Series temporais

Modelos de suavização exponencial

Os métodos de previsões em séries temporais, em geral, baseia-se na ideia de que observações passadas contêm informações sobre o padrão de comportamento da série temporal. Uma técnica utilizada para entender esse comportamento da série ao longo do tempo são os modelos de suavização exponencial.

Essas técnicas buscam suavizar os valores de uma série baseado no comportamento mais recente da série.

Modelos para séries localmente constantes

Vamos considerar inicialmente o caso de uma série temporal $Z_1, \ldots, Z_N$, localmente composta de seu nível mais um ruído aleatório, isto é,

$$Z_t = \mu_t + a_t, \quad t = 1, \ldots, N,$$

onde $E(a_t) = 0, \operatorname{Var}(a_t) = \sigma_a^3$ e μ_t é um paràmetro desconhecido, que pode variar lentamente com o tempo.

Médias móveis simples

- **Procedimento**

A técnica de média móvel consiste em calcular a média aritmética das r observações mais recentes, isto é.

$$M_t = \frac{Z_t + Z_{t-1} + \cdots + Z_{t-r+1}}{r}$$

ou

$$M_t = M_{t-1} + \frac{Z_t - Z_{t-r}}{r}$$

Assim, M_t é uma estimativa do nivel μ_t que não leva em conta (ou não pondera) as observações mais antigas, o que é razoável devido ao fato do parâmetro variar suavemente com o tempo.

O nome média móvel é utilizado porque, a cada período, a observação mais antiga é substituída pela mais recente, calculando-se uma média nova.

- **Previsão**

A previsão de todos os valores futuros é dada pela última média móvel calculada, isto é,

$$\hat{Z}_t(h) = \hat{Z}_{t-1}(h+1) + \frac{Z_t - Z_{t-r}}{r}, \quad \text{,h } > 0.$$

A equação acima pode ser interpretada como um mecanismo de atualização de previsão, pois a cada instante (ou a cada nova observação) corrige a estimativa prévia de Z_{t+h}.

- **Determinação de r (tamanho do janelamento)**

As propriedades do método dependem do número de observações utilizadas na média (valor de r). Um valor grande de r faz com que a previsão acompanhe lentamente as mudanças do parâmetro μ_t : um valor pequeno implica numa reação mais rápida. Existem dois casos extremos:

(i) se $r = 1$, então o valor mais recente da série é utiliządo como previsão de todos os valores futuros (este é o tipo de previsão mais simples que existe e é denominado “método ingênuo”);

(ii) se $r = N$, então a previsão será igual à média aritmética de todos os dados observados.

Entretanto, como podemos encontrar um r ‘ótimo’ para o nosso problema ? Não existe uma resposta exata para esta pergunta, porém o que podemos fazer é escolher, por exemplo, o valor de r que forneça a melhor previsão.

- **Vantagens e desvantagens do método**

As principais **vantagens** são:

(i) simples aplicação;

(ii) é aplicável quando se tem um número pequeno de observações;

(iii) permite uma flexibilidade grande devido à variação de r de acordo com o padrão da série:

e as **desvantagens** são:

(i) deve ser utilizado somente para prever séries estacionárias, caso contrário a precisão das previsões obtidas será muito pequena, pois os pesos atribuídos às r observações são todos iguais e nenhum peso é dado às observações anteriores a esse período;

(ii) necessidade de armazenar pelo menos $(r-1)$ observações; e

(iv) dificuldade em determinar o valor de r.

Na prática, o método de médias móveis não é utilizado frequentemente, pois o Método de Suavização Exponencial Simples, que veremos logo a seguir, possui todas as vantagens anteriores e mais algumas, que o tornam mais atraente.

Suavização exponencial simples (SES)

Em séries temporais é razoável pensar que as informações mais recentes tenham um peso maior que as informações mais antigas. Este é exatamente o conceito por trás da suavização exponencial simples. As previsões são calculadas usando médias ponderadas, onde os pesos diminuem à medida que as observações ficam mais longe do tempo atual.

- **Procedimento**

Suavização exponencial simples pode ser descrita matematicamente por

$$\hat{y}_{T+1|T} = \alpha\, y_T + \alpha(1-\alpha)\, y_{T-1} + \alpha(1-\alpha)^2\, y_{T-2} + \cdots,$$

Onde $0 \leq \alpha \leq 1$ é o parâmetro de suavização.. A previsão um passo à frente para o tempo $T+1$ é uma média ponderada de todas as observações na série $y_1, \ldots, y_T$. A taxa na qual os pesos diminuem é controlada pelo parâmetro α.

A tabela abaixo mostra os pesos anexados às observações para quatro valores diferentes de α ao prever usando suavização exponencial simples. Observe que a soma dos pesos, mesmo para um pequeno valor de α será aproximadamente um para qualquer tamanho de amostra razoável.

	$\alpha = 0,2$	$\alpha = 0,4$	$\alpha = 0,6$	$\alpha = 0,8$
y_T	$0,2000$	$0,4000$	$0,6000$	$0,8000$
y_{T-1}	$0,1600$	$0,2400$	$0,2400$	$0,1600$
y_{T-2}	$0,1280$	$0,1440$	$0,0960$	$0,0320$
y_{T-3}	$0,1024$	$0,0864$	$0,0384$	$0,0064$
y_{T-4}	$0,0819$	$0,0518$	$0,0154$	$0,0013$
y_{T-5}	$0,0655$	$0,0311$	$0,0061$	$0,0003$

Para qualquer α entre 0 e 1 , os pesos atribuídos às observações diminuem exponencialmente à medida que voltamos no tempo, daí o nome "suavização exponencial". Se α é pequeno (isto é, próximo de 0), mais peso é dado às observações do passado mais distante. Se α é grande (ou seja, próximo de 1), mais peso é dado às observações mais recentes.

- **Determinação da constante α**

Quanto menor for o valor de α mais estáveis serão as previsões finais, uma vez que a utilização de baixo valor de α implica que pesos maiores serão dados às observações passadas e, conseqüentemente, qualquer flutuação aleatória, no presente, exercerá um peso menor no cálculo da previsão.

Brown (1962) faz alguns comentários sobre a determinação dessa constante, de acordo com alguns critérios, tais como tipo de autocorrelação entre os dados e custo de previsão. Um procedimento mais objetivo é selecionar o valor que fornece a "melhor previsão" das observações já obtidas.

Representação alternativa

Podemos representar o modelo de suavização exponencial simples de outra maneira, através da **forma de componentes**. Para SES, o único componente incluído é o nível, ℓ_t. (Outros métodos que são considerados posteriormente também podem incluir uma tendência b_t e um componente sazonal s_t). As representações em forma de componentes dos métodos de suavização exponencial compreendem uma **equação de previsão** e uma **equação de suavização** para cada um dos componentes incluídos no método. A forma componente da suavização exponencial simples é dada por:

$$\begin{aligned} &\text{Equação de previsão:} \quad \hat{y}_{t+h|t} = \ell_t \\ &\text{Equação de suavização:} \quad \ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}, \end{aligned}$$

Onde ℓ_t é o nível (ou o valor suavizado) da série no tempo t.

Contexto $h = 1$ fornece os valores ajustados, enquanto configura $t = T$ fornece as previsões verdadeiras além dos dados de treinamento.

A equação de previsão mostra que o valor da previsão no tempo $t + 1$ é o nível estimado no instante se nós substituíssemos ℓ_t com $\hat{y}_{t+1|t}$ e ℓ_{t-1} com $\hat{y}_{t|t-1}$ na equação de suavização, recuperaremos a forma média ponderada da suavização exponencial simples.

Previsões fixas

A suavização exponencial simples tem uma função de previsão constante:

$$\hat{y}_{T+h|T} = \hat{y}_{T+1|T} = \ell_T, \quad h = 2, 3, \ldots$$

Ou seja, todas as previsões assumem o mesmo valor, igual ao componente de último nível. Dessa forma, podemos imaginar que se a série possui tendência ou sazonalidade as previsões não serão boas.

Método com tendência

Método de tendência linear de Holt

Holt (1957) estendeu a suavização exponencial simples para permitir a previsão de dados com tendência. Este método envolve uma equação de previsão e duas equações de suavização (uma para o nível e outra para a tendência):

$$\text{Equação de previsão: } \hat{y}_{t+h|t} = \ell_t + hb_t \tag{1}$$

$$\text{Equação de nível: } \ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1}) \tag{2}$$

$$\text{Equação de tendência: } b_t = \beta^* (\ell_t - \ell_{t-1}) + (1-\beta^*) b_{t-1} \tag{3}$$

em que ℓ_t denota uma estimativa do nível da série no instante t, b_t denota uma estimativa da tendência (inclinação) da série no instante t, α é o parâmetro de suavização para o nível, $0 \leq \alpha \leq 1$, e β^* é o parâmetro de suavização para a tendência, $0 \leq \beta^* \leq 1$.

Assim como na suavização exponencial simples, a equação de nível aqui mostra que ℓ_t é uma média ponderada de observação y_t e a previsão de treinamento um passo à frente para o tempo t, aqui dado por $\ell_{t-1} + b_{t-1}$. A equação de tendência mostra que b_t é uma média ponderada da tendência estimada no instante t baseado em $\ell_t - \ell_{t-1}$ e b_{t-1}, a estimativa anterior da tendência.

A função de previsão não é mais constante. A previsão h passos a frente é igual ao último nível estimado mais h vezes o último valor de tendência estimado. Portanto, as previsões são uma função linear de h.

Métodos de tendência amortecida

As previsões geradas pelo método linear de Holt apresentam uma tendência constante (aumentando ou diminuindo) indefinidamente no futuro. Evidências empíricas indicam que esses métodos tendem a superestimar, especialmente para horizontes de previsão mais longos. Motivados por esta observação, **Gardner & McKenzie (1985)** introduziram um parâmetro que "amortece" a tendência para uma linha plana em algum momento no futuro.

Em conjunto com os parâmetros de suavização α e β^* (com valores entre 0 e 1 como no método de Holt), este método também inclui um parâmetro de amortecimento $0 < \phi < 1$:

$$
\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t + \left(\phi + \phi^2 + \cdots + \phi^h\right) b_t \\
\ell_t &= \alpha y_t + (1 - \alpha)\left(\ell_{t-1} + \phi b_{t-1}\right) \\
b_t &= \beta^*\left(\ell_t - \ell_{t-1}\right) + (1 - \beta^*)\, \phi b_{t-1}
\end{aligned}
$$

se $\phi = 1$, o método é idêntico ao método linear de Holt. Para valores entre 0 e 1, ϕ amortece a tendência para que ela se aproxime de uma constante em algum momento no futuro. De fato, as previsões convergem para $\ell_T + \phi b_T/(1 - \phi)$ Como $h \to \infty$ para qualquer valor $0 < \phi < 1$. Isso significa que as previsões de curto prazo são tendenciosas, enquanto as previsões de longo prazo são constantes.

Na prática, ϕ raramente é inferior a 0,8 , pois o amortecimento tem um efeito muito forte para valores menores. Valores de ϕ próximo de 1 significa que um modelo amortecido não pode ser distinguido de um modelo não amortecido. Por essas razões, geralmente restringimos ϕ mínimo de 0,8 e máximo de 0,98.

Métodos com sazonalidade

Holt (1957) e Winters (1960) estenderam o método de Holt para capturar a sazonalidade. O método sazonal de HoltWinters compreende a equação de previsão e três equações de suavização - uma para o nível ℓ_t, um para a tendência b_t, e um para o componente sazonal s_t, com parâmetros de suavização correspondentes α, β^* e γ. Usamos m para denotar o período da sazonalidade, ou seja, o número de estações em um ano. Por exemplo, para dados trimestrais $m = 4$, e para dados mensais $m = 12$.

Existem duas variações para este método que diferem na natureza do componente sazonal. O **método aditivo** é preferido quando as variações sazonais são aproximadamente constantes ao longo da série, enquanto o **método multiplicativo** é preferido quando as variações sazonais estão mudando proporcionalmente ao nível da série. Com o método aditivo, o componente sazonal é expresso em termos absolutos na escala da série observada, e na equação de nível a série é ajustada sazonalmente subtraindo-se o componente sazonal. Dentro de cada ano, o componente sazonal somará aproximadamente expresso em termos relativos (porcentagens), e a série é ajustada sazonalmente dividindo-se pelo componente sazonal. Dentro de cada ano, m.

Método aditivo de Holt-Winters

A forma de componente para o método aditivo é:

$$\begin{aligned}
\hat{y}_{t+h|t} &= \ell_t + hb_t + s_{t+h-m(k+1)} \\
\ell_t &= \alpha\,(y_t - s_{t-m}) + (1-\alpha)\,(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta^*\,(\ell_t - \ell_{t-1}) + (1-\beta^*)\,b_{t-1} \\
s_t &= \gamma\,(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma) s_{t-m}
\end{aligned}$$

em que k é a parte inteira de $(h-1)/m$, que garante que as estimativas dos índices sazonais utilizados para a previsão sejam provenientes do último ano da amostra. A equação de nível mostra uma média ponderada entre a observação ajustada sazonalmente $(y_t - s_{t-m})$ e a previsão não sazonal $(\ell_{t-1} + b_{t-1})$ por instante t. A equação de tendência é idêntica ao método linear de Holt. A equação sazonal mostra uma média ponderada entre o índice sazonal atual, $(y_t - \ell_{t-1} - b_{t-1})$, e o índice sazonal da mesma estação do ano passado (ou seja, m períodos de tempo atrás).

Método multiplicativo de Holt-Winters

A forma de componente para o método multiplicativo é:

$$\begin{aligned}
\hat{y}_{t+h|t} &= (\ell_t + hb_t)\, s_{t+h-m(k+1)} \\
\ell_t &= \alpha \frac{y_t}{s_{t-m}} + (1-\alpha)\,(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta^*\,(\ell_t - \ell_{t-1}) + (1-\beta^*)\,b_{t-1} \\
s_t &= \gamma \frac{y_t}{(\ell_{t-1} + b_{t-1})} + (1-\gamma) s_{t-m}.
\end{aligned}$$

Exemplo: viagens domésticas noturnas na Austrália

Aplicamos o método de Holt-Winters com sazonalidade aditiva e multiplicativa para prever turismo doméstico noturno na Austrália. A Figura seguinte mostra os dados de 1998 – 2017 e as previsões para 2018 e 2020. Os dados mostram um padrão sazonal evidente, com picos observados no trimestre de março de cada ano, correspondendo ao verão australiano.

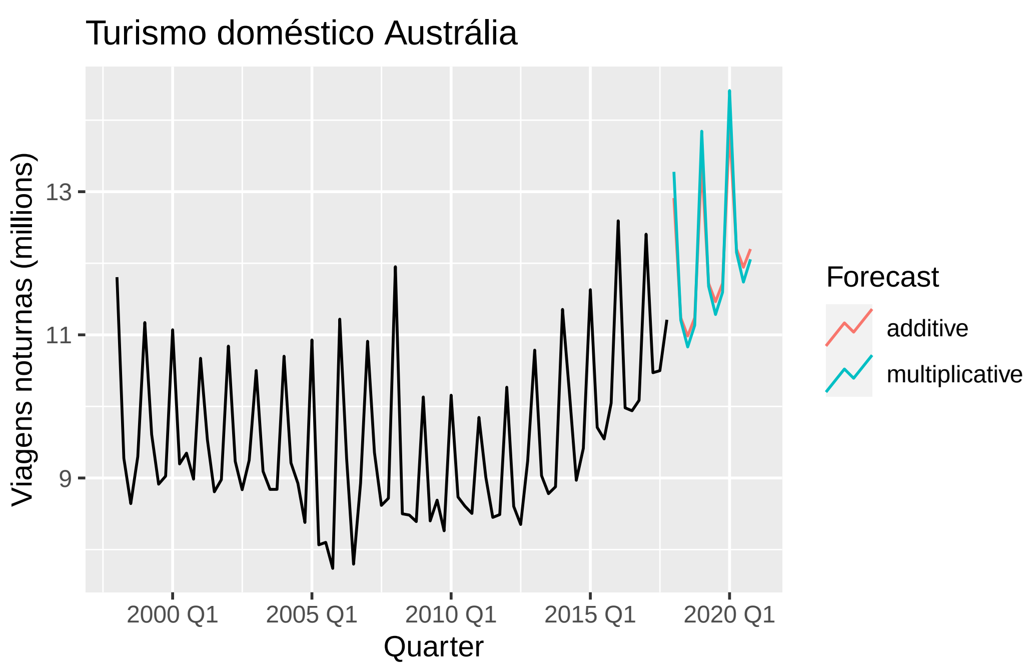
```
library(forecast)
```

```
Registered S3 method overwritten by 'quantmod':
  method            from
  as.zoo.data.frame zoo
```

```r
library(fpp3)
```

```
Warning: package 'fpp3' was built under R version 4.1.3

-- Attaching packages -------------------------------------------- fpp3 0.4.0 --

v tibble      3.1.6      v tsibble     1.1.2
v dplyr       1.0.9      v tsibbledata 0.4.1
v tidyr       1.2.0      v feasts      0.3.0
v lubridate   1.7.10     v fable       0.3.2
v ggplot2     3.3.5

Warning: package 'tibble' was built under R version 4.1.2

Warning: package 'dplyr' was built under R version 4.1.3

Warning: package 'tidyr' was built under R version 4.1.2

Warning: package 'tsibble' was built under R version 4.1.3

Warning: package 'tsibbledata' was built under R version 4.1.3

Warning: package 'feasts' was built under R version 4.1.3

Warning: package 'fabletools' was built under R version 4.1.3

Warning: package 'fable' was built under R version 4.1.3

-- Conflicts ------------------------------------------------- fpp3_conflicts --
x lubridate::date()      masks base::date()
x dplyr::filter()        masks stats::filter()
x fabletools::forecast() masks forecast::forecast()
x tsibble::intersect()   masks base::intersect()
x tsibble::interval()    masks lubridate::interval()
x dplyr::lag()           masks stats::lag()
x tsibble::setdiff()     masks base::setdiff()
x tsibble::union()       masks base::union()
```

```r
aus_holidays <- tourism %>%
  filter(Purpose == "Holiday") %>%
  summarise(Trips = sum(Trips)/1e3)

fit <- aus_holidays %>%
  model(
    additive = ETS(Trips ~ error("A") + trend("A") +
                                                season("A")),
    multiplicative = ETS(Trips ~ error("M") + trend("A") +
                                                season("M"))
  )

fc <- fit %>% forecast(h = "3 years")

fc %>%
  autoplot(aus_holidays, level = NULL) +
  labs(title="Turismo doméstico Austrália",
       y="Viagens noturnas (millions)") +
  guides(colour = guide_legend(title = "Forecast"))
```



Taxonomia dos métodos de suavização exponencial

Trend Component	Seasonal Component		
	N	A	M
	(None)	(Additive)	(Multiplicative)
N (None)	(N, N)	(N, A)	(N, M)
A (Additive)	(A, N)	(A, A)	(A, M)
A_d (Additive damped)	$(\mathrm{A}_d,\ \mathrm{N})$	$(\mathrm{A}_d,\ \mathrm{A})$	$(\mathrm{A}_d, \mathrm{M})$

Alguns desses métodos são resumidos a seguir:

Short hand	Method
(N, N)	Simple exponential smoothing
(A, N)	Holt's linear method
$(\mathrm{A}_d,\ \mathrm{N})$	Additive damped trend method
(A, A)	Additive Holt-Winters' method
(A, M)	Multiplicative Holt-Winters' method
$(\mathrm{A}_d, \mathrm{M})$	Holt-Winters' damped method