SHIFT

FIAP

PYTHON JOURNEY

MACHINE & DEEP LEARNING

ADELAIDE ALVES DE OLIVEIRA

PROFESSORA

profadelaide.alves@fiap.com.br

Formação Acadêmica

- Bacharel em Estatística – UNICAMP
- Mestre em Ciências – FSP/USP

Atividades Profissionais

- Diretora Técnica Estatística da empresa **SD&W: www.sdw.com.br**
- Professora de Fundamentos Estatísticos, DataMining, Análise Preditiva e Machine Learning na FIAP dos cursos MBA Big Data, MBA Business Intelligence & Analytics, Data Science e IA & ML e Shift em People Analytics e Python Journey

SÉRIES TEMPORAIS

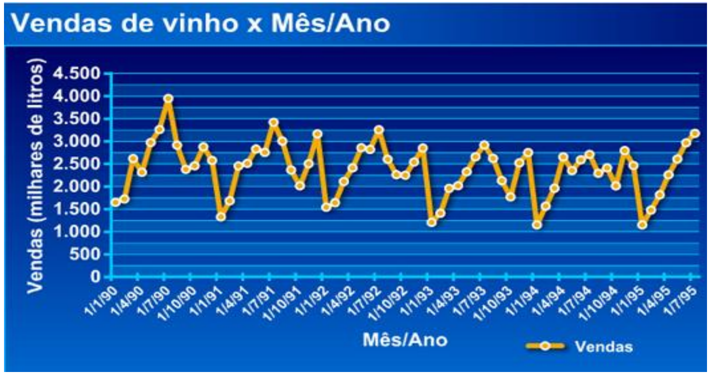
TÉCNICAS SUPERVISIONADAS
PREVISÃO E ESTIMAÇÃO

TÉCNICAS DE PREVISÃO:

TÉCNICAS QUANTITATIVAS

Essas técnicas podem ser agrupadas em:

- Modelos de séries temporais: Enfoca os padrões e suas mudanças, desenvolvido por meio de sua série histórica.



- Modelos causais: Utiliza informações refinadas e específicas sobre relações entre elementos do sistema.

$$Qualidade\ do\ Vinho = b0 + b_1x_1 + b_2x_2 + \ldots + b_nx_n$$

- Variáveis preditoras como: tipo do vinho, acidez, ph, açúcar, ...

PREVISÃO

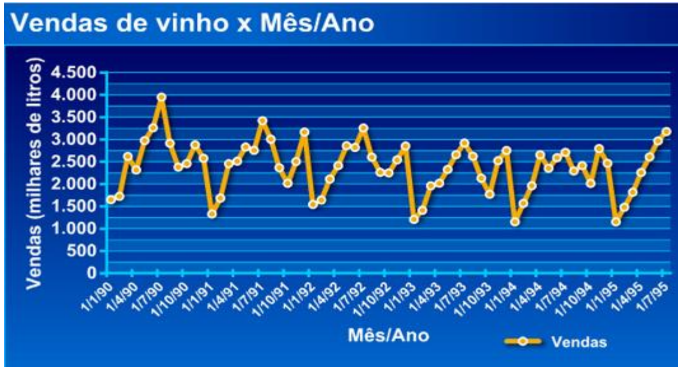
A escolha da técnica adequada depende:

- Horizonte de previsão.
  - Curto, médio e longo prazo.
- Acuracidade desejada.
- Relevância e disponibilidade de dados.
- Custo/benefício da previsão.
- Tempo disponível para modelagem.

TÉCNICAS DE PREVISÃO

MODELOS DE SÉRIES TEMPORAIS

O objetivo é identificar os padrões e suas mudanças, desenvolvido através de sua série histórica.



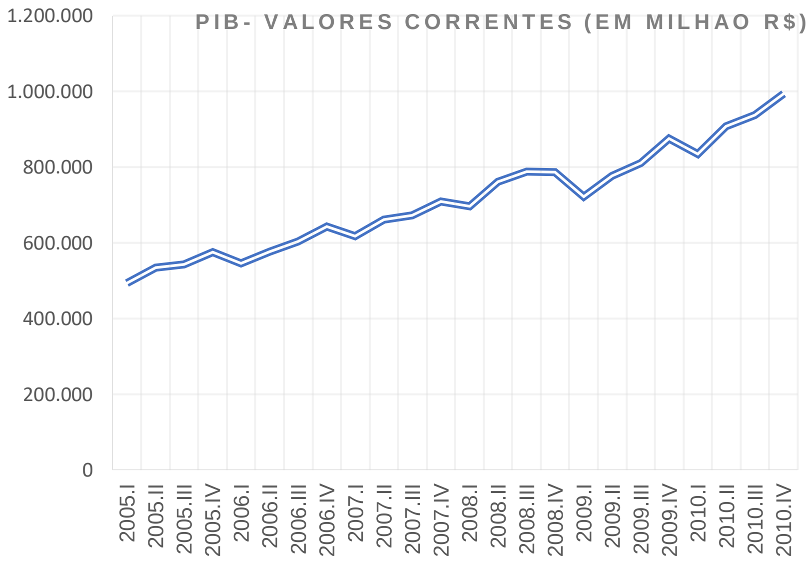
Utilização: As técnicas quantitativas são aplicadas nas condições:

- Informações históricas de pelo menos dois anos disponíveis;
- Informações quantificáveis em forma numérica;
- Assumir a hipótese de que algo dos padrões do passado irá se repetir no futuro (hipótese de continuidade).

SÉRIES TEMPORAIS

Considerações gerais:

- Uma série temporal é qualquer conjunto de observações ordenadas no tempo.
- Exemplos:
  - faturamento da campanha
  - número de pedidos
  - produção mensal
  - estoque mensal
  - PIB trimestral

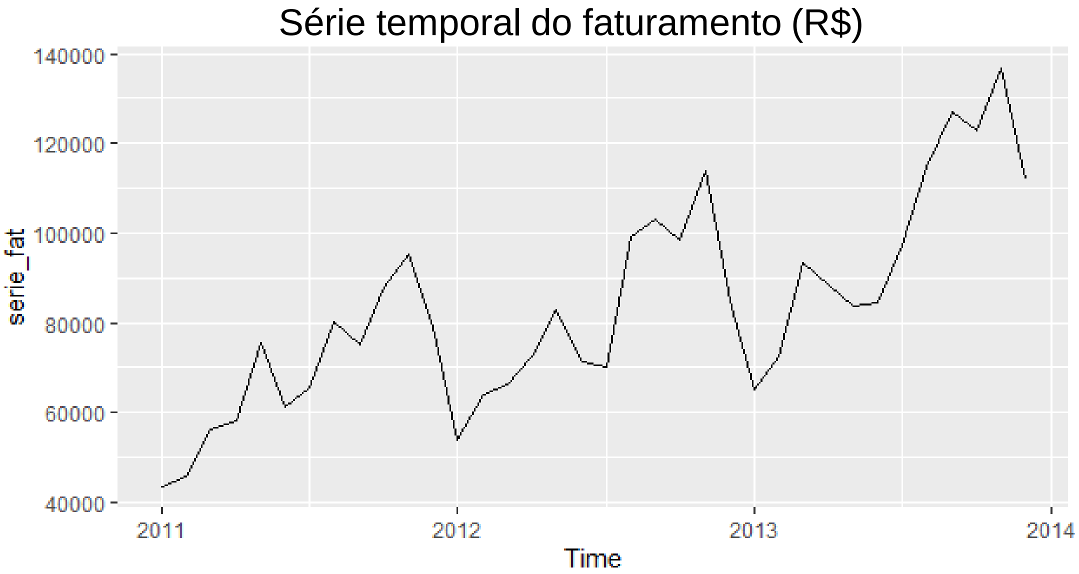


SÉRIES TEMPORAIS

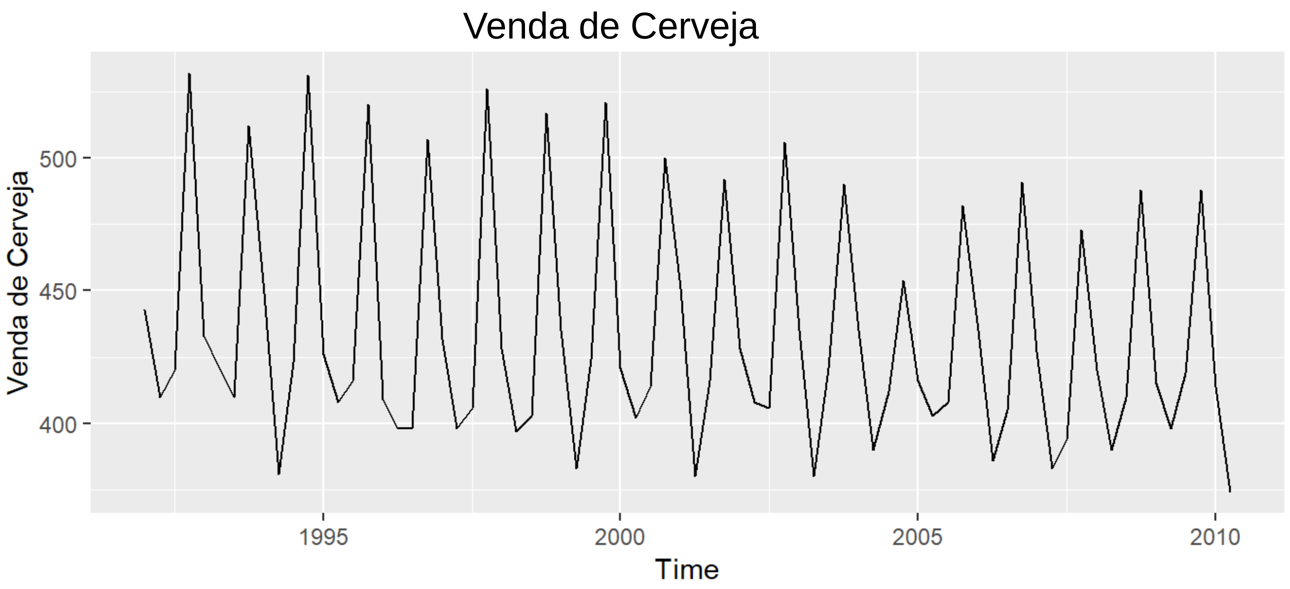
Principais objetivos ao analisar uma série temporal:

- Investigar o mecanismo gerador da série temporal; por exemplo, analisando uma série de altura de ondas, queremos saber como estas ondas foram geradas;
- Fazer previsões de valores futuros (curto ou longo prazo) da série;
- Descrever apenas o comportamento da série;
- Procurar periodicidade relevante nos dados.

SÉRIES TEMPORAIS
EXEMPLOS



SÉRIES TEMPORAIS
EXEMPLOS



SÉRIES TEMPORAIS
FREQUÊNCIA DA SÉRIE

UNIDADE DE ANÁLISE	FREQUÊNCIA
Anual	1
Mensal	12
Diária	365
Trimestral	4
Semanal	52

SÉRIES TEMPORAIS
MODELOS TRADICIONAIS vs MACHINE LEARNING

Modelos Tradicionais

- Não permite predição em períodos mais a frente(uma predição atrás da outra);
- Modelo para no tempo t está pronto para predizer o tempo t+1;
- Difícil de "tunar", precisa conhecer mais a fundo sobre séries temporais;
- Não pode adicionar atributos que variam no tempo

Machine Learning

- Permite fazer predição diretamente para qualquer tempo (daqui a 1, 5 ou 8 meses por exemplo);
- Modelos diferentes, treinar tudo de novo (1 modelo para 1 mês, outro modelo para 2 meses (exceto seq2seq);
- Mais fácil de acertar;
- Pode adicionar atributos que variam no tempo
- criando "features" autorregressivas, ou seja, valores históricos, máximos, mínimos, média, lags.

SÉRIES TEMPORAIS
MODELOS TRADICIONAIS vs MACHINE LEARNING

TIPOS DE MODELOS

➔ Tradicional: Exemplos
  Univariado: ARIMA, SARIMAX, Prophet, Neural Prophet
  Multivariado: Modelo Vetorial Autorregressivo (VAR).

➔ Machine Learning : Qualquer algoritmo Regressor:
  NeuralNetwork Regressor (MLP),CatBoost Regressor, XGBoosting Regressor, RandomForest Regressor , SVM, etc

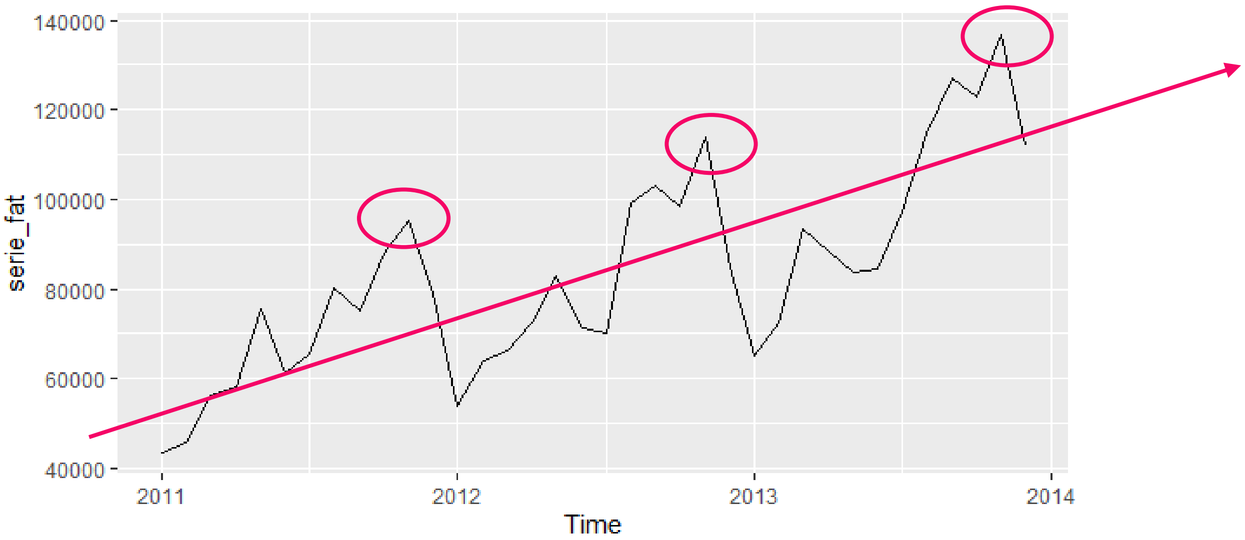
➔ Modelos “Avançados”:
  Redes neurais recursivas (já nasceram temporais)
  LSTM: Long Short-Term Memory
  GRU: Gated Recurrent Memory
  TCN: CNN Redes Neurais Convolucionais para séries temporais(kernels identificam “características” no tempo)
  Transformers: também em séries temporais

SÉRIES TEMPORAIS

Quando o comportamento de uma série temporal está sujeita a fatores desconhecidos, ou seja, a efeitos aleatórios, pode-se encontrar um modelo que é obtido com base no cálculo de probabilidade. Tal modelo é chamado de estocástico. Um processo estocástico é caracterizado por uma família de variáveis aleatórias que descrevem a evolução de um fenômeno de interesse, neste caso a evolução temporal da série em estudo.

SÉRIES TEMPORAIS

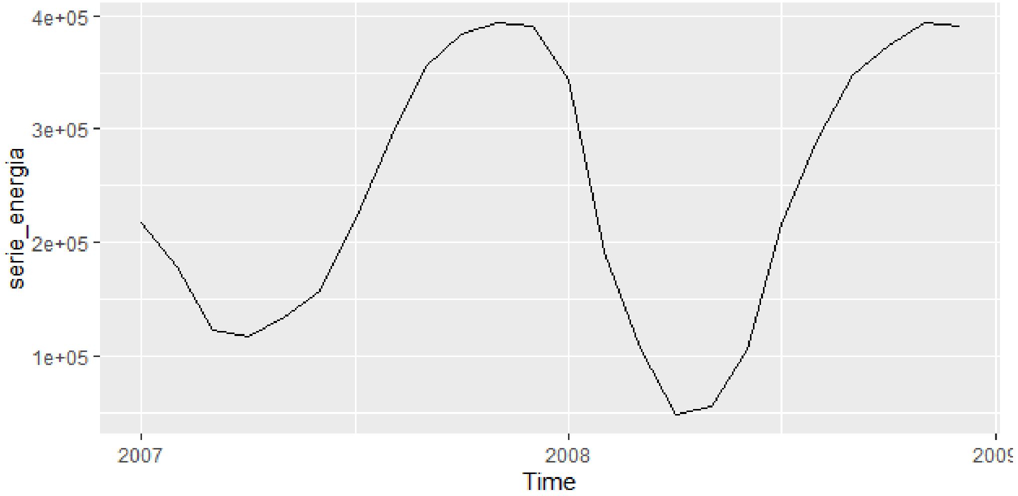
Exemplo: Série temporal do faturamento (R$)



A série apresenta tendência? Sazonalidade?

SÉRIES TEMPORAIS

Exemplo: Consumo de energia elétrica (Kw/h) de empresas do setor Agricultura



A série apresenta tendência? Sazonalidade?

SÉRIES TEMPORAIS
CORRELAÇÃO

A autocorrelação é a medida de correlação entre uma variável e valores passados da mesma.

Autocorrelação total, a correlação entre observação X_t e a observação X_{t-p} é calculada levando em consideração a dependência linear das observações intermediárias X_{t-1}. X_{t-2},.....,X_{t-p+1}.

Autocorrelação parcial, por outro lado, calcula a correlação entre a observação X_t e X_{t-p} eliminando a dependência linear das observações intermediárias.

SÉRIES TEMPORAIS

A análise de séries temporais visa identificar e explicar:

Série = T + S + C + a

T: Tendência

S: Sazonalidade

C: Ciclo

A: Aleatório

Tendência – evolução do fenômeno de interesse. Verifica o sentido de deslocamento da série ao longo de vários anos.

Sazonalidade – regularidade ou variação sistemática na série de dados;

Padrões Cíclicos – repetição de padrão num prazo superior a 2 anos;

Aleatório – comportamento não explicável pelos três componentes anteriores. Erro Aleatório.

SÉRIES TEMPORAIS

COMPONENTE: TENDÊNCIA ESTIMATIVA

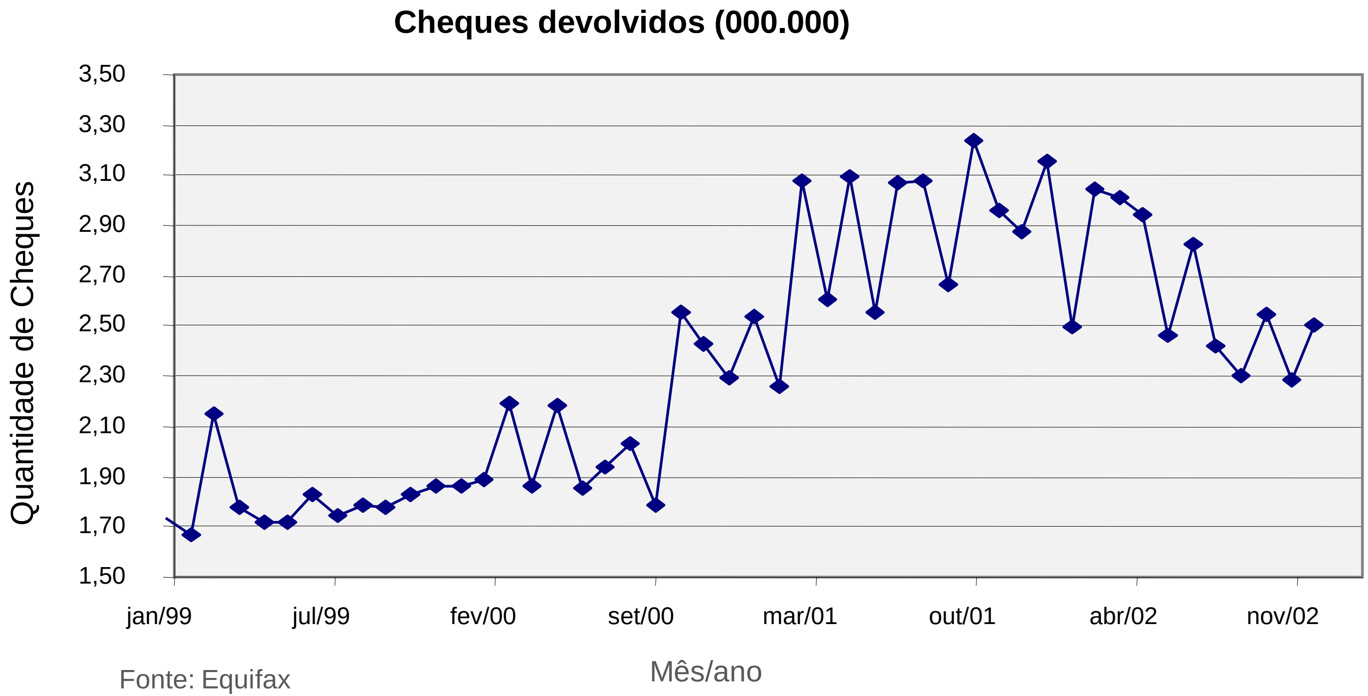
Representa a evolução da série ao longo do período.

A tendência de uma série temporal pode ser estimada por vários métodos:

- Ajuste de uma função do tempo (**Regressão)**
- Suavização dos valores ao redor de um ponto, estimando assim a tendência (**médias móveis**)

SÉRIES TEMPORAIS

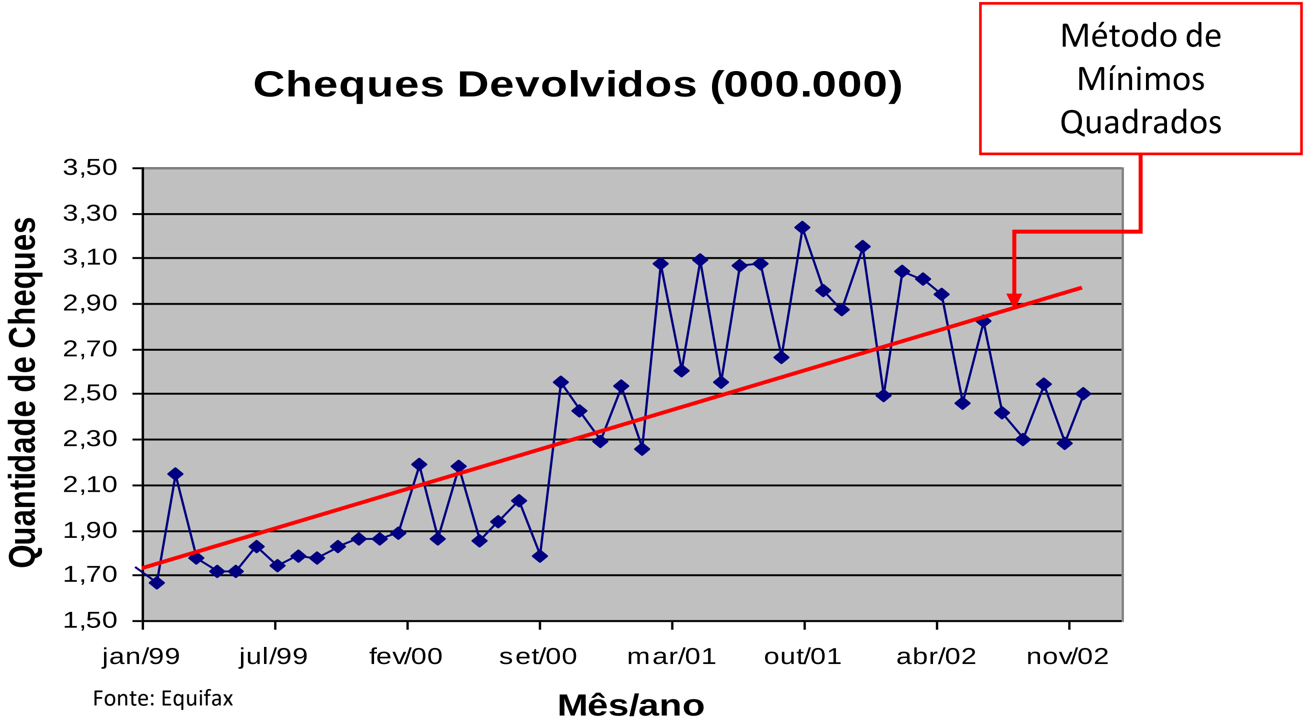
COMPONENTE: TENDÊNCIA ESTIMATIVA



Fonte: Equifax

SÉRIES TEMPORAIS

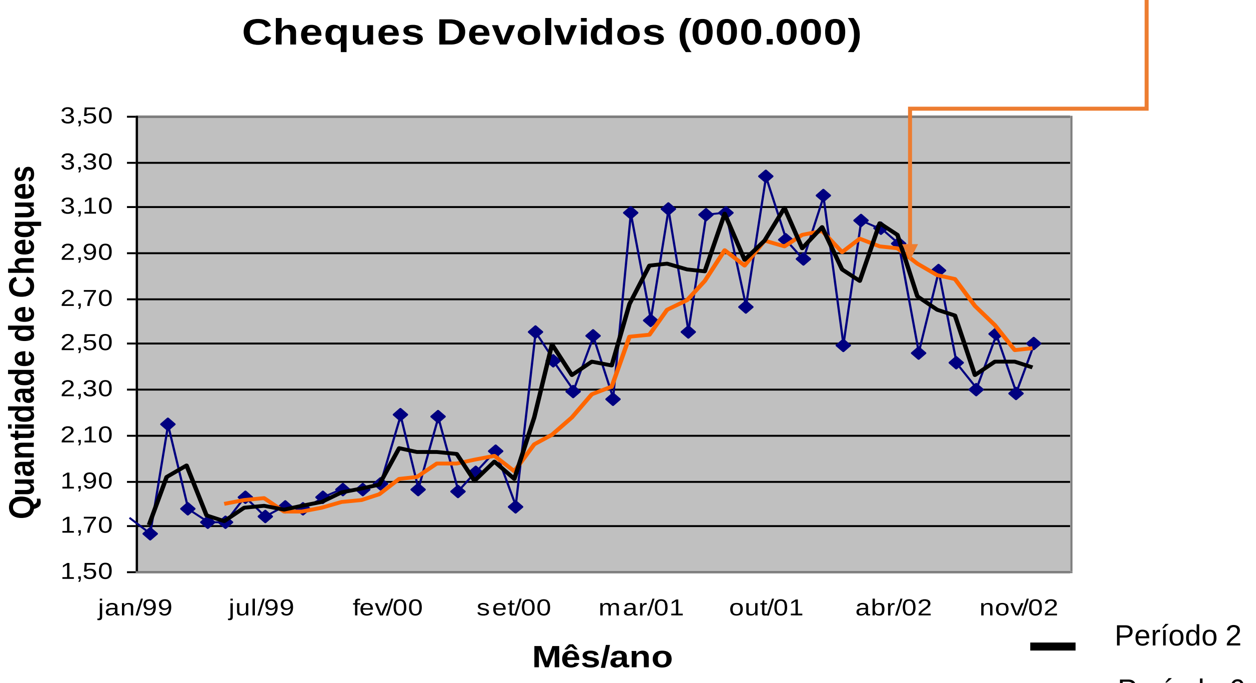
COMPONENTE: TENDÊNCIA ESTIMATIVA



Fonte: Equifax

SÉRIES TEMPORAIS
COMPONENTE: TENDÊNCIA ESTIMATIVA

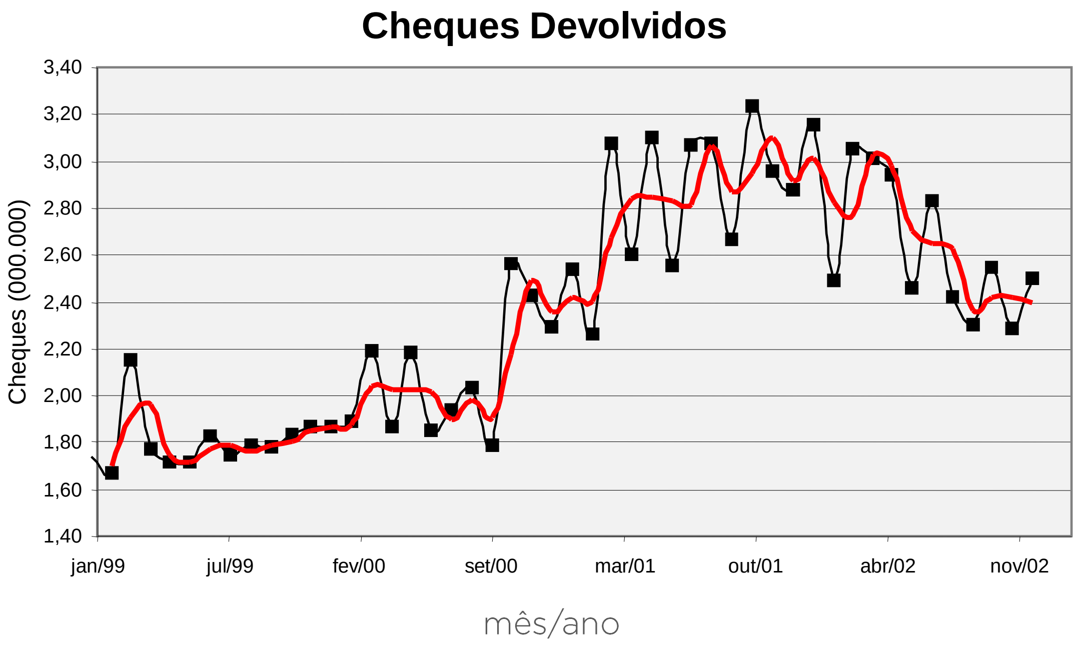
Método de Médias Móveis



SÉRIES TEMPORAIS

COMPONENTE: TENDÊNCIA ESTIMATIVA

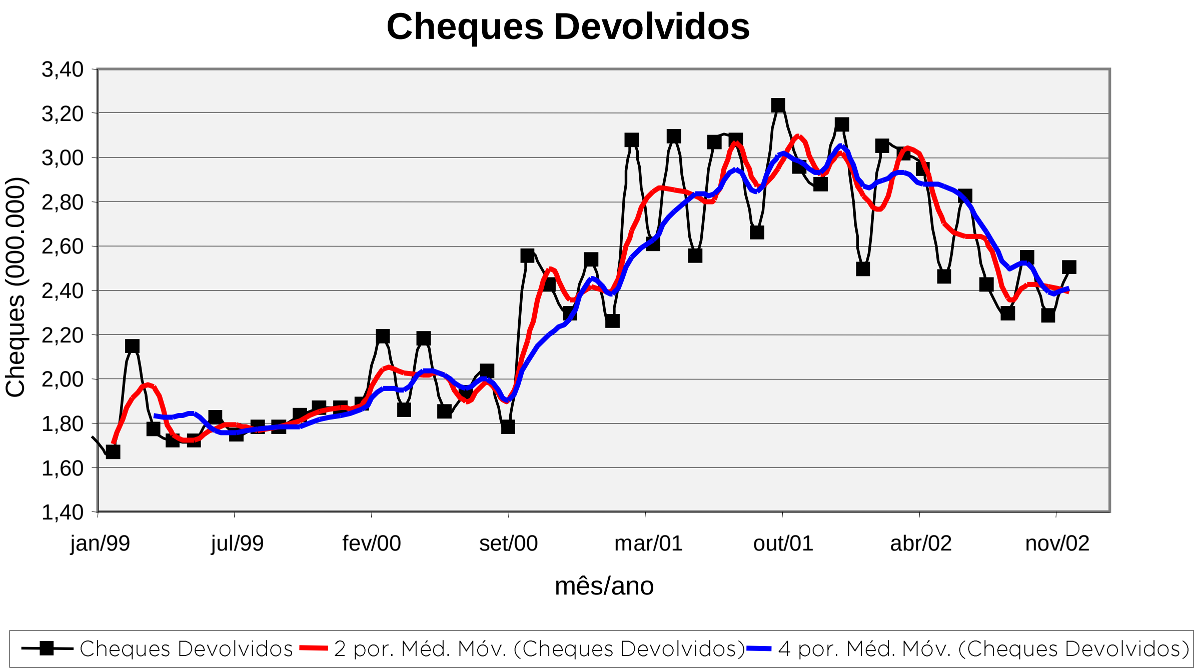
Médias Móveis – PERÍODO 2



SÉRIES TEMPORAIS

COMPONENTE: TENDÊNCIA ESTIMATIVA

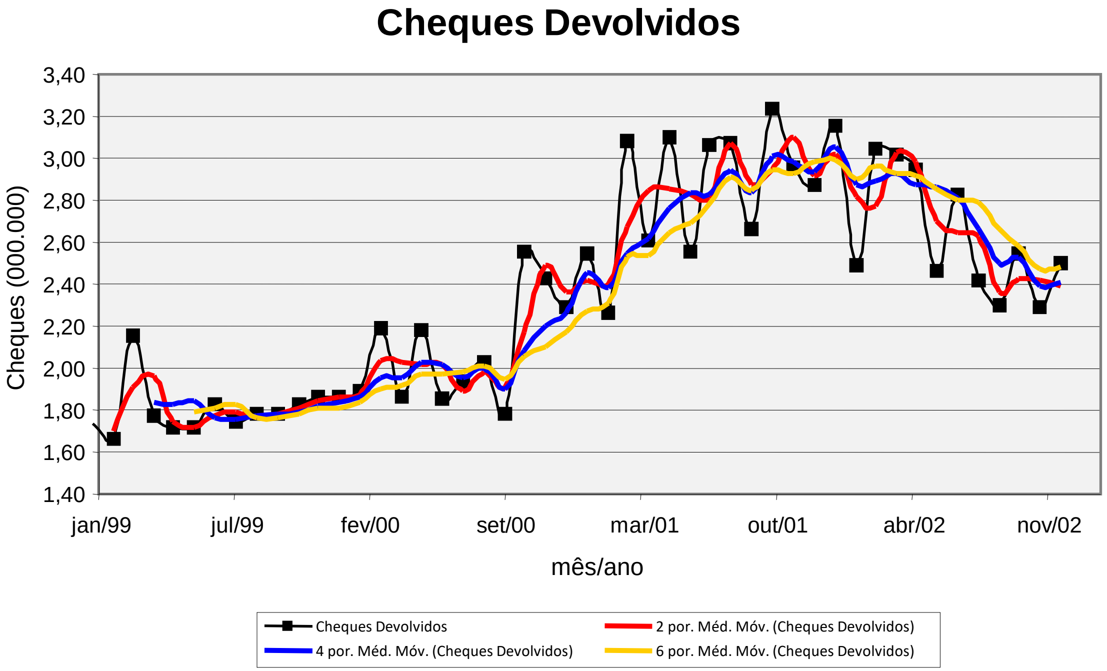
Médias Móveis – PERÍODOS 2 E 4



SÉRIES TEMPORAIS

COMPONENTE: TENDÊNCIA ESTIMATIVA

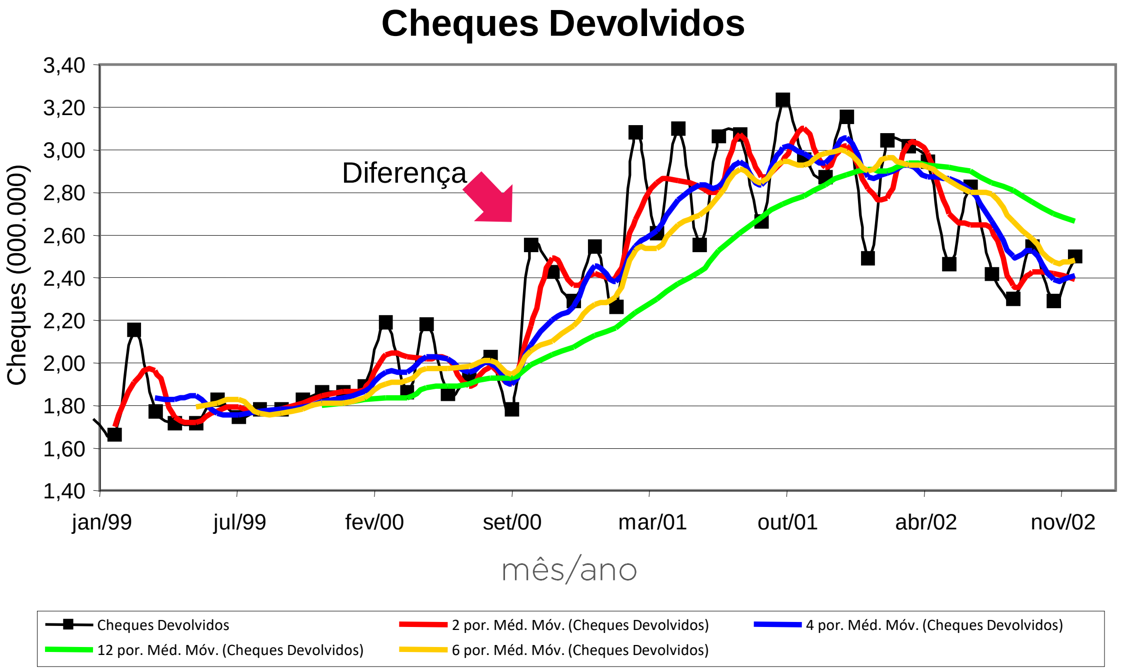
Médias Móveis – PERÍODOS 2, 4 E 6



SÉRIES TEMPORAIS

COMPONENTE: TENDÊNCIA ESTIMATIVA

Médias Móveis – PERÍODOS 2, 4, 6 e 12



SÉRIES TEMPORAIS

COMPONENTE: TENDÊNCIA ESTIMATIVA

Média Móvel

Cada ponto de uma média móvel é definido como a média aritmética ou ponderada de um número de pontos consecutivos.

O número de pontos a ser escolhido depende dos efeitos da sazonalidade e/ou irregularidade, que deverão ser por ele eliminados.

A determinação do tamanho do período deverá ser criteriosa.

- um valor grande o acompanhamento é muito lento
- um valor pequeno implica numa reação muito rápida

O tamanho escolhido deverá ser o que minimiza o erro quadrático médio.

SÉRIES TEMPORAIS

COMPONENTE: TENDÊNCIA ESTIMATIVA

Média Móvel

A Média Móvel consiste em calcular a média das r observações mais recentes.

O nome de Média Móvel é utilizado pois, a cada período, a observação mais antiga é substituída pela mais recente, calculando-se nova média.

A previsão é dada pela última média calculada.

Exemplos:

série	10	12	13	14	16	18
média móvel ordem 2		11	12.5	13.5	15	17

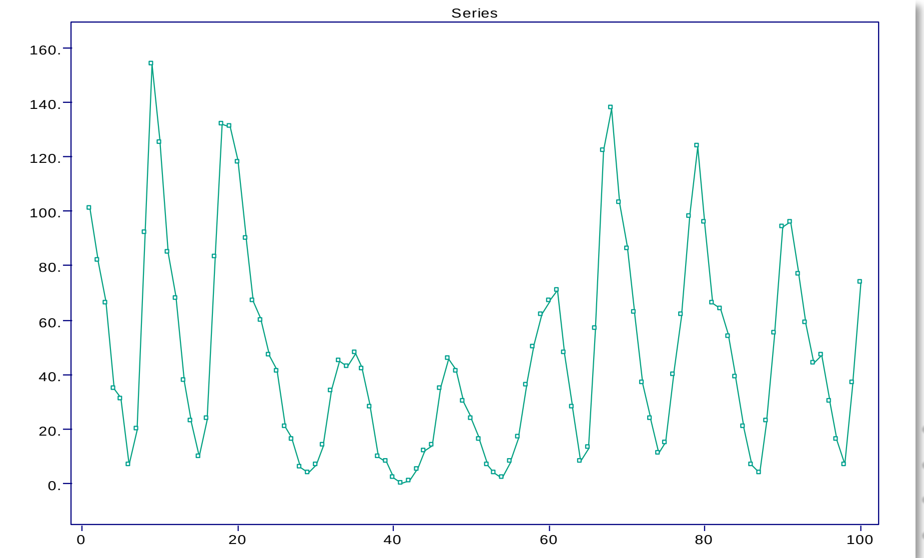
série	10	12	13	14	16	18
média móvel ordem 3			11.7	13.0	14.3	16.0

SÉRIES TEMPORAIS

COMPONENTE: CICLO

São os movimentos repetidos em período superior a um ano, sendo ou não periódicos.

Para sua estimativa pode-se utilizar as médias móveis ou outros métodos de correlação (auto).



SÉRIES TEMPORAIS

COMPONENTE: SAZONALIDADE

Considera-se a sazonalidade a existência de fenômenos que ocorrem regularmente de ano para ano, semana a semana dentro mês, ou ainda de dia para dia, dentro da semana.

As causas principais de flutuações sazonais podem ser:

Calendário – Natal, Carnaval, Páscoa,...

Decisões sobre um tempo pre-definido – férias,recebimento de salário,..

Temperatura – Estação

Expectativa – Liquidação de inverno em Séries de Preço

SÉRIES TEMPORAIS
COMPONENTE: SAZONALIDADE

Calendário

Quando houver série de dados afetada por dias dentro do mês, é importante que antes de qualquer análise seja corrigida, pois poderá afetar o movimento sazonal, além de causar correlações espúrias entre séries.

Neste sentido poderá ser utilizado o seguinte fator de correção:

$$\text{Fator}_t = \frac{\text{No. de dias efetivos no mês médio}}{\text{No. de dias efetivos no mês } t}$$

SÉRIES TEMPORAIS

COMPONENTE: SAZONALIDADE

Calendário - **Exemplo:**

$$\text{Fator}_{\text{março}} = \frac{\text{No. de dias efetivos no mês médio}}{\text{No. de dias efetivos no mês março}} = \frac{20}{18} = 1{,}1$$

$$\text{Fator}_{\text{agosto}} = \frac{\text{No. de dias efetivos no mês médio}}{\text{No. de dias efetivos no mês agosto}} = \frac{20}{22} = 0{,}99$$

SÉRIES TEMPORAIS
COMPONENTE: SAZONALIDADE

ÍNDICES SAZONAIS

A definição de um Índice Sazonal tem por objetivo se eliminar ou neutralizar o efeito sazonal de uma série.

Os métodos aqui apresentados sempre se baseiam na decomposição, ou seja,

1. Calcula-se a tendência
2. Elimina-se a tendência, restando a sazonalidade e aleatório
3. Calcula-se o Índice Sazonal, que calculado sobre a série integral, tem-se a série desazonalizada.

SÉRIES TEMPORAIS

COMPONENTE: SAZONALIDADE

ÍNDICES SAZONAIS

Método da Porcentagem Média

1. os dados de cada mês são expressos em percentagens da média anual;

2. as percentagens dos meses correspondentes, para diferentes anos, são balanceadas mediante o emprego de uma nova média;

3. as 12 percentagens resultantes dão os índices de sazonalidade.

SÉRIES TEMPORAIS
COMPONENTE: SAZONALIDADE

Obtenção da componente de Sazonalidade para o modelo aditivo:

- Após remover Tendência da série temporal, resta apenas as componentes de Sazonalidade e os Ruídos Aleatórios: $Y_t - T_t = S_t + \varepsilon_t$
- Para estimar a Sazonalidade, calcula-se as diferenças entre o valor observado (Yt) e as médias móveis centradas para cada mês do ano, conforme a equação acima.

Obtenção da componente de Sazonalidade para o modelo multiplicativo:

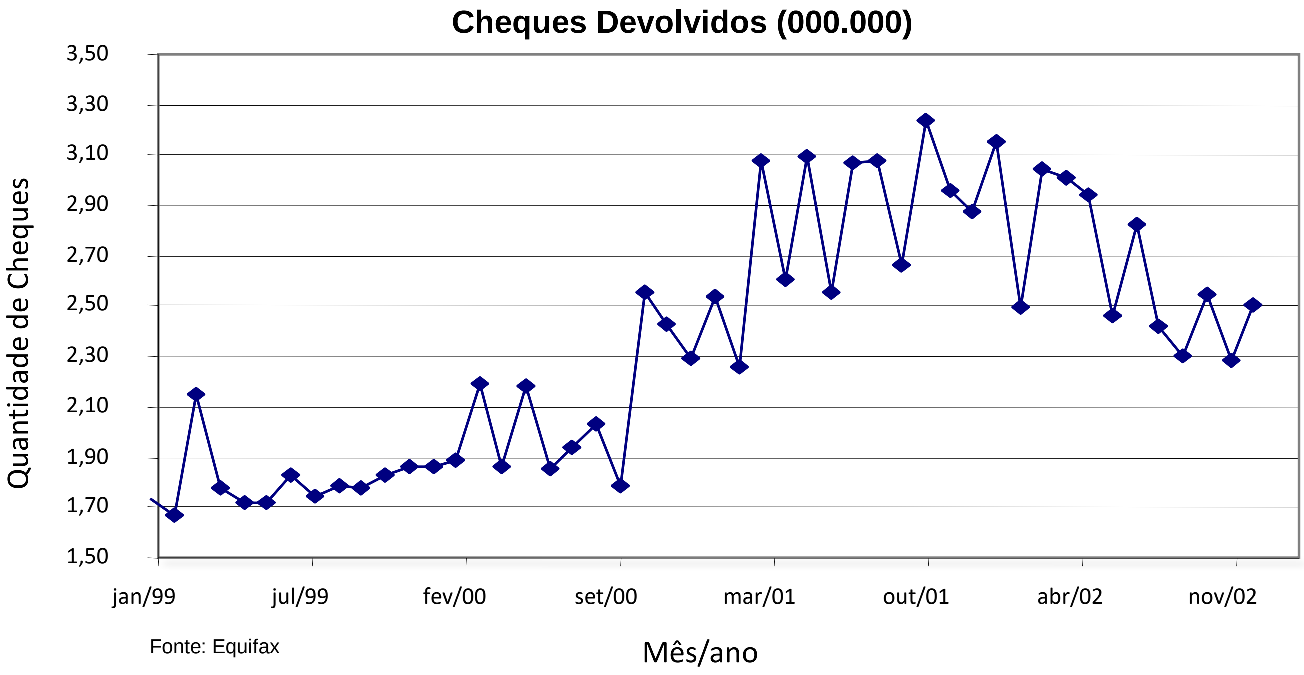
- Após remover Tendência da série temporal, resta apenas as componentes de Sazonalidade e os Ruídos Aleatórios:

$$\frac{Y_t}{T_t} = \frac{T_t \times S_t \times \varepsilon_t}{T_t} = S_t \times \varepsilon_t$$

- As componentes de Sazonalidade são obtidas através da técnica da razão de médias móveis. Esta técnica consiste em fazer a razão entre o valor observado na série e o valor da média móvel.

SÉRIES TEMPORAIS

COMPONENTE: SAZONALIDADE



Fonte: Equifax

SÉRIES TEMPORAIS

ÍNDICES SAZONAIS: MÉTODO DA PORCENTAGEM MÉDIA

Exemplo

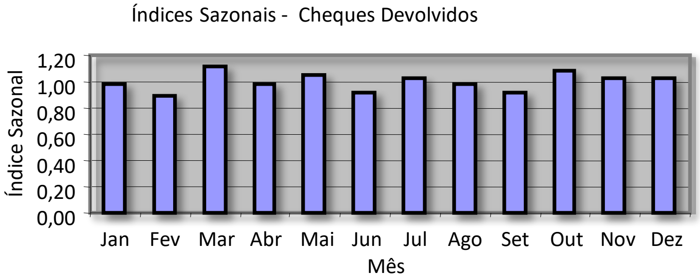
Evolução dos Cheques Devolvidos (Jan/99 a Dez/03)

	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Média
1999	1,74	1,67	2,15	1,78	1,72	1,72	1,83	1,75	1,79	1,78	1,83	1,87	1,80
2000	1,87	1,89	2,19	1,86	2,18	1,85	1,94	2,03	1,78	2,56	2,42	2,29	2,07
2001	2,54	2,26	3,08	2,60	3,10	2,56	3,07	3,07	2,66	3,24	2,96	2,88	2,83
2002	3,15	2,49	3,05	3,01	2,94	2,46	2,83	2,42	2,30	2,55	2,29	2,50	2,67

Índices de Sazonalidade para a Evolução dos Cheques Devolvidos

	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Média
1999	0,96	0,93	1,19	0,99	0,95	0,95	1,02	0,97	0,99	0,99	1,02	1,04	1,00
2000	0,90	0,91	1,06	0,90	1,05	0,89	0,94	0,98	0,86	1,24	1,17	1,11	1,00
2001	0,90	0,80	1,09	0,92	1,09	0,90	1,08	1,08	0,94	1,14	1,04	1,01	1,00
2002	1,18	0,93	1,14	1,13	1,10	0,92	1,06	0,91	0,86	0,96	0,86	0,94	1,00
Sazonalidade	0,99	0,89	1,12	0,98	1,05	0,92	1,02	0,99	0,91	1,08	1,02	1,02	1,00

Jan/2002
=3,15/2,67



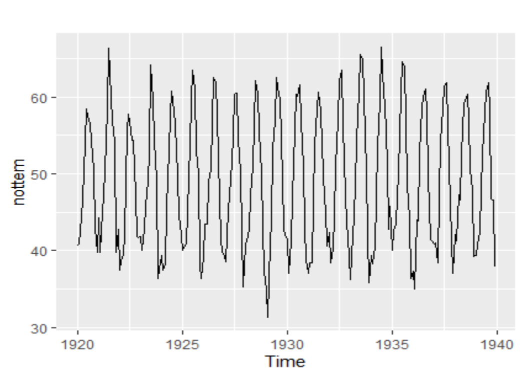
SÉRIES TEMPORAIS

CONCEITO DE ESTACIONARIEDADE

Uma série temporal estacionária é aquela cujas propriedades estatísticas, como a média, a variância e a auto correlação, *<u>são constantes ao longo do tempo</u>*. E, uma série não estacionária é aquela cujas propriedades estatísticas mudam com o tempo.

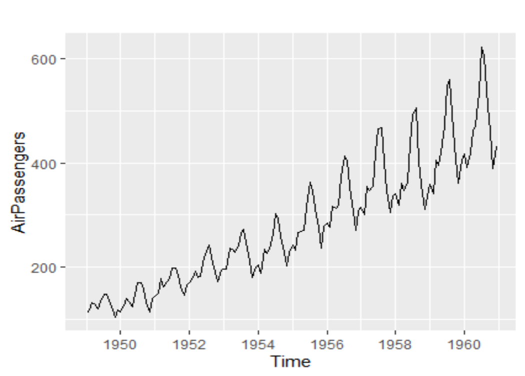
SÉRIE ESTACIONÁRIA

Average Monthly Temperatures at Nottingham, 1920–1939



SÉRIE NÃO ESTACIONÁRIA

Monthly Airline Passenger Numbers, 1949-1960

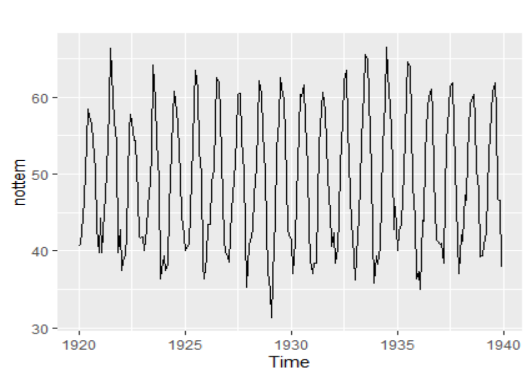


SÉRIES TEMPORAIS

SÉRIE ESTACIONÁRIA

SÉRIE ESTACIONÁRIA

Average Monthly Temperatures at Nottingham, 1920–1939



série sem tendência e sazonalidade

Um processo estocástico Z_t é estacionário quando as propriedades estatísticas* de qualquer sequência finita $z_1, z_2, .. z_k$ de componentes de Z_t são semelhantes às da sequência $z_{1+h}, z_{2+h}, .. z_{k+h}$ para qualquer número inteiro h

* Propriedades

$$E[z_t] = E[z_{t+k}] = \mu$$

$$Var[z_t] = E[(z_t - \mu)^2] = \sigma^2$$

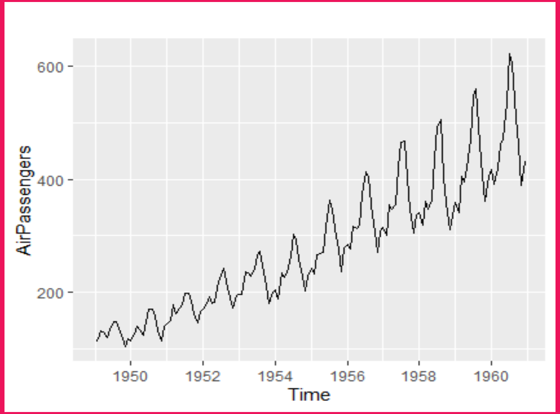
$$Cov[z_t, z_{t+k}] = E[(z_t - \mu) \cdot (z_{t+k} - \mu)]$$

SÉRIES TEMPORAIS

SÉRIE NÃO ESTACIONÁRIA

SÉRIE NÃO ESTACIONÁRIA

Monthly Airline Passenger Numbers, 1949-1960



série com tendência e sazonalidade

Um processo estocástico Z_t é NÃO estacionário quando as propriedades estatísticas de ao menos uma sequência finita $z_1, z_2, .. z_k$ de componentes de Z_t são diferentes das de sequência $z_{1+h}, z_{2+h}, .. z_{k+h}$ para ao menos um número inteiro h .

SÉRIES TEMPORAIS

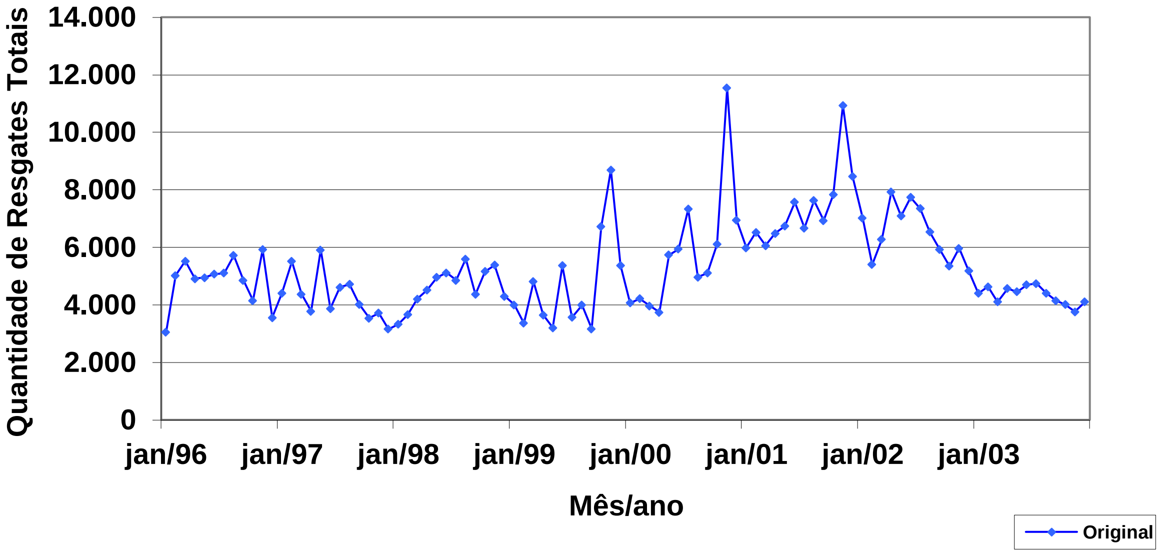
SÉRIE ESTACIONÁRIA

É indicado utilizar testes estatísticos para confirmar se a séries é estacionária. Um teste utilizado com esse objetivo é o **teste de Dickey Fuller.**
Utilizando um alfa de 5%, por exemplo, caso o valor p-value esteja abaixo desses 5% significa que a série é estatisticamente estacionária.

SÉRIES TEMPORAIS

SÉRIE NÃO ESTACIONÁRIA

Exemplo



SÉRIES TEMPORAIS
SÉRIE NÃO ESTACIONÁRIA

ELIMINAÇÃO: Usar uma transformação da série

A partir da estimativa da tendência, pode-se eliminá-la através de:

$$X^* = X - T$$

ou

$$X^* = X / T$$

SÉRIES TEMPORAIS
SÉRIE NÃO ESTACIONÁRIA

ELIMINAÇÃO

Ainda se elimina a tendência através de diferenças, ou seja, utiliza-se:

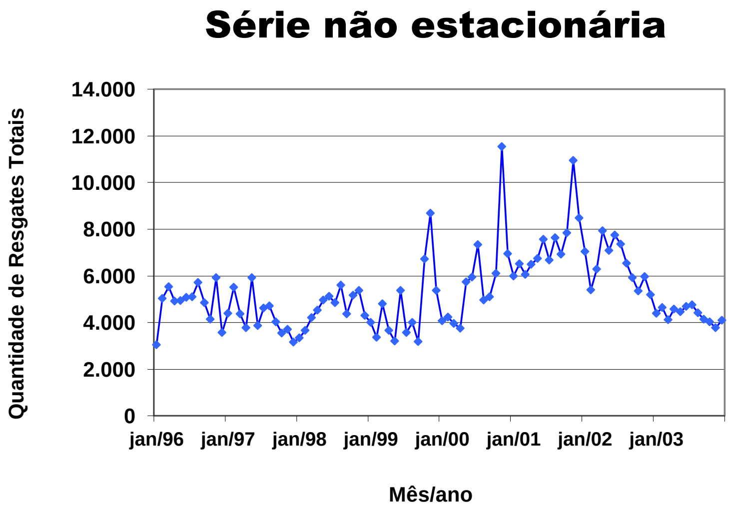
Primeira diferenciação: $Y_t = X_{t+1} - X_t$

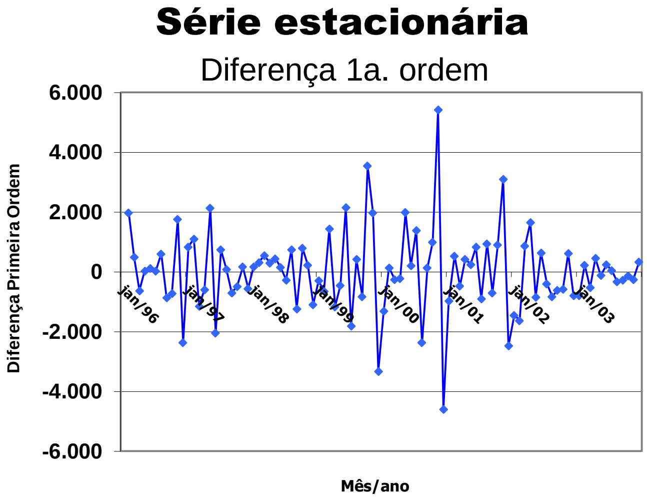
Ou, Diferença de ordem superior (2a, 3a..)

SÉRIES TEMPORAIS
SÉRIE NÃO ESTACIONÁRIA

ELIMINAÇÃO POR DIFERENÇA

Exemplo:





SÉRIES TEMPORAIS

ALGUNS TIPOS DE MODELOS

- Método de Decomposição
- Alisamento Exponencial ou Suavização Exponencial
- Modelos de Box Jenkins

No universo dos métodos de previsão de séries temporais encontram-se os modelos Autorregressivos e Médias Móveis (AR, MA e ARMA), modelos Autorregressivos Integrado de Médias Móveis (ARIMA), Filtros de Kalman e AEP, modelos ARARMA de Parzen, modelos ARMA Multivariáveis (MARMA), e da análise desses padrões derivam uma série de modelo SARIMA, SARIMAX, VARIMA, VARIMAX, entre outros.

SÉRIES TEMPORAIS
MODELO DE DECOMPOSIÇÃO

Modelos de Decomposição consistem em, como o próprio nome diz, decompor o modelo que descreve o comportamento da série temporal através de suas componentes:

- Valor da série temporal = padrão temporal + erro
- Valor da série temporal = ***função***(Tendência, Ciclo, Sazonalidade., Ruído Aleatório)

  ou Y_t = $\boldsymbol{f}$ $(T_t, C_t, S_t, \varepsilon_t)$ onde:

  - Y_t valor da série temporal
  - T_t componente Tendência no período t
  - C_t componente Ciclo no período t
  - S_t componente de Sazonalidade no período t
  - ε_t componente de erro ou Ruído Aleatório no período t

SÉRIES TEMPORAIS
MODELO DE DECOMPOSIÇÃO

Uma observação:

É comum incluir a componente Ciclo no modelo para representar movimentos com períodos longos. Contudo, como em geral os períodos observados são pequenos quando comparados com o tamanho do Ciclo, muitas vezes o que se observa ao analisar o efeito da Tendência é parte do Ciclo. Este fato nos leva a confundir o efeito do componente Ciclo com a componente Tendência. Por isto, alguns autores* utilizam para representar o Modelo de Decomposição a seguinte expressão: $Y_t = \boldsymbol{f}\,(T_t, S_t, \varepsilon_t)$

*Makridakis et al. (1998) e Morettin e Toloi (1987)

SÉRIES TEMPORAIS
MODELO DE DECOMPOSIÇÃO

Encontramos dois tipos de métodos de composição: aditivo ou multiplicativo.

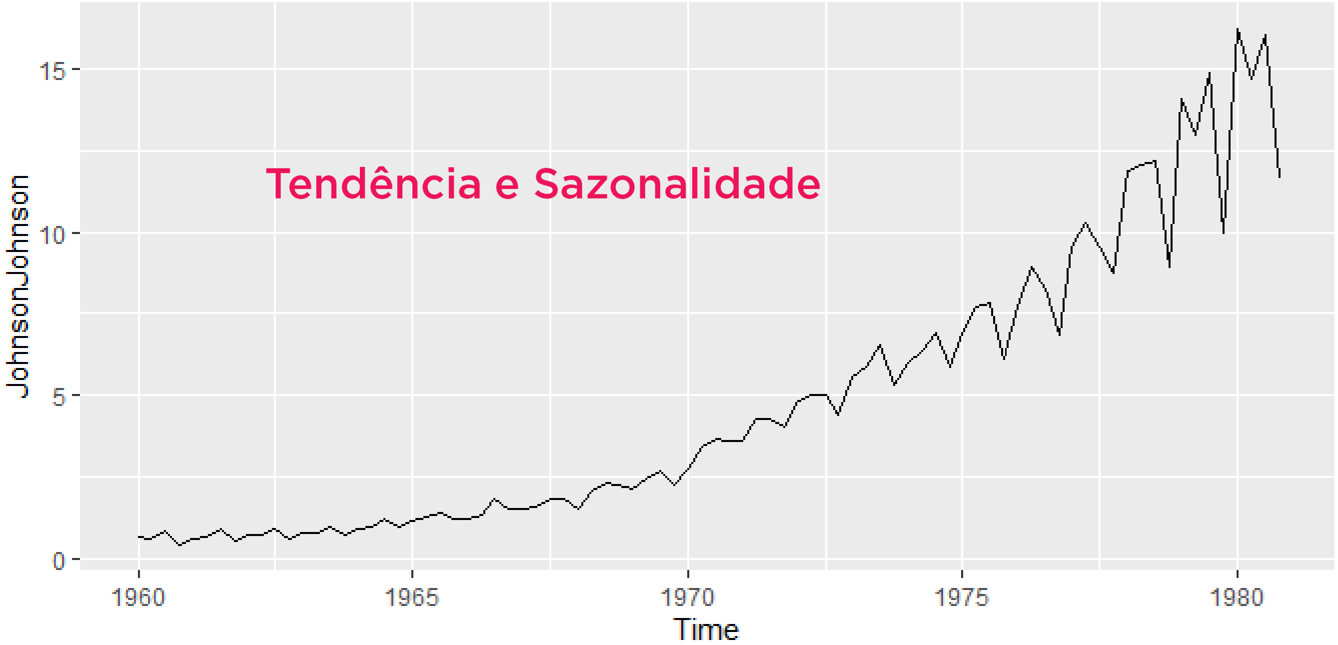
O Modelo de Decomposição Aditivo considera que a série temporal é resultante da soma das componente: $Y_t = T_t+S_t+\varepsilon_t$

O Modelo de Decomposição Multiplicativo considera que a série temporal é resultante do produto das componentes: $Y_t = T_t*S_t*\varepsilon_t$

SÉRIES TEMPORAIS

MODELO DE DECOMPOSIÇÃO

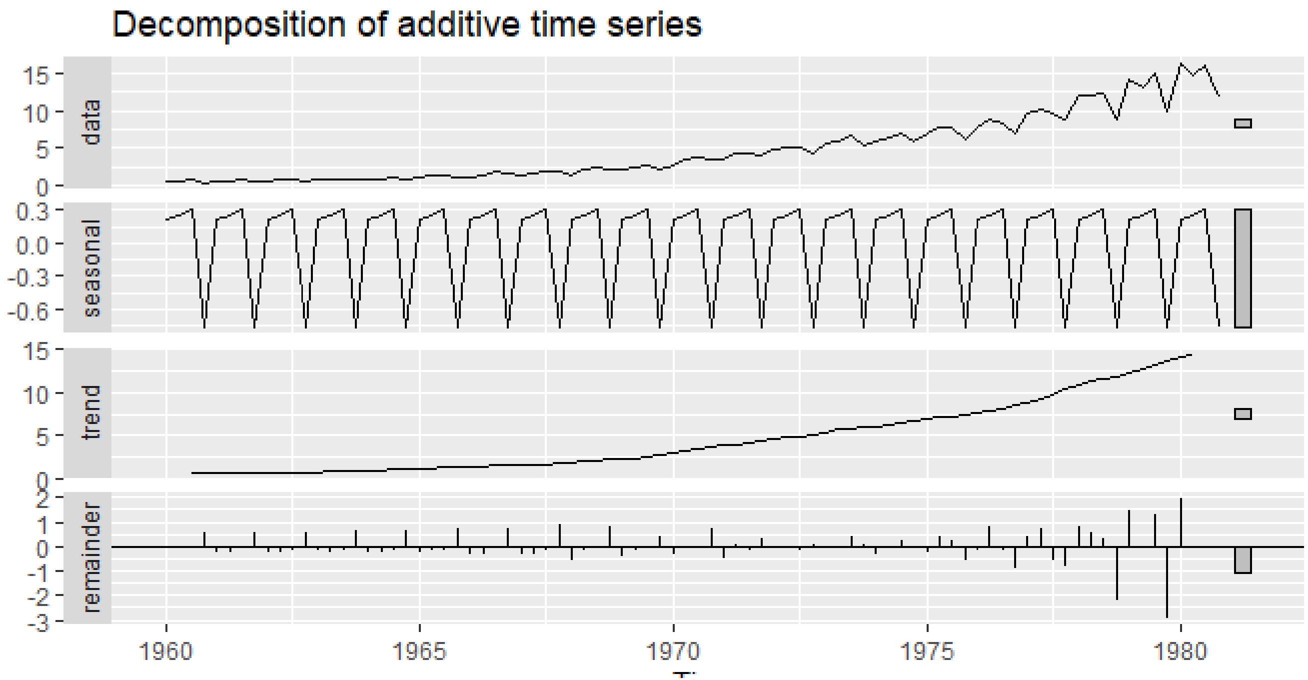
Exemplo: Quarterly Earnings per Johnson & Johnson Share



SÉRIES TEMPORAIS

MODELO DE DECOMPOSIÇÃO

Exemplo: Quarterly Earnings per Johnson & Johnson Share



SÉRIES TEMPORAIS

MODELO DE DECOMPOSIÇÃO

A componente Tendência consiste em um movimento superior a um ano e indica qual a direção de deslocamento da série, se esta ocorrendo um aumento global ou uma diminuição global.

O isolamento da componente Tendência, tem duas finalidades:

◦ identificá-la para auxiliar no processo de previsão ➔ usa-se análise de regressão

e

◦ removê-la, para que as demais componentes da série temporal sejam conhecidas ➔ usa-se médias móveis.

SÉRIES TEMPORAIS

ALISAMENTO EXPONENCIAL SIMPLES

A princípio, o método conhecido como Alisamento Exponencial Simples se assemelha ao da Média Móvel por extrair das observações da série temporal o comportamento aleatório pelo alisamento dos dados históricos. Entretanto, a inovação introduzida pelo Alisamento Exponencial Simples advém do fato de este método atribuir pesos diferentes a cada observação da série. Enquanto que na Média Móvel as observações usadas para encontrar a previsão do valor futuro contribuem em igual proporção para o cálculo dessa previsão, no Alisamento Exponencial Simples as informações mais recentes são evidenciadas pela aplicação de um fator que determina essa importância. Este método corresponde a uma média ponderada onde se dá peso maior para as observações mais recentes.

SÉRIES TEMPORAIS

ALISAMENTO EXPONENCIAL SIMPLES

O método Alisamento Exponencial Simples pode ser representado

através da equação:

$$\bar{Z}_t = \alpha Z_t + (1 - \alpha) \bar{Z}_{t-1}$$

Onde $\bar{Z}$ representa a série suavizada no tempo *t e t-1* e α é o peso atribuído à série, $0 \leq \alpha \leq 1$.

A previsão do valor futuro é dada pelo último valor exponencialmente alisado.

SÉRIES TEMPORAIS

ALISAMENTO EXPONENCIAL HOLT WINTERS

Este método produz resultados similares ao Alisamento Exponencial Linear, sendo, no entanto, capaz de manipular séries temporais que além de apresentarem tendência nos dados, apresentam também sazonalidade.

O método de Holt-Winters é baseado em três equações, uma para cada componente: nível, tendência e sazonalidade.

$$Y_{t+m} = (L_t + b_t m) S_{t-s+m}$$

SÉRIES TEMPORAIS

ALISAMENTO EXPONENCIAL HOLT WINTERS

O método Alisamento Exponencial Holt Winters

- Tendência Linear – Bi Paramétrico HW
- Sazonal – HW
    - Aditivo (variação sazonal constante)
    - Multiplicativo (variação sazonal não é constante)

SÉRIES TEMPORAIS

ALISAMENTO EXPONENCIAL HOLT WINTERS

Nível

$$L_t = \boxed{\alpha} \frac{X_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + b_{t-1})$$

Tendência

$$b_t = \boxed{\beta}(L_t - L_{t-1}) + (1 - \gamma)b_{t-1}$$

Sazonalidade

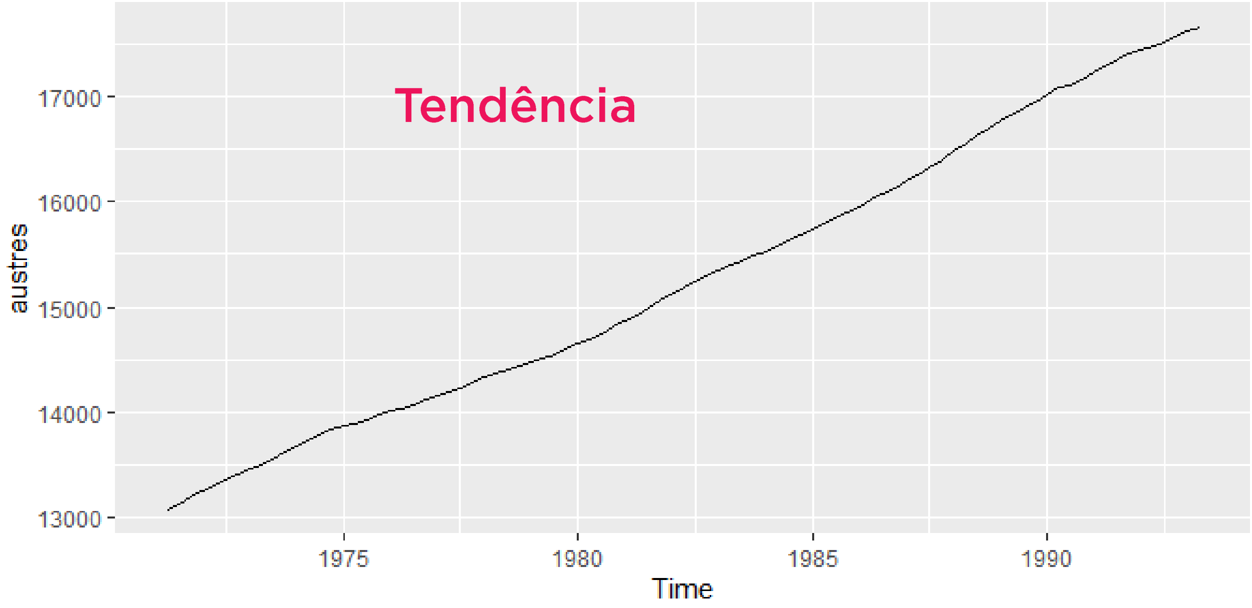
$$S_t = \boxed{\boldsymbol{\gamma}} \frac{X_t}{L_t} + (1 - \beta)S_{t-s}$$

s = comprimento da sazonalidade (trimestre, semana ou mês)

SÉRIES TEMPORAIS

ALISAMENTO EXPONENCIAL HOLT WINTERS

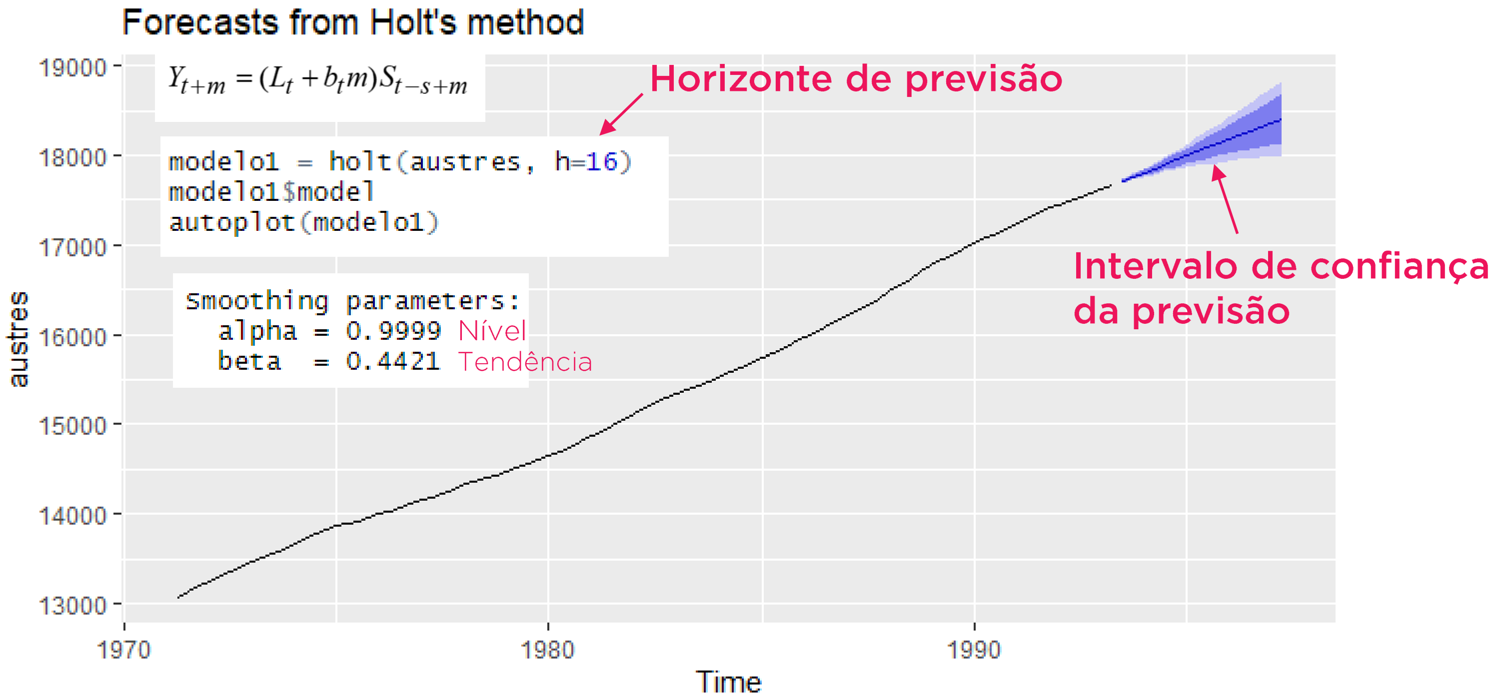
Exemplo: Quarterly Time Series of the Number of Australian Residents



SÉRIES TEMPORAIS

ALISAMENTO EXPONENCIAL HOLT WINTERS

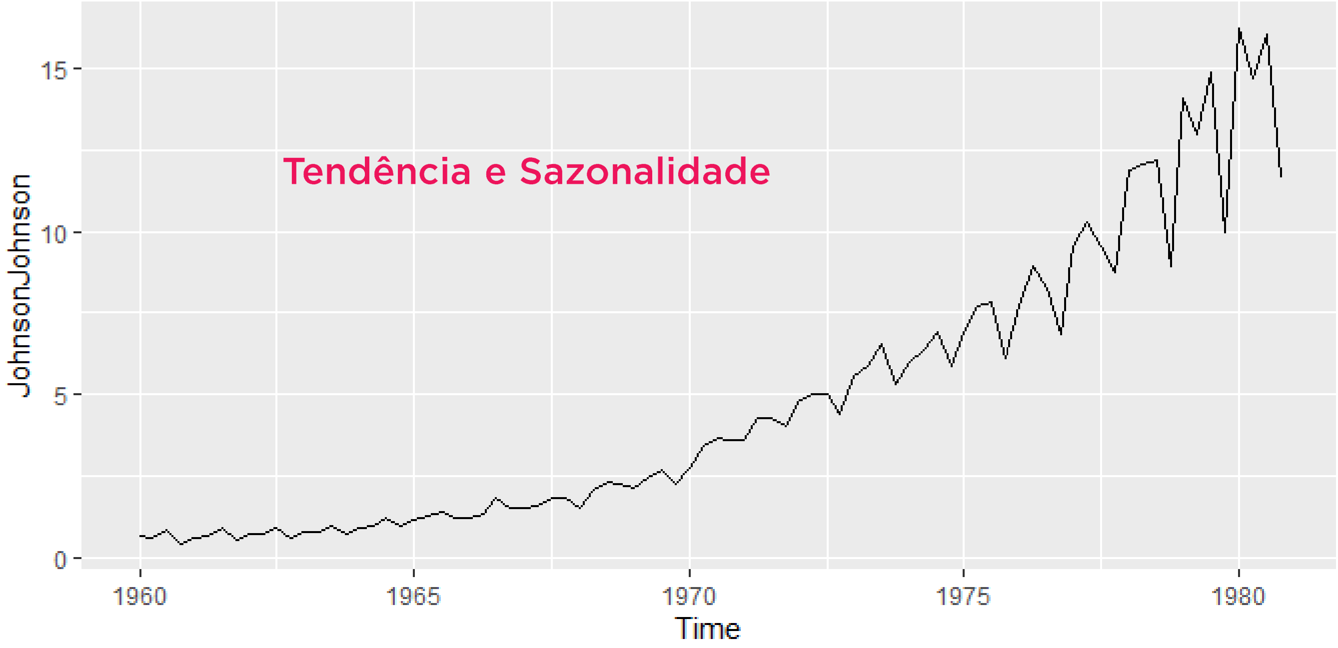
Exemplo: Quarterly Time Series of the Number of Australian Residents



SÉRIES TEMPORAIS

ALISAMENTO EXPONENCIAL HOLT WINTERS

Exemplo: Quarterly Earnings per Johnson & Johnson Share

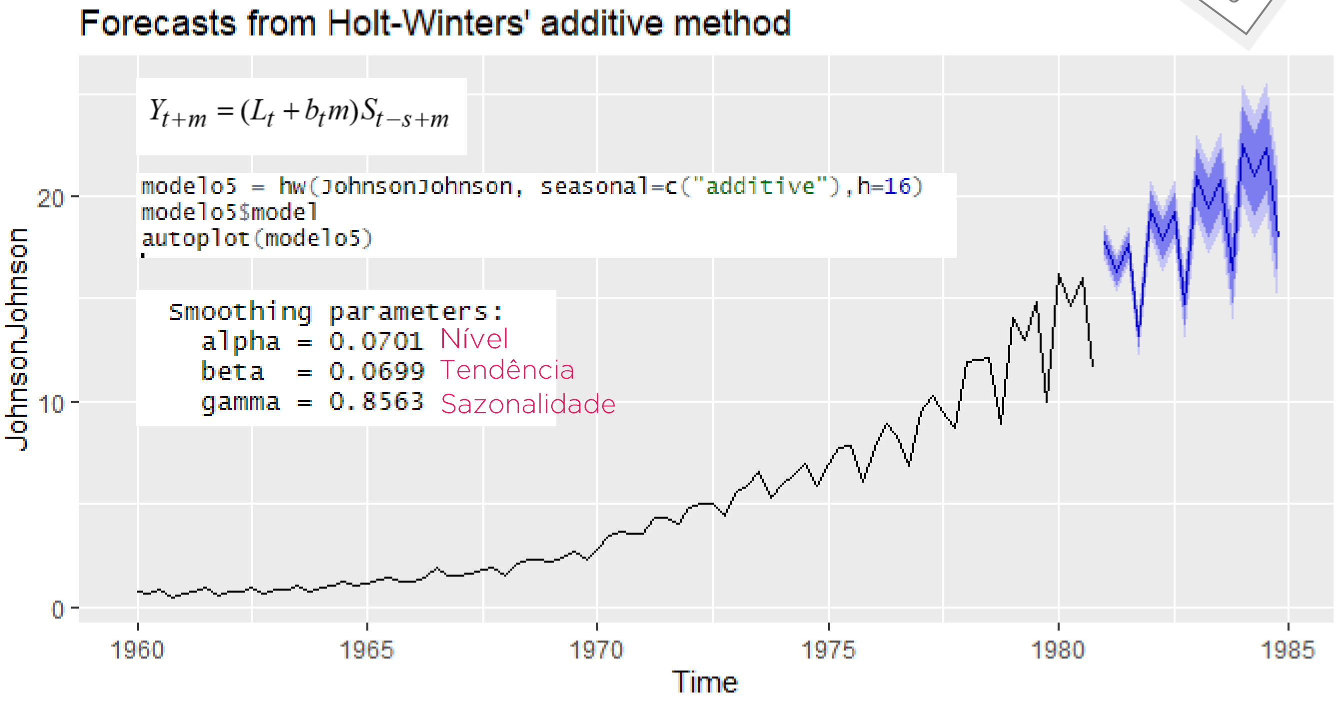


SÉRIES TEMPORAIS

ALISAMENTO EXPONENCIAL HOLT WINTERS

Exemplo
Saída RStudio

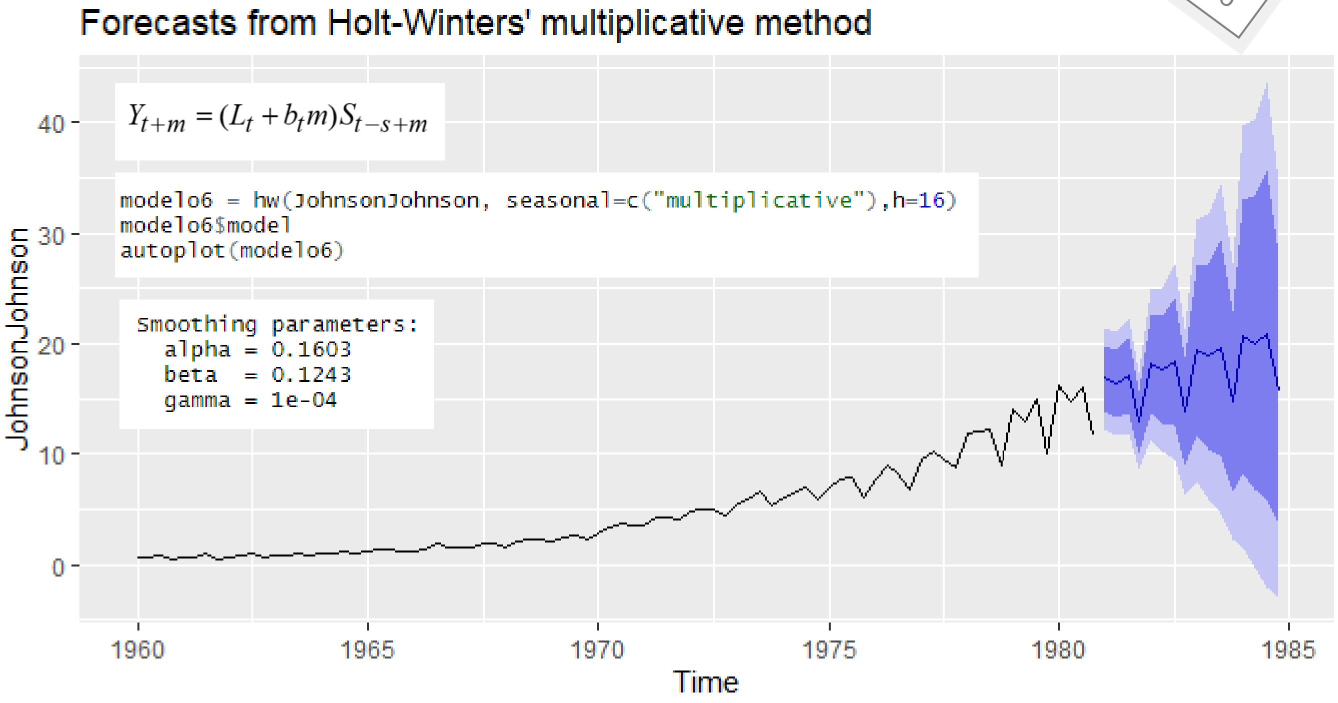
Exemplo: Quarterly Earnings per Johnson & Johnson Share



SÉRIES TEMPORAIS

ALISAMENTO EXPONENCIAL HOLT WINTERS

Exemplo: Quarterly Earnings per Johnson & Johnson Share

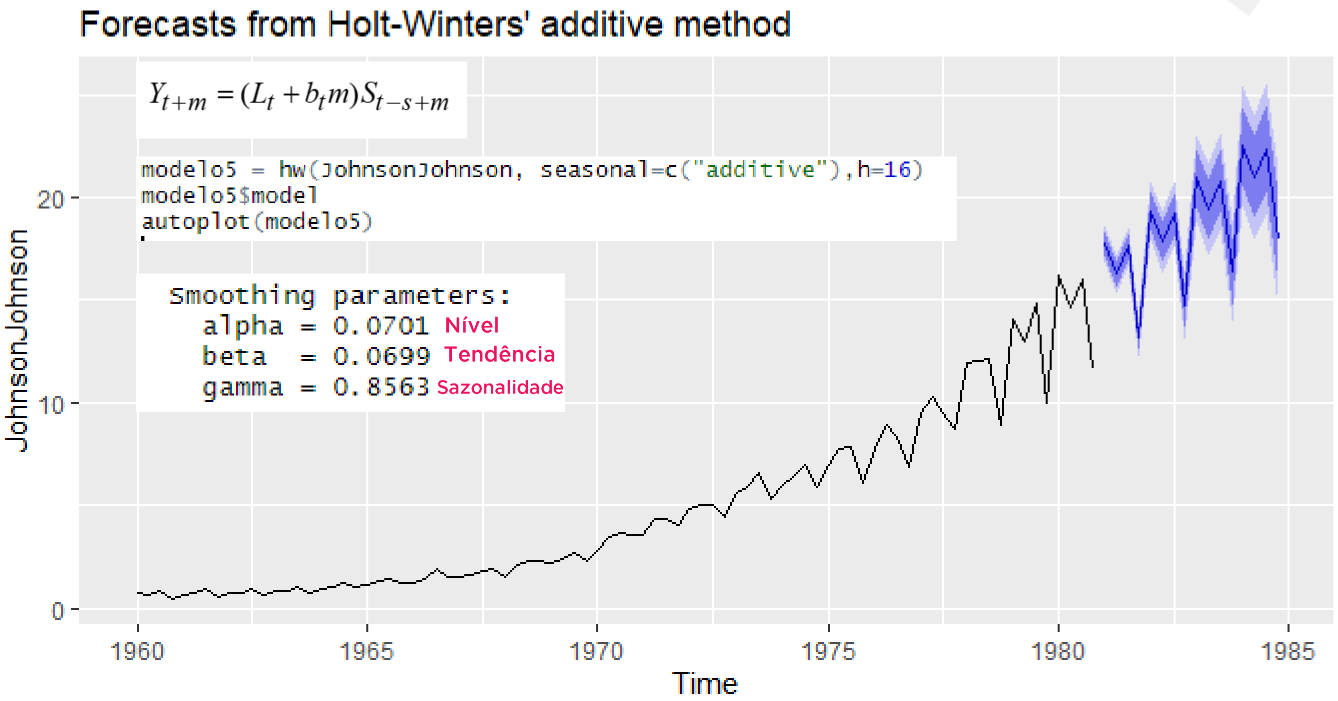


SÉRIES TEMPORAIS

ALISAMENTO EXPONENCIAL HOLT WINTERS

Exemplo
Saída RStudio

Exemplo: Quarterly Earnings per Johnson & Johnson Share



$$Y_{t+m} = (L_t + b_t m) S_{t-s+m}$$

```
modelo5 = hw(JohnsonJohnson, seasonal=c("additive"),h=16)
modelo5$model
autoplot(modelo5)
```

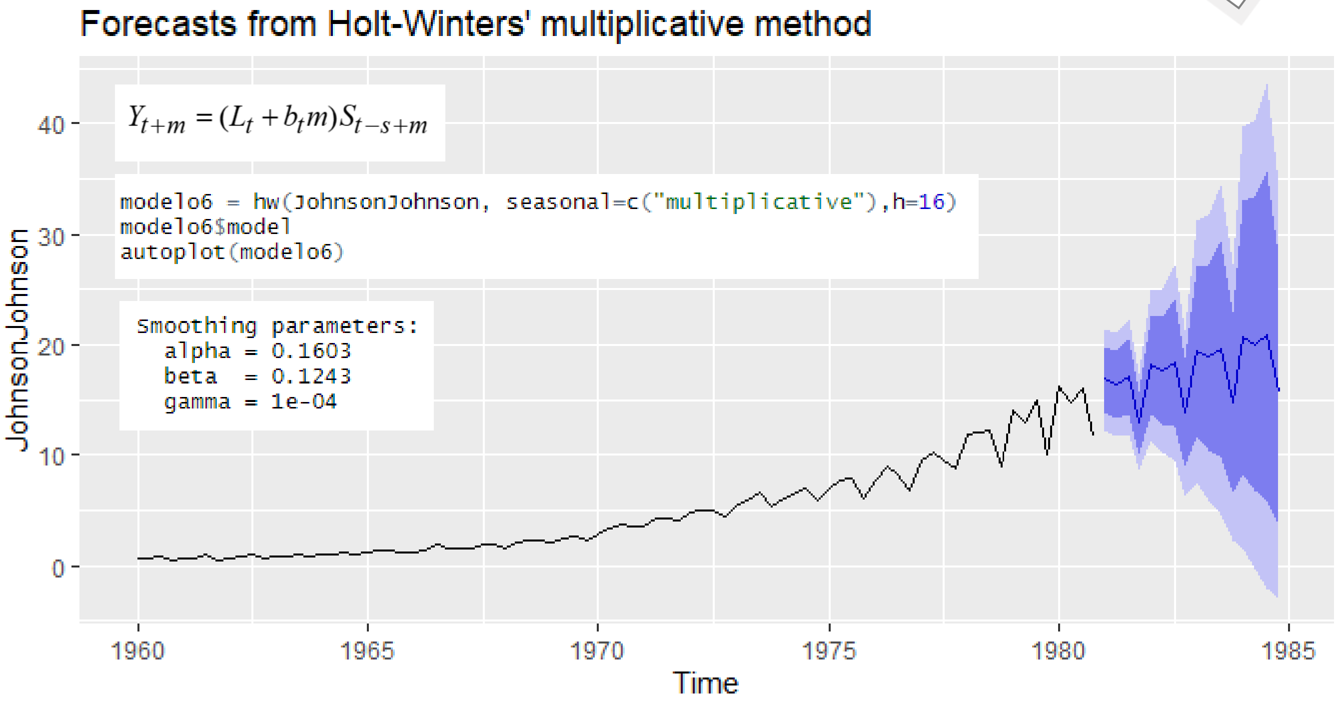
```
  Smoothing parameters:
    alpha = 0.0701  Nível
    beta  = 0.0699  Tendência
    gamma = 0.8563  Sazonalidade
```

SÉRIES TEMPORAIS

ALISAMENTO EXPONENCIAL HOLT WINTERS

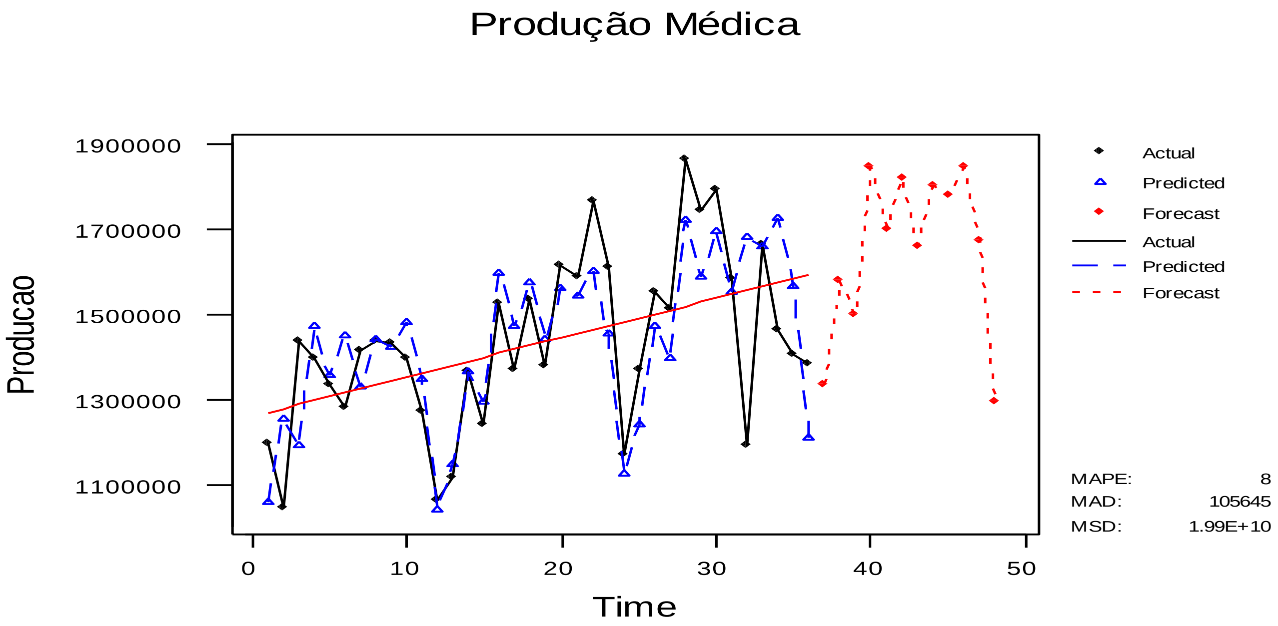
Exemplo
Saída RStudio

Exemplo: Quarterly Earnings per Johnson & Johnson Share



SÉRIES TEMPORAIS

ALISAMENTO EXPONENCIAL HOLT WINTERS



SÉRIES TEMPORAIS

MODELOS DE BOX JENKINS

Para os modelos de decomposição e os modelos de suavização exponencial é usual assumir que os erros t (t =1,...,T) são não correlacionados, o que implica que as observações também não estão correlacionadas. Esta suposição raramente ocorre na prática. Correlação serial é esperada uma vez que os dados são coletados sequencialmente no tempo.

Os modelos que capturam esta estrutura de correlação são uma classe especial de modelos chamados de modelos Autoregressivos Integrados de Médias Móveis (ARIMA). Estes modelos apresentam uma variedade de diferentes estruturas de correlação, se esta estrutura de correlação for bem modelada, ele fornecerá boas previsões.

Os modelos ARIMA foram popularizados por George Box e Gwilym Jenkins no início dos anos 70, e seus nomes tem sido usado como sinônimo destes modelos. Box e Jenkins colocaram de forma compreensiva a informação necessária para entender e usar os modelos ARIMA para séries temporais univariadas.

SÉRIES TEMPORAIS

MODELO ARIMA

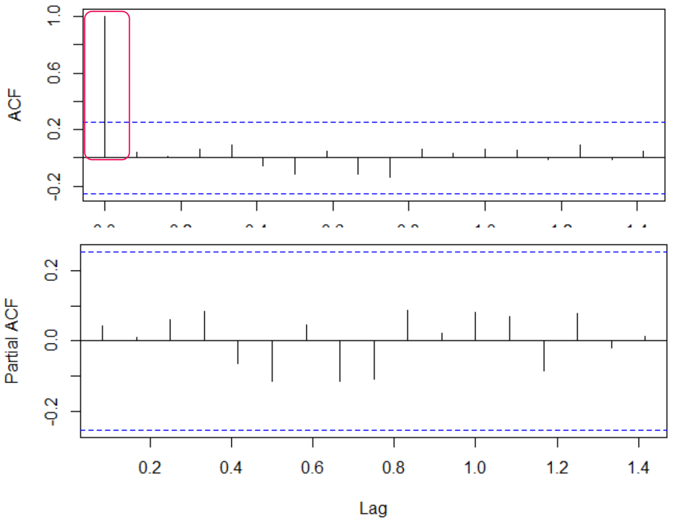
Pressuposto do modelo ⇨ série estacionária

A característica de série temporal estacionária:

- a série se desenvolve em torno de média e variância constante.

SÉRIES TEMPORAIS
MODELO ARIMA

Função de autocorrelação e Função autocorrelação parcial é utilizada para identificar o tipo de modelo.



A saída do gráfico ACF do R sempre estima o valor 1 para o lag igual a zero. Desconsidere o lag igual a 0 (zero) na sua análise.

O intervalo de confiança é representado pelas linhas tracejadas em azul.

SÉRIES TEMPORAIS

MODELO ARIMA

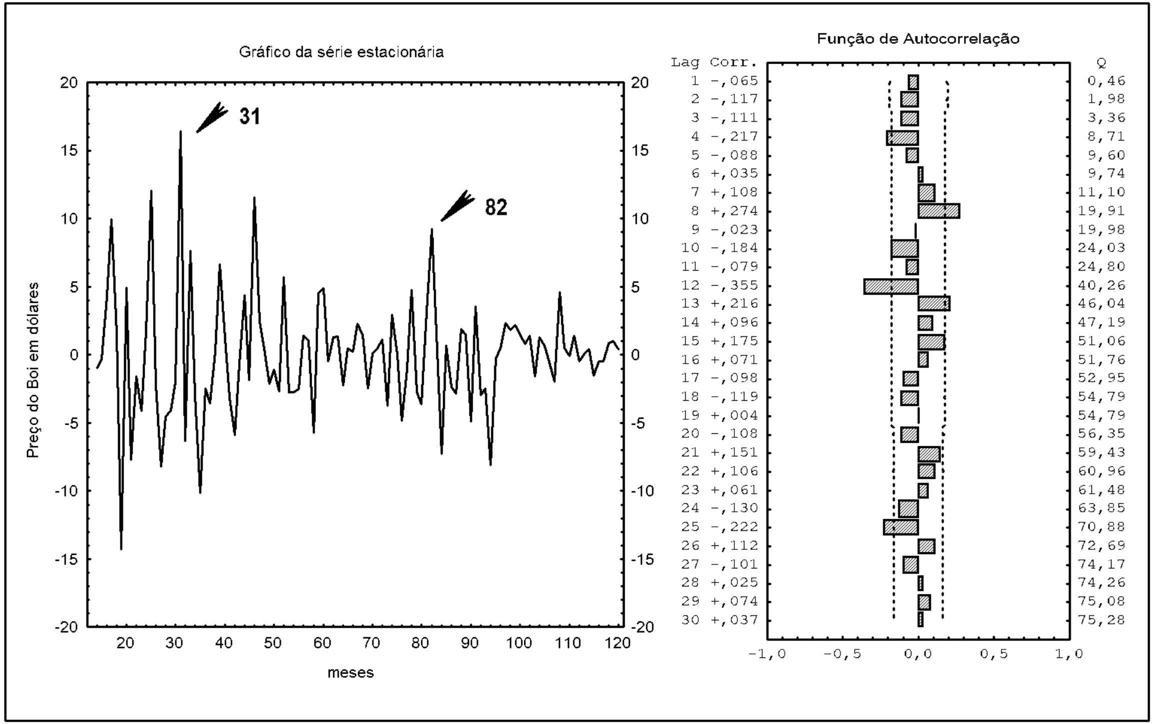
Autocorrelação: correlação entre membros de uma série de observações ordenadas no tempo. Assim, a Função de Autocorrelação (FAC) que calcula a correlação entre x_t e x_{t-k} é calculada por

$$r_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}$$

O gráfico que contempla as autocorrelações denomina-se correlograma.

SÉRIES TEMPORAIS

MODELO ARIMA



SÉRIES TEMPORAIS

MODELO ARIMA

A análise de séries temporais visa identificar e explicar:

$$\text{Série} = T + S + C \boxed{+} a$$

Modelos ARIMA

T: Tendência

S: Sazonalidade

C: Ciclo

A: Aleatório

SÉRIES TEMPORAIS

MODELO ARIMA

Método estatístico que utiliza autoregressão e médias móveis para previsão de séries temporais. Um modelo linear é construído incluindo um número especificado de termos e os dados são preparados por um nível de diferenciação afim de tornar este estacionário.

Modelos autorregressivos integrados de médias móveis: ARIMA (p,q,d)

p → ordem do modelo Autorregressimo (AR)

d → Integrado (I)

q → ordem do modelo Médias Móveis (MA)

SÉRIES TEMPORAIS

MODELO ARIMA

AR I MA(p, d, q).

AR: Autoregression : é a parte auto-regressiva do modelo. Isso nos permite incorporar o efeito de valores passados (lags) em nosso modelo.

I: Integrated: a parte integrada do modelo. é a quantidade de diferenciação (ou seja, o número de pontos no tempo passados a serem subtraídos do valor atual) a serem aplicados à série temporal. O objetivo é transformar a série temporal em estacionária.

MA: Moving Average : é a parte média móvel do modelo. Isso nos permite definir o erro de nosso modelo como uma combinação linear dos valores de erro observados em pontos de tempo anteriores.

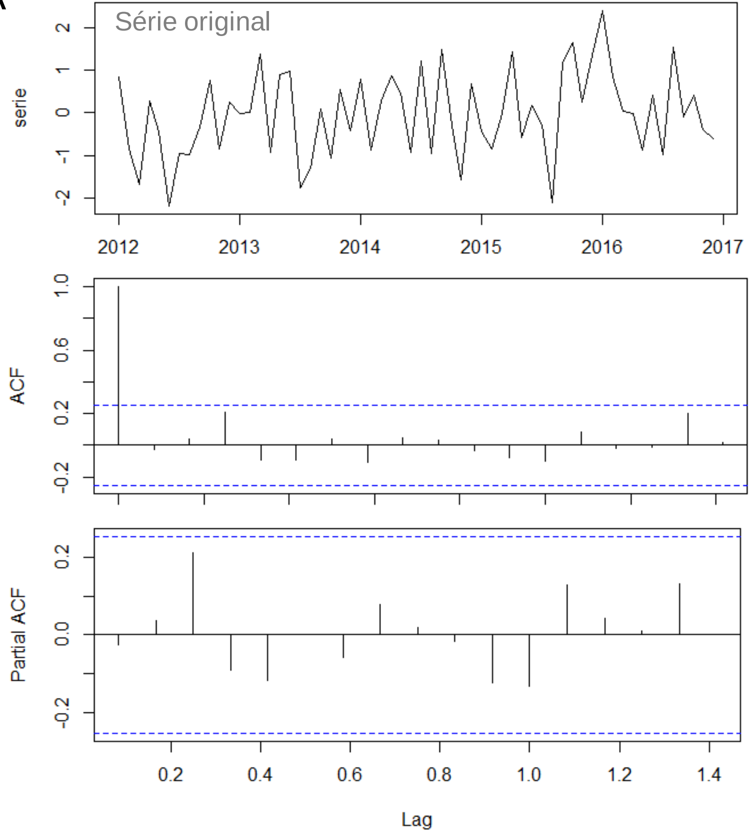
Erros Residuais: = Valores Observados – Valores Estimados

- Erros residuais contém estruturas temporais que podem ser modeladas. Existem sinais complexos nos erros residuais. Um modelo que prever o erro residual pode ser usado para ajustar os próximos erros e melhorar um modelo que aprende com o histórico.

SÉRIES TEMPORAIS

MODELO ARIMA

Exemplo de modelos ARIMA(0,0,0)



SÉRIES TEMPORAIS

MODELO ARIMA

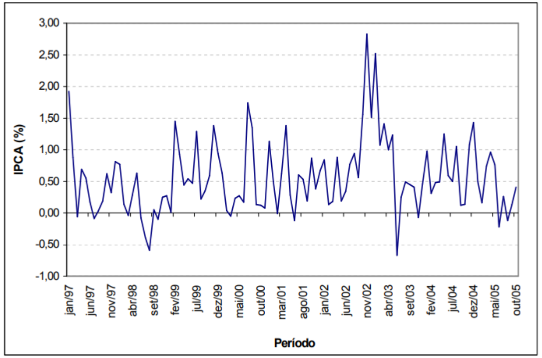
Podemos usar um valor 0 para desligar um parâmetro, dessa forma, aquela função em questão não será feita, por exemplo, se no parâmetro d definirmos 0 não será realizada uma diferenciação nos dados. Neste exemplo teríamos um modelo ARMA.

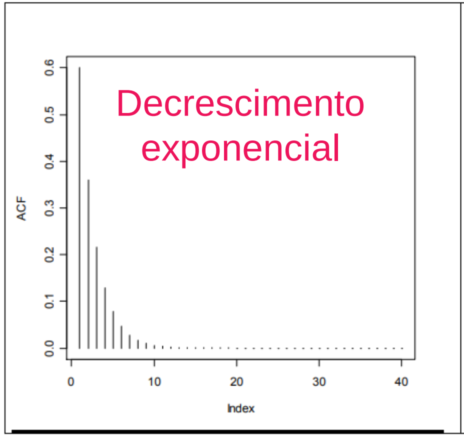
Modelo Auto-Regressivo de Médias Móveis (ARMA)
O modelo ARMA é a junção AR(p) e MA(q)
AR(p): tenta explicar o efeito de momentum da série
MA(q): tenta capturar o efeito do ruído da série.
Esse efeito pode ser interpretado como eventos inesperados que afetam a observação.

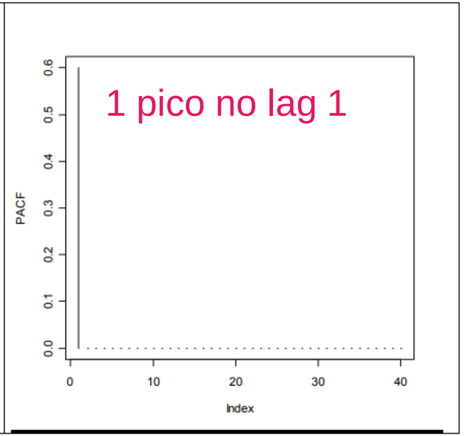
SÉRIES TEMPORAIS

MODELO ARIMA

Exemplo de modelo ARIMA(1,0,0) ou AR(1)





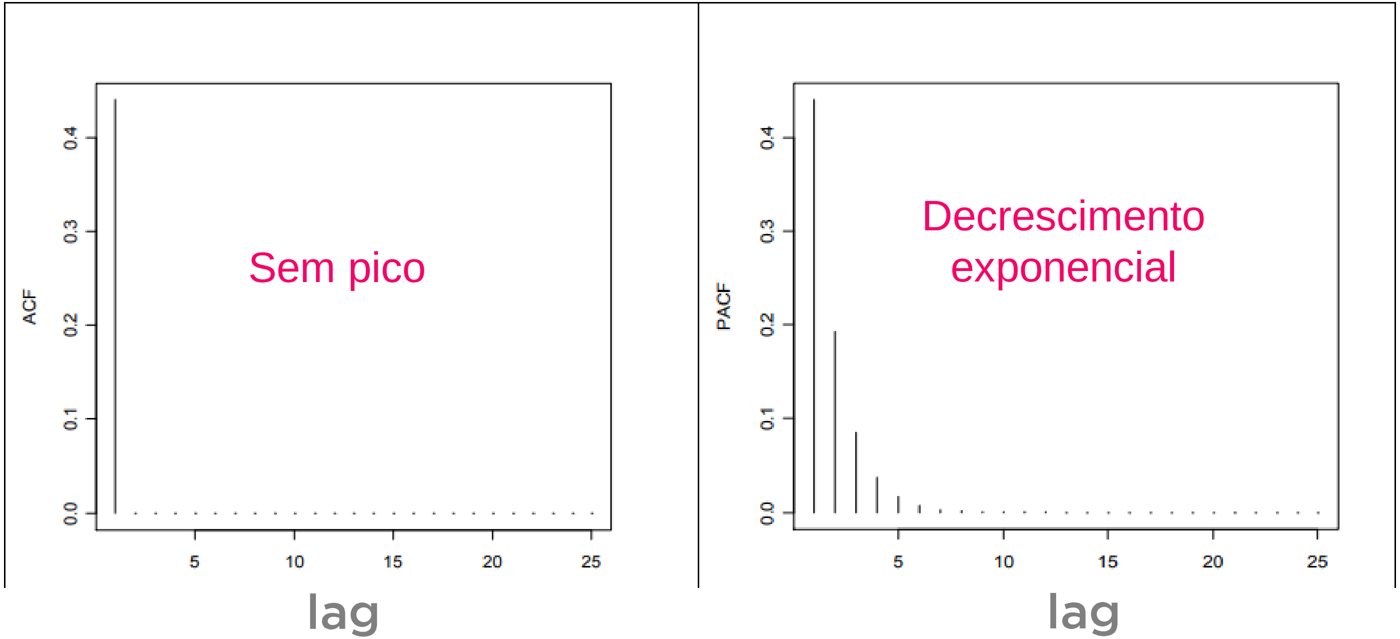


SÉRIES TEMPORAIS

MODELO ARIMA

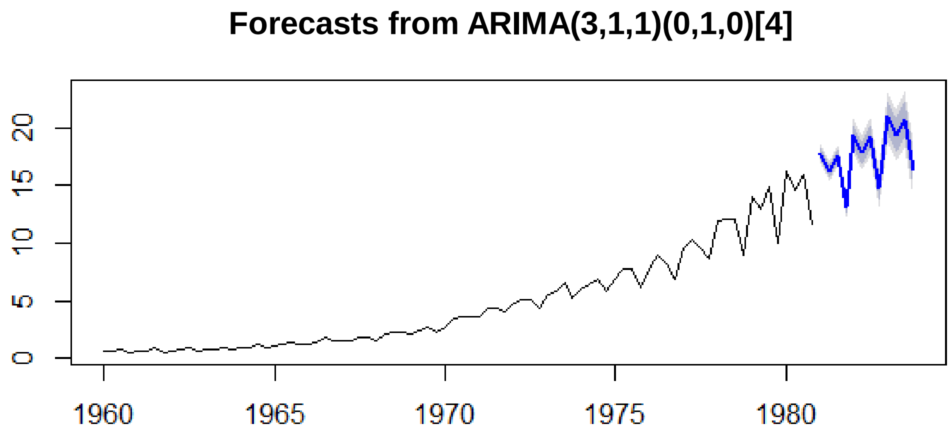
Exemplo de modelo ARIMA(0,1,0) ou MA(1)

Médias móveis



SÉRIES TEMPORAIS

MODELO ARIMA




```
> modelo1_arima = auto.arima(JohnsonJohnson)
> print(modelo1_arima)
Series: JohnsonJohnson
ARIMA(3,1,1)(0,1,0)[4]

Coefficients:
          ar1     ar2     ar3      ma1
      -0.1712  0.1387  -0.208  -0.6636
s.e.   0.1769  0.1701   0.121   0.1542

sigma^2 estimated as 0.1808:  log likelihood=-43.01
AIC=96.02   AICc=96.84   BIC=107.86
```

SÉRIES TEMPORAIS:
PROCESSO DE MODELAGEM

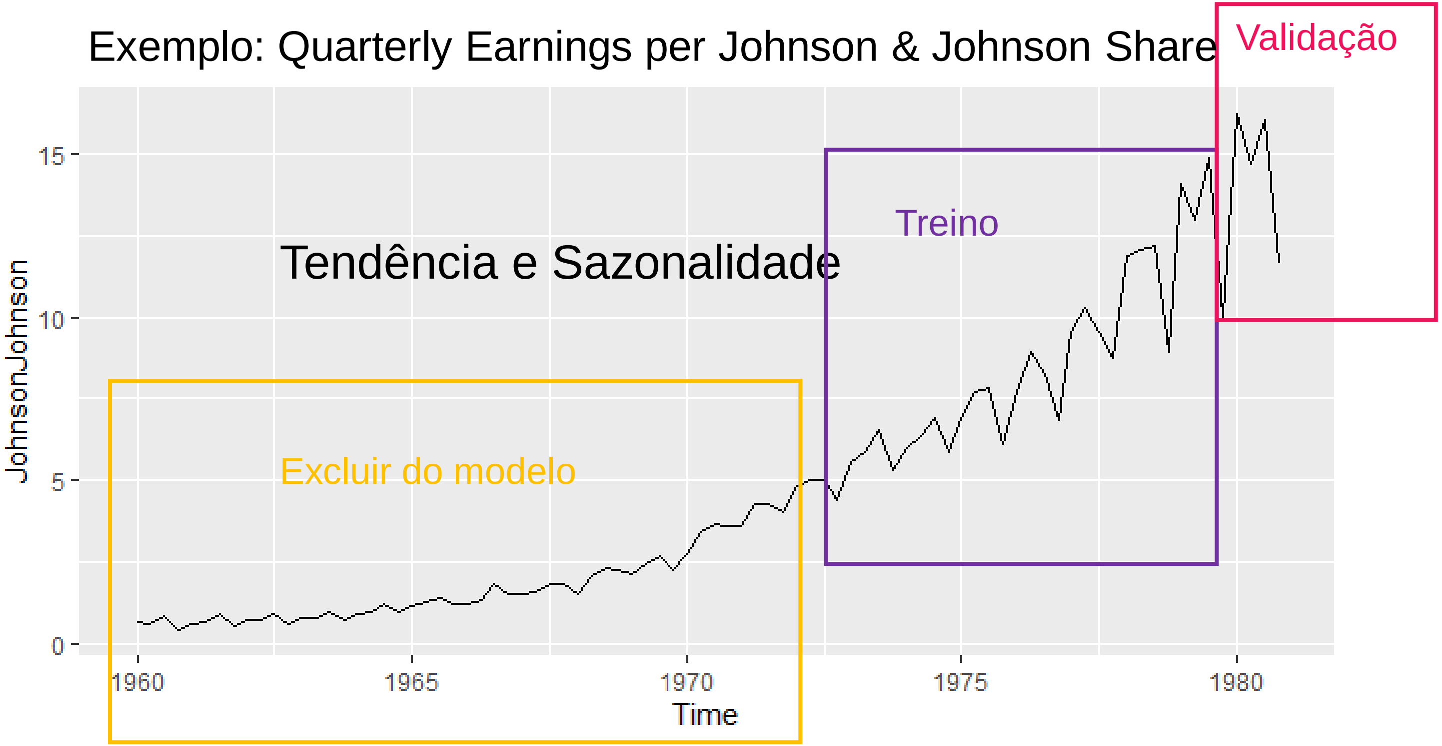
Construir a série temporal → Gráfico da série temporal → Identificar os componentes da série

Dividir a série em duas partes

→ Seleção dos modelos → Acurácia do modelo → Análise de resíduos

→ Validação do modelo → Acurácia do modelo

SÉRIES TEMPORAIS:
PROCESSO DE MODELAGEM

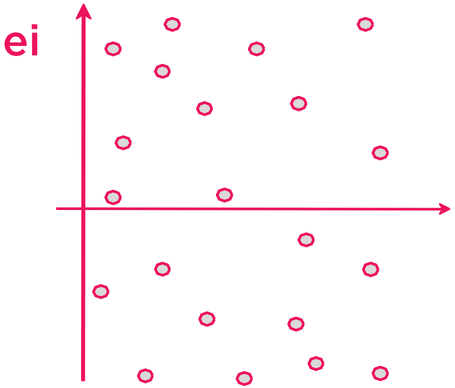


SÉRIES TEMPORAIS:
ANÁLISE DE RESÍDUOS

NORMALIDADE

Pelo histograma dos resíduos padronizados, pode-se analisar a suposição de normalidade.

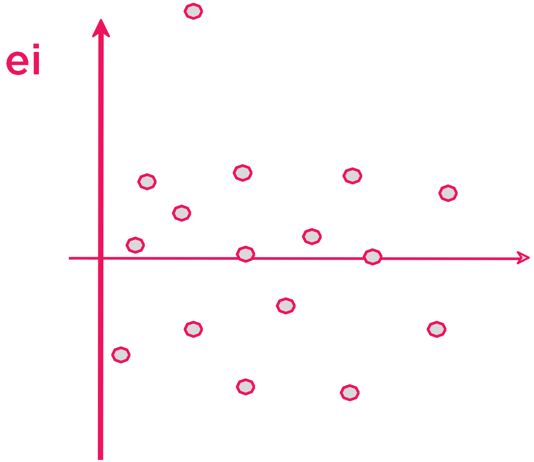
Então, a suposição de igualdade da variância está violada.



SÉRIES TEMPORAIS:
ANÁLISE DE RESÍDUOS

LOCALIZANDO OS OUTLIERS:

Em geral, resíduos padronizados com valores maiores que 2 são considerados outliers.



SÉRIES TEMPORAIS:
CRITÉRIOS PARA SELEÇÃO DE MODELOS

COMPETIÇÃO ENTRE MODELOS

Critério	Descrição
Akaike Information Criterion (AIC)	Fornece uma medida da qualidade do modelo obtida pela simulação da situação em que este é testado em um conjunto de dados diferente. Isto é, amostra, treino e validação.
Bayesian Information Criterion (BIC)	Fornece uma medida da qualidade do modelo dentro do contexto Bayesiano.
Cross Validation (CV)	Não utiliza em modelos tradicionais

SÉRIES TEMPORAIS

MEDIDAS DE DESEMPENHO DOS MODELOS

SÉRIES TEMPORAIS:
MEDIDAS DE DESEMPENHO DOS MODELOS

Erro Médio (Mean error-ME): $\text{ME}=\frac{\sum_{i=1}^{n} y_i-\hat{y}_i}{n}$

Erro Médio Absoluto (Mean Absolut Error-MAE): $\text{MAE}=\frac{\sum_{i=1}^{n} |y_i-\hat{y}_i|}{n}$

Raiz do Erro Quadrático Médio (Root Mean Squared Error-RMSE): $\text{RMSE}=\sqrt{\frac{\sum_{i=}^{n}(y_i-\hat{y}_i)^2}{n}}$

Erro Percentual Médio (Mean Percent Error-MPE): $MPE=\frac{\sum_{i=1}^{n}\frac{y_i-\hat{y}_i}{y_i*100}}{n}$

Erro Percentual Absoluto Médio (Mean Absolut Percent Error-MAPE):

$$MAPE=\frac{\sum_{i=1}^{n}\frac{|y_i-\hat{y}_i|}{y_i*100}}{n}$$

ROOT MEAN SQUARE ERROR (RMSE)

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 = \frac{(\hat{y}_1 - y_1)^2 + \cdots + (\hat{y}_n - y_n)^2}{n},$$

n é o número de observações,

y_i é o valor real e

$\hat{y}_i$ é a predição do modelo.

Utilizamos a Raiz quadrada desse valor RMSE

➔ Nesse caso, um modelo bom é aquele que possui o **menor erro quadrático médio**.

ROOT MEAN SQUARE ERROR (RMSE)

- A medida de erro normalmente utilizada para avaliar a qualidade do ajuste de um modelo é a chamada RAIZ DO ERRO MÉDIO QUADRÁTICO.
- Ela é a raiz do erro médio quadrático da diferença entre a predição e o valor real.
- Podemos pensar nela como sendo uma medida análoga ao desvio-padrão.
- A medida RMSE tem a mesma unidade que os valores de **y**.
- RMSE é uma boa medida, porque geralmente ela representa explicitamente o que vários métodos tendem a minimizar.

SÉRIES TEMPORAIS:
MEDIDAS DE DESEMPENHO DOS MODELOS

Exemplo:

Data	Vendas (y) A	Budget (x)	Vendas estimadas (B)	Erro (A-B)	Erro absoluto |A-B|	Erro^2	%erro	% erro absoluto
jan/18	207	13	236	-28.7	28.7	822.8	-13.9	13.9
fev/18	289	22	265	24.0	24.0	576.1	8.3	8.3
mar/18	285	24	272	13.5	13.5	181.9	4.7	4.7
abr/18	292	26	278	14.0	14.0	195.2	4.8	4.8
mai/18	269	28	285	-15.5	15.5	241.6	-5.8	5.8
jun/18	291	32	298	-6.6	6.6	43.2	-2.3	2.3
jul/18	331	34	304	26.9	26.9	724.4	8.1	8.1
ago/18	283	35	307	-24.3	24.3	592.6	-8.6	8.6
set/18	364	44	337	27.3	27.3	747.6	7.5	7.5
out/18	345	45	340	5.1	5.1	25.9	1.5	1.5
nov/18	370	53	366	4.0	4.0	16.2	1.1	1.1
dez/18	310	48	350	-39.7	39.7	1575.1	-12.8	12.8
soma				0.0	229.7	5742.3	-7.3	79.3

MEDIDA	SOMA	n	RESULTADO
ME	**0.0**	**12**	**0.0**
MAE	229.7	12	19.14
RMSE	5742.3	12	21.88
MPE	-7.3	12	-1.04
MAPE	79.3	12	6.61

REFERÊNCIAS BIBLIOGRÁFICAS

- HYNDMAN, R.J.; ATHANASOPOULOS, G.; Forecasting: Principles and Practice. 3.ed. Monash University, Australia https://otexts.com/fpp3/
- MORETTIN, P.A. ECONOMETRIA FINANCEIRA UM CURSO EM SÉRIES TEMPORAIS FINANCEIRAS.

Tks!

Adelaide Alves

linkedin.com/adelaide-alves