

Analytics e Inteligência Artificial

Tema da aula
Séries Temporais



BUSINESS SCHOOL

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



CONSULTING

Consultoria personalizada
que oferece soluções
baseada 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**.



Profª Drª Alessandra Montini

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

+1000 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 a estação de metrô (com estacionamento)



Profª Dra.
Alessandra Montini

Diretora do LABDATA-FIA, apaixonada por dados e pela arte de lecionar. Têm 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. Membro do Conselho Curador da FIA, Coordenadora de Grupos de Pesquisa no CNPQ, Parecerista da FAPESP e Colunista de grandes Portais de Tecnologia.



[linkedin.com/in/alessandramontini/](https://www.linkedin.com/in/alessandramontini/)



Prof. Dr.
Adolpho Walter Canton

Diretor do LABDATA-FIA. Consultor em Projetos de *Analytics*, *Big Data* e Inteligência Artificial. Professor FEA – USP. PhD em Estatística Aplicada pela *University of North Carolina at Chapel Hill*, Estados Unidos.

CONTEÚDO PROGRAMÁTICO



ANÁLISE
EXPLORATÓRIA

TÉCNICAS DE
PROJEÇÃO

TÉCNICAS DE
CLASSIFICAÇÃO

TÉCNICAS DE
SEGMENTAÇÃO

TÉCNICAS DE
ANALYTICS

LINGUAGEM



PYTHON



PROJETO



PROJETO ANALYTICS

Conteúdo da Aula

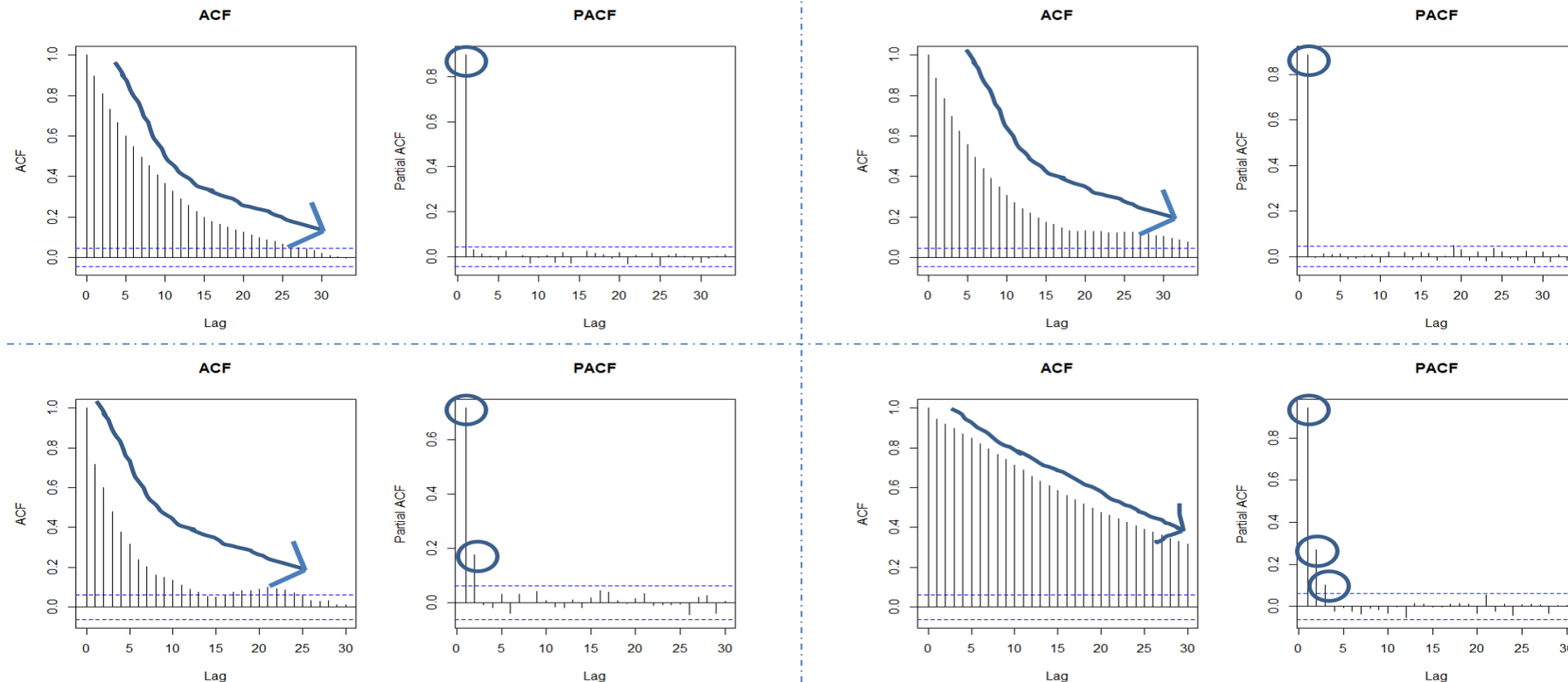
- 1. Introdução
- 2. Modelo Auto Regressivo e Médias-móveis: Ordens 1, 2, p,q
 - i. Identificação
 - ii. Estimação
- 3. Exercícios para Fixação
- 4. Medidas de qualidade do modelo
 - i. Exercícios

1. Introdução



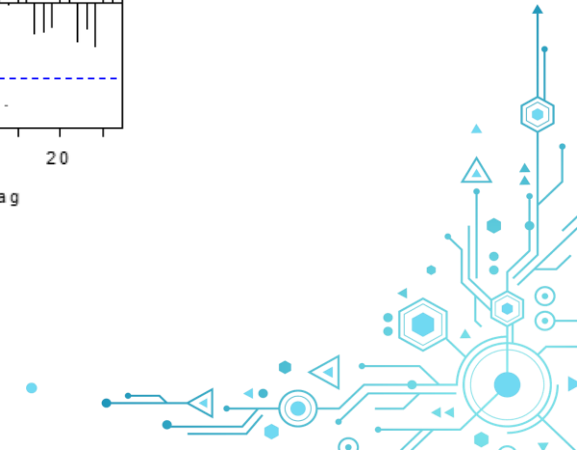
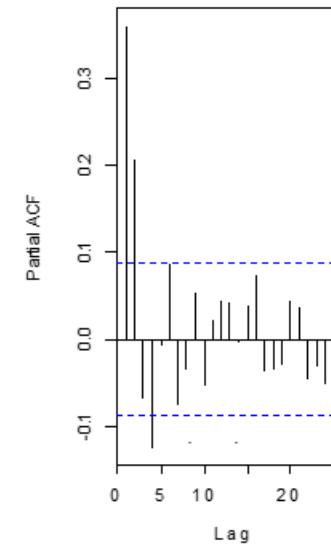
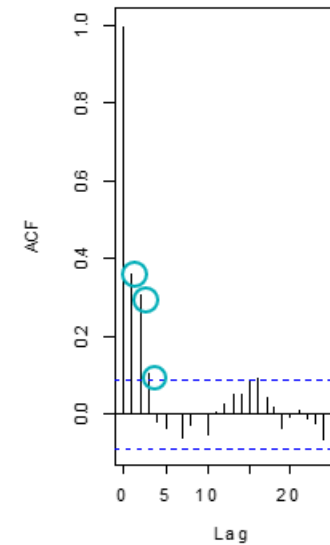
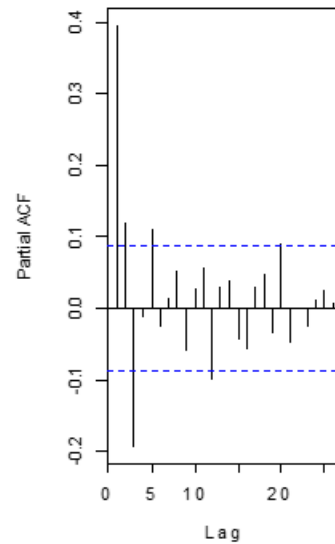
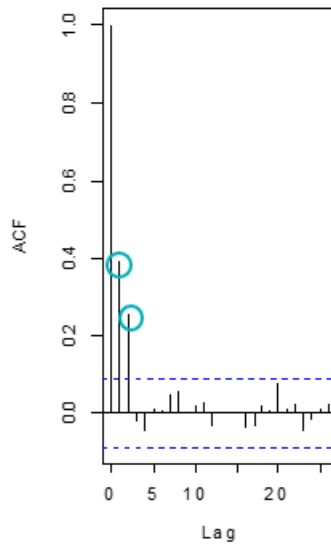
Até este momento, estudamos dois tipos de modelos da família de séries temporais:

Modelo AR: Uma série temporal segue um processo auto-regressivo quando o valor da série no tempo t depende do que aconteceu em, por exemplo, $t-1$, $t-2$, $t-3$, $t-4$, etc.



Até este momento, estudamos dois tipos de modelos da família de séries temporais:

Modelo MA: Uma série temporal segue um processo de médias móveis quando o valor da série no tempo t depende somente dos valores dos erros do modelo em $t-1$, $t-2$, $t-3$, etc.



2. Modelo Auto Regressivo e de Médias Móveis



Modelo Auto Regressivo e de Médias Móveis

2. MODELOS ARMAS | SÉRIE TEMPORAL

25

Uma série temporal segue um processo auto regressivo e de médias móveis quando o valor da série no tempo t depende do valor da série em $t-1, t-2, t-3$, etc e do valor do erro do modelo em $t-1, t-2, t-3$, etc.



Modelo Auto Regressivo e de Médias Móveis

2. MODELOS ARMAS | SÉRIE TEMPORAL

25

Modelo auto-regressivo e de médias móveis de ordem p,q – ARMA(p,q)

$$\hat{Y}_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$



Exemplo: ARMA(1)

2. MODELOS ARMAS | SÉRIE TEMPORAL

26

Uma série temporal segue um processo auto regressivo e de médias móveis quando o valor da série no tempo t depende do valor da série em $t-1, t-2, t-3$, etc e do valor do erro do modelo em $t-1, t-2, t-3$, etc.

$$Y_t = 600 + 0,15 * Y_{t-1} + 0,8 * \varepsilon_{t-1}$$

Y_t : valor previsto da série no instante t

Y_{t-1} : valor da série no instante $t-1$

ε_{t-1} : valor do erro do modelo no instante $t-1$

Seja Y , a variável vendas mensais de uma empresa de Cosméticos, o gestor da área gostaria de saber qual será a venda de abril sabendo que a venda de março foi de 700 unidades e o erro do modelo em março foi de 50 unidades.



Exemplo: ARMA(1)

2. MODELOS ARMAS | SÉRIE TEMPORAL

26

Uma série temporal segue um processo auto regressivo e de médias móveis quando o valor da série no tempo t depende do valor da série em $t-1, t-2, t-3$, etc e do valor do erro do modelo em $t-1, t-2, t-3$, etc.

$$Y_t = 600 + 0,15 * Y_{t-1} + 0,8 * \varepsilon_{t-1}$$

Y_t : valor previsto da série no instante t

Y_{t-1} : valor da série no instante $t-1$

ε_{t-1} : valor do erro do modelo no instante $t-1$

Seja Y , a variável vendas mensais de uma empresa de Cosméticos, o gestor da área gostaria de saber qual será a venda de abril sabendo que a venda de março foi de 700 unidades e o erro do modelo em março foi de 50 unidades.

$$Y_t = 600 + 0,15 * Y_{t-1} + 0,8 * \varepsilon_{t-1}$$

$$Y_t = 600 + 0,15 * 700 + 0,8 * 50$$

$$Y_t = 745$$



Exercício: ARMA(1)

2. MODELOS ARMAS | SÉRIE TEMPORAL

26

Uma série temporal segue um processo auto regressivo e de médias móveis quando o valor da série no tempo t depende do valor da série em $t-1, t-2, t-3$, etc e do valor do erro do modelo em $t-1, t-2, t-3$, etc.

$$Y_t = 24 + 0,823 * Y_{t-1} + 0,5 * \varepsilon_{t-1}$$

Y_t : valor previsto da série no instante t

Y_{t-1} : valor da série no instante $t-1$

ε_{t-1} : valor do erro do modelo no instante $t-1$



Seja Y , a variável quantidade de chamadas em uma central de atendimento de uma empresa de serviços que dá suporte para seus usuários, o gestor da área gostaria de saber qual será a quantidade de chamadas previstas para amanhã, dado que hoje o dia fechou com 321 chamadas e o erro do modelo foi de 40 chamadas.



Metodologia de Box e Jenkins

2.1 MODELO AUTO-REGRESSIVO | SÉRIE TEMPORAL

Uma vez confirmado pelo teste de Dickey-Fuller Aumentado que a série é estacionária, pode-se utilizar a metodologia de Box e Jenkins, que consiste na realização das seguintes fases:

- i. Identificação
- ii. Estimação
- iii. Previsão



1. Fase de Identificação

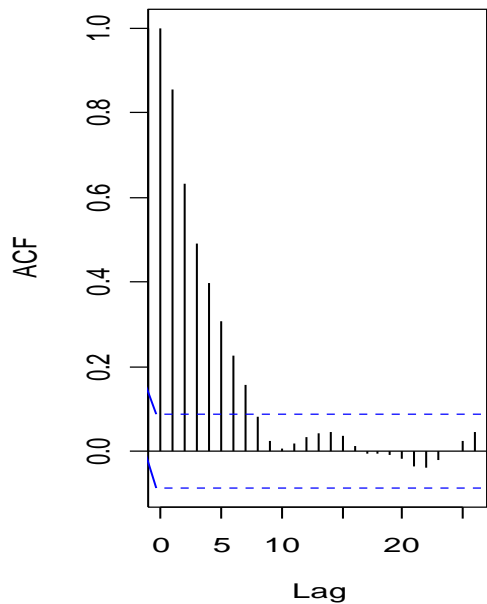
2.i. IDENTIFICAÇÃO | METODOLOGIA DE BOX JENKINS

33

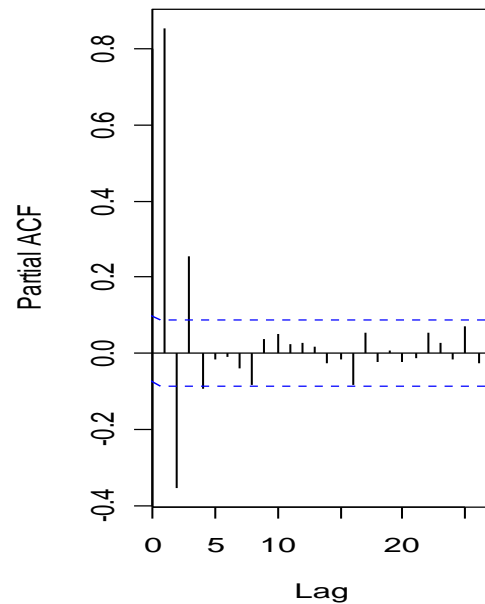
Um modelo ARMA não é de simples identificação.

Inicia-se com o modelo AR ou MA e, na análise dos resíduos, identifica-se a necessidade de inclusão dos demais parâmetros.

Função Autocorrelação



Função de Autocorrelação Parcial



Dado o conteúdo apresentado até o momento e a análise dos gráficos ACF e PACF, qual ordem de modelo vocês indicariam?

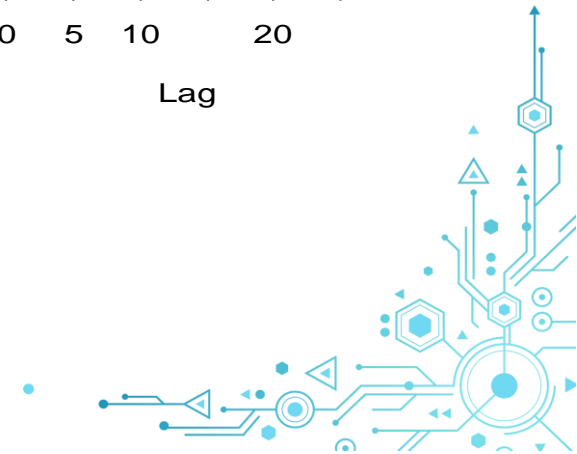
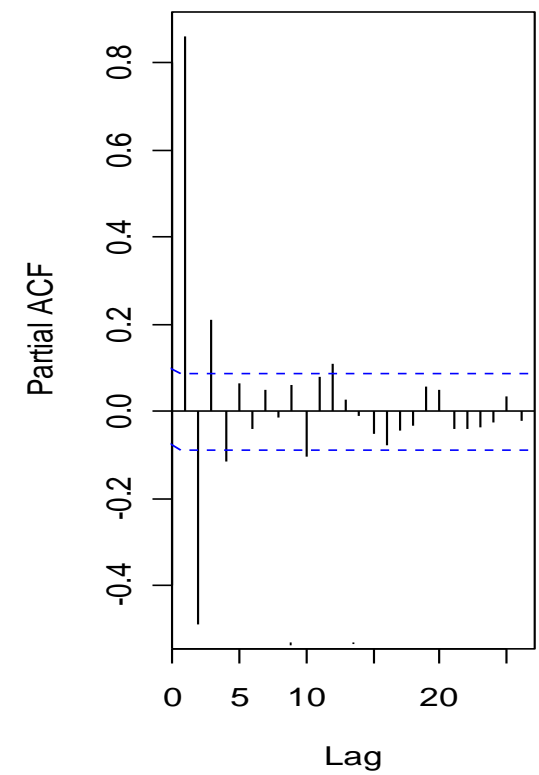
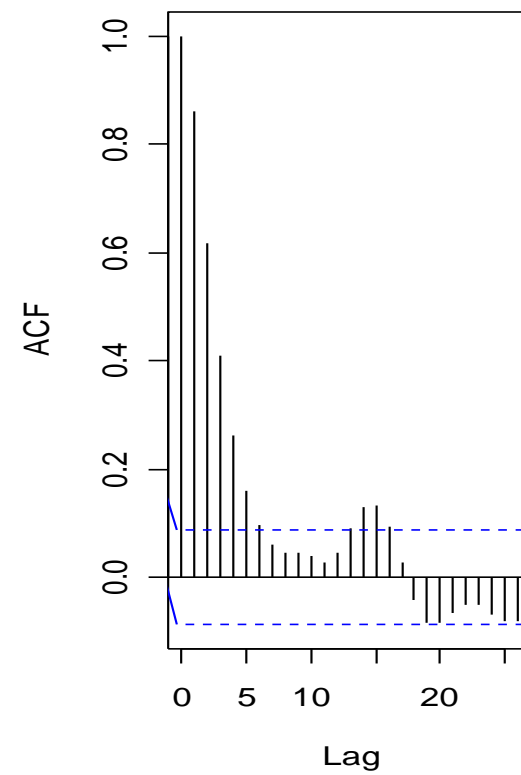
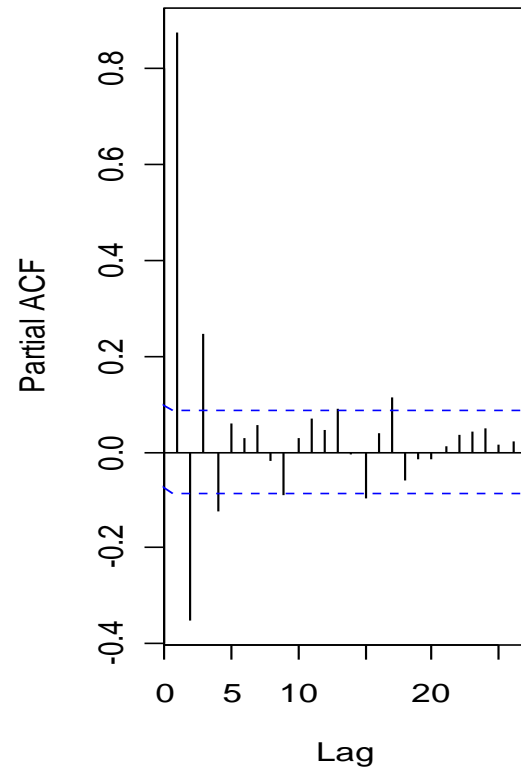
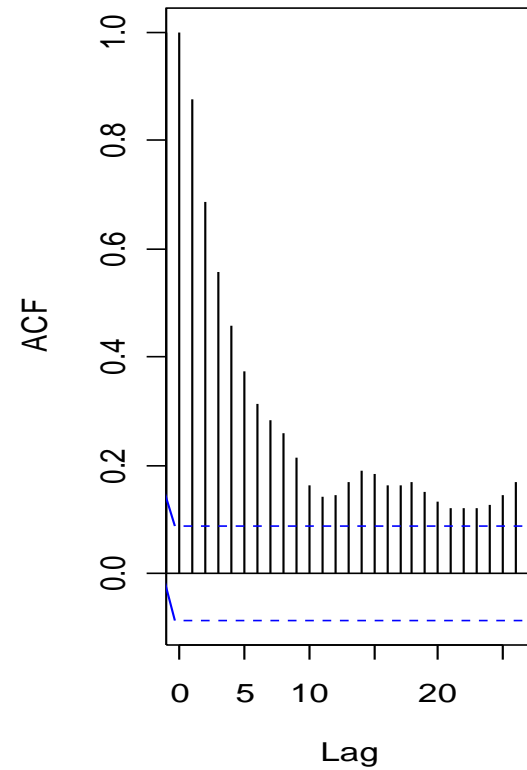
Note que a **ACF na lag zero será igual a um**, pois é a correlação da série no mesmo instante t contra ela mesma, ou seja, sem *lag* de defasagem.



Exemplos: ARMA(1,1)

2.i. IDENTIFICAÇÃO | METODOLOGIA DE BOX JENKINS

36



Case Serie 12: Realize as fase de identificação e estimação da série

2.ii. ESTIMAÇÃO | METODOLOGIA DE BOX EJENKINS

39

Um investidor deseja saber se aconselhável investir em uma determinada carteira de ações. Ele tem disponível os dados históricos das variações da carteira. Realize uma análise desta série de retornos e avalie a possibilidade do ajuste de um modelo ARMA.



- (a) Faça a análise exploratória da série, e comente seu comportamento.
- (b) Construa o gráfico descritivo da série.
- (c) Teste se a série é estacionária.
- (d) Apresente graficamente a função de autocorrelação (ACF) e a função de autocorrelação parcial (PACF) e discuta sobre a ordem a ser testada no modelo.
- (e) Avalie a significância dos parâmetros.
- (f) Apresente graficamente a função de autocorrelação (ACF) e a função de autocorrelação parcial (PACF) dos resíduos do modelo e discuta se a ordem do AR adotada está adequada.
- (g) Escreva a equação do modelo

Vamos fazer
juntos?

R Studio®

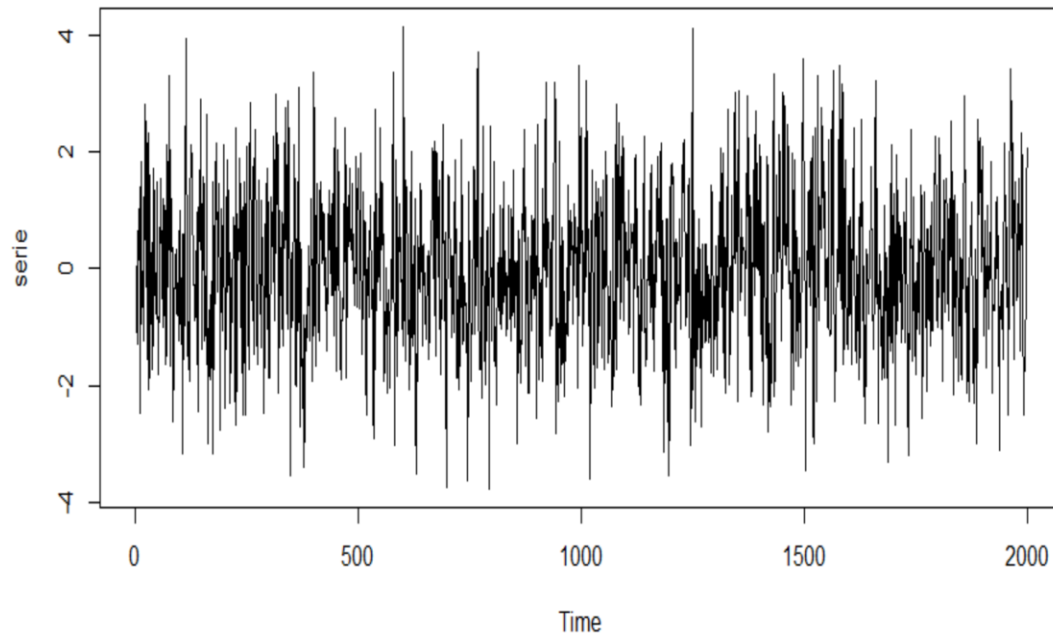


Case Serie 12 : Análise Exploratória

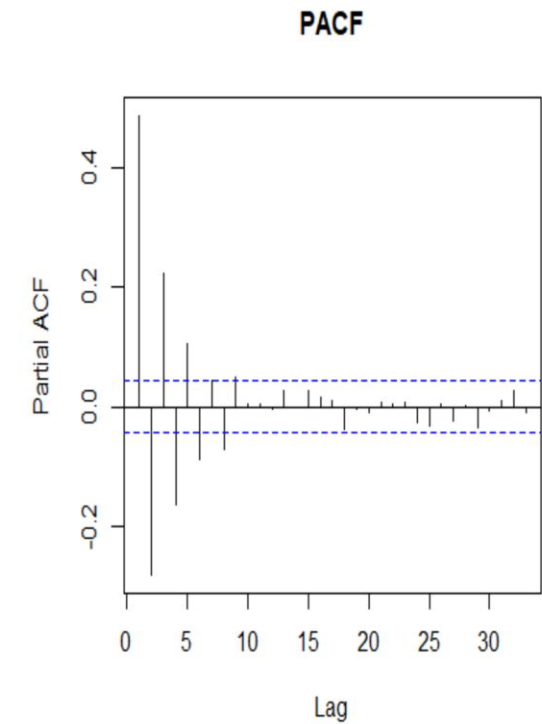
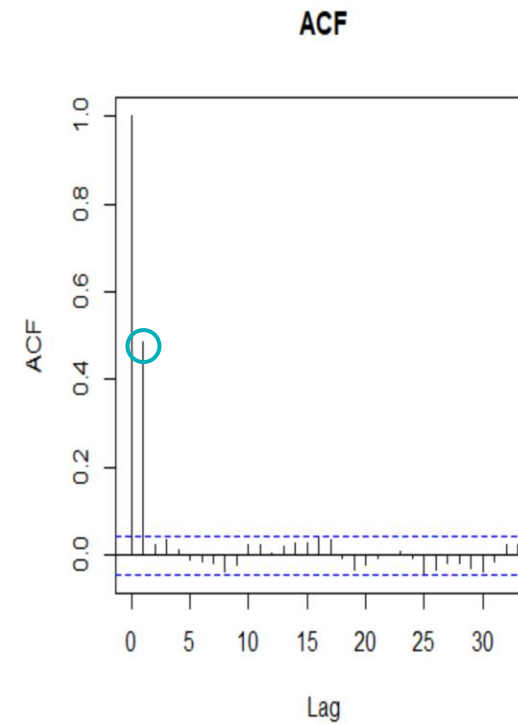
2.ii. ESTIMAÇÃO | METODOLOGIA DE BOX EJENKINS

40

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-3.768	-0.8898	-0.0127	-0.0026	0.8518	4.1476



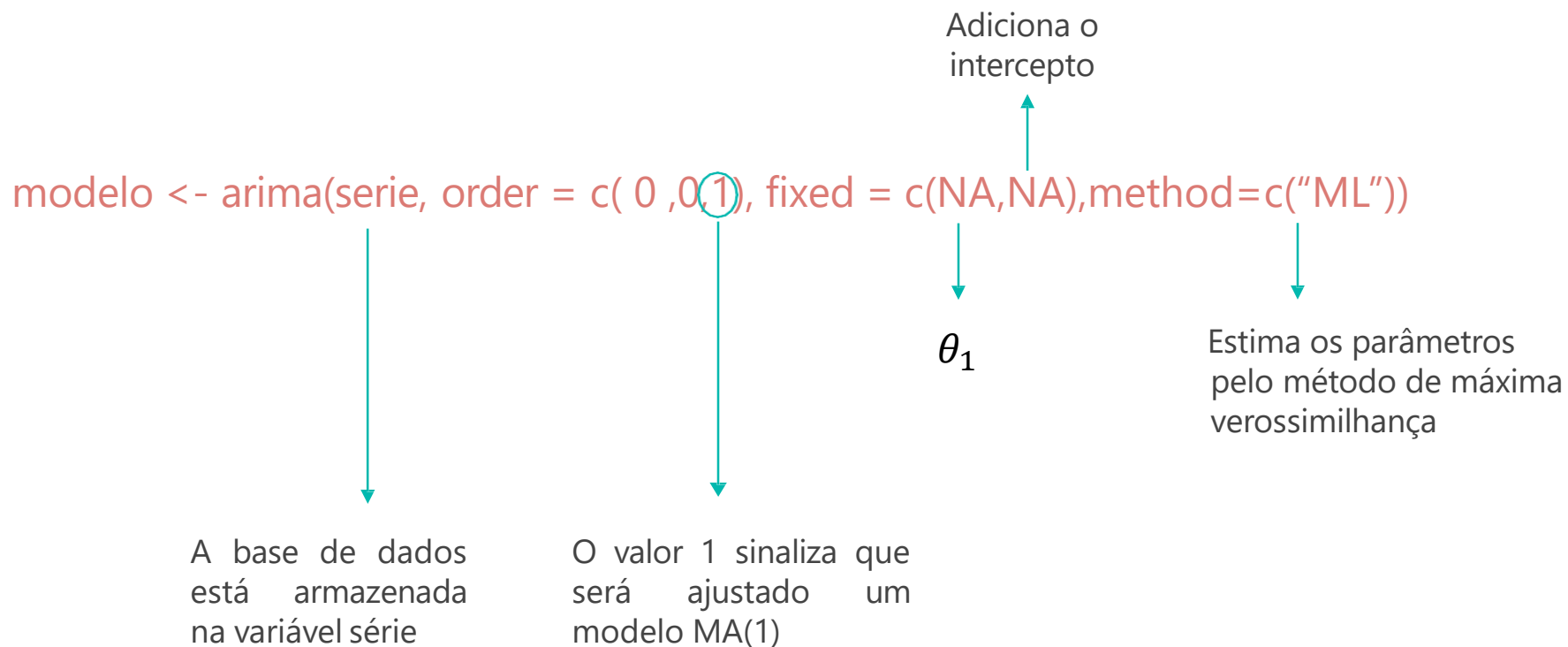
Augmented Dickey-Fuller Testdata:
p-value = 0.01
alternative hypothesis: stationary



Case Serie 12 : Ajuste do modelo no R

2.ii. ESTIMAÇÃO | METODOLOGIA DE BOX EJENKINS

42



Case Serie 12 : Ajuste do modelo no R

2.ii. ESTIMAÇÃO | METODOLOGIA DE BOX EJENKINS

41

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
ma1	0.7607385	0.0144410	52.6792	<0.00000000000000002	***
intercept	-0.0026656	0.0402744	-0.0662	0.9472	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

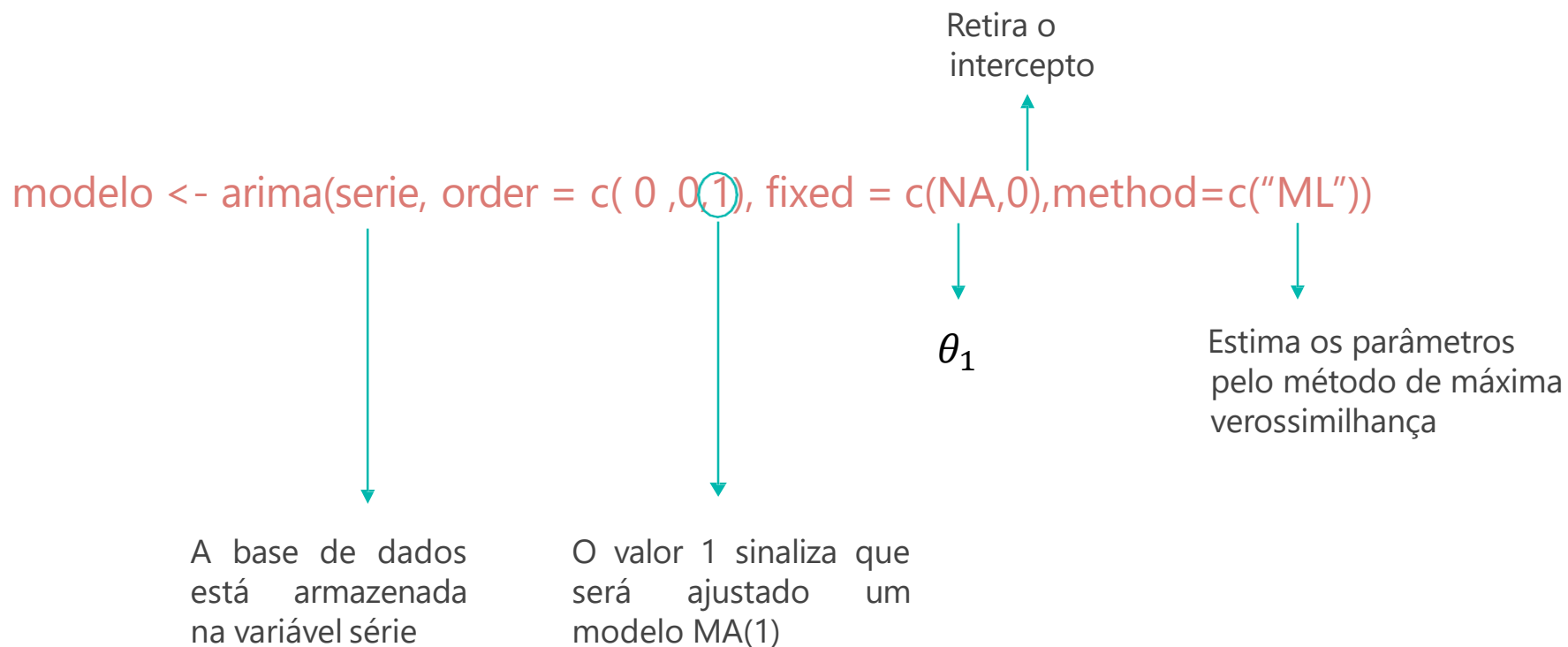
Como o **p-valor associado ao intercepto é maior que 0,10***, há evidência que é igual a 0 e deve ser retirado do modelo.



Case Serie 12 : Ajuste do modelo no R

2.ii. ESTIMAÇÃO | METODOLOGIA DE BOX EJENKINS

42



Case Serie 12 : Ajuste do modelo no R

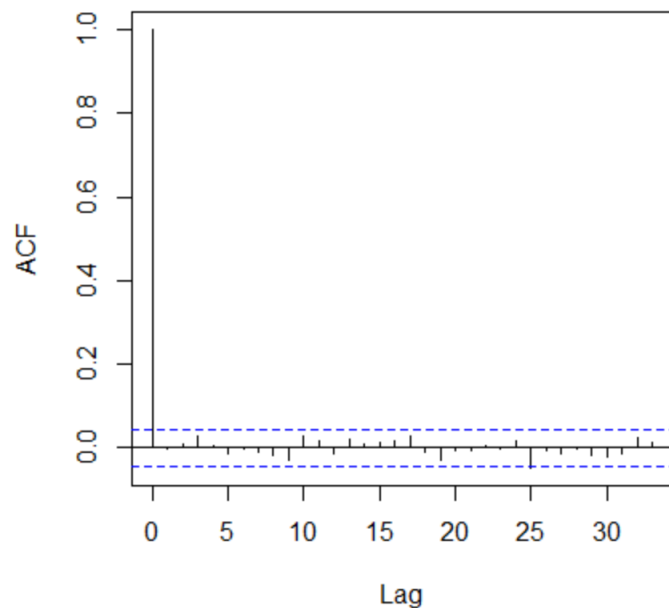
2.ii. ESTIMAÇÃO | METODOLOGIA DE BOX EJENKINS

41

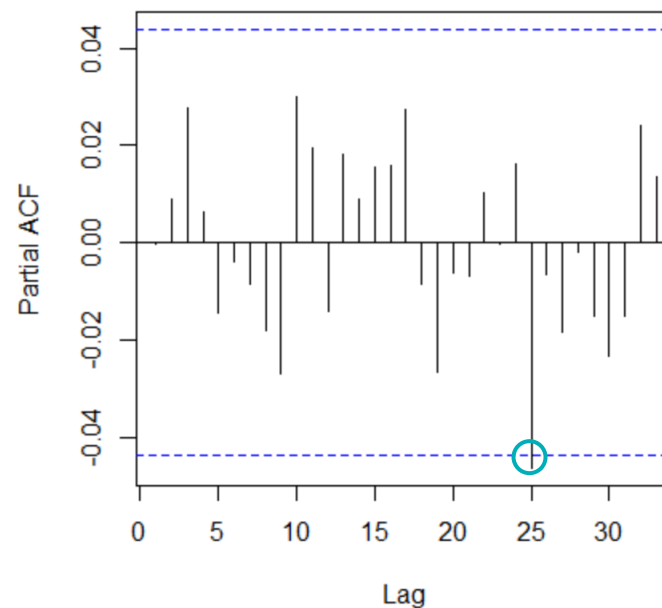
z test of coefficients:

```
      Estimate Std. Error z value      Pr(>|z|)
ma1  0.76075    0.01444  52.681 < 0.00000000000000022 ***
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ACF dos Resíduos



PACF dos Resíduos



Nota-se que apenas a autocorrelação parcial de ordem 25 é diferente de zero. Desta forma, o modelo não está adequado e é necessário adicionar mais parâmetros no modelo.

Quando tiverem autocorrelações parciais diferentes de zero no resíduo, deve-se adicionar ordens autorregressivas. Neste caso, deve-se adicionar um AR de ordem 25, AR(25).



2.ii. ESTIMAÇÃO | METODOLOGIA DE BOX EJENKINS

42

O valor 1 sinaliza que será ajustado um modelo MA(1)

 $\phi_{26} \quad \theta_{25}$

Retira o intercepto



Case Serie 12 : Ajuste do modelo no R

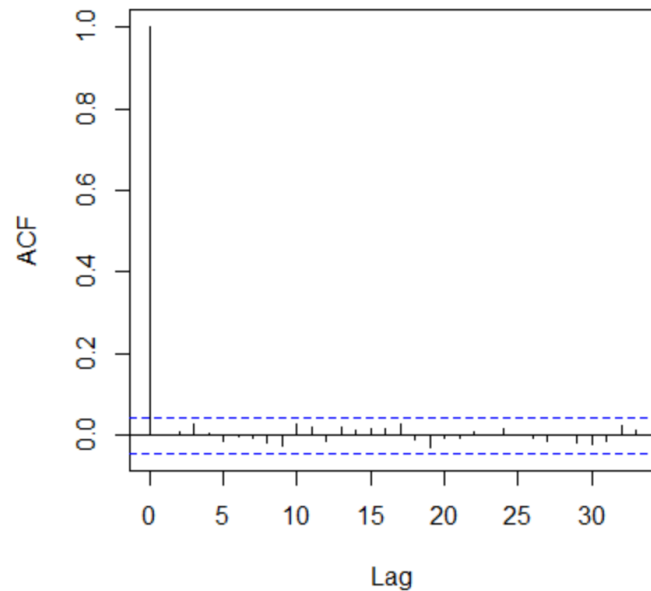
2.ii. ESTIMAÇÃO | METODOLOGIA DE BOX EJENKINS

41

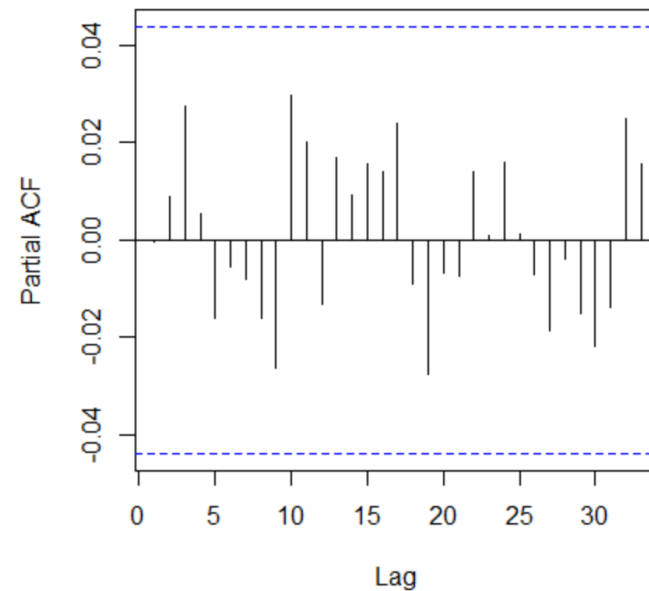
z test of coefficients:

```
      Estimate Std. Error z value      Pr(>|z|)
ar25 -0.047351   0.022448 -2.1094    0.03491 *
ma1   0.761481   0.014431 52.7657 < 0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ACF dos Resíduos



PACF dos Resíduos



Nota-se que todos os lags de defasagem das autocorrelações e as autocorrelações parciais dos resíduos são estatisticamente iguais a zero. Dessa forma, o modelo ARMA (25,1) para o modelo adequado.

Note que a **ACF na lag zero será igual a um**, pois é a correlação da série no mesmo instante t contra ela mesma, ou seja, sem *lag* de defasagem.



Case Serie 14 : Ajuste do modelo no R

2.ii. ESTIMAÇÃO | METODOLOGIA DE BOX EJENKINS

41

Equação do Modelo:

$$\hat{Y}_t = -0,0474 * Y_{t-1} + 0,7615 * \varepsilon_{t-1}$$



3. Exercícios para fixação



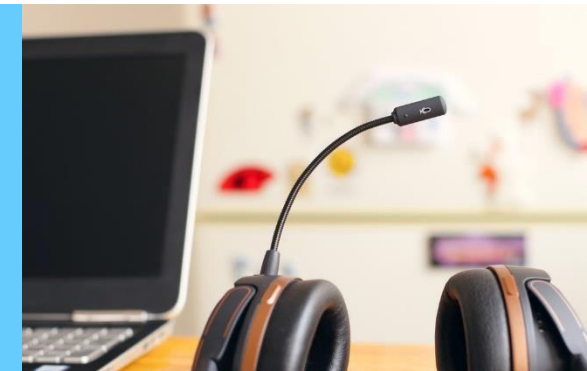
Exercícios para Fixação

3. EXERCÍCIOS | SERIES TEMPORAIS

39

Bases de dados:

- Série 3
- Série 4
- Série 5



- Faça a análise exploratória da série, e comente seu comportamento.
- Construa o gráfico descritivo da série.
- Teste se a série é estacionária.
- Apresente graficamente a função de autocorrelação (ACF) e a função de autocorrelação parcial (PACF) e discuta sobre a ordem a ser testada no modelo.
- Avalie a significância dos parâmetros.
- Apresente graficamente a função de autocorrelação (ACF) e a função de autocorrelação parcial (PACF) dos resíduos do modelo e discuta se a ordem do AR adotada está adequada.
- Escreva a equação do modelo

Vamos fazer
juntos?

R Studio®



4. Medidas de Qualidade do Modelo



Medidas de qualidade do modelo

4. MEDIDAS DE QUALIDADE | SÉRIE TEMPORAL

25

Existem várias medidas de qualidade do ajuste de um modelo de séries temporais.

Vamos apresentar as três mais usuais: SSE, MAE e MAPE.

Recomendamos o uso de, ao menos, duas medidas para avaliação de um modelo.



Medidas de qualidade do modelo

4. MEDIDAS DE QUALIDADE | SÉRIE TEMPORAL

25

	Nomenclatura	Fórmula
SSE	Sum of Squared Errors	$SSE = \sum (observado - estimado)^2$
MAE	Mean Absolute Error	$MAE = \frac{\sum (observado - estimado) }{n}$
MAPE	Mean Absolute Percent Error	$MAPE = \frac{\sum (observado - estimado) / observado}{n}$

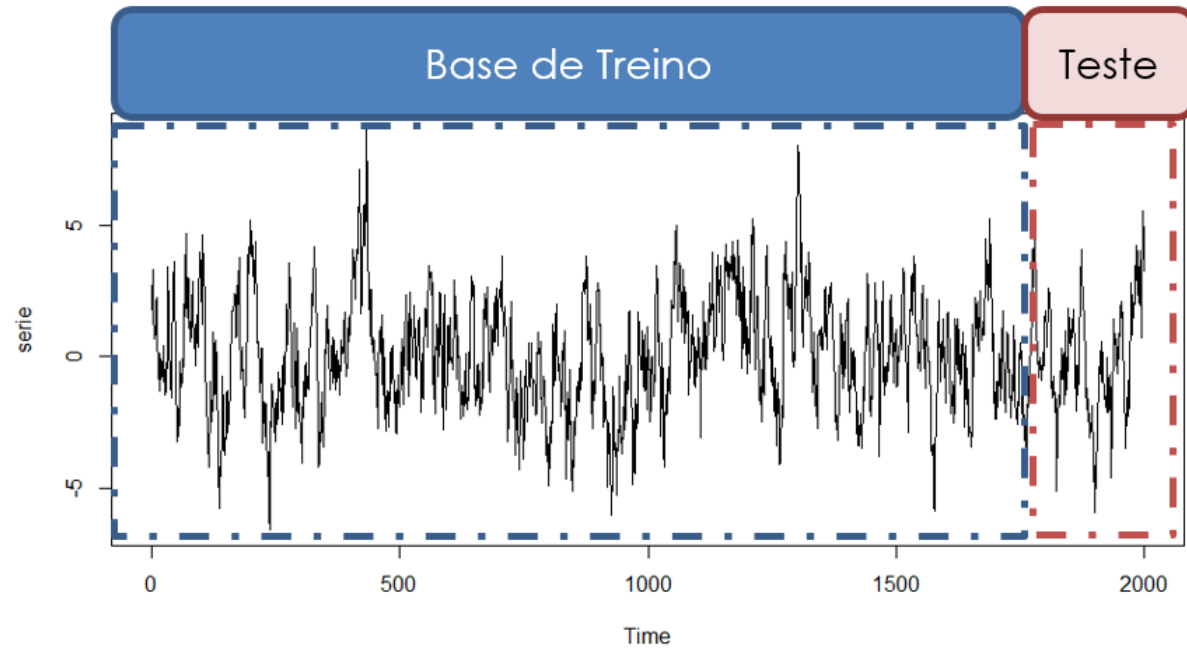


Medidas de qualidade do modelo

4. MEDIDAS DE QUALIDADE | SÉRIE TEMPORAL

18

Todas as medidas apresentadas têm como objetivo comparar o valor observado com o valor predito pelo modelo.



Exercícios

4. MEDIDAS DE QUALIDADE | SÉRIE TEMPORAL

39

Bases de dados:

- Série 22
- Série 31
- Série 14

- Divida a base em treino e teste (12 observações para teste);
- Estime o modelo;
- Calcule o SSE, MAE e MAPE da base de teste.

Vamos fazer
juntos?

R Studio®



- Box, G. E. P. & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.
- Dickey, D. & Fuller, W. A. (1979). *Distribution of the estimates for autoregressive time series with a unit root*. Journal of the American Statistical Association, 74(2), 427-431.
- Morettin, P. A. & Toloi, C. M. de C. (2004). *Análise de séries temporais*. São Paulo: Edgard Blücher.
- Morettin, P. A. (2008). *Econometria financeira: um curso em séries temporais financeiras*. São Paulo: Edgard Blücher.

