	5,96
Set	9	6,35	6,31	6,13	6,00
Out	10	5,94	5,98	6,04	6,00
Nov	11	6,20	6,18	6,12	6,02
Dez	12	6,29	6,28	6,20	6,04
Jan			6,28	6,20	6,04

A figura abaixo mostra as diferentes curvas para cada um dos diferentes valores de alfa.



Os valores estimados do nível da série (N_t) são apresentados na coluna 4 da Tabela 2 e são obtidos utilizando-se a equação:

$$N_t = \alpha\left(N_{t-1} + T_{t-1}\right) + (1-\alpha)\left(y_t\right), \quad 0 < \alpha < 1$$

Para o início do processo tem-se que $N_2 = y_2$, ou seja , N_2=5,99 e $T_2 = y_2 - y_1$= 5,99-6,05 = -0,06.

A previsão para março do nível da série é dada por :

$$N_3 = 0{,}2.\left(N_2 + T_2\right) + (0{,}8).\left(y_t\right) = 0{,}2.(5{,}99 - 0{,}06) + 0{,}8.(5{,}62) = 5{,}68$$

$$N_4 = 0{,}2.\left(N_3 + T_3\right) + (0{,}8).\left(y_4\right) = 0{,}2.(5{,}68 - 0{,}18) + 0{,}8.(5{,}59) = 5{,}57$$

As previsões do nível para os demais meses são obtidas analogamente.

A previsão para a tendência da série é obtida por meio da expressão:
$T_t = \beta T_{t-1} + (1-\beta)(N_t - N_{t-1})$, $0 < \beta < 1$

Para o mês de março a tendência é dada por:
$T_3 = \beta T_{3-1} + (1-\beta)(N_3 - N_{3-1}) = 0{,}5.T_2 + 0{,}5.(N_3 - N_2) = 0{,}5.(-0{,}06) + 0{,}5.(5{,}68 - 5{,}99) = -0{,}18$
$T_4 = \beta T_{4-1} + (1-\beta)(N_4 - N_{4-1}) = 0{,}5.T_3 + 0{,}5.(N_4 - N_3) = 0{,}5.(-0{,}18) + 0{,}5.(5{,}57 - 5{,}68) = -0{,}15$
As previsões da tendência para os demais meses são obtidas de forma similar e apresentadas na coluna 5 (Tt) da Tabela 2.

As previsões para os meses de janeiro a dezembro são dadas por $\hat{y}_t = N_t + T_t$ e apresentadas na coluna 6 da Tabela 1.

Tabela 1 – Vendas, Nt , Tt , vendas previstas ($N_t + T_t$), erro absoluto médio e erro quadrático médio

Mês	t	Y_t	N_t	T_t	$N_t + T_t$	$\lvert y_t - \hat{y}_t \rvert$	$(y_t - \hat{y}_t)^2$
Jan	1	6,05					
Fev	2	5,99	5,99	-0,06	5,93	0,06	0,00
Mar	3	5,62	5,68	-0,18	5,50	0,12	0,01
Abr	4	5,59	5,57	-0,15	5,42	0,17	0,03
Mai	5	5,8	5,72	0,00	5,73	0,07	0,01
Jun	6	6,45	6,31	0,29	6,60	0,15	0,02
Jul	7	5,72	5,90	-0,06	5,84	0,12	0,01
Ago	8	5,92	5,90	-0,03	5,88	0,04	0,00
Set	9	6,35	6,26	0,16	6,42	0,07	0,00
Out	10	5,94	6,04	-0,03	6,01	0,07	0,00
Nov	11	6,2	6,16	0,05	6,21	0,01	0,00
Dez	12	6,29	6,27	0,08	6,35	0,06	0,00
Jan							
Soma						**0,9372**	**0,1018**
Média						**0,0852**	**0,0093**

As previsões para os meses de janeiro a dezembro são dadas por $\hat{y}_t = N_t + T_t$ e apresentadas na coluna 6 da Tabela 1.

A previsão para o próximo janeiro é dada por: $\hat{y}_{n+h} = N_n + hT_n$, onde h=1, ou seja,

$\hat{y}_{13} = N_{12} + (1)T_{12} = 6{,}27 + (1)(0{,}08) = 6{,}35$.

Erro quadrático médio é dado por:

$$EQM = \sum_{t=1}^{n} \frac{(y_t - \hat{y}_t)^2}{n}$$

Erro absoluto médio é dado por:

$$EAM = \sum_{t=1}^{n} \frac{|y_t - \hat{y}_t|}{n}$$

Dessa forma, tem-se que:

$$EQM = \frac{0,1018}{11} = 0,0093$$

$$EAM = \frac{0,9372}{11} = 0,0852$$

O gráfico, das vendas originais e previstas utilizando o método de Holt-Winters, é mostrado abaixo.