Analytics e Inteligência Artificial Data Science

Tema da aula
Séries Temporais

FIA BUSINESS SCHOOL lab.data

FIA Business School

NOSSOS DIFERENCIAIS | QUEM SOMOS

BUSINESS SCHOOL
Graduação, pós-graduação, MBA, Pós- MBA, Mestrado Profissional, Curso In *Company* e EAD

CONSULTING
Consultoria personalizada que oferece soluções baseadas em seu problema de negócio

RESEARCH
Atualização dos conhecimentos e do material didático oferecidos nas atividades de ensino

Líder em Educação Executiva, referência de ensino nos cursos de graduação, pós-graduação e MBA, tendo excelência nos programas de educação. Uma das principais escolas de negócio do mundo, possuindo convênios internacionais com Universidades nos EUA, Europa e Ásia. +8.000 projetos de consultorias em organizações públicas e privadas.

Único curso de graduação em administração a receber as notas máximas

A primeira escola brasileira a ser finalista da maior competição de MBA do mundo

Única *Business School* brasileira a figurar no *ranking* LATAM

Signatária do Pacto Global da ONU

Membro fundador da ANAMBA - Associação Nacional MBAs

Credenciada pela AMBA - *Association of MBAs*

Credenciada ao *Executive MBA Council*

Filiada a AACSB - *Association to Advance Collegiate Schools of Business*

Filiada a EFMD - *European Foundation for Management Development*

Referência em cursos de MBA nas principais mídias de circulação

O **Laboratório de Análise de Dados** – LABDATA é um Centro de Excelência que atua nas áreas de ensino, pesquisa e consultoria em análise de informação utilizando técnicas de ***Big Data***, ***Analytics*** e **Inteligência Artificial**.

O LABDATA é um dos pioneiros no lançamento dos cursos de *Big Data* e *Analytics* no Brasil.

Os diretores foram professores de grandes especialistas do mercado.

+10 anos de atuação.

+9.000 alunos formados.

Docentes

- Sólida formação acadêmica: doutores e mestres em sua maioria;
- Larga experiência de mercado na resolução de *cases;*
- Participação em congressos nacionais e internacionais;
- Professor assistente que acompanha o aluno durante todo o curso.

Estrutura

- 100% das aulas realizadas em laboratórios;
- Computadores para uso individual durante as aulas;
- 5 laboratórios de alta qualidade (investimento +R$2MM);
- 2 unidades próximas à estação de metrô (com estacionamento).

PROFA. DRA. ALESSANDRA DE ÁVILA MONTINI

Diretora do LABDATA-FIA, apaixonada por dados e pela arte de lecionar. Tem muito orgulho de ter criado na FIA cinco laboratórios para as aulas de Big Data e Inteligência Artificial. Possui mais de 20 anos de trajetória nas áreas de Data Mining, Big Data, Inteligência Artificial e Analytics. Cientista de dados com carreira realizada na Universidade de São Paulo. Graduada e mestra em Estatística Aplicada pelo IME-USP e doutora pela FEA-USP. Com muita dedicação chegou ao cargo de professora e pesquisadora na FEA-USP, ganhou mais de 30 prêmios de excelência acadêmica pela FEA-USP e mais de 30 prêmios de excelência acadêmica como professora dos cursos de MBA da FIA. Orienta alunos de mestrado e de doutorado na FEA-USP. Parecerista da FAPESP e colunista de grandes portais de tecnologia.

PROF. ÂNGELO CHIODE, MSc

Bacharel, mestre e candidato ao PhD em Estatística (IME-USP), atua como professor de Estatística Aplicada para turmas de especialização, pós-graduação e MBA na FIA. Trabalha como consultor nas áreas de Analytics e Ciência de Dados há 13 anos, apoiando empresas na resolução de desafios de negócio nos contextos de finanças, adquirência, seguros, varejo, tecnologia, aviação, telecomunicações, entretenimento e saúde. Nos últimos 5 anos, tem atuado na gestão corporativa de times de Analytics, conduzindo projetos que envolviam análise estatística, modelagem preditiva e *machine learning*. É especializado em técnicas de visualização de dados e design da informação (Harvard) e foi indicado ao prêmio de Profissional do Ano na categoria Business Intelligence, em 2019, pela Associação Brasileira de Agentes Digitais (ABRADi).

Conteúdo Programático

DISCIPLINAS

- IA E TRANSFORMAÇÃO DIGITAL
- ANALYTICS
- INTELIGÊNCIA ARTIFICIAL: MACHINE LEARNING
- INTELIGÊNCIA ARTIFICIAL: DEEP LEARNING
- EMPREENDEDORISMO E INOVAÇÃO
- COMPORTAMENTO HUMANO E SOFT SKILLS

TEMAS: ANALYTICS E MACHINE LEARNING

- ANÁLISE EXPLORATÓRIA DE DADOS
- INFERÊNCIA ESTATÍSTICA
- TÉCNICAS DE PROJEÇÃO
- TÉCNICAS DE CLASSIFICAÇÃO
- TÓPICOS DE MODELAGEM
- TÉCNICAS DE SEGMENTAÇÃO
- TÓPICOS DE ANALYTICS
- MANIPULAÇÃO DE BASE DE DADOS
- AUTO ML

TEMAS: DEEP LEARNING

- REDES DENSAS
- REDES CONVOLUCIONAIS
- REDES RECORRENTES
- MODELOS GENERATIVOS

FERRAMENTAS

- LINGUAGEM R
- LINGUAGEM PYTHON
- DATABRICKS

Conteúdo da Aula

1. Introdução

Case: Vendas

Exemplo:
Prever a quantidade de vendas de um determinado produto no próximo mês, com base no histórico mensal de vendas, a fim de dimensionar o estoque necessário e ajustar estratégias de precificação.

Aplicação:
Área comercial

Case: Tráfego Virtual

Exemplo:
Predizer a quantidade de acessos que um site de *e-commerce* receberá nas próximas semanas, a fim de otimizar a capacidade dos servidores, planejar recursos e garantir a boa experiência do usuário, caso haja alta demanda.

Aplicação:
Área de tecnologia

Case: Consumo de Água

1. INTRODUÇÃO | SÉRIES TEMPORAIS

Exemplo:
Prever o consumo de água potável de um determinado bairro ou região nos próximos meses, para realizar um bom planejamento do processo de tratamento e distribuição do volume de água demandado pelos habitantes.

Aplicação:
Área de indústria de saneamento

Case: Ocupação Hoteleira

Exemplo:
Predizer o percentual de ocupação de vagas em hotéis no próximo final de ano, tomando como referência os dados dos finais de anos anteriores, a fim de planejar a gestão de reservas e ajustar preços de forma mais adequada.

Aplicação:
Área de hotelaria

O Que é uma Série Temporal?

Uma **série temporal** consiste em um conjunto de observações de uma variável, em geral quantitativa, indexadas pelo **tempo**.

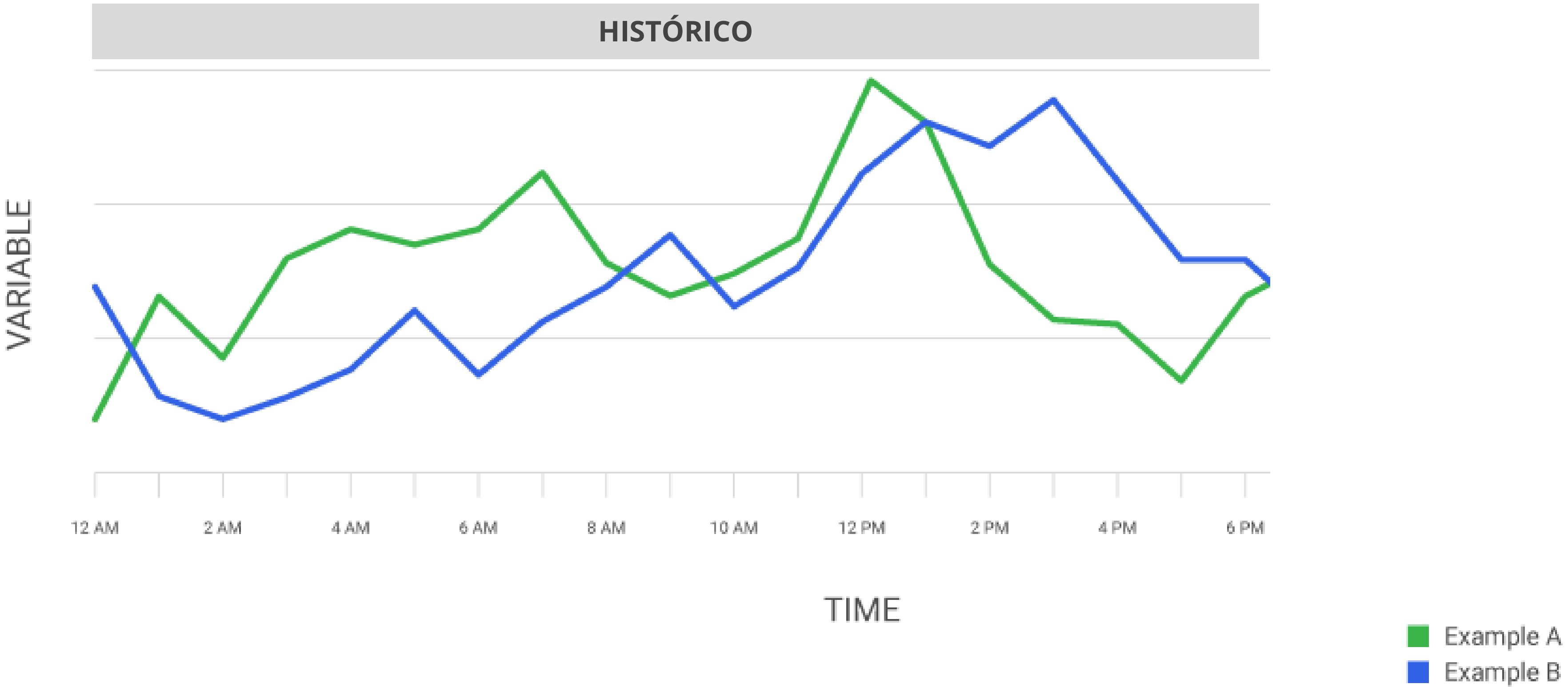
Exemplos:

- Cotações diárias do dólar comercial, entre 01/01/2015 e 31/12/2023.
- Faturamento mensal bruto da empresa, entre janeiro e dezembro de 2023.
- Condição climática diária em uma cidade (ensolarado, chuvoso ou nublado), nos últimos 14 dias.

Diferentemente de bases de dados em formato de **seção transversal**, com as quais trabalhamos nos modelos de regressão e árvore de decisão, as observações de uma série temporal possuem **alto grau de dependência** entre si. Por isso, as técnicas que vimos anteriormente não são adequadas para dados temporais.

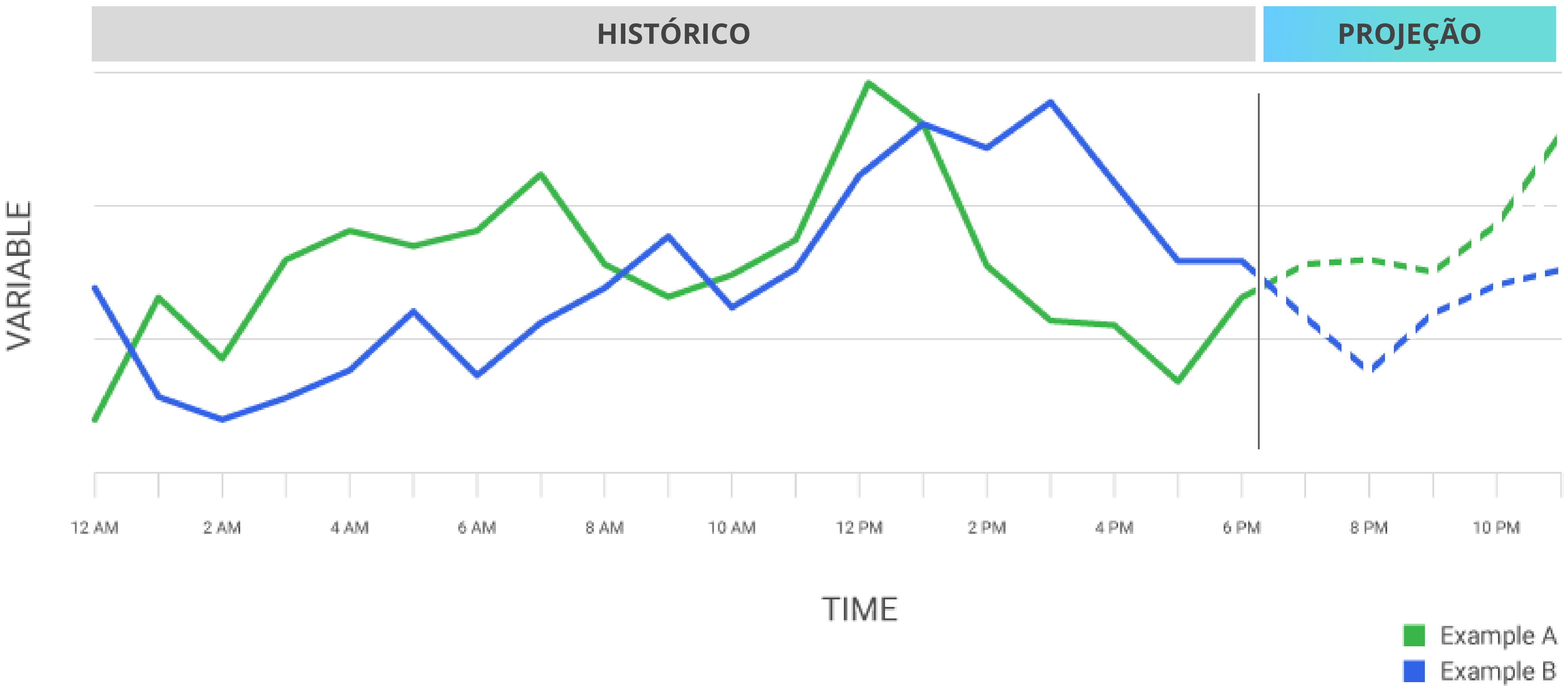
2. Objetivo

Objetivo



Nosso objetivo principal em modelagem de séries temporais consiste em **projetar** os valores mais plausíveis de uma série para momentos **futuros**, a partir do comportamento demonstrado no passado.

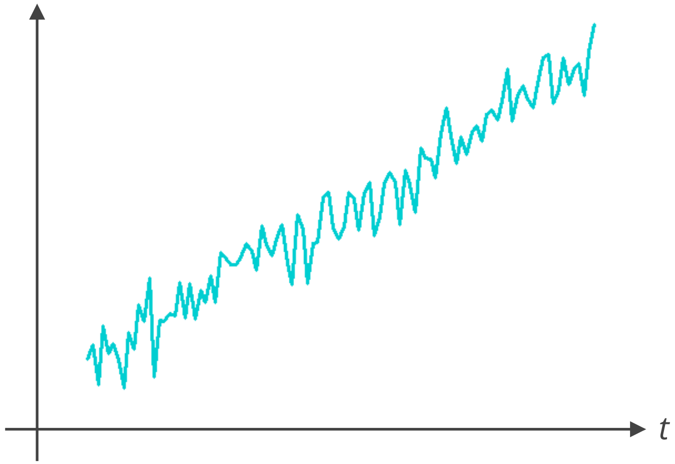
Objetivo



Nosso objetivo principal em modelagem de séries temporais consiste em **projetar** os valores mais plausíveis de uma série para momentos **futuros**, a partir do comportamento demonstrado no passado.

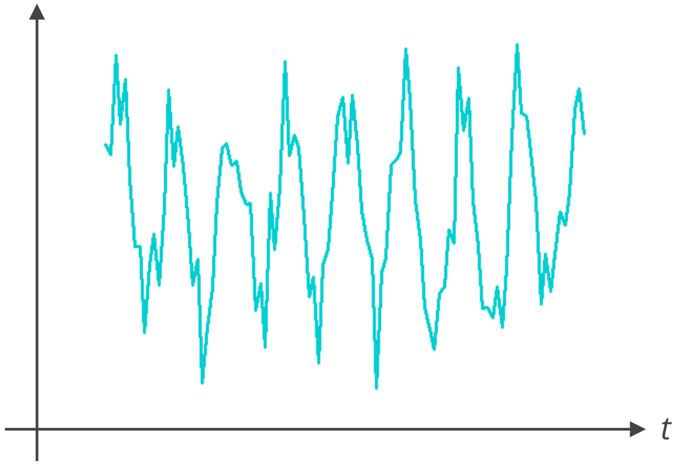
3. Componentes da Série

Tendência



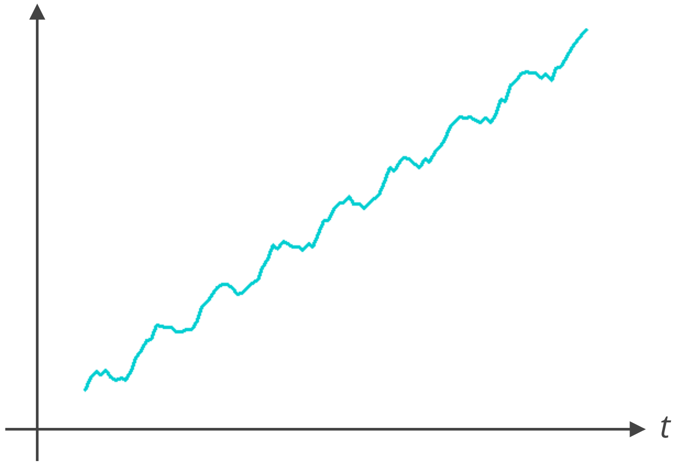
Esta série temporal apresenta **tendência crescente**, pois os valores crescem ao longo do tempo. A tendência também pode ser **decrescente**.

Sazonalidade



Esta série temporal apresenta **sazonalidade**, pois ocorrem padrões de oscilação em instantes **igualmente espaçados** no tempo.

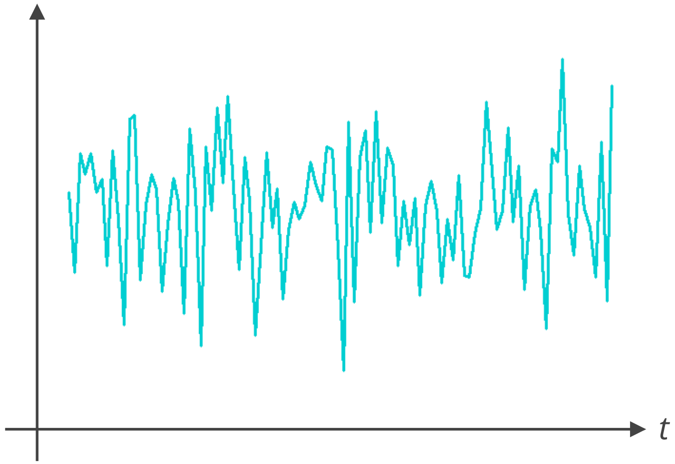
Tendência e Sazonalidade



Esta série temporal apresenta **tendência** e **sazonalidade**, pois os valores crescem ao longo do tempo e apresentam picos igualmente espaçados.

Estacionariedade

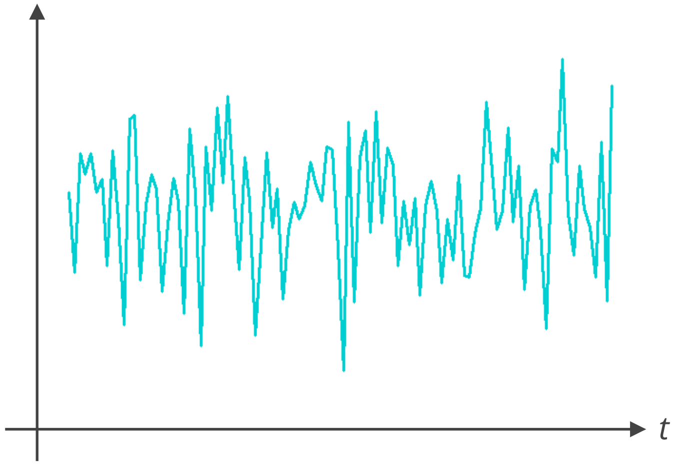
Estacionaridade



- Quanto a série temporal **não** apresenta tendência nem sazonalidade, dizemos que ela é **estacionária**.
- Em uma série estacionária, os valores oscilam em torno de uma **média constante**, com **variabilidade constante**.
- Alguns dos **modelos** que vamos estudar só podem ser utilizados para séries temporais estacionárias. Por isso, é importante avaliar se a série atende ou não a este requisito.
- Podemos fazer isso por meio de um teste de hipóteses chamado **Teste de Dickey-Fuller Generalizado (DF-GLS)**.

Teste de Dickey-Fuller Generalizado (DF-GLS)

Estacionaridade

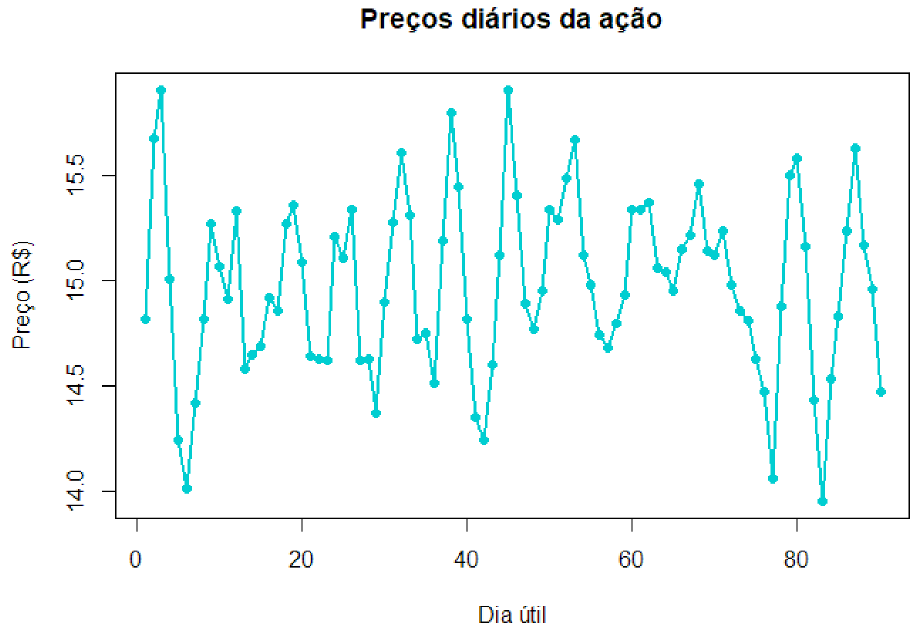


- Os cálculos por trás da estatística do **Teste de Dickey-Fuller Generalizado (DF-GLS)** se baseiam na tentativa de explicar os valores de uma série temporal a partir de diferenças em relação aos valores anteriores, como em uma **regressão**.
- As hipóteses testadas são:
  - H: a série é **não-estacionária**
  - A: a série é **estacionária**
- Caso o valor da estatística do teste seja inferior aos valores críticos (em geral, 5% ou 10%), **rejeitamos a hipótese H** e concluímos que a série é estacionária.

Case: Preços Diários de Ação

Um investidor deseja prever quais serão os **preços** de uma determinada ação nos próximos **3 dias úteis**, com base no histórico dos últimos 90 dias úteis.

DIA_UTIL	PRECO_ACAO
1	14,82
2	15,68
3	15,91
4	15,01
5	14,24
6	14,01
7	14,42
8	14,82
9	15,27
10	15,07
...	...



A série é estacionária?

Arquivo: Precos_Acao (.txt)

Um investidor deseja prever quais serão os **preços** de uma determinada ação nos próximos **3 dias úteis**, com base no histórico dos últimos 90 dias úteis.

```
> summary(ur.ers(serie))

###############################################
# Elliot, Rothenberg and Stock Unit Root Test #
###############################################

Test of type DF-GLS
detrending of series with intercept

Value of test-statistic is: -3.3382

Critical values of DF-GLS are:
                 1pct  5pct 10pct
critical values -2.59 -1.94 -1.62
```

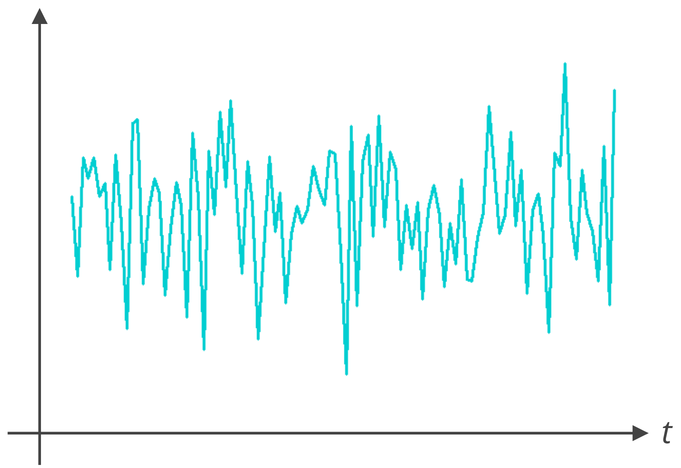
O teste DF-GLS forneceu uma estatística de teste de **-3,3382**, inferior ao valor crítico de 1% **(-2,59)**. Isso indica que o ***p*-valor** do teste seria **inferior a 1%**. Assim, concluímos que os preços da ação oscilaram em torno de uma média constante, com variabilidade constante, no período analisado.

Arquivo: Precos_Acao (.txt)

4. Modelos para Séries Estacionárias

Como Modelar uma Série Estacionária?

- Quando uma série é estacionária, a **modelagem** do comportamento histórico e **previsão** do comportamento futuro são mais simples, dado que não precisamos modelar tendências e sazonalidades.
- Os modelos mais comuns para séries estacionárias são:
  - **AR**: modelo autorregressivo *(autoregressive)*
  - **MA**: modelo de médias móveis *(moving average)*
  - **ARMA**: modelo autorregressivo de médias móveis
- Todos esses modelos são chamados **lineares** pois modelam os valores da série como uma combinação linear de comportamentos observados no passado.



Modelo Autorregressivo (AR)

O **modelo autorregressivo** (AR) consiste em estabelecer que o valor da série no instante de tempo $\boldsymbol{t}$ depende linearmente dos valores em instantes de tempo anteriores: $\boldsymbol{t-1}$, $\boldsymbol{t-2}$, $\boldsymbol{t-3}$ etc.

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \epsilon_t$$

- Y_t é o valor da série no instante t
- $Y_{t-1}, \dots, Y_{t-p}$ são os valores da série nos instantes anteriores $t-1, \dots, t-p$
- $\phi_1, \dots, \phi_p$ são os parâmetros associados aos valores $Y_{t-1}, \dots, Y_{t-p}$
- ϵ_t é o erro aleatório associado ao modelo no instante t, com média 0 e variância constante

* Letra grega ϕ: lê-se como '*fi*'

* Letra grega ϵ: lê-se como '*épsilon*'

Modelo Autorregressivo (AR)

O **modelo autorregressivo** (AR) consiste em estabelecer que o valor da série no instante de tempo $\boldsymbol{t}$ depende linearmente dos valores em instantes de tempo anteriores: $\boldsymbol{t-1}$, $\boldsymbol{t-2}$, $\boldsymbol{t-3}$ etc.

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \epsilon_t$$

A quantidade de instantes passados é chamada **ordem do modelo AR**, e denotada pela letra ***p***.

Exemplos:

- Modelo AR(p = 1) ou AR(1): $Y_t = \phi_1 Y_{t-1} + \epsilon_t$
- Modelo AR(p = 2) ou AR(2): $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t$
- Modelo AR(p = 5) ou AR(5): $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \phi_4 Y_{t-4} + \phi_5 Y_{t-5} + \epsilon_t$

* Letra grega ϕ: lê-se como *'fi'*
* Letra grega ϵ: lê-se como *'épsilon'*

Modelo de Médias Móveis (MA)

O **modelo de médias móveis** (MA) consiste em estabelecer que o valor da série no instante de tempo $\boldsymbol{t}$ corresponde a uma média constante μ, acrescida de uma combinação linear de erros ao redor de μ em instantes de tempo anteriores: $\boldsymbol{t-1}$, $\boldsymbol{t-2}$, $\boldsymbol{t-3}$ etc.

$$Y_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q}$$

- Y_t é o valor da série no instante t
- μ é a média constante da série
- ϵ_t é o erro aleatório em torno de μ no instante t, com média 0 e variância constante
- $\epsilon_{t-1}, \dots, \epsilon_{t-q}$ são os erros aleatórios em torno de μ nos instantes $t-1, \dots, t-q$, com média 0 e variância constante
- $\theta_1, \dots, \theta_q$ são os parâmetros associados aos erros $\epsilon_{t-1}, \dots, \epsilon_{t-q}$

* Letra grega μ: lê-se como '*mi*'
* Letra grega θ: lê-se como '*téta*'
* Letra grega ϵ: lê-se como '*épsilon*'

Modelo de Médias Móveis (MA)

O **modelo de médias móveis** (MA) consiste em estabelecer que o valor da série no instante de tempo $\boldsymbol{t}$ corresponde a uma média constante μ, acrescida de uma combinação linear de erros ao redor de μ em instantes de tempo anteriores: $\boldsymbol{t-1}$, $\boldsymbol{t-2}$, $\boldsymbol{t-3}$ etc.

$$Y_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q}$$

A quantidade de instantes passados é chamada **ordem do modelo MA**, e denotada pela letra $\boldsymbol{q}$.

Exemplos:

- Modelo MA(q = 1) ou MA(1): $Y_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1}$
- Modelo MA(q = 2) ou MA(2): $Y_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2}$
- Modelo MA(q = 5) ou MA(5): $Y_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \theta_3\epsilon_{t-3} + \theta_4\epsilon_{t-4} + \theta_5\epsilon_{t-5}$

* Letra grega μ: lê-se como '*mi*'
* Letra grega θ: lê-se como '*téta*'
* Letra grega ϵ: lê-se como '*épsilon*'

Modelo Autorregressivo de Médias Móveis (ARMA)

O **modelo autorregressivo de médias móveis** (ARMA) consiste em estabelecer que o valor da série no instante de tempo $\boldsymbol{t}$ depende linearmente dos valores em instantes de tempo anteriores, acrescidos de uma combinação linear de erros em instantes de tempo anteriores: $\boldsymbol{t-1}$, $\boldsymbol{t-2}$, $\boldsymbol{t-3}$ etc.

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

- Y_t é o valor da série no instante t
- $Y_{t-1}, \dots, Y_{t-p}$ são os valores da série nos instantes anteriores $t-1, \dots, t-p$ (cumprem o papel de μ)
- $\phi_1, \dots, \phi_p$ são os parâmetros associados aos valores $Y_{t-1}, \dots, Y_{t-p}$
- ϵ_t é o erro aleatório associado à parte autorregressiva no instante t, com média 0 e variância constante
- $\epsilon_{t-1}, \dots, \epsilon_{t-q}$ são os erros aleatórios em torno da parte autorregressiva nos instantes $t-1, \dots, t-q$, com média 0 e variância constante
- $\theta_1, \dots, \theta_q$ são os parâmetros associados aos erros $\epsilon_{t-1}, \dots, \epsilon_{t-q}$

* Letra grega ϕ: lê-se como '*fi*'

* Letra grega θ: lê-se como '*téta*'

* Letra grega ϵ: lê-se como '*épsilon*'

Modelo Autorregressivo de Médias Móveis (ARMA)

O **modelo autorregressivo de médias móveis** (ARMA) consiste em estabelecer que o valor da série no instante de tempo $\boldsymbol{t}$ depende linearmente dos valores em instantes de tempo anteriores, acrescidos de uma combinação linear de erros em instantes de tempo anteriores: $\boldsymbol{t-1}$, $\boldsymbol{t-2}$, $\boldsymbol{t-3}$ etc.

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

Ou seja, o modelo ARMA é uma junção dos modelos AR e MA. As **ordens do modelo** são denotadas por ***p*** e ***q***.

Exemplos:

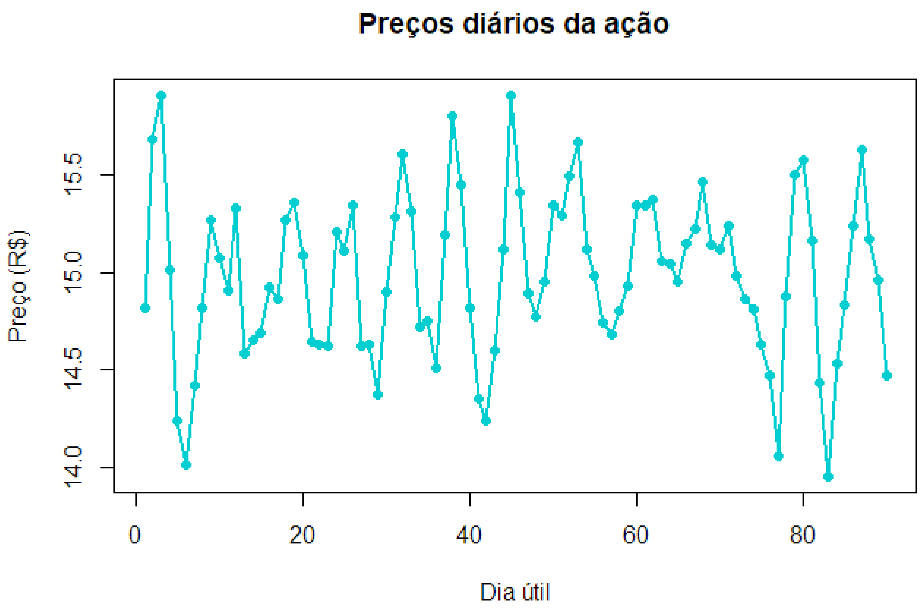
- Modelo ARMA$(p = 1, q = 1)$ ou ARMA$(1,1)$: $Y_t = \phi_1 Y_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1}$
- Modelo ARMA$(p = 1, q = 2)$ ou ARMA$(1,2)$: $Y_t = \phi_1 Y_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$
- Modelo ARMA$(p = 2, q = 2)$ ou ARMA$(2,2)$: $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$

* Letra grega ϕ: lê-se como '*fi*'

* Letra grega θ: lê-se como '*téta*'

* Letra grega ϵ: lê-se como '*épsilon*'

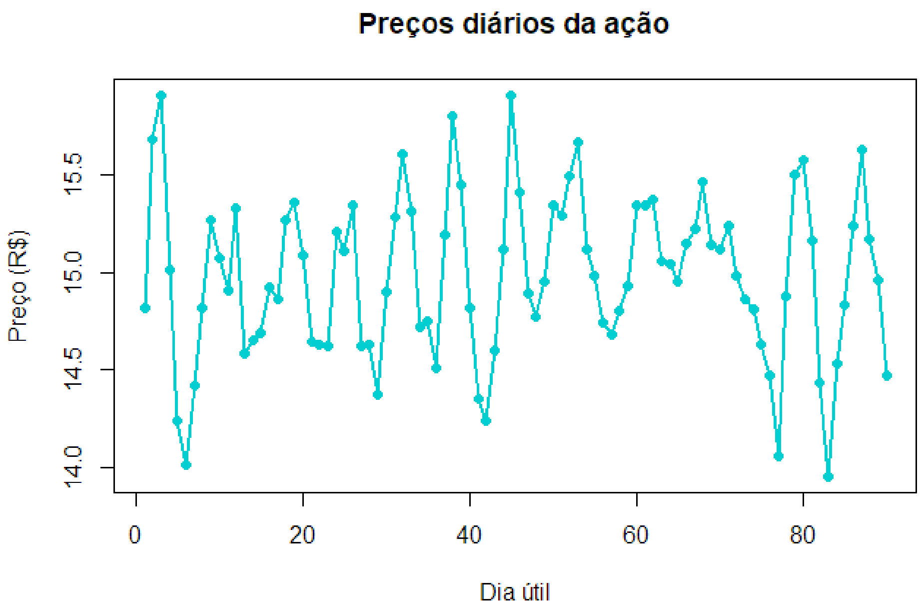
Um investidor deseja prever quais serão os **preços** de uma determinada ação nos próximos **3 dias úteis**, com base no histórico dos últimos 90 dias úteis.



- Qual o melhor modelo para esta série?

Arquivo: Precos_Acao (.txt)

Um investidor deseja prever quais serão os **preços** de uma determinada ação nos próximos **3 dias úteis**, com base no histórico dos últimos 90 dias úteis.



- Qual o melhor modelo para esta série?

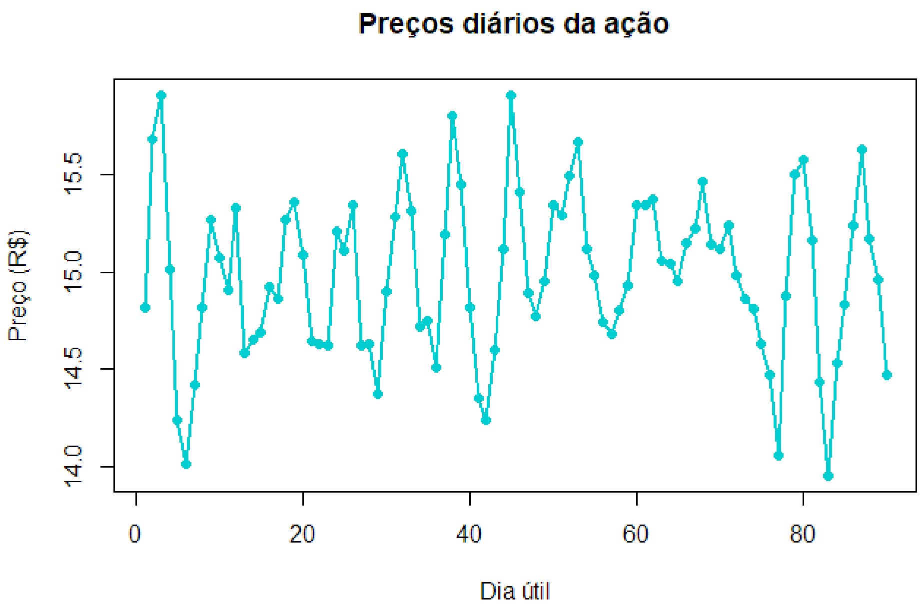
A função ***auto.arima*** do pacote ***forecast*** do R nos ajuda a identificar o melhor modelo para uma série estacionária (AR, MA ou ARMA), bem como a sua ordem.

A identificação do melhor modelo é feita por meio de testes baseados no padrão de **correlação** entre os valores históricos, a depender da distância *(lag)* entre eles.

Arquivo: Precos_Acao (.txt)

Case: Preços Diários de Ação

Um investidor deseja prever quais serão os **preços** de uma determinada ação nos próximos **3 dias úteis**, com base no histórico dos últimos 90 dias úteis.



- Qual o melhor modelo para esta série?

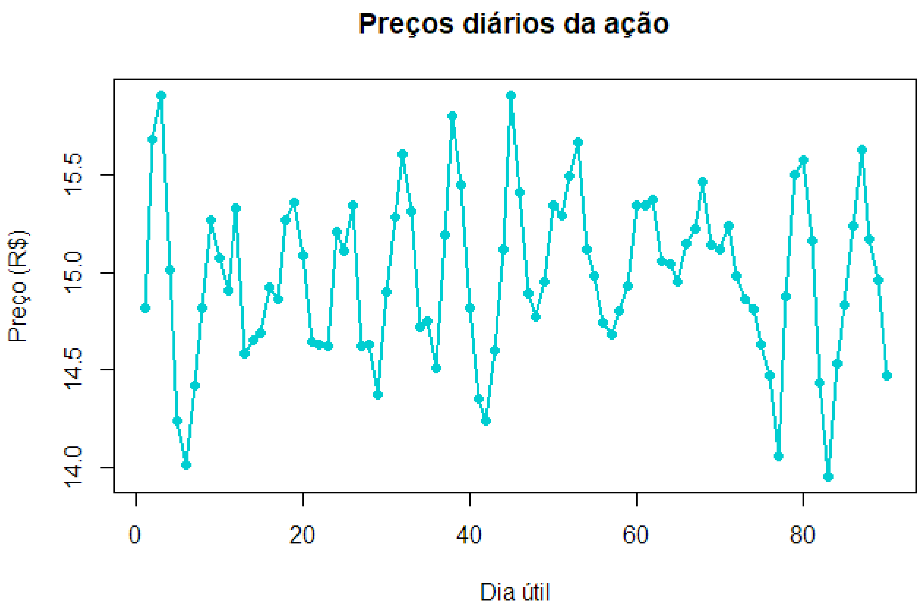
```
> modelo <- auto.arima(base_precos$PRECO_ACAO)
> print(modelo)
Series: base_precos$PRECO_ACAO
ARIMA(2,0,2) with non-zero mean

Coefficients:
         ar1      ar2      ma1     ma2     mean
      1.1288  -0.8777  -0.4721  0.3888  14.9700
s.e.  0.0758   0.0655   0.1329  0.1257   0.0336

sigma^2 = 0.071:  log likelihood = -7.13
AIC=26.26   AICc=27.27   BIC=41.26
```

Arquivo: Precos_Acao (.txt)

Case: Preços Diários de Ação

4. MODELOS PARA SÉRIES ESTACIONÁRIAS | SÉRIES TEMPORAIS

Um investidor deseja prever quais serão os **preços** de uma determinada ação nos próximos **3 dias úteis**, com base no histórico dos últimos 90 dias úteis.



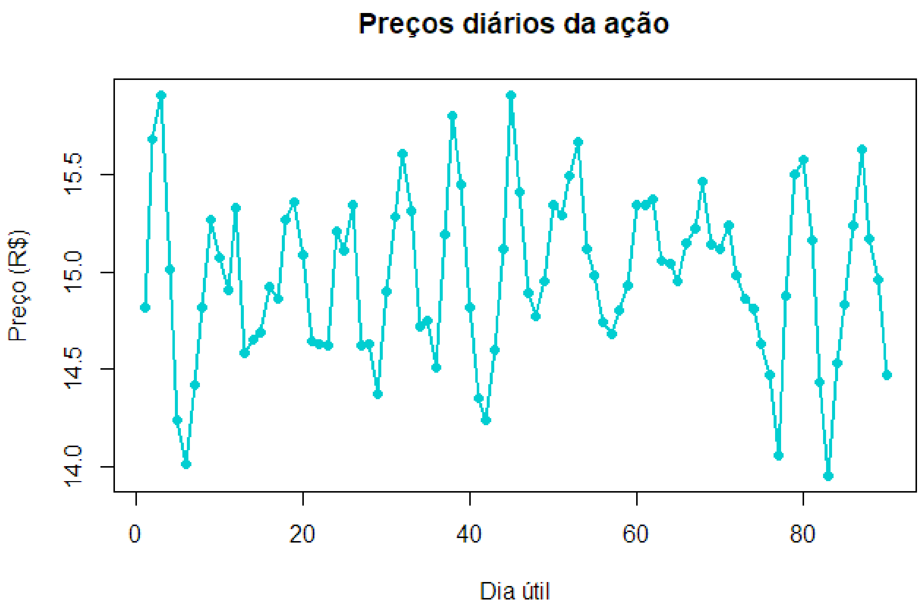
➢ Qual o melhor modelo para esta série?

```
> modelo <- auto.arima(base_precos$PRECO_ACAO)
> print(modelo)
Series: base_precos$PRECO_ACAO
ARIMA(2,0,2) with non-zero mean

Coefficients:
         ar1      ar2      ma1     ma2     mean
      1.1288  -0.8777  -0.4721  0.3888  14.9700
s.e.  0.0758   0.0655   0.1329  0.1257   0.0336

sigma^2 = 0.071:  log likelihood = -7.13
AIC=26.26   AICc=27.27   BIC=41.26
```

2 termos AR

Arquivo: Precos_Acao (.txt)

Um investidor deseja prever quais serão os **preços** de uma determinada ação nos próximos **3 dias úteis**, com base no histórico dos últimos 90 dias úteis.



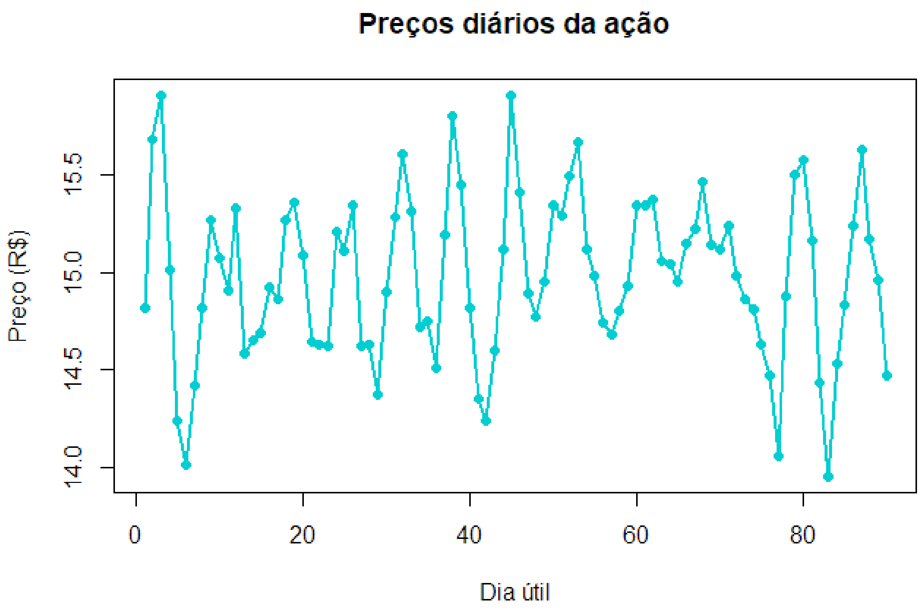
- Qual o melhor modelo para esta série?

```
> modelo <- auto.arima(base_precos$PRECO_ACAO)
> print(modelo)
Series: base_precos$PRECO_ACAO
ARIMA(2,0,2) with non-zero mean

Coefficients:
         ar1      ar2      ma1     ma2     mean
      1.1288  -0.8777  -0.4721  0.3888  14.9700
s.e.  0.0758   0.0655   0.1329  0.1257   0.0336

sigma^2 = 0.071:  log likelihood = -7.13
AIC=26.26   AICc=27.27   BIC=41.26
```

2 termos MA

Arquivo: Precos_Acao (.txt)

Um investidor deseja prever quais serão os **preços** de uma determinada ação nos próximos **3 dias úteis**, com base no histórico dos últimos 90 dias úteis.



➢ Qual o melhor modelo para esta série?

```
> modelo <- auto.arima(base_precos$PRECO_ACAO)
> print(modelo)
Series: base_precos$PRECO_ACAO
ARIMA(2,0,2) with non-zero mean

Coefficients:
         ar1      ar2      ma1     ma2     mean
      1.1288  -0.8777  -0.4721  0.3888  14.9700
s.e.  0.0___   0.0___   0.1___  0.1___   0.0336

sigma^2 = 0.071:  log likelihood = -7.13
AIC=26.26   AICc=27.27   BIC=41.26
```

ϕ_1 ϕ_2 θ_1 θ_2

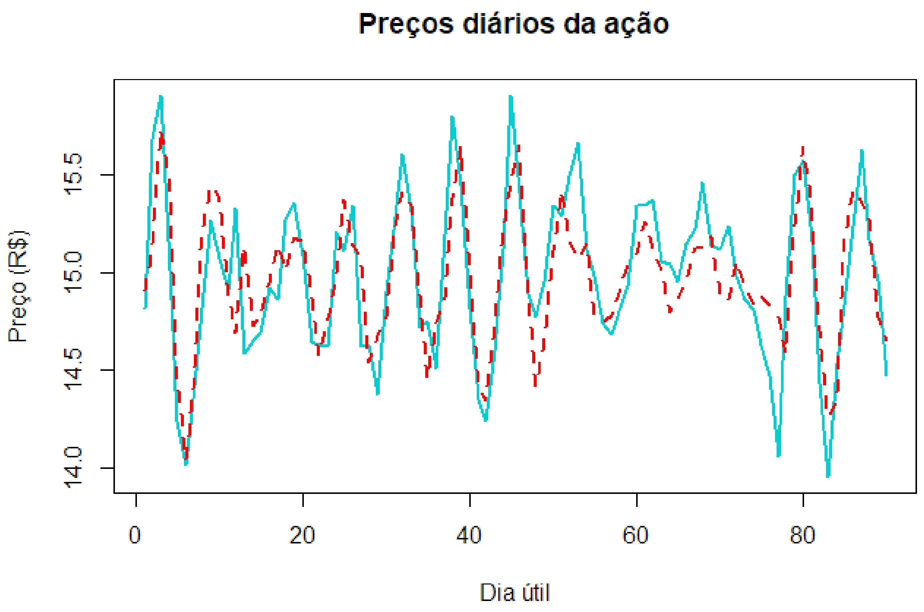
Estimativas do modelo ARMA, com *p = 2 e q = 2*

Arquivo: Precos_Acao (.txt)

FIA BUSINESS SCHOOL lab.data

Case: Preços Diários de Ação

Um investidor deseja prever quais serão os **preços** de uma determinada ação nos próximos **3 dias úteis**, com base no histórico dos últimos 90 dias úteis.



Aplicando o modelo na base de dados, podemos construir um **gráfico comparativo** dos valores reais observados *versus* valores preditos pelo modelo (ao lado).

Para avaliar a qualidade do modelo, é comum utilizar o índice **REQM**, que é a raiz quadrada do erro quadrático médio:

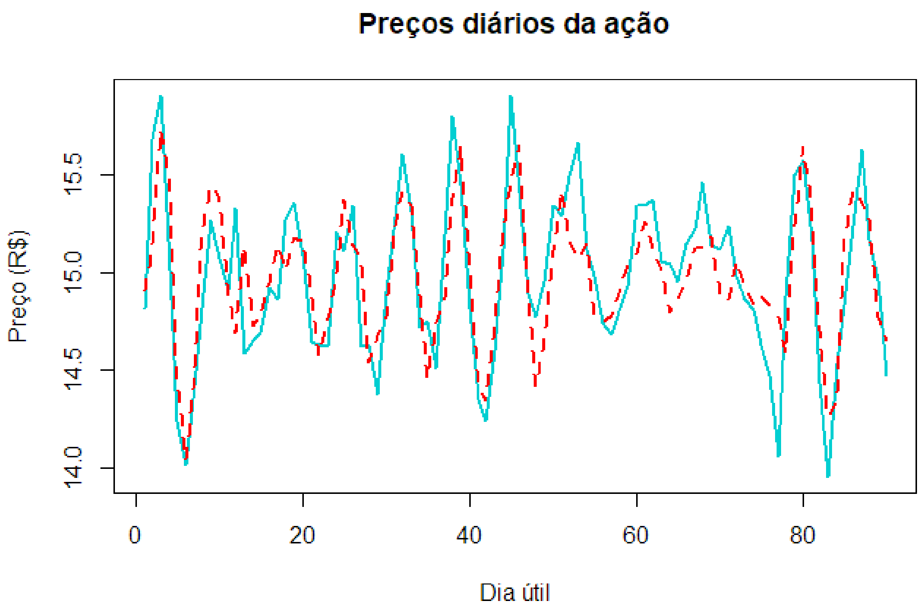
$$\text{REQM} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2}$$

Valores preditos

Valores reais

Arquivo: Precos_Acao (.txt)

Um investidor deseja prever quais serão os **preços** de uma determinada ação nos próximos **3 dias úteis**, com base no histórico dos últimos 90 dias úteis.



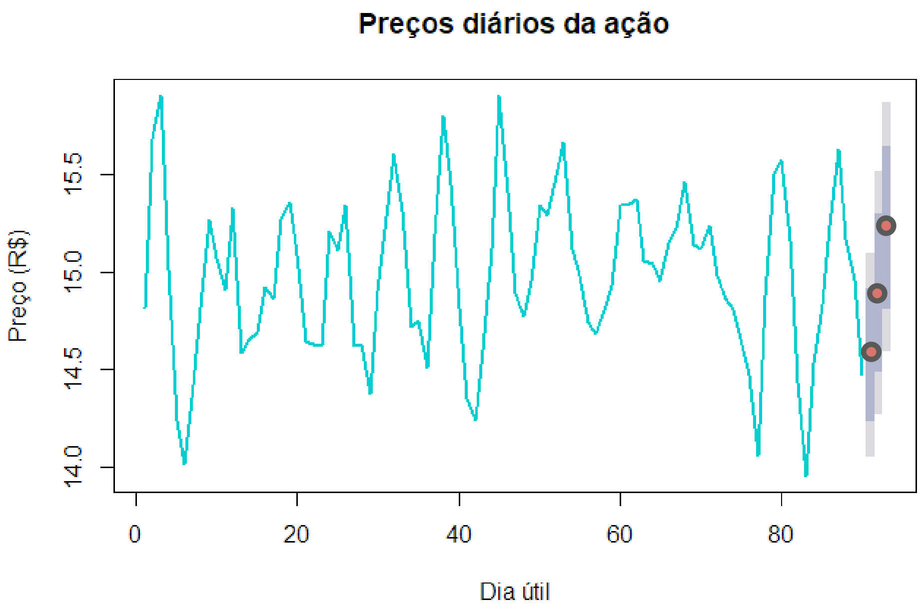
Neste *case*, **REQM = 0,259**.

Ou seja, os valores preditos pelo modelo desviam-se dos valores reais em cerca de **26 centavos**, em média.

Arquivo: Precos_Acao (.txt)

Case: Preços Diários de Ação

Um investidor deseja prever quais serão os **preços** de uma determinada ação nos próximos **3 dias úteis**, com base no histórico dos últimos 90 dias úteis.



Projetando a fórmula do modelo estimado para os próximos 3 dias úteis, obtemos as seguintes **previsões** (destacadas como pontos vermelhos no gráfico ao lado):

- ✓ Dia 91: **R$ 14,57**
- ✓ Dia 92: **R$ 14,89**
- ✓ Dia 93: **R$ 15,22**

Ou seja, estimamos que haverá **aumento** no preço da ação nos próximos dias.

Arquivo: Precos_Acao (.txt)

Funções de Autocorrelação e Autocorrelação Parcial (tópico extra)

Uma outra forma de identificar modelos para séries estacionárias é por meio da **função de autocorrelação** (ACF) e da **função de autocorrelação parcial** (PACF) associadas a uma série temporal.

Funções de Autocorrelação e Autocorrelação Parcial (tópico extra)

- **Autocorrelação**: Índice que mede a associação linear entre os valores da série em um instante t e os valores em um instante $t - k$. O valor k é chamado defasagem, ou simplesmente *lag*.

  Exemplos:

  - A autocorrelação para *lag* $k = 1$ corresponde ao grau de associação linear entre os pares $(x_t, x_{t-1}), (x_{t-1}, x_{t-2}), \ldots$
  - A autocorrelação para *lag* $k = 4$ corresponde ao grau de associação linear entre os pares $(x_t, x_{t-4}), (x_{t-1}, x_{t-5}), \ldots$

  A **função de autocorrelação** (ACF) é uma função que, para cada valor de *lag* k, associa o valor do índice de autocorrelação associado aos pares de valores da série defasados em k instantes.

  É comum representar a função de autocorrelação por meio de um **gráfico de linhas verticais**, como veremos a seguir.

- **Autocorrelação parcial**: Índice que mede a associação linear remanescente entre os valores da série em um instante t e os valores em um instante $t - k$, dados os valores de instantes intermediários $t - 1, t - 2, \ldots, t - (k - 1)$.

  Pode ser compreendida como o grau de informação adicional que o valor de k instantes anteriores fornece sobre o valor de um determinado instante, que não é explicado pelos valores dos $k - 1$ instantes mais recentes.

  Exemplos:

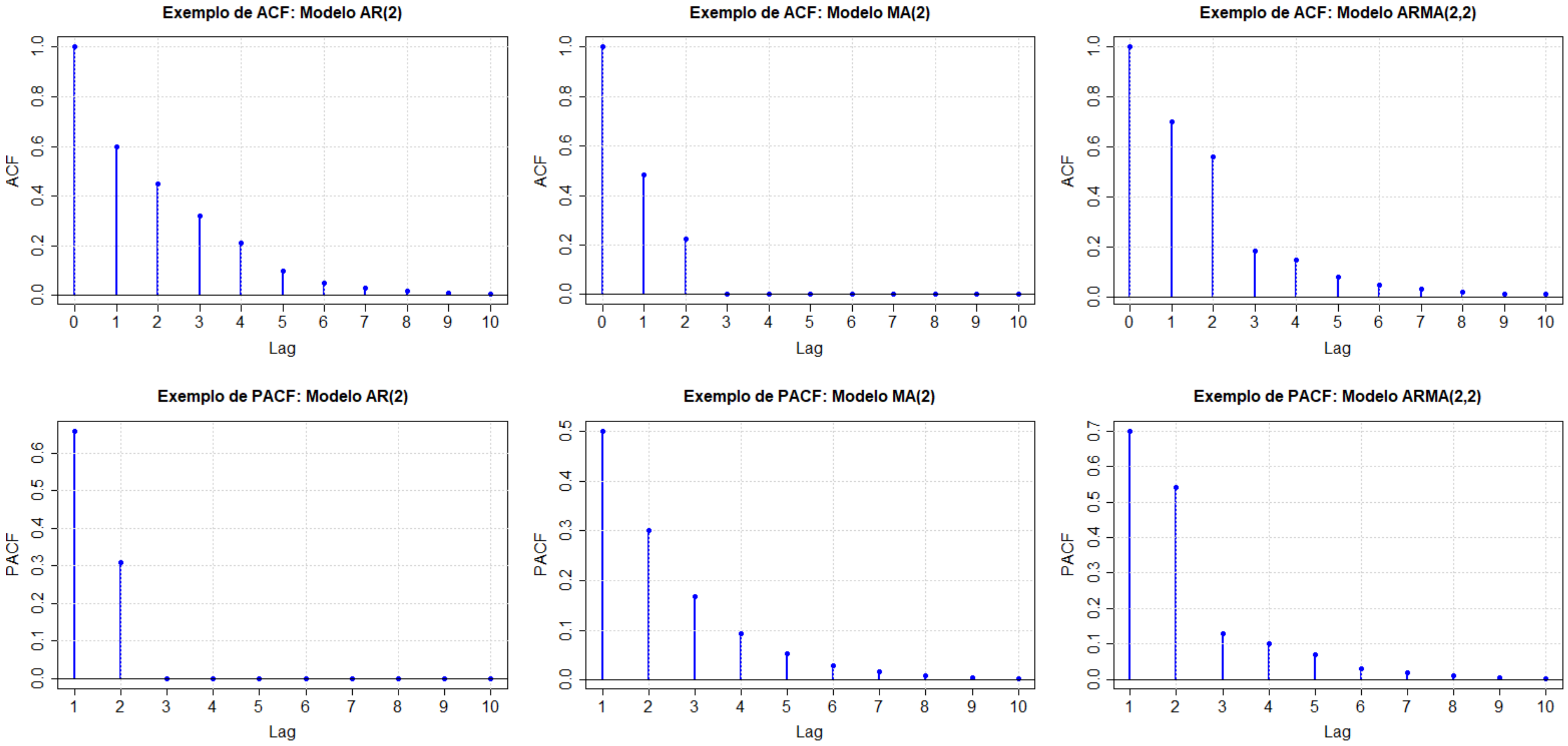
  - A autocorrelação para *lag* $k = 2$ corresponde ao grau de associação linear entre os pares $(x_t, x_{t-2}), (x_{t-1}, x_{t-3}), \ldots$, sem considerar a influência já embutida no valor de 1 instante atrás (*lag* $k = 1$).
  - A autocorrelação para *lag* $k = 5$ corresponde ao grau de associação linear entre os pares $(x_t, x_{t-5}), (x_{t-1}, x_{t-6}), \ldots$, sem considerar a influência já embutida nos valores de 1 a 4 instantes atrás (*lags* $k = 1, 2, 3, 4$).

  A **função de autocorrelação parcial** (PACF) é uma função que, para cada valor de *lag* k, associa o valor do índice de autocorrelação parcial associado aos pares de valores da série defasados em k instantes.

  Também é comum representar a função de autocorrelação parcial por meio de um **gráfico de linhas verticais**, como veremos a seguir.

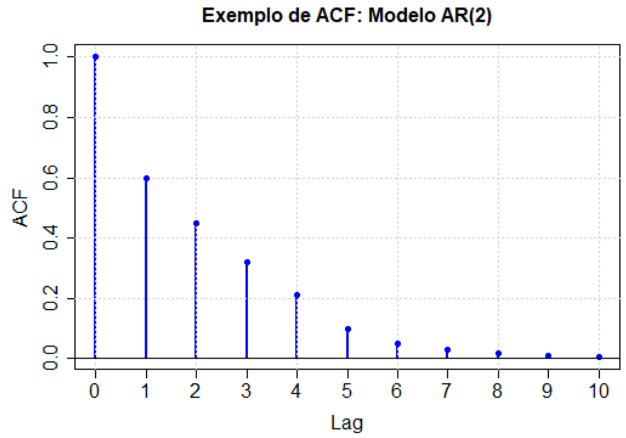
Funções de Autocorrelação e Autocorrelação Parcial (tópico extra)

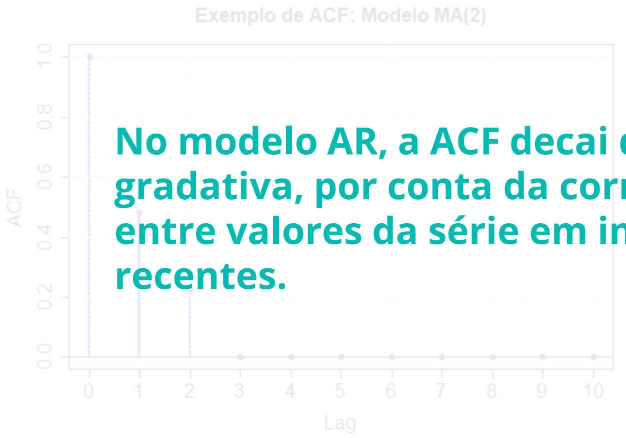
Comportamento típico dos gráficos de ACF e PACF para modelos AR (2), MA (2) e ARMA (2,2).



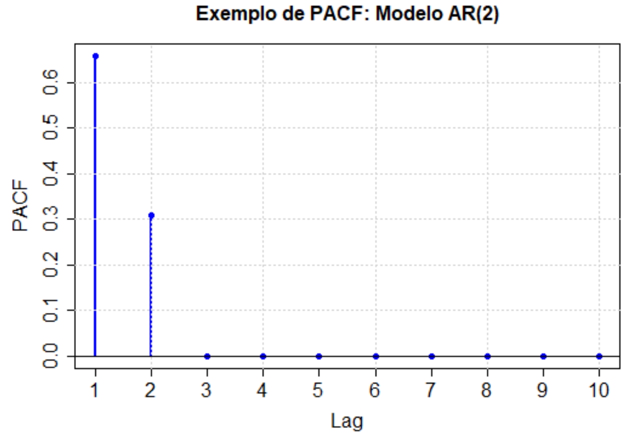
Funções de Autocorrelação e Autocorrelação Parcial (tópico extra)

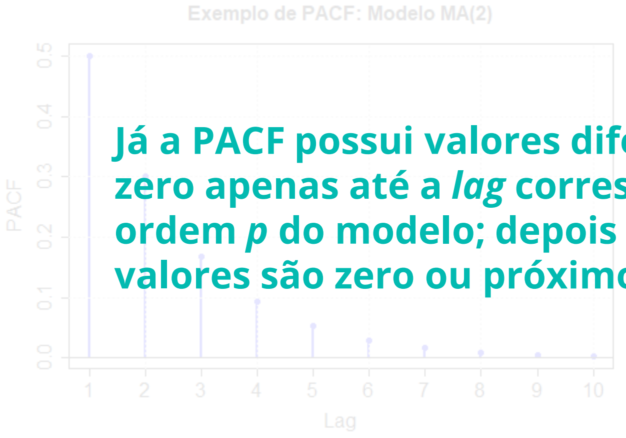
Comportamento típico dos gráficos de ACF e PACF para modelos AR(2), MA(2) e ARMA(2,2).

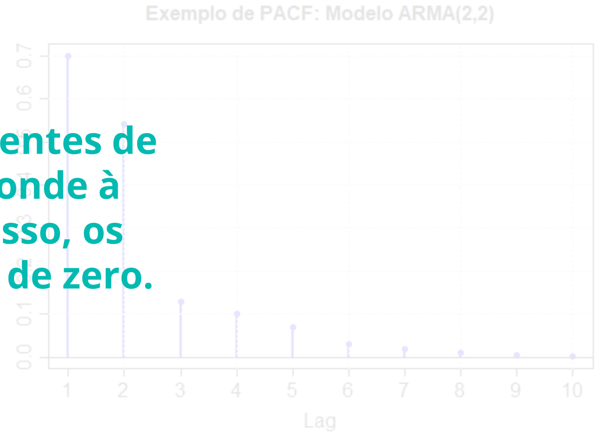




No modelo AR, a ACF decai de forma gradativa, por conta da correlação entre valores da série em instantes recentes.





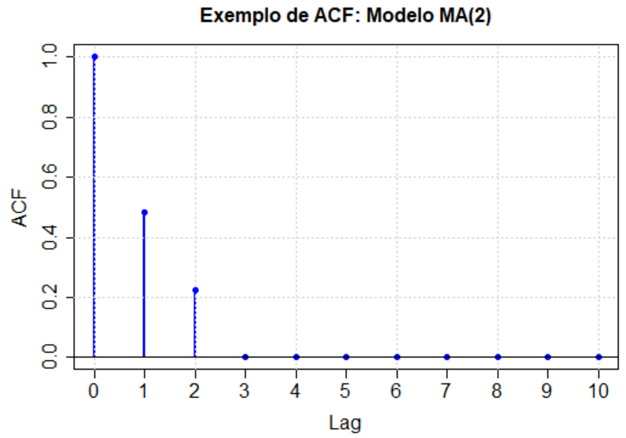


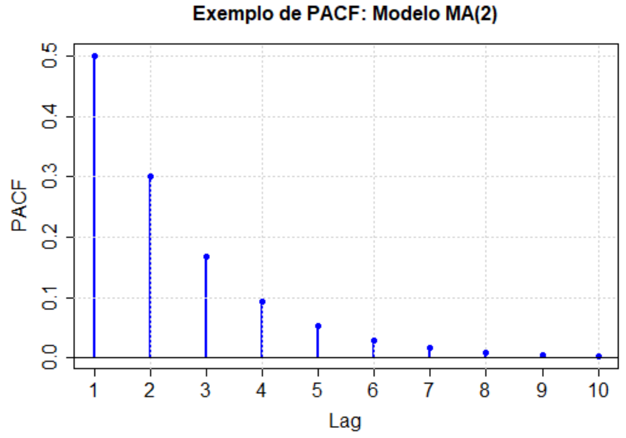
Já a PACF possui valores diferentes de zero apenas até a *lag* corresponde à ordem *p* do modelo; depois disso, os valores são zero ou próximos de zero.

BUSINESS SCHOOL lab.data

Funções de Autocorrelação e Autocorrelação Parcial (tópico extra)

Comportamento típico dos gráficos de ACF e PACF para modelos AR(2), MA(2) e ARMA(2,2).

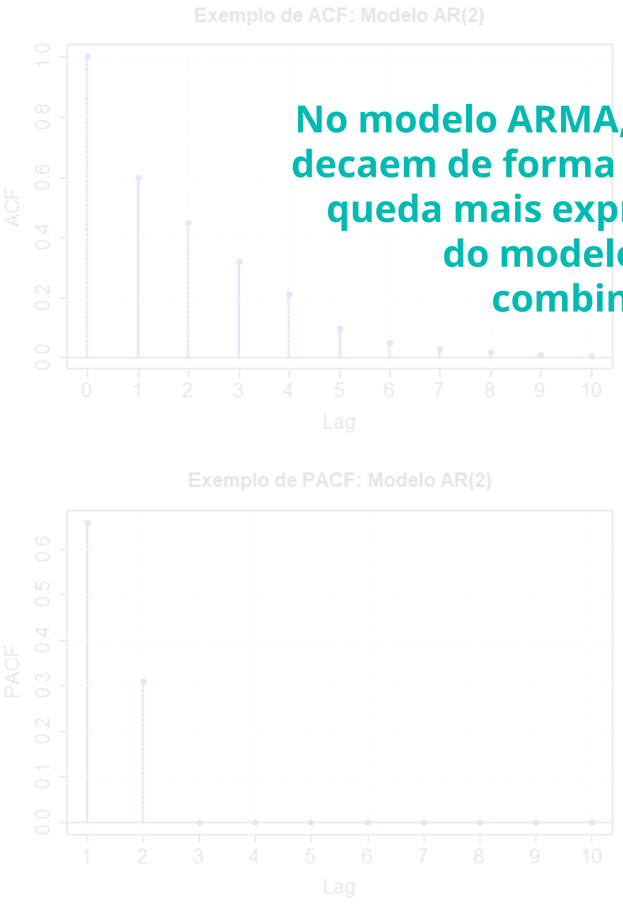




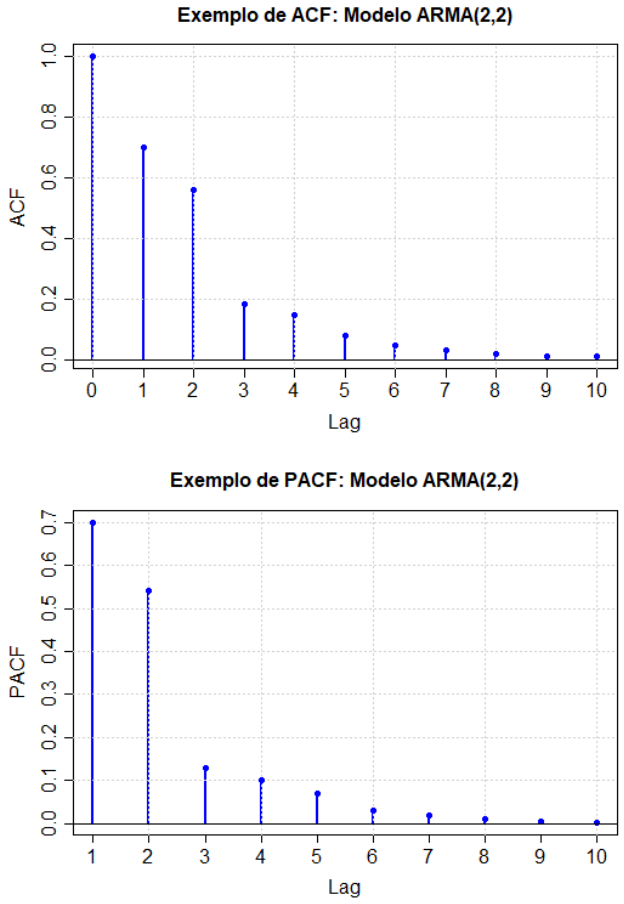
No modelo MA, a ACF possui valores diferentes de zero apenas até a *lag* corresponde à ordem *q* do modelo; depois disso, os valores são zero ou próximos de zero.

Já a PACF decai de forma gradativa, por conta da correlação entre valores da série e os erros de instantes recentes.

Comportamento típico dos gráficos de ACF e PACF para modelos AR(2), MA(2) e ARMA(2,2).



No modelo ARMA, tanto a ACF quanto a PACF decaem de forma gradativa, com ênfase para queda mais expressiva após as ordens *p* e *q* do modelo, visto que se trata de uma combinação dos modelos AR e MA.

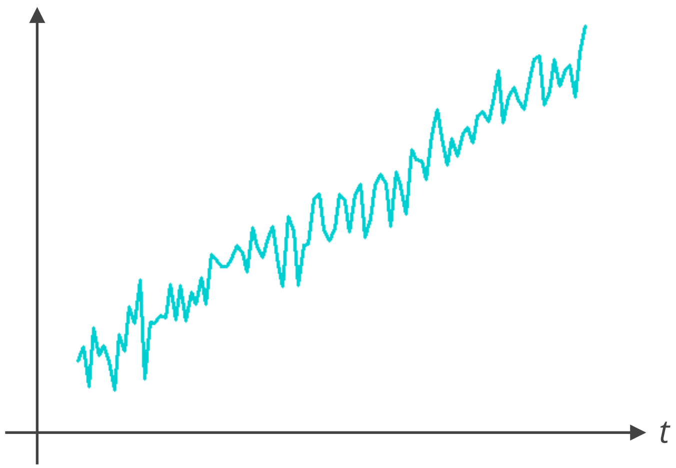


FIA BUSINESS SCHOOL lab.data

5. Modelos para Séries Não Estacionárias

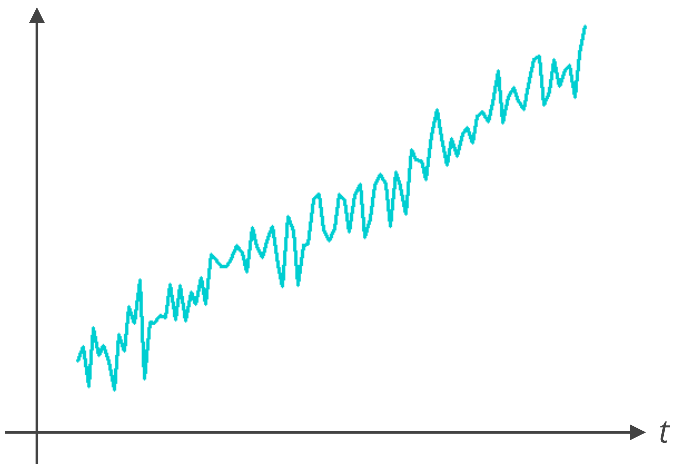
Como Modelar uma Série Não Estacionária?

- Quando uma série não é estacionária, a **modelagem** do comportamento histórico e **previsão** do comportamento futuro envolvem aplicar transformações para que a série se torne estacionária e possa, então, ser explicada pelos modelos que já estudamos: AR, MA, ARMA.
- Os modelos mais comuns para séries não estacionárias são:
  - **ARIMA**: modelo autorregressivo de média móveis integrado
  - **SARIMA**: modelo autorregressivo de média móveis integrado sazonal



Como Modelar uma Série Não Estacionária?

- Quando uma série não é estacionária, a **modelagem** do comportamento histórico e **previsão** do comportamento futuro envolvem aplicar transformações para que a série se torne estacionária e possa, então, ser explicada pelos modelos que já estudamos: AR, MA, ARMA.
- Os modelos mais comuns para séries não estacionárias são:
  - **ARIMA**: modelo autorregressivo de média móveis integrado
  - **SARIMA**: modelo autorregressivo de média móveis integrado sazonal



Estes modelos são generalizações do modelo **ARMA** para séries que possuem tendência e/ou sazonalidade

Modelo Autorregressivo de Médias Móveis Integrado (ARIMA)

O **modelo autorregressivo de médias móveis integrado** (ARIMA) consiste em um modelo ARMA aplicado após um tratamento da série original, realizado para remover o componente de **tendência**. Este tratamento é chamado ***integração*** (ou *diferenciação*), cuja inicial motiva o acréscimo da letra I à sigla.

A integração consiste em definir uma nova série Z_t cujos valores são as **diferenças** entre valores consecutivos da série original Y_t, ou seja:

$$Z_t = Y_t - Y_{t-1}$$

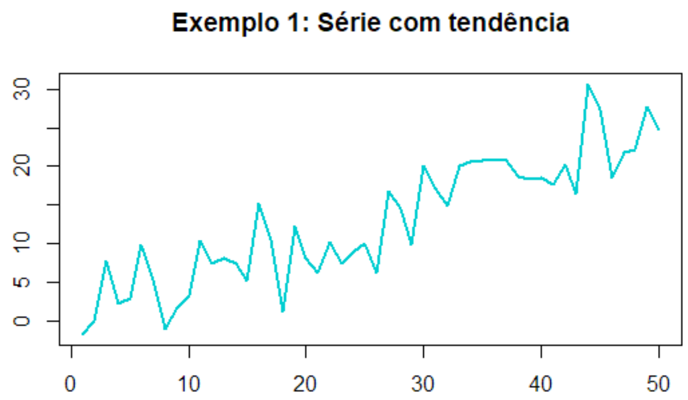
Em muitos casos, realizar este tratamento já é suficiente para **remover as tendências** e tornar Z_t uma série **estacionária**, sobre a qual o modelo ARMA poderá ser aplicado.

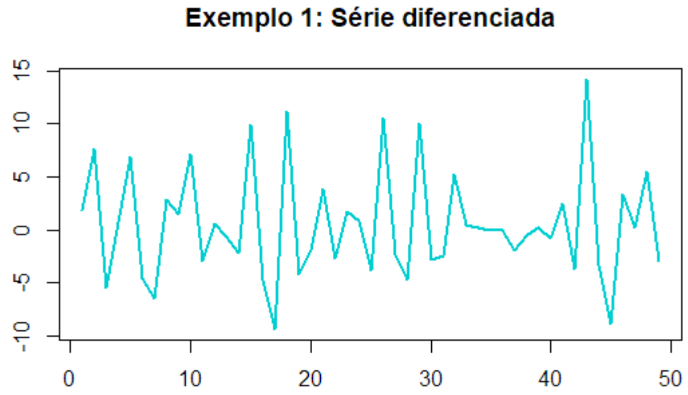
Caso as tendências da série original possuam um comportamento mais complexo, pode ocorrer de uma única integração não ser suficiente para tornar a série estacionária. Neste caso, **novas integrações** podem ser realizadas, de forma cumulativa. Por exemplo, a segunda integração seria:

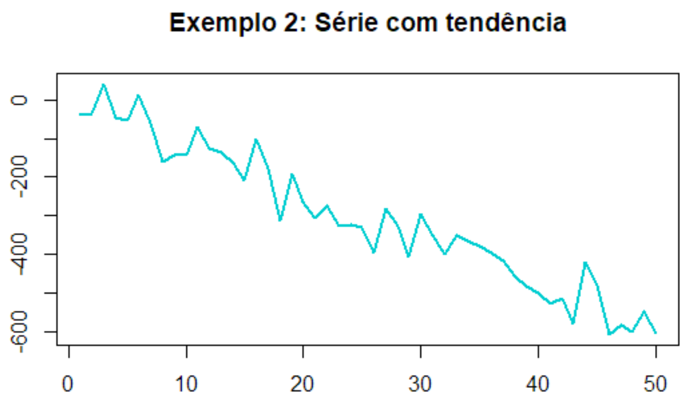
$$Z_t - Z_{t-1}$$

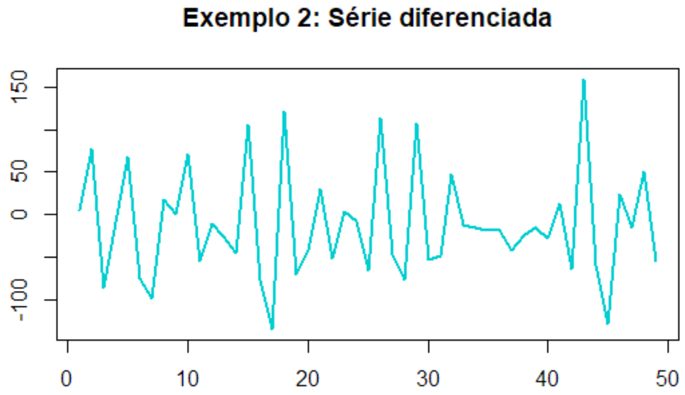
Modelo Autorregressivo de Médias Móveis Integrado (ARIMA)

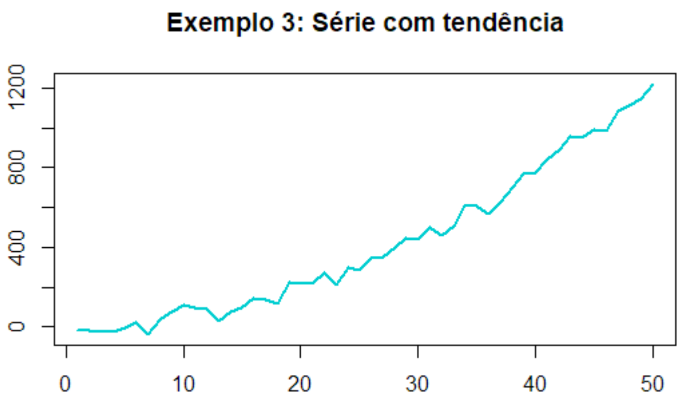
Exemplos:

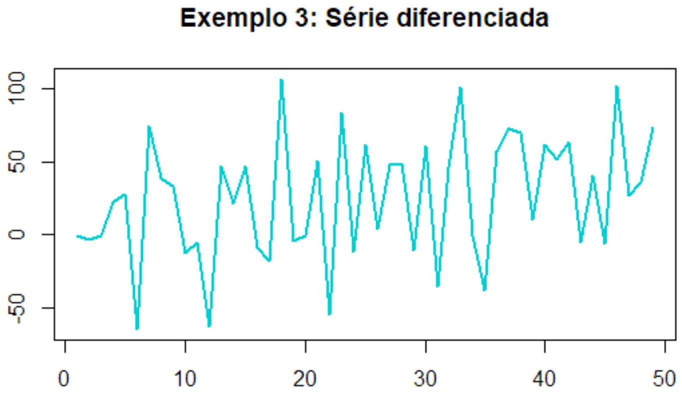






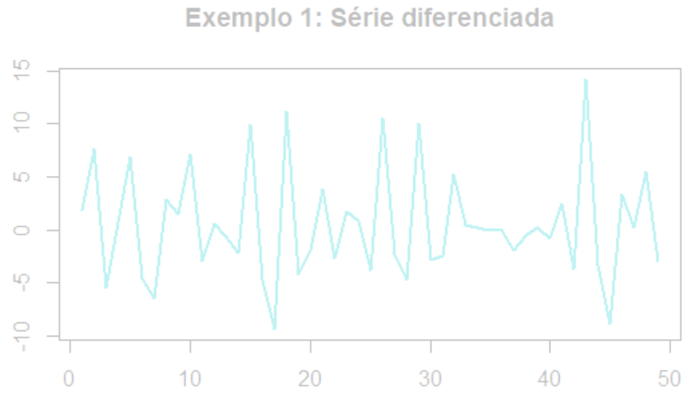


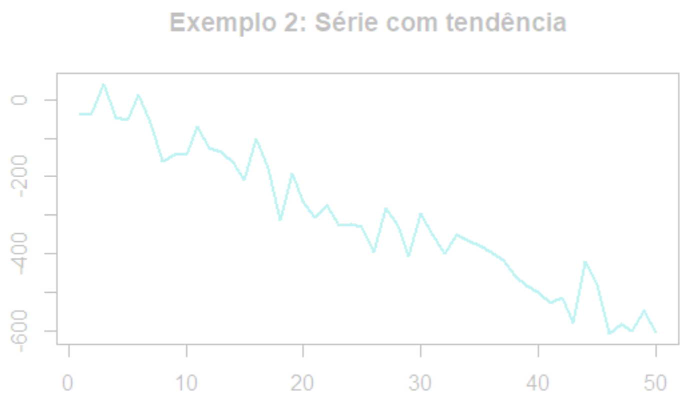


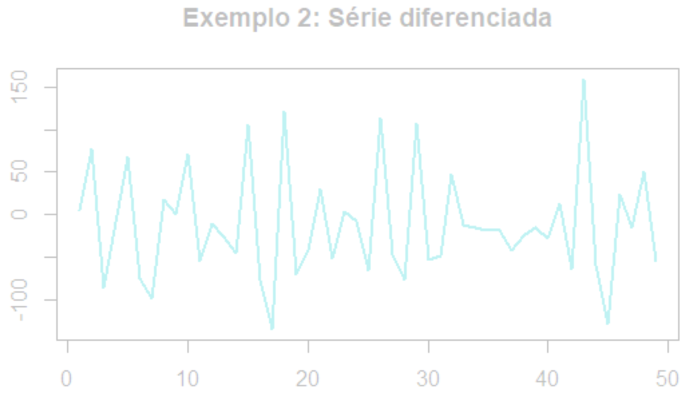


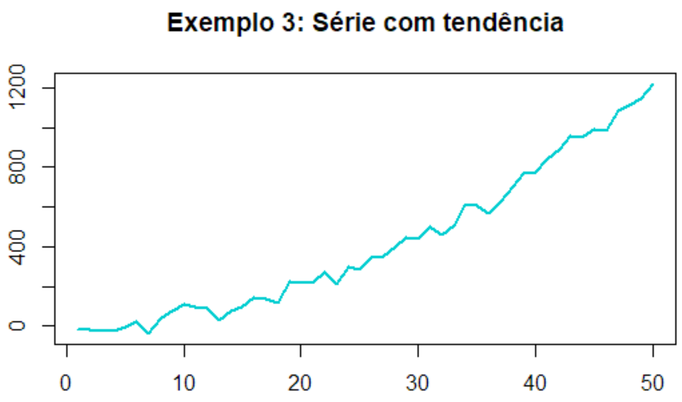
Modelo Autorregressivo de Médias Móveis Integrado (ARIMA)

Exemplos:

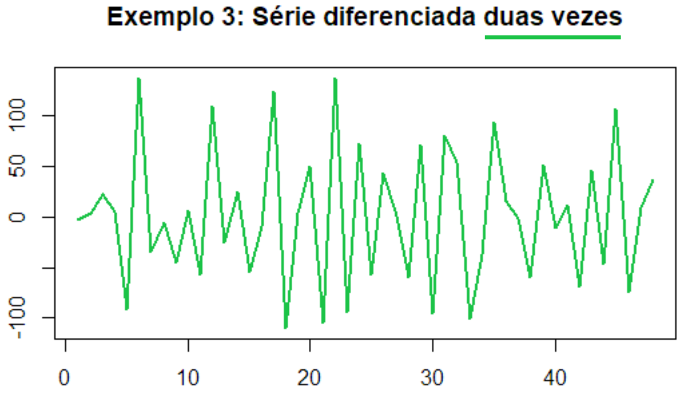








A segunda diferenciação parece necessária aqui



Dessa forma, o **modelo autorregressivo de médias móveis integrado** (ARIMA) é dado por:

$$Z_t = \phi_1 Z_{t-1} + \cdots + \phi_p Z_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

sendo:

$$Z_t = (1 - B)^d Y_t$$

onde $(1 - B)$ é um símbolo que corresponde ao tratamento de **integração**, realizado ***d*** vezes cumulativas sobre a série original Y_t até que ela se torne estacionária.

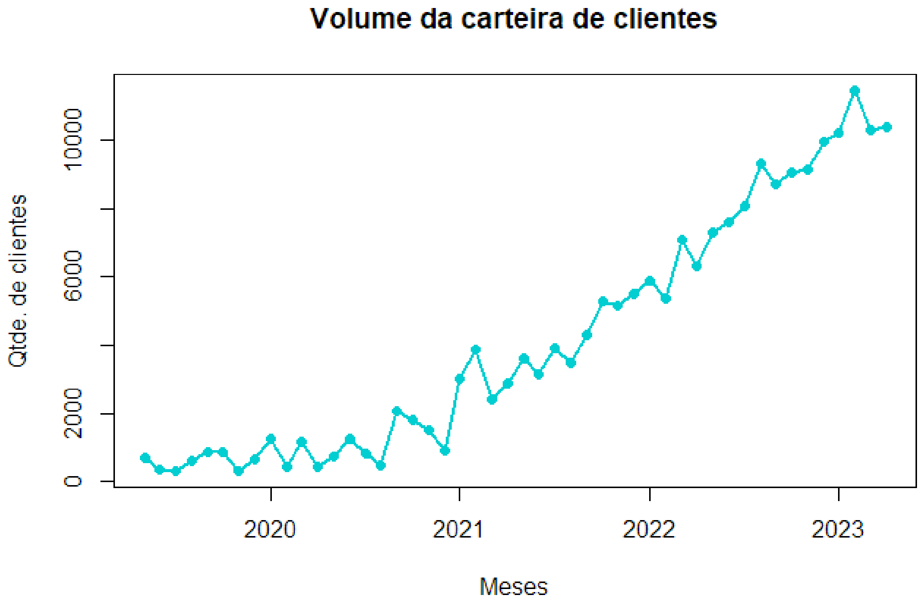
As **ordens do modelo** são denotadas por ***p***, ***d*** e ***q***. O modelo *ARIMA (**p**,**d**,**q**)* é um modelo *ARMA (**p**,**q**)* com ***d*** operações de integração.

Exemplos:

- Modelo *ARIMA (p = 1, d = 1, q = 2)* ou *ARIMA (1,1,2)*: $Z_t = \phi_1 Z_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$, onde $Z_t = (1 - B)\, Y_t$
- Modelo *ARIMA (p = 2, d = 2, q = 1)* ou *ARIMA (2,2,1)*: $Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \epsilon_t + \theta_1 \epsilon_{t-1}$, onde $Z_t = (1 - B)^2\, Y_t$

Case: Carteira de Clientes

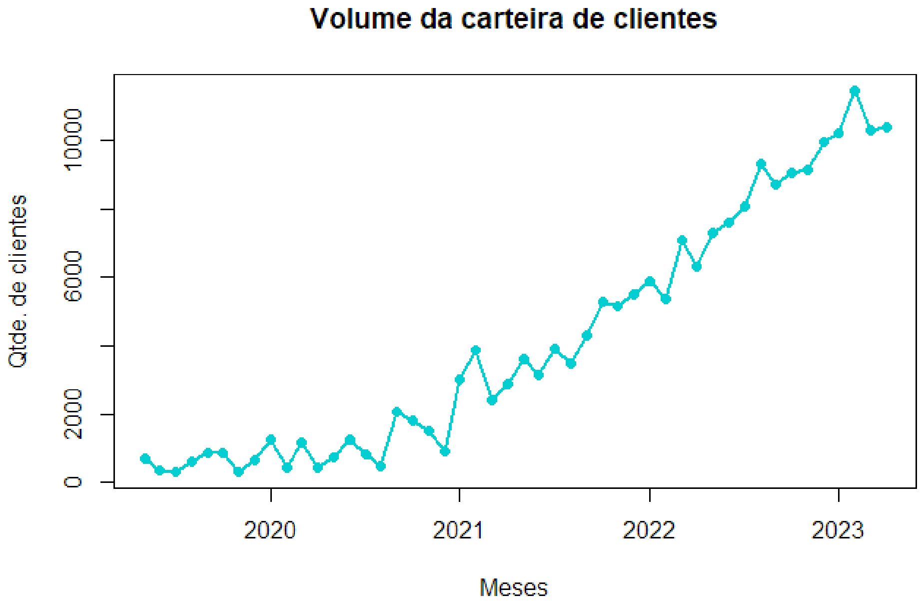
Uma empresa de serviços financeiros deseja projetar o volume de sua **carteira de clientes** para o mês seguinte, com base no histórico de 48 meses anteriores (mai/19 a abr/23).



- Qual o melhor modelo para esta série?

Arquivo: Carteira_Clientes (.txt)

Uma empresa de serviços financeiros deseja projetar o volume de sua **carteira de clientes** para o mês seguinte, com base no histórico de 48 meses anteriores (mai/19 a abr/23).



- Qual o melhor modelo para esta série?

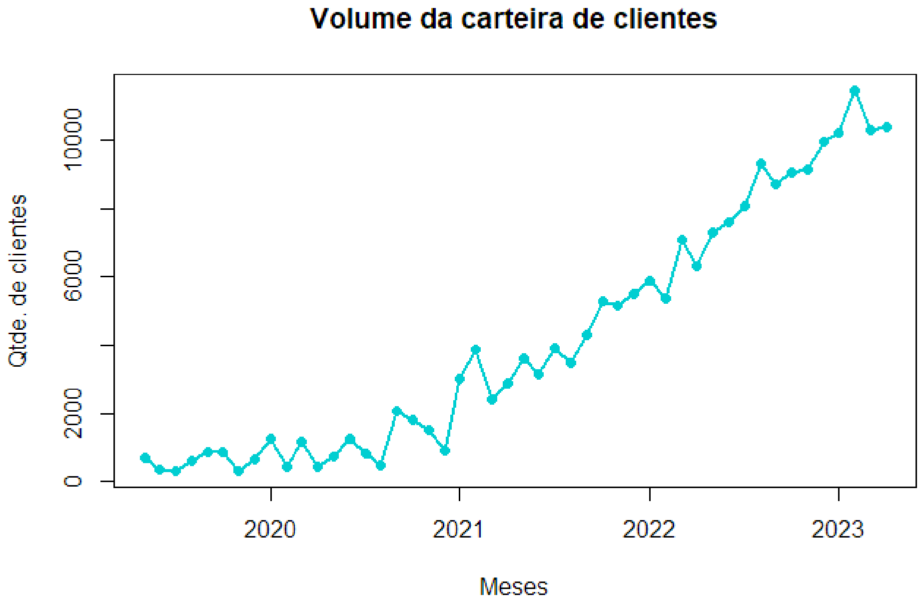
Novamente, a função ***auto.arima*** do pacote ***forecast*** do R nos ajuda a identificar o melhor modelo para uma série não estacionária (ARIMA), bem como a sua ordem.

A identificação do melhor modelo é feita por meio de testes baseados no padrão de **correlação** entre os valores históricos, a depender da distância *(lag)* entre eles.

Arquivo: Carteira_Clientes (.txt)

Case: Carteira de Clientes

Uma empresa de serviços financeiros deseja projetar o volume de sua **carteira de clientes** para o mês seguinte, com base no histórico de 48 meses anteriores (mai/19 a abr/23).



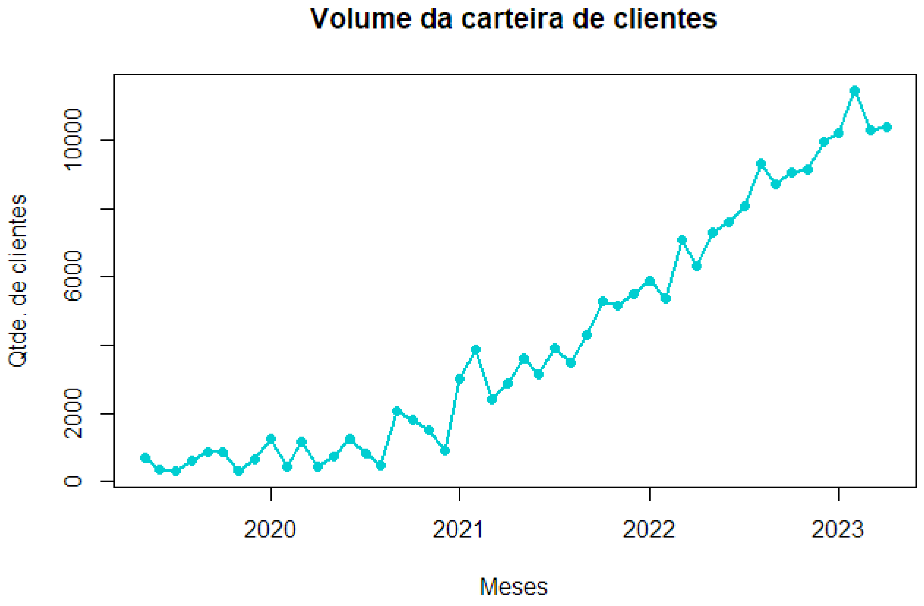
- Qual o melhor modelo para esta série?

```
> modelo <- auto.arima(base_clientes$QTDE_CLIENTES)
> print(modelo)
Series: base_clientes$QTDE_CLIENTES
ARIMA(1,1,2)

Coefficients:
         ar1      ma1     ma2
      0.9736  -1.6951  0.7852
s.e.  0.0391   0.1266  0.0977

sigma^2 = 391955:  log likelihood = -369.02
AIC=746.05   AICc=747   BIC=753.45
```

Arquivo: Carteira_Clientes (.txt)

BUSINESS SCHOOL

Uma empresa de serviços financeiros deseja projetar o volume de sua **carteira de clientes** para o mês seguinte, com base no histórico de 48 meses anteriores (mai/19 a abr/23).



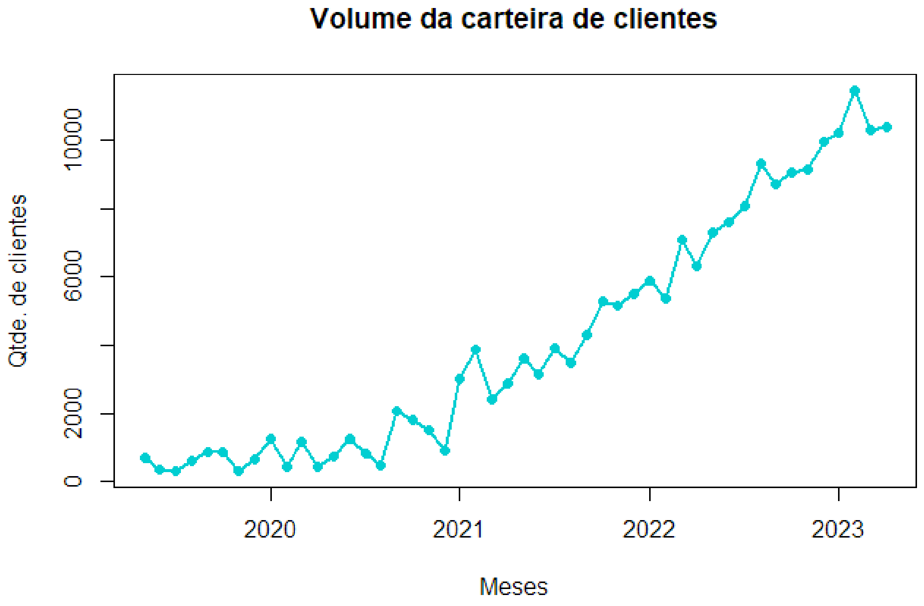
- Qual o melhor modelo para esta série?

```
> modelo <- auto.arima(base_clientes$QTDE_CLIENTES)
> print(modelo)
Series: base_clientes$QTDE_CLIENTES
ARIMA(1,1,2)

Coefficients:
         ar1      ma1     ma2
      0.9736  -1.6951  0.7852
s.e.  0.0391   0.1266  0.0977

sigma^2 = 391955:  log likelihood = -369.02
AIC=746.05   AICc=747   BIC=753.45
```

1 termo AR

Arquivo: Carteira_Clientes (.txt)

Uma empresa de serviços financeiros deseja projetar o volume de sua **carteira de clientes** para o mês seguinte, com base no histórico de 48 meses anteriores (mai/19 a abr/23).



- Qual o melhor modelo para esta série?

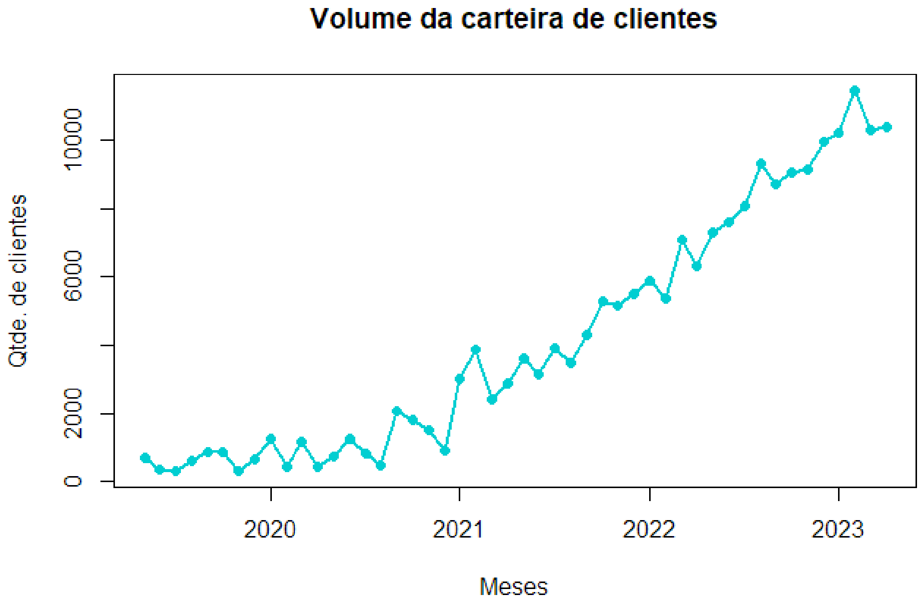
```
> modelo <- auto.arima(base_clientes$QTDE_CLIENTES)
> print(modelo)
Series: base_clientes$QTDE_CLIENTES
ARIMA(1,1,2)

Coefficients:
         ar1      ma1     ma2
      0.9736  -1.6951  0.7852
s.e.  0.0391   0.1266  0.0977

sigma^2 = 391955:  log likelihood = -369.02
AIC=746.05   AICc=747   BIC=753.45
```

2 termos MA

Arquivo: Carteira_Clientes (.txt)

BUSINESS SCHOOL

Uma empresa de serviços financeiros deseja projetar o volume de sua **carteira de clientes** para o mês seguinte, com base no histórico de 48 meses anteriores (mai/19 a abr/23).



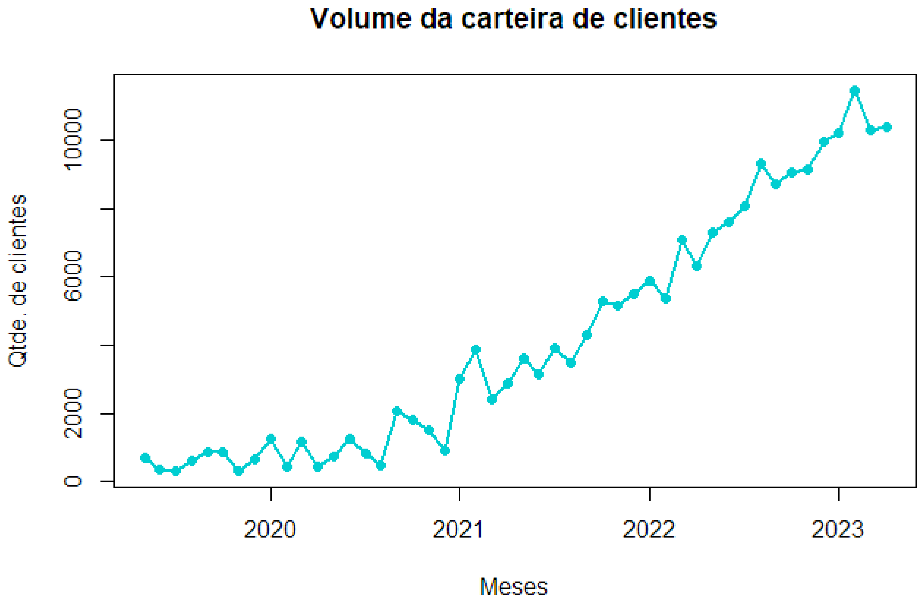
- Qual o melhor modelo para esta série?

```
> modelo <- auto.arima(base_clientes$QTDE_CLIENTES)
> print(modelo)
Series: base_clientes$QTDE_CLIENTES
ARIMA(1,1,2)

Coefficients:
         ar1      ma1     ma2
      0.9736  -1.6951  0.7852
s.e.  0.0391   0.1266  0.0977

sigma^2 = 391955:  log likelihood = -369.02
AIC=746.05   AICc=747   BIC=753.45
```

1 integração

Arquivo: Carteira_Clientes (.txt)

Uma empresa de serviços financeiros deseja projetar o volume de sua **carteira de clientes** para o mês seguinte, com base no histórico de 48 meses anteriores (mai/19 a abr/23).



- Qual o melhor modelo para esta série?

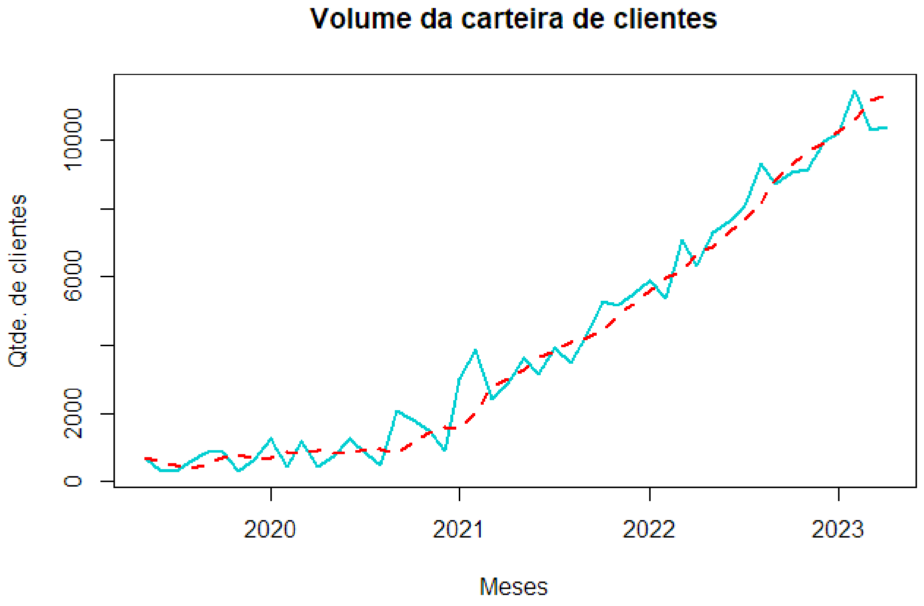
```
> modelo <- auto.arima(base_clientes$QTDE_CLIENTES)
> print(modelo)
Series: base_clientes$QTDE_CLIENTES
ARIMA(1,1,2)

Coefficients:
         ar1      ma1     ma2
      0.9736  -1.6951  0.7852
s.e.  0.0      0.1      0.0

sigma^2 = 391955:  log likelihood = -369.02
AIC=746.05   AICc=747   BIC=753.45
```

ϕ_1 θ_1 θ_2

Estimativas do modelo ARIMA, com *p = 1 , d = 1* e *q = 2*

Arquivo: Carteira_Clientes (.txt)

Uma empresa de serviços financeiros deseja projetar o volume de sua **carteira de clientes** para o mês seguinte, com base no histórico de 48 meses anteriores (mai/19 a abr/23).



Aplicando o modelo na base de dados, podemos construir um **gráfico comparativo** dos valores reais observados *versus* valores preditos pelo modelo (ao lado).

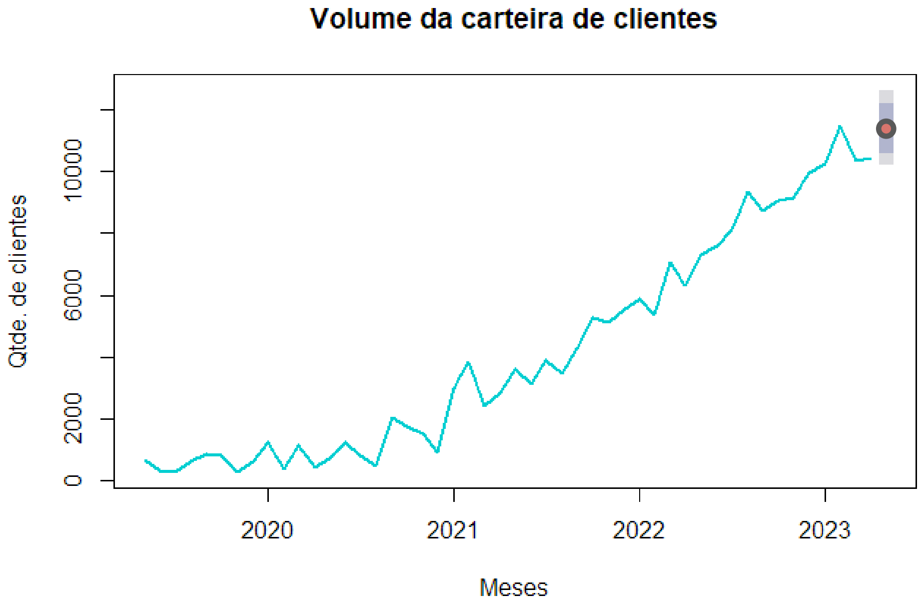
Neste *case*, **REQM = 599,4**.

Ou seja, os valores preditos pelo modelo desviam-se dos valores reais em cerca de **599 clientes**, em média.

Arquivo: Carteira_Clientes (.txt)

Case: Carteira de Clientes

Uma empresa de serviços financeiros deseja projetar o volume de sua **carteira de clientes** para o mês seguinte, com base no histórico de 48 meses anteriores (mai/19 a abr/23).



Projetando a fórmula do modelo estimado para o próximo mês, obtemos a seguinte **previsão** (destacada como um ponto vermelho no gráfico ao lado):

✓ mai/23: **11.403 clientes**

Ou seja, estimamos que haverá **aumento** no volume de clientes no próximo mês, um pouco abaixo do recorde histórico de 11.494 clientes (observado no mês de fev/23).

Arquivo: Carteira_Clientes (.txt)

O **modelo autorregressivo de médias móveis integrado sazonal** (SARIMA) consiste em um modelo ARMA aplicado após um tratamento da série original, realizado para remover os componentes de **sazonalidade** e **tendência**. O tratamento adicional em relação ao ARIMA é chamado ***integração sazonal*** (ou *diferenciação sazonal*), cuja inicial motiva o acréscimo da letra **S** à sigla.

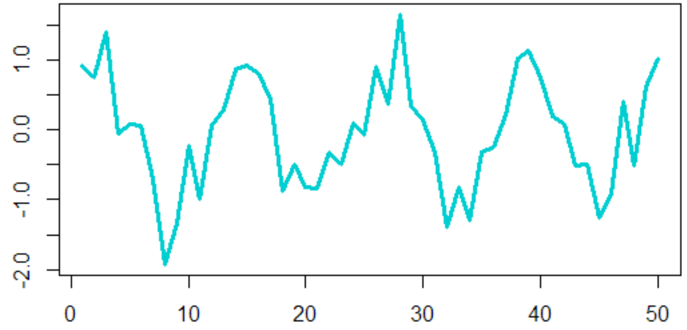
A integração sazonal consiste em definir uma nova série Z_t cujos valores são as **diferenças** entre valores da série original Y_t defasados em ***m*** instantes de tempo; em geral, *m* = 7 dias, *m* = 12 meses etc. Ou seja:

$$Z_t = Y_t - Y_{t-m}$$

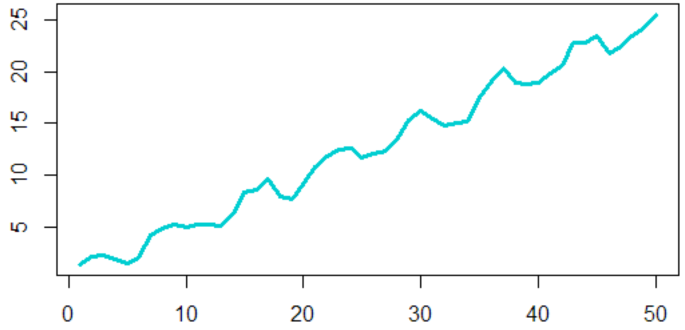
Em muitos casos, realizar a integração simples e a integração sazonal uma única vez (cada) já é suficiente para **remover sazonalidade e tendência** e tornar Z_t uma série **estacionária**, sobre a qual o modelo ARMA poderá ser aplicado. Porém, de forma geral, diremos que podem ser necessárias ***d*** integrações simples e ***D*** integrações sazonais.

Exemplos:

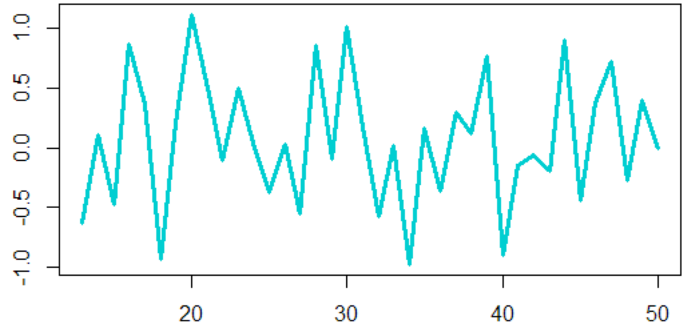
Exemplo 1: Série com sazonalidade



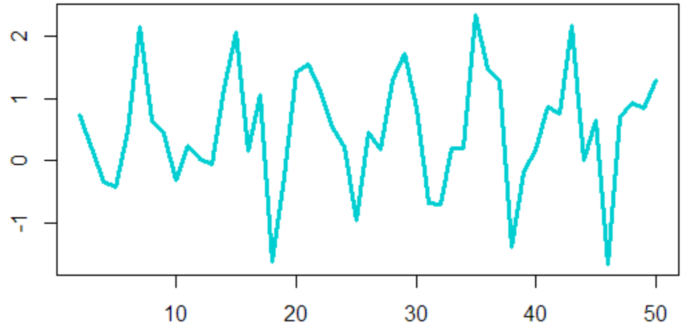
Exemplo 2: Série com tendência e sazonalidade



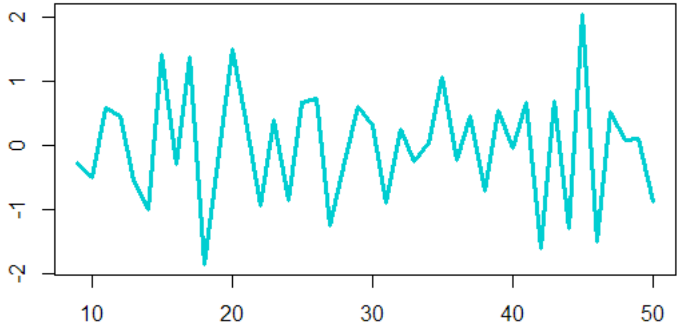
Exemplo 1: Série dessazonalizada (m = 12)



Exemplo 2: Série diferenciada



Exemplo 2: Série dessazonalizada (m = 7)



A equação do **modelo autorregressivo de médias móveis integrado sazonal** (SARIMA) é mais complexa que o ARIMA, devido ao fato de envolver integrações simples e sazonais de forma cumulativa. A série modelada será:

$$Z_t = (1 - B_m)^D (1 - B)^d Y_t$$

onde $(1 - B)$ é um símbolo que corresponde ao tratamento de **integração**, realizado ***d*** vezes cumulativas sobre a série original Y_t até remover a tendência, e $(1 - B_m)$ corresponde ao tratamento de **integração sazonal**, realizado ***D*** vezes cumulativas sobre a série diferenciada até remover a sazonalidade, com periodicidade *m*.

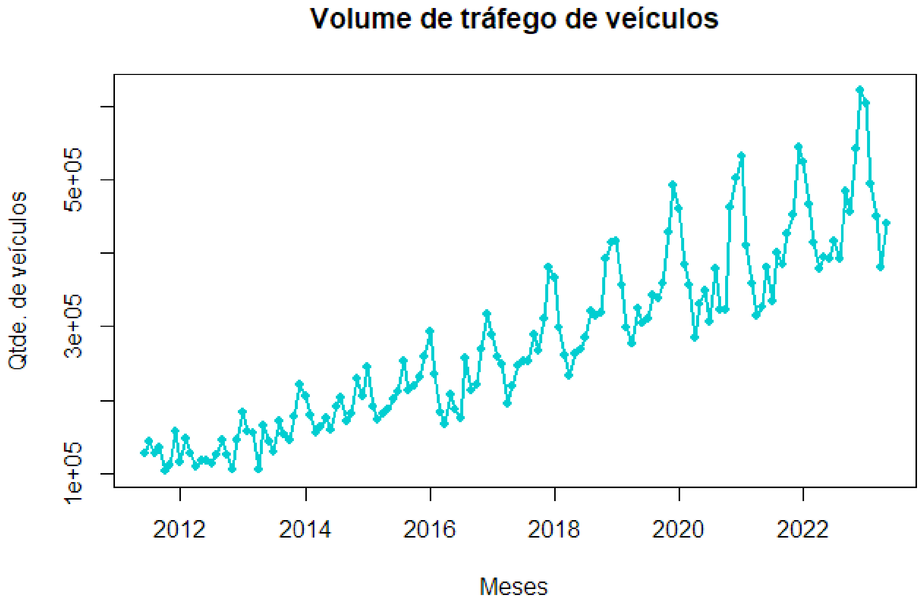
As **ordens do modelo** são denotadas por ***p***, ***d*** e ***q***, como no *ARIMA*, e ***P***, ***D*** *e* ***Q***, que equivalem aos respectivos componentes sazonais de autorregressão, integração e médias móveis.

Exemplos:

- Modelo *SARIMA* *(p = 1, d = 1, q = 2) x (P = 1, D = 1, Q = 3)*^{12} ou *SARIMA (1,1,2) x (1,1,3)*^{12}:

  é semelhante ao modelo *ARIMA* *(1,1,2)*, com acréscimo do termo *(1,1,3)*^{12}, que indica que a série original Y_t recebeu uma integração sazonal *(D = 1)*, depende do valor de Y_{t-12} *(P = 1)* e depende dos erros de estimação de Y_{t-12}, Y_{t-24} e Y_{t-36} *(Q = 3)*.

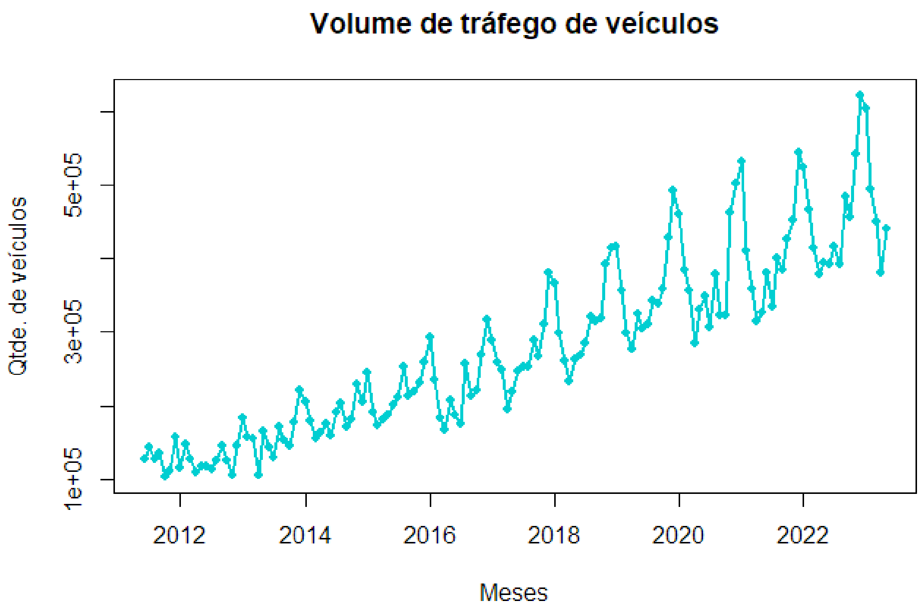
A concessionária de uma rodovia precisa estimar a quantidade mensal de **veículos** que trafegarão em determinado trecho no ano seguinte, com base em um histórico de 12 anos (jun/11 a mai/23).



- Qual o melhor modelo para esta série?

Arquivo: Veiculos (.txt)

A concessionária de uma rodovia precisa estimar a quantidade mensal de **veículos** que trafegarão em determinado trecho no ano seguinte, com base em um histórico de 12 anos (jun/11 a mai/23).



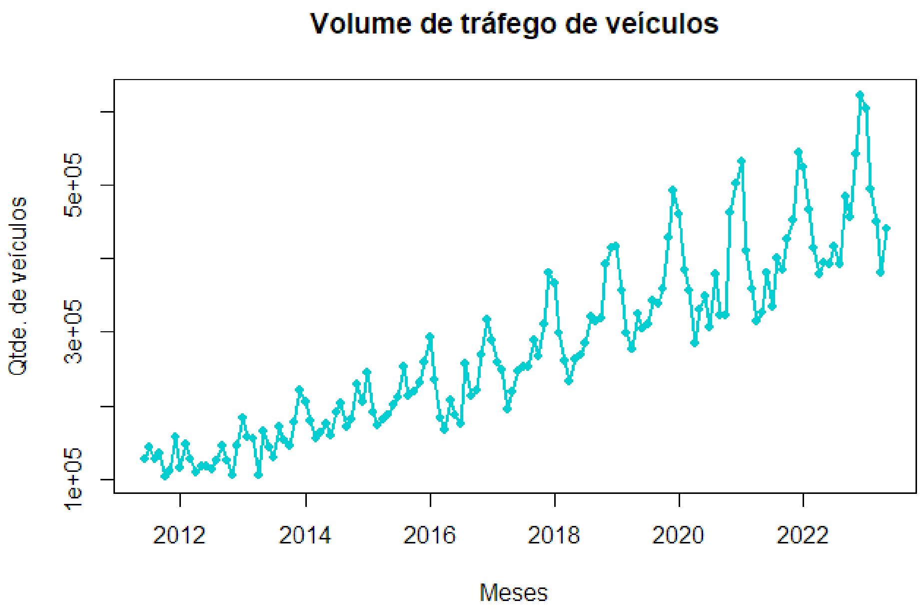
- Qual o melhor modelo para esta série?

Novamente, a função ***auto.arima*** do pacote ***forecast*** do R nos ajuda a identificar o melhor modelo para uma série não estacionária sazonal (SARIMA), bem como a sua ordem.

A identificação do melhor modelo é feita por meio de testes baseados no padrão de **correlação** entre os valores históricos, a depender da distância *(lag)* entre eles.

Arquivo: Veiculos (.txt)

A concessionária de uma rodovia precisa estimar a quantidade mensal de **veículos** que trafegarão em determinado trecho no ano seguinte, com base em um histórico de 12 anos (jun/11 a mai/23).

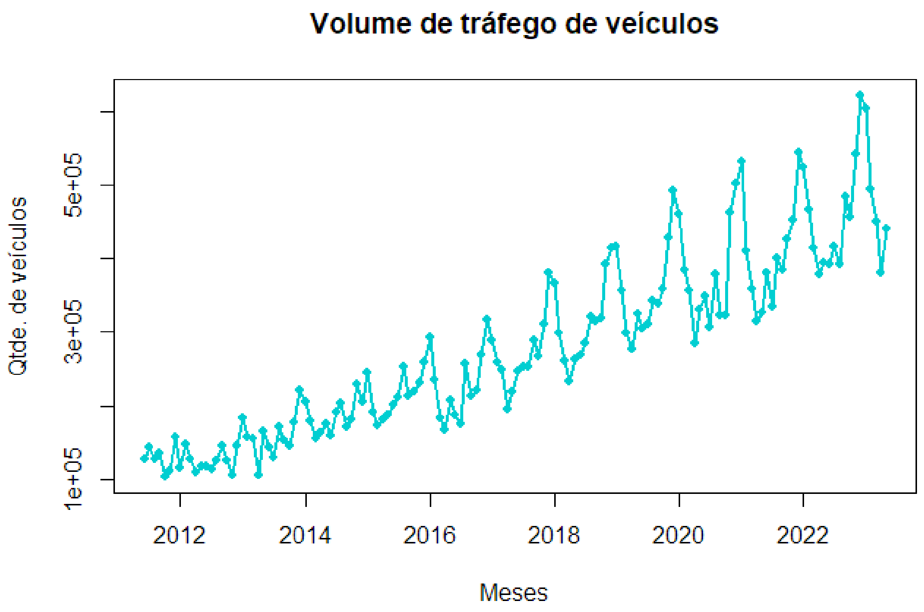


- Qual o melhor modelo para esta série?

```
> modelo <- auto.arima(serie,
+                      seasonal = TRUE)
> print(modelo)
Series: serie
ARIMA(0,1,1)(0,1,1)[12]

Coefficients:
          ma1     sma1
      -0.8203  -0.4183
s.e.   0.0641   0.0751

sigma^2 = 6.79e+08:  log likelihood = -1518.64
AIC=3043.29   AICc=3043.48   BIC=3051.91
```

Arquivo: Veiculos (.txt)

A concessionária de uma rodovia precisa estimar a quantidade mensal de **veículos** que trafegarão em determinado trecho no ano seguinte, com base em um histórico de 12 anos (jun/11 a mai/23).



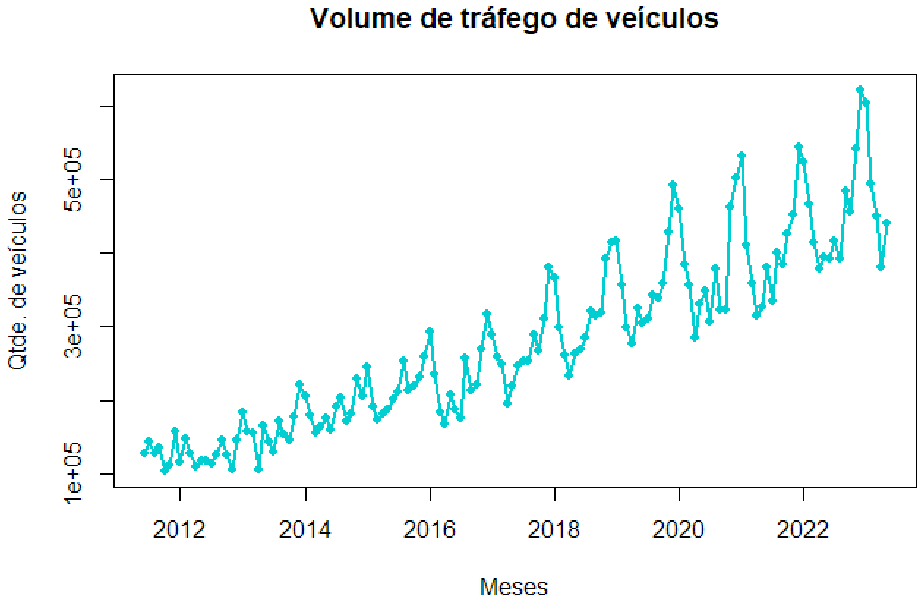
- Qual o melhor modelo para esta série?

```
> modelo <- auto.arima(serie,
+                      seasonal = TRUE)
> print(modelo)
Series: serie
ARIMA(0,1,1)(0,1,1)[12]

Coefficients:
          ma1     sma1
      -0.8203  -0.4183
s.e.   0.0641   0.0751

sigma^2 = 6.79e+08:  log likelihood = -1518.64
AIC=3043.29   AICc=3043.48   BIC=3051.91
```

1 termo MA

Arquivo: Veiculos (.txt)

A concessionária de uma rodovia precisa estimar a quantidade mensal de **veículos** que trafegarão em determinado trecho no ano seguinte, com base em um histórico de 12 anos (jun/11 a mai/23).



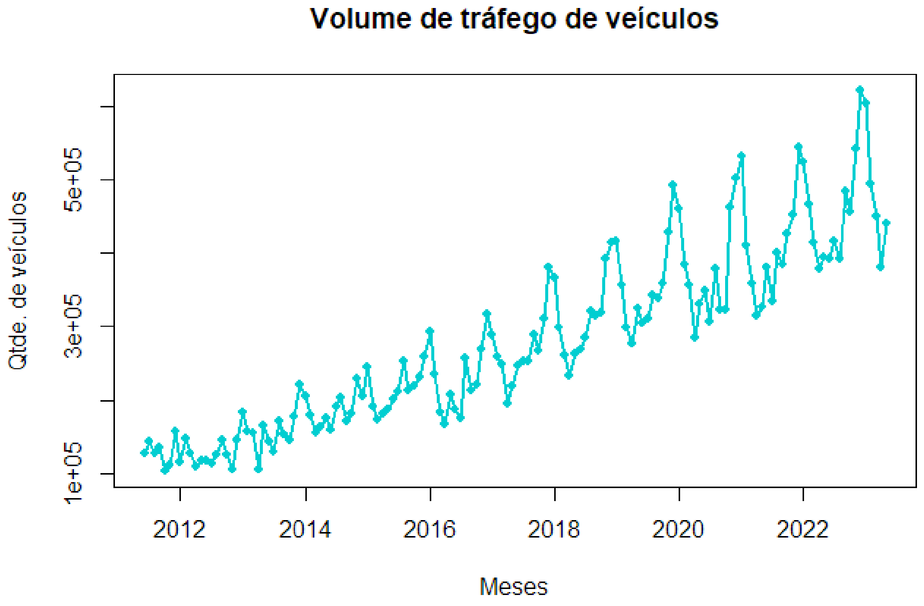
➢ Qual o melhor modelo para esta série?

```
> modelo <- auto.arima(serie,
+                      seasonal = TRUE)
> print(modelo)
Series: serie
ARIMA(0,1,1)(0,1,1)[12]

Coefficients:
          ma1     sma1
      -0.8203  -0.4183
s.e.   0.0641   0.0751

sigma^2 = 6.79e+08:  log likelihood = -1518.64
AIC=3043.29   AICc=3043.48   BIC=3051.91
```

1 termo MA sazonal

Arquivo: Veiculos (.txt)

A concessionária de uma rodovia precisa estimar a quantidade mensal de **veículos** que trafegarão em determinado trecho no ano seguinte, com base em um histórico de 12 anos (jun/11 a mai/23).



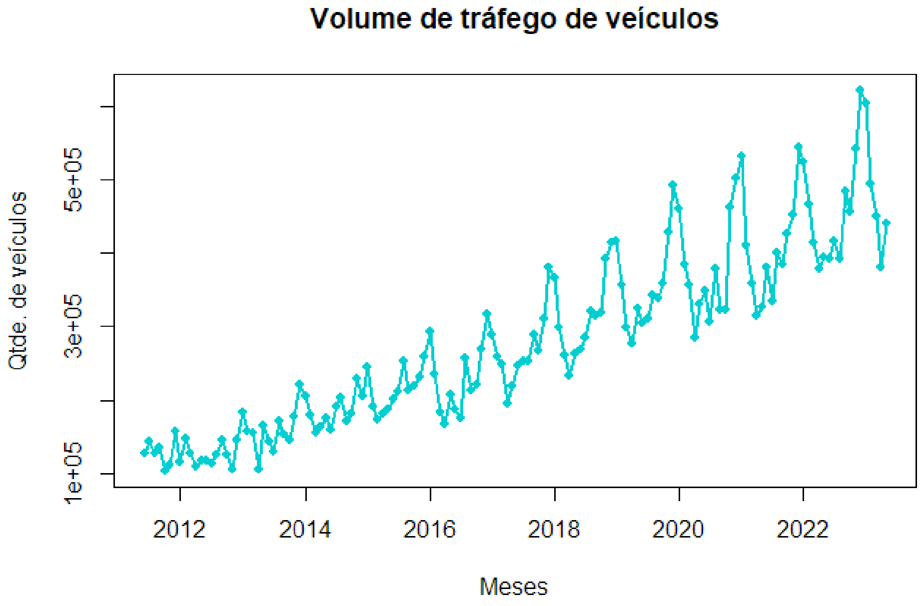
- Qual o melhor modelo para esta série?

```
> modelo <- auto.arima(serie,
+                      seasonal = TRUE)
> print(modelo)
Series: serie
ARIMA(0,1,1)(0,1,1)[12]

Coefficients:
          ma1     sma1
      -0.8203  -0.4183
s.e.   0.0641   0.0751

sigma^2 = 6.79e+08:  log likelihood = -1518.64
AIC=3043.29   AICc=3043.48   BIC=3051.91
```

1 integração

Arquivo: Veiculos (.txt)

A concessionária de uma rodovia precisa estimar a quantidade mensal de **veículos** que trafegarão em determinado trecho no ano seguinte, com base em um histórico de 12 anos (jun/11 a mai/23).



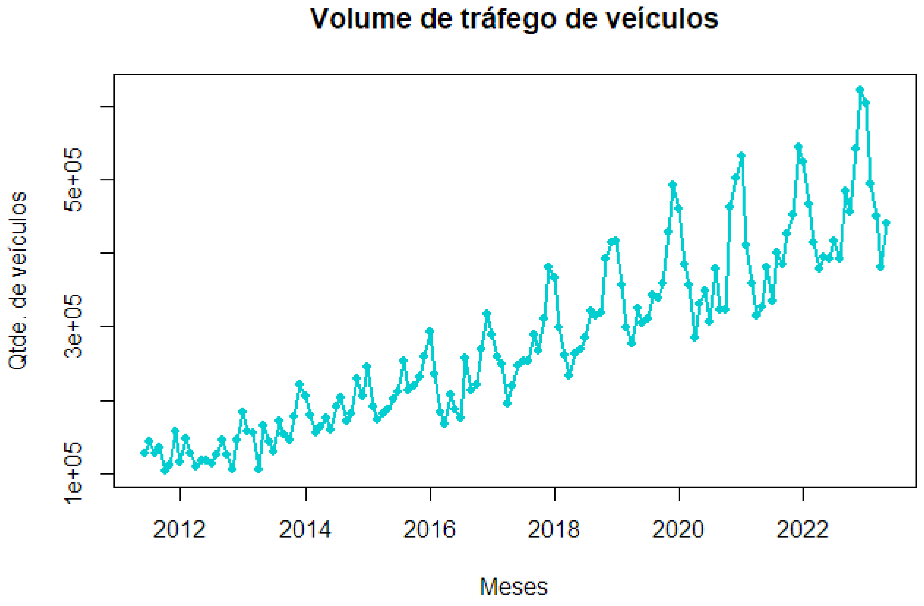
- Qual o melhor modelo para esta série?

```
> modelo <- auto.arima(serie,
+                      seasonal = TRUE)
> print(modelo)
Series: serie
ARIMA(0,1,1)(0,1,1)[12]

Coefficients:
          ma1     sma1
      -0.8203  -0.4183
s.e.   0.0641   0.0751

sigma^2 = 6.79e+08:  log likelihood = -1518.64
AIC=3043.29   AICc=3043.48   BIC=3051.91
```

1 integração sazonal, de período 12

Arquivo: Veiculos (.txt)

A concessionária de uma rodovia precisa estimar a quantidade mensal de **veículos** que trafegarão em determinado trecho no ano seguinte, com base em um histórico de 12 anos (jun/11 a mai/23).



- Qual o melhor modelo para esta série?

```
> modelo <- auto.arima(serie,
+                      seasonal = TRUE)
> print(modelo)
Series: serie
ARIMA(0,1,1)(0,1,1)[12]

Coefficients:
          ma1        sma1
      -0.   3    -(
s.e.   0.    1    0.
```

θ_1 θ_1 *saz.*

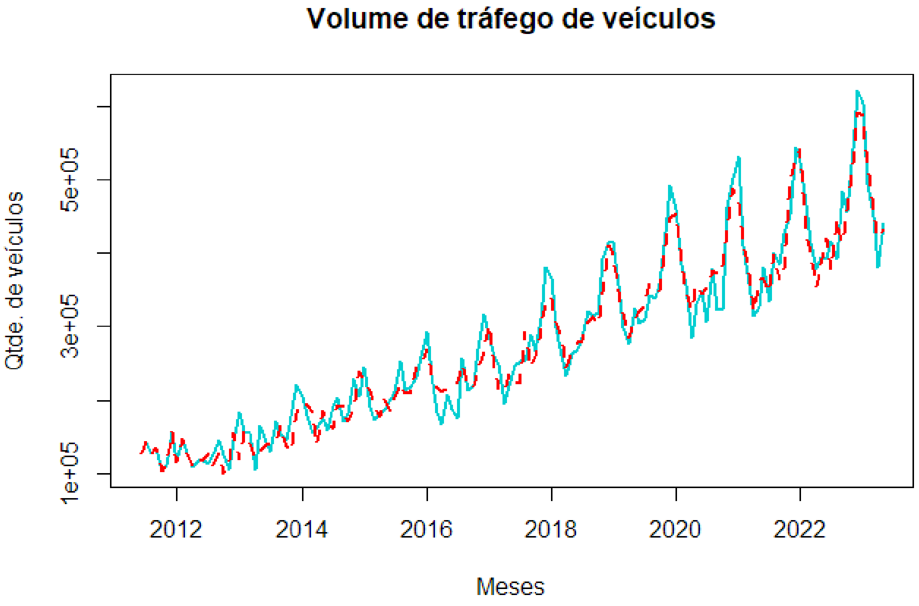
```
sigma^2 = 6.79e+08:  log likelihood = -1518.64
AIC=3043.29   AICc=3043.48   BIC=3051.91
```

Estimativas do modelo SARIMA, com *p = 0, d = 1, q = 1, P = 0, D = 1, Q = 1*

Arquivo: Veiculos (.txt)

FIA BUSINESS SCHOOL lab.data

A concessionária de uma rodovia precisa estimar a quantidade mensal de **veículos** que trafegarão em determinado trecho no ano seguinte, com base em um histórico de 12 anos (jun/11 a mai/23).



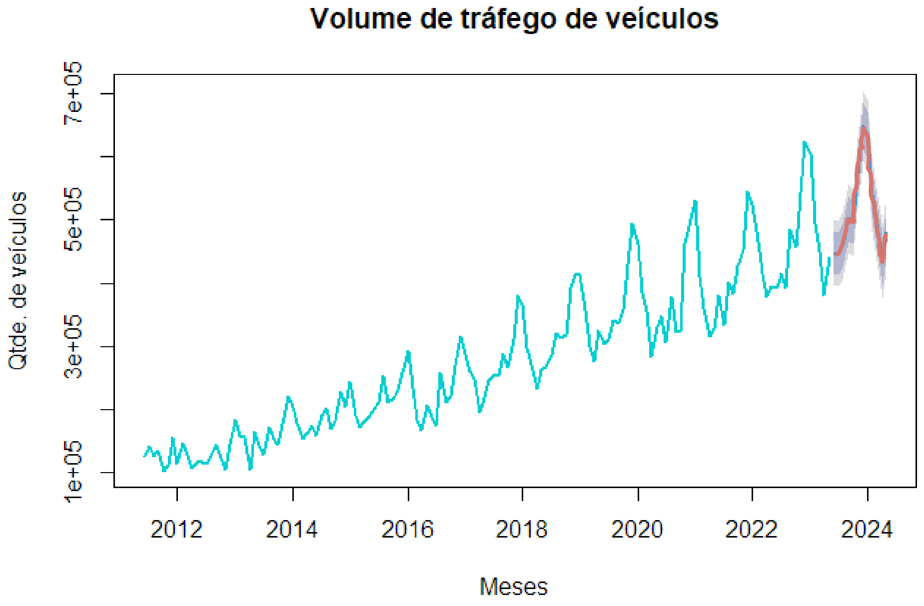
Aplicando o modelo na base de dados, podemos construir um **gráfico comparativo** dos valores reais observados *versus* valores preditos pelo modelo (ao lado).

Neste *case*, **REQM = 24.664**.

Ou seja, os valores preditos pelo modelo desviam-se dos valores reais em cerca de **24.664 veículos**, em média.

Arquivo: Veiculos (.txt)

A concessionária de uma rodovia precisa estimar a quantidade mensal de **veículos** que trafegarão em determinado trecho no ano seguinte, com base em um histórico de 12 anos (jun/11 a mai/23).



Projetando a fórmula do modelo estimado para os próximos 12 meses, obtemos as seguintes **previsões** (destacadas em vermelho no gráfico ao lado):

- ✓ jun/23: **448.165** veículos
- ✓ jul/23: **446.676** veículos
- ✓ ago/23: **458.870** veículos
- ✓ set/23: **502.402** veículos
- ✓ out/23: **496.468** veículos
- ✓ nov/23: **572.177** veículos
- ✓ dez/23: **649.553** veículos
- ✓ jan/24: **634.999** veículos
- ✓ fev/24: **540.303** veículos
- ✓ mar/24: **494.211** veículos
- ✓ abr/24: **435.552** veículos
- ✓ mai/24: **479.388** veículos

Ou seja, estimamos que os padrões de tendência e sazonalidade se manterão nos próximos 12 meses.

Arquivo: Veiculos (.txt)

Case: Inadimplência

Um banco deseja prever o % de clientes **inadimplentes** no cartão de crédito (ou seja, que atrasam o pagamento de sua fatura) para os 12 meses seguintes, com base em um histórico de 5 anos.

a) Analise o gráfico da série. Ela parece estacionária, ou possui tendências/sazonalidade?
b) Realize o teste de Dickey-Fuller generalizado e afirme formalmente se a série é estacionária ou não.
c) Ajuste o melhor modelo (S)ARIMA para a série. Qual foi o modelo identificado? Escreva a sua equação.
d) Construa um gráfico único comparando a evolução da série com os valores preditos pelo modelo. Analisando visualmente, o modelo parece adequado?
e) Avalie a qualidade do modelo a partir do índice REQM.
f) Projete o percentual de clientes inadimplentes para os próximos 12 meses.

Arquivo: Carteira_Clientes (.txt)

6. Modelo com Variáveis Exógenas

Como Predizer a Série por Meio de Outras Variáveis?

- Note que, nos modelos das classes (S)ARIMA, uma variável era projetada apenas a partir de dados históricos dela mesma. Porém, é comum que se tenha interesse em modelar uma série temporal a partir de dados de **outras variáveis explicativas**. Estas são chamadas, geralmente, de variáveis **exógenas**, ou **externas**.
- Uma das alternativas para isso é o modelo de **regressão linear temporal**. Diferentemente da regressão linear tradicional, na qual as observações e erros eram **independentes**, agora, temos que considerar que as observações possuem um grau potencial de **dependência** entre si.

O modelo de **regressão linear temporal** para uma resposta quantitativa indexada pelo tempo Y_t, a partir de k variáveis explicativas também indexadas pelo tempo $X_{1,t}, \ldots, X_{k,t}$, pode ser escrito como:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \cdots + \beta_k X_{k,t} + \eta_t$$

A diferença é que, agora, os resíduos η_t **não são independentes**, e sim, seguem um processo ***ARIMA*** (p, d, q). Ou seja:

$$(1 - B)^d \eta_t = \phi_1 (1 - B)^d \eta_{t-1} + \cdots + \phi_p (1 - B)^d \eta_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

Exemplos:

- Modelo com 1 explicativa e erros ARIMA (1,0,1): $Y_t = \beta_0 + \beta_1 X_{1,t} + \eta_t$, onde $\eta_t = \phi_1 \eta_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1}$
- Modelo com 2 explicativas e erros ARIMA (1,1,0): $Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \eta_t$, onde $(1 - B)\eta_t = \phi_1 (1 - B)\eta_{t-1} + \epsilon_t$

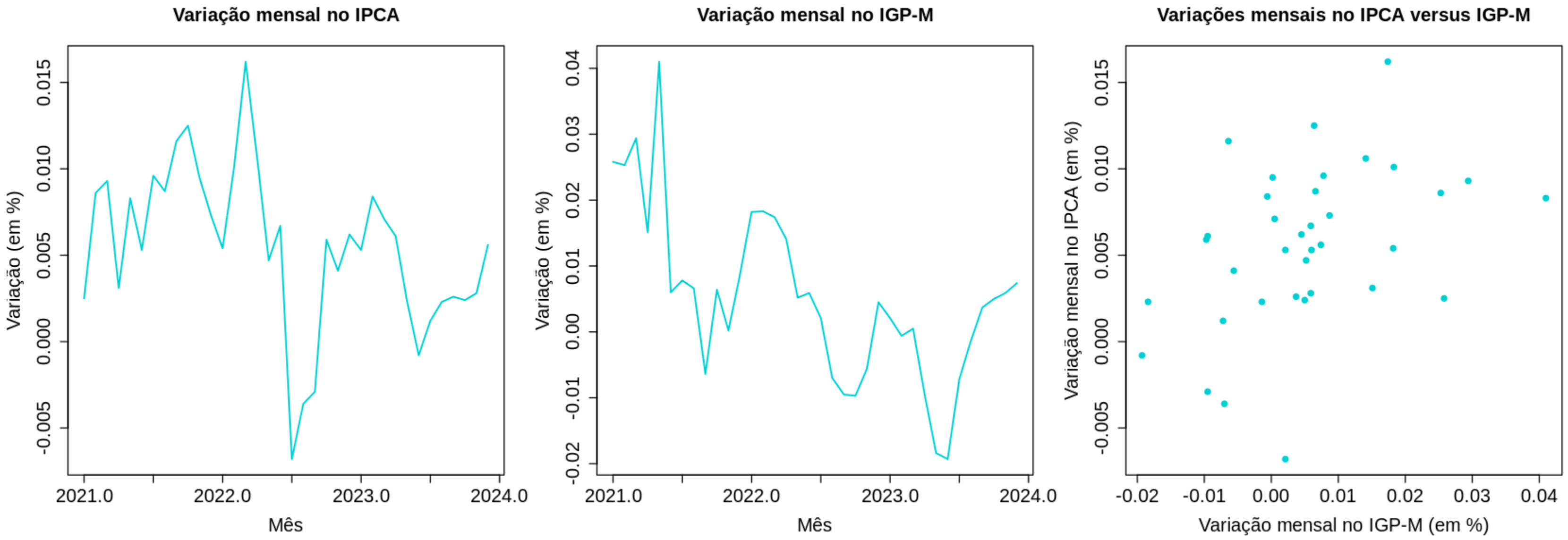
Caso haja comportamento sazonal, pode-se generalizar os resíduos para um processo ***SARIMA*** $(p, d, q) \times (P, D, Q)^m$.

* Letra grega β: lê-se como '*béta*'
* Letra grega η: lê-se como '*êta*'
* Letra grega ϕ: lê-se como '*fi*'
* Letra grega θ: lê-se como '*téta*'
* Letra grega ϵ: lê-se como '*épsilon*'

Case: Inflação

Uma consultoria financeira deseja projetar o valor do Índice de Preços ao Consumidor Amplo (IPCA), principal índice para cálculo da inflação oficial do país, em função do Índice Geral de Preços de Mercado (IGP-M), também utilizado como indexador de inflação, especialmente nos setores imobiliário e energético.

Vamos utilizar dados históricos das variações mensais (em %) dos índices IPCA e IGP-M ao longo de 36 meses (jan/21 a dez/23), a fim de projetar o mês seguinte (jan/24).



Arquivo: Inflacao (.txt)

Uma consultoria financeira deseja projetar o valor do Índice de Preços ao Consumidor Amplo (IPCA), principal índice para cálculo da inflação oficial do país, em função do Índice Geral de Preços de Mercado (IGP-M), também utilizado como indexador de inflação, especialmente nos setores imobiliário e energético.

Vamos utilizar dados históricos das variações mensais (em %) dos índices IPCA e IGP-M ao longo de 36 meses (jan/21 a dez/23), a fim de projetar o mês seguinte (jan/24).

```
Call:
lm(formula = serie_ipca ~ serie_igpm)
```

Modelo de regressão linear simples

```
Coefficients:
              Estimate Std. Error t value    Pr(>|t|)
(Intercept) 0.0047093  0.0007738   6.086 0.000000666 ***
serie_igpm  0.1524677  0.0559545   2.725      0.0101 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.004299 on 34 degrees of freedom
Multiple R-squared:  0.1792,    Adjusted R-squared:  0.1551
```

Arquivo: Inflacao (.txt)

Uma consultoria financeira deseja projetar o valor do Índice de Preços ao Consumidor Amplo (IPCA), principal índice para cálculo da inflação oficial do país, em função do Índice Geral de Preços de Mercado (IGP-M), também utilizado como indexador de inflação, especialmente nos setores imobiliário e energético.

Vamos utilizar dados históricos das variações mensais (em %) dos índices IPCA e IGP-M ao longo de 36 meses (jan/21 a dez/23), a fim de projetar o mês seguinte (jan/24).

```
Call:
lm(formula = serie_ipca ~ serie_igpm)

Coefficients:
            Estimate Std. Error t value    Pr(>|t|)
(Intercept) 0.0047093  0.0007738   6.086 0.000000666 ***
serie_igpm  0.1524677  0.0559545   2.725      0.0101 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.004299 on 34 degrees of freedom
Multiple R-squared:  0.1792     Adjusted R-squared:  0.1551
```

- Ajustando um modelo de **regressão linear simples** para avaliar a associação entre a variação de IPCA e a variação de IGP-M, vemos que os parâmetros β_0 e β_1 são estatisticamente **diferentes de zero**, com alta confiança.
- Logo, apesar do **baixo R^2** (0,18), é possível afirmar que existe associação linear estatisticamente significativa entre as variações mensal nos índices IPCA e IGP-M.

Arquivo: Inflacao (.txt)

Uma consultoria financeira deseja projetar o valor do Índice de Preços ao Consumidor Amplo (IPCA), principal índice para cálculo da inflação oficial do país, em função do Índice Geral de Preços de Mercado (IGP-M), também utilizado como indexador de inflação, especialmente nos setores imobiliário e energético.

Vamos utilizar dados históricos das variações mensais (em %) dos índices IPCA e IGP-M ao longo de 36 meses (jan/21 a dez/23), a fim de projetar o mês seguinte (jan/24).

```
Call:
lm(formula = serie_ipca ~ serie_igpm)

Coefficients:
             Estimate Std. Error t value    Pr(>|t|)
(Intercept) 0.0047093  0.0007738   6.086 0.000000666 ***
serie_igpm  0.1524677  0.0559545   2.725      0.0101 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.004299 on 34 degrees of freedom
Multiple R-squared:  0.1792     Adjusted R-squared:  0.1551
```

- Entretanto, ao ajustar este modelo, estamos **ignorando** a estrutura de **correlação temporal** existente entre as observações.

Arquivo: Inflacao (.txt)

Uma consultoria financeira deseja projetar o valor do Índice de Preços ao Consumidor Amplo (IPCA), principal índice para cálculo da inflação oficial do país, em função do Índice Geral de Preços de Mercado (IGP-M), também utilizado como indexador de inflação, especialmente nos setores imobiliário e energético.

Vamos utilizar dados históricos das variações mensais (em %) dos índices IPCA e IGP-M ao longo de 36 meses (jan/21 a dez/23), a fim de projetar o mês seguinte (jan/24).

```
Series: serie_ipca
Regression with ARIMA(1,0,0) errors
```

Modelo de regressão temporal

```
Coefficients:
         ar1  intercept    xreg
      0.5507     0.0046  0.1347
s.e.  0.1393     0.0013  0.0623

sigma^2 = 0.0000133:  log likelihood = 152.41
AIC=-296.81   AICc=-295.52   BIC=-290.48
```

Arquivo: Inflacao (.txt)

Uma consultoria financeira deseja projetar o valor do Índice de Preços ao Consumidor Amplo (IPCA), principal índice para cálculo da inflação oficial do país, em função do Índice Geral de Preços de Mercado (IGP-M), também utilizado como indexador de inflação, especialmente nos setores imobiliário e energético.

Vamos utilizar dados históricos das variações mensais (em %) dos índices IPCA e IGP-M ao longo de 36 meses (jan/21 a dez/23), a fim de projetar o mês seguinte (jan/24).

```
Series: serie_ipca
Regression with ARIMA(1,0,0) errors

Coefficients:
         ar1  intercept    xreg
      0.5507     0.0046  0.1347
s.e.  0.1393     0.0013  0.0623

sigma^2 = 0.0000133:  log likelihood = 152.41
AIC=-296.81   AICc=-295.52   BIC=-290.48
```

- Agora, ajustamos uma **regressão linear temporal** com a função *auto.arima* do R.
- O termo **xreg** corresponde à variável explicativa (IGP-M).
- Os resíduos do modelo de regressão linear seguem um modelo temporal **AR (1)**.

Arquivo: Inflacao (.txt)

Uma consultoria financeira deseja projetar o valor do Índice de Preços ao Consumidor Amplo (IPCA), principal índice para cálculo da inflação oficial do país, em função do Índice Geral de Preços de Mercado (IGP-M), também utilizado como indexador de inflação, especialmente nos setores imobiliário e energético.

Vamos utilizar dados históricos das variações mensais (em %) dos índices IPCA e IGP-M ao longo de 36 meses (jan/21 a dez/23), a fim de projetar o mês seguinte (jan/24).

```
Series: serie_ipca
Regression with ARIMA(1,0,0) errors

Coefficients:
         ar1  intercept    xreg
      0.5507     0.0046  0.1347
s.e.  0.1393     0.0013  0.0623

sigma^2 = 0.0000133:  log likelihood = 152.41
AIC=-296.81   AICc=-295.52   BIC=-290.48
```

- **Observação**: A função *auto.arima* não fornece o *p*-valor do teste de nulidade do parâmetro β_1, associado à variável explicativa *(xreg).* Para avaliar se ela é significativa no modelo, temos que calcular o intervalo de confiança de 95%:

$$IC(\beta_1; 95\%) = \left[\hat{\beta}_1 - 2 \cdot erro\left(\hat{\beta}_1\right); \hat{\beta}_1 + 2 \cdot erro\left(\hat{\beta}_1\right)\right]$$

Se o intervalo **não incluir o valor zero**, a variável deve permanecer no modelo.

Arquivo: Inflacao (.txt)

Uma consultoria financeira deseja projetar o valor do Índice de Preços ao Consumidor Amplo (IPCA), principal índice para cálculo da inflação oficial do país, em função do Índice Geral de Preços de Mercado (IGP-M), também utilizado como indexador de inflação, especialmente nos setores imobiliário e energético.

Vamos utilizar dados históricos das variações mensais (em %) dos índices IPCA e IGP-M ao longo de 36 meses (jan/21 a dez/23), a fim de projetar o mês seguinte (jan/24).

```
Series: serie_ipca
Regression with ARIMA(1,0,0) errors

Coefficients:
         ar1  intercept    xreg
      0.5507     0.0046  0.1347
s.e.  0.1393     0.0013  0.0623

sigma^2 = 0.0000133:  log likelihood = 152.41
AIC=-296.81   AICc=-295.52   BIC=-290.48
```

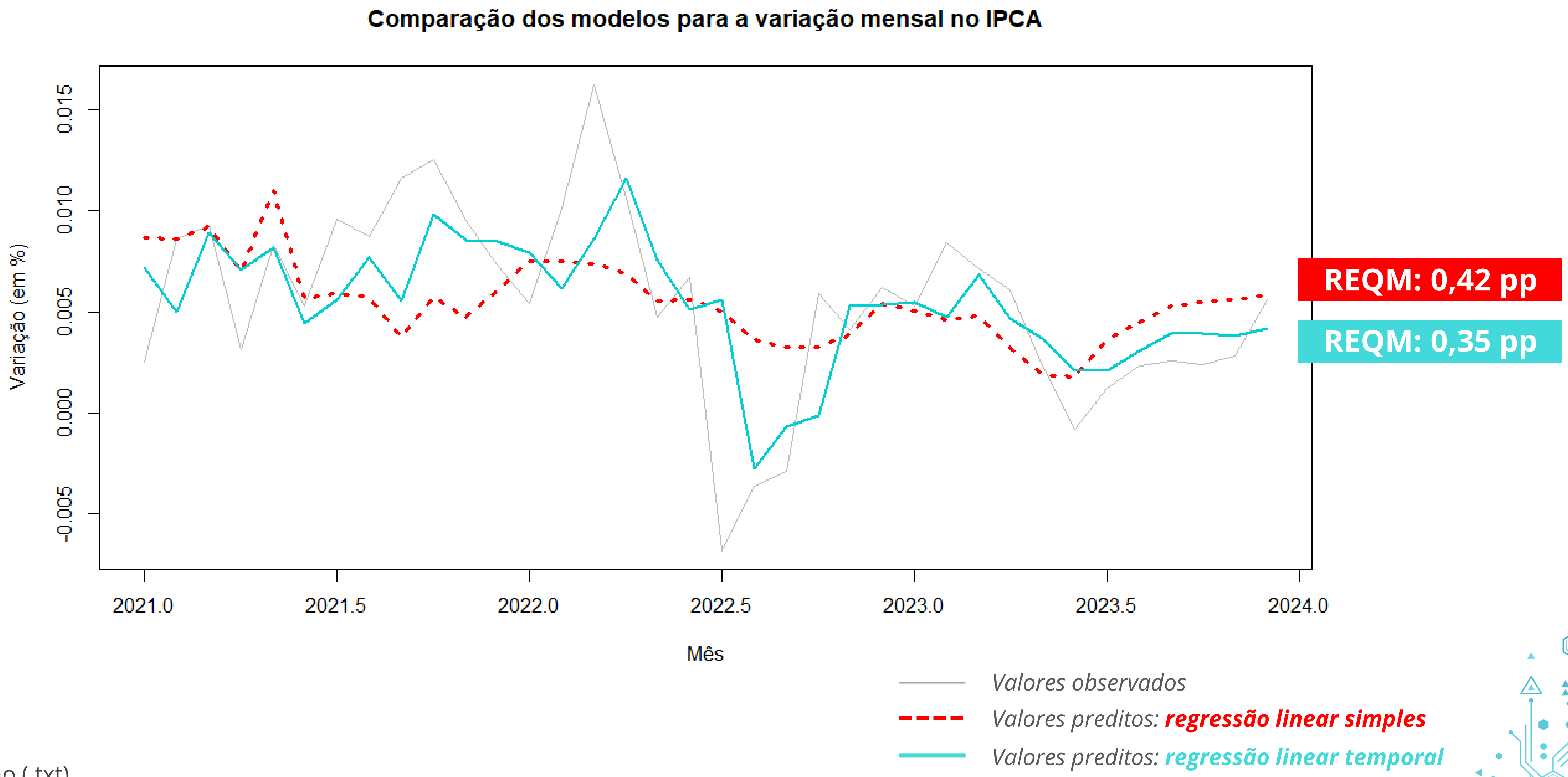
Arquivo: Inflacao (.txt)

$$IC(\beta_1; 95\%) = [\hat{\beta}_1 - 2 \cdot erro(\hat{\beta}_1); \hat{\beta}_1 + 2 \cdot erro(\hat{\beta}_1)]$$

$$IC(\beta_1; 95\%) = [0{,}1347 - 2 \cdot 0{,}0623; 0{,}1347 + 2 \cdot 0{,}0623]$$

$$= [0{,}0101; 0{,}2593]$$

- A amostra nos fornece 95% de confiança de que o parâmetro populacional β_1 está entre **0,0101** e **0,2593**. Logo, assumimos que ele é diferente de zero, e que a variação mensal no IPCA está linearmente associada à variação mensal no IGP-M.



Arquivo: Inflacao (.txt)

Uma consultoria financeira deseja projetar o valor do Índice de Preços ao Consumidor Amplo (IPCA), principal índice para cálculo da inflação oficial do país, em função do Índice Geral de Preços de Mercado (IGP-M), também utilizado como indexador de inflação, especialmente nos setores imobiliário e energético.

Vamos utilizar dados históricos das variações mensais (em %) dos índices IPCA e IGP-M ao longo de 36 meses (jan/21 a dez/23), a fim de projetar o mês seguinte (jan/24).

O **REQM** da regressão linear foi reduzido em cerca de **17%** (de 0,42 pp para 0,35 pp) a partir da incorporação de erros correlacionados, segundo o modelo AR(1).

Logo, o modelo de **regressão linear temporal** é mais adequado do que a **regressão linear simples** para este *case*, conforme esperado.

Arquivo: Inflacao (.txt)

Case: Inflação

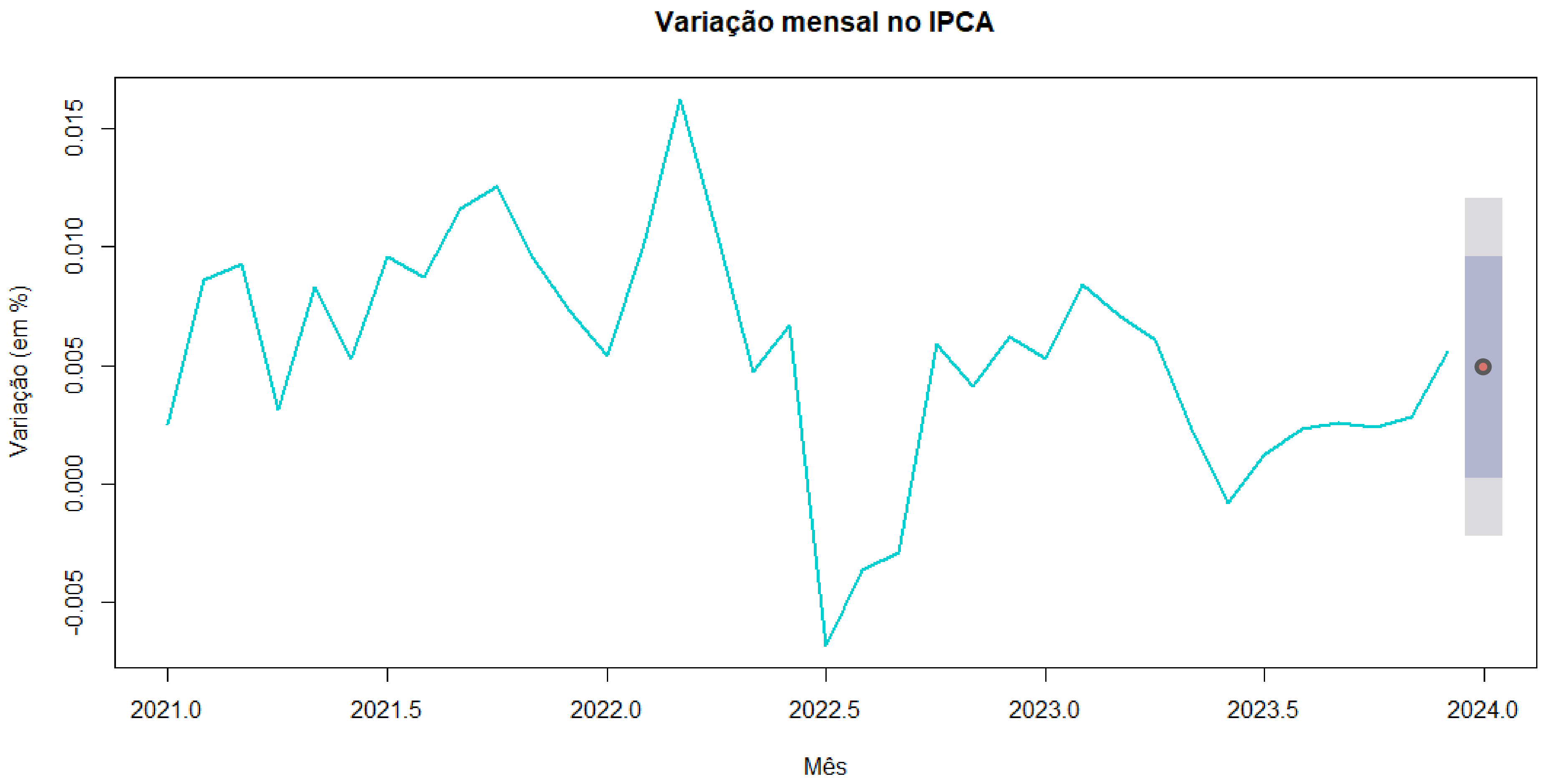
Uma consultoria financeira deseja projetar o valor do Índice de Preços ao Consumidor Amplo (IPCA), principal índice para cálculo da inflação oficial do país, em função do Índice Geral de Preços de Mercado (IGP-M), também utilizado como indexador de inflação, especialmente nos setores imobiliário e energético.

Vamos utilizar dados históricos das variações mensais (em %) dos índices IPCA e IGP-M ao longo de 36 meses (jan/21 a dez/23), a fim de projetar o mês seguinte (jan/24).

Para **previsão** da série da variável resposta (variação de IPCA) para o mês de jan/24, é necessário estipular um valor para a série da variável explicativa (variação de IGP-M) neste mesmo mês.

- O valor médio das variações mensais de IGP-M nos últimos 6 meses de histórico (de jul/23 a dez/23) foi de 0,0022 ,ou seja, 0,22%. Assumindo esse valor como uma estimativa para a variação de IGP-M no mês de jan/24, a previsão de variação no IPCA neste mês é de **0,49%**.

Arquivo: Inflacao (.txt)



Previsão para jan/24: **0,0049**

Arquivo: Inflacao (.txt)

Case: Inflação

Uma consultoria financeira deseja projetar o valor do Índice de Preços ao Consumidor Amplo (IPCA), principal índice para cálculo da inflação oficial do país, em função do Índice Geral de Preços de Mercado (IGP-M), também utilizado como indexador de inflação, especialmente nos setores imobiliário e energético.

Vamos utilizar dados históricos das variações mensais (em %) dos índices IPCA e IGP-M ao longo de 36 meses (jan/21 a dez/23), a fim de projetar o mês seguinte (jan/24).

Para **previsão** da série da variável resposta (variação de IPCA) para o mês de jan/24, é necessário estipular um valor para a série da variável explicativa (variação de IGP-M) neste mesmo mês.

- O valor histórico real de variação do IGP-M em jan/24 foi de 0,07%. Caso esse valor fosse conhecido a priori e utilizado no modelo, a estimativa de variação do **IPCA** neste mês diminuiria de 0,49% para **0,47%**.
- O valor histórico real de variação do IPCA em jan/24 foi de **0,42%**, um pouco abaixo da estimativa.

Arquivo: Inflacao (.txt)

A prefeitura de um município está viabilizando gradativamente para os habitantes a **distribuição de energia** a partir de fontes **renováveis**. Temos uma base com dados quinzenais de um período de 2 anos para as seguintes variáveis:

- **Capacidade instalada** para geração de energia renovável, em MW *(megawatt)*
- **Consumo quinzenal** de energia elétrica renovável, em GWh *(gigawatt/hora)*

Deseja-se avaliar se o consumo de energia renovável está relacionado linearmente com a capacidade instalada, bem como predizer o consumo para o mês seguinte, supondo que a capacidade instalada seja de 200MW na quinzena seguinte, e 230MW na quinzena subsequente.

a) Analise o gráfico de evolução das séries de capacidade instalada e de consumo quinzenal. Elas parecem estacionárias, ou possuem tendências/sazonalidade?
b) Analise o gráfico de dispersão entre as duas séries. Parece existir correlação linear entre capacidade instalada e consumo?
c) Realize o teste de Dickey-Fuller generalizado e afirme formalmente se as séries são estacionárias ou não.
d) Ajuste o melhor modelo ARIMA para o consumo quinzenal, sem considerar a capacidade instalada. Qual foi o modelo identificado? Escreva a sua equação.

Arquivo: Energia (.txt)

A prefeitura de um município está viabilizando gradativamente para os habitantes a **distribuição de energia** a partir de fontes **renováveis**. Temos uma base com dados quinzenais de um período de 2 anos para as seguintes variáveis:

- **Capacidade instalada** para geração de energia renovável, em MW *(megawatt)*
- **Consumo quinzenal** de energia elétrica renovável, em GWh *(gigawatt/hora)*

Deseja-se avaliar se o consumo de energia renovável está relacionado linearmente com a capacidade instalada, bem como predizer o consumo para o mês seguinte, supondo que a capacidade instalada seja de 200MW na quinzena seguinte, e 230MW na quinzena subsequente.

e) Ajuste um modelo de regressão linear simples para o consumo quinzenal, utilizando a capacidade instalada como variável explicativa. Existe associação estatisticamente significativa entre as variáveis, com 95% de confiança? Obs.: mantenha o intercepto no modelo, mesmo que não seja significativo.

f) Ajuste um modelo de regressão linear temporal para o consumo quinzenal, utilizando a capacidade instalada e erros (S)ARIMA. Qual foi o melhor modelo identificado para os erros? Escreva a sua equação.

g) No modelo do item (f), a presença da variável de capacidade instalada foi significativa? Justifique por meio do cálculo do intervalo de confiança associado ao coeficiente que multiplica esta variável.

Arquivo: Energia (.txt)

A prefeitura de um município está viabilizando gradativamente para os habitantes a **distribuição de energia** a partir de fontes **renováveis**. Temos uma base com dados quinzenais de um período de 2 anos para as seguintes variáveis:

- **Capacidade instalada** para geração de energia renovável, em MW *(megawatt)*
- **Consumo quinzenal** de energia elétrica renovável, em GWh *(gigawatt/hora)*

Deseja-se avaliar se o consumo de energia renovável está relacionado linearmente com a capacidade instalada, bem como predizer o consumo para o mês seguinte, supondo que a capacidade instalada seja de 200MW na quinzena seguinte, e 230MW na quinzena subsequente.

h) Construa um gráfico único comparando a evolução da série de consumo quinzenal com os valores preditos dos três modelos construídos nos itens anteriores. Analisando visualmente, qual dos três modelos parece se ajustar melhor aos dados ao longo do tempo?

i) Justifique qual é, de fato, o modelo que melhor se adequa aos dados, a partir dos índices REQM.

j) Utilizando o melhor modelo identificado no item (i), projete o consumo de energia renovável para as próximas 2 quinzenas, utilizando as informações do enunciado. Faça um gráfico da evolução quinzenal do consumo, incluindo dois novos pontos referentes às projeções.

Arquivo: Energia (.txt)

- Morettin, P. A., Toloi, C. M. C. *Análise de Séries Temporais*. 2ª edição. Blucher, 2006.
- Hyndman, R. J., Athanasopoulos, G. *Forecasting: Principles and Practice*. 3ª edição. Otexts, 2021.
- Shumway, R. H., Stoffer, D. S. *Time Series Analysis and Its Apllications*. Springer, 2006.

FIA BUSINESS SCHOOL

lab.data

http://labdata.fia.com.br
Instagram: @labdatafia
Facebook: @LabdataFIA