Analytics e Inteligência Artificial

Tema da aula
Séries Temporais

LABDATA | FIA FUNDAÇÃO INSTITUTO DE ADMINISTRAÇÃO BUSINESS SCHOOL

FIA Business School

NOSSOS DIFERENCIAIS | QUEMSOMOS

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 domaterial didático oferecidos nas atividades de ensino

Únicocursode graduação em administração a receber as notas máximas

A primeira escola brasileira a ser finalistadamaior competição de MBA domundo

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

Signatária do Pacto Global daONU

Membro fundador da ANAMBA- Associação Nacional MBAs

Credenciada pelaAMBA - *Association of MBAs*

Credenciada ao *Executive MBA Council*

Filiada a AACSB - *Associationto Advance Collegiate Schoolsof Business*

FiliadaaEFMD - *European Foundationfor Management Development*

Referência em cursosdeMBA nas principais mídiasde circulação

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 eprivadas.

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 (comestacionamento)

Corpo Diretivo

COORDENADORES DO LABDATA | ATUAÇÃO ACADÊMCIA E PROFISSIONAL

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/

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

ANALYTICS

MACHINE LEARNING

DEEP LEARNING

ANÁLISE EXPLORATÓRIA

TÉCNICAS DE PROJEÇÃO

TÉCNICAS DE CLASSIFICAÇÃO

TÉCNICAS DE SEGMENTAÇÃO

TÉCNICAS DE ANALYTICS

LINGUAGEM

PYTHON

R

PROJETO

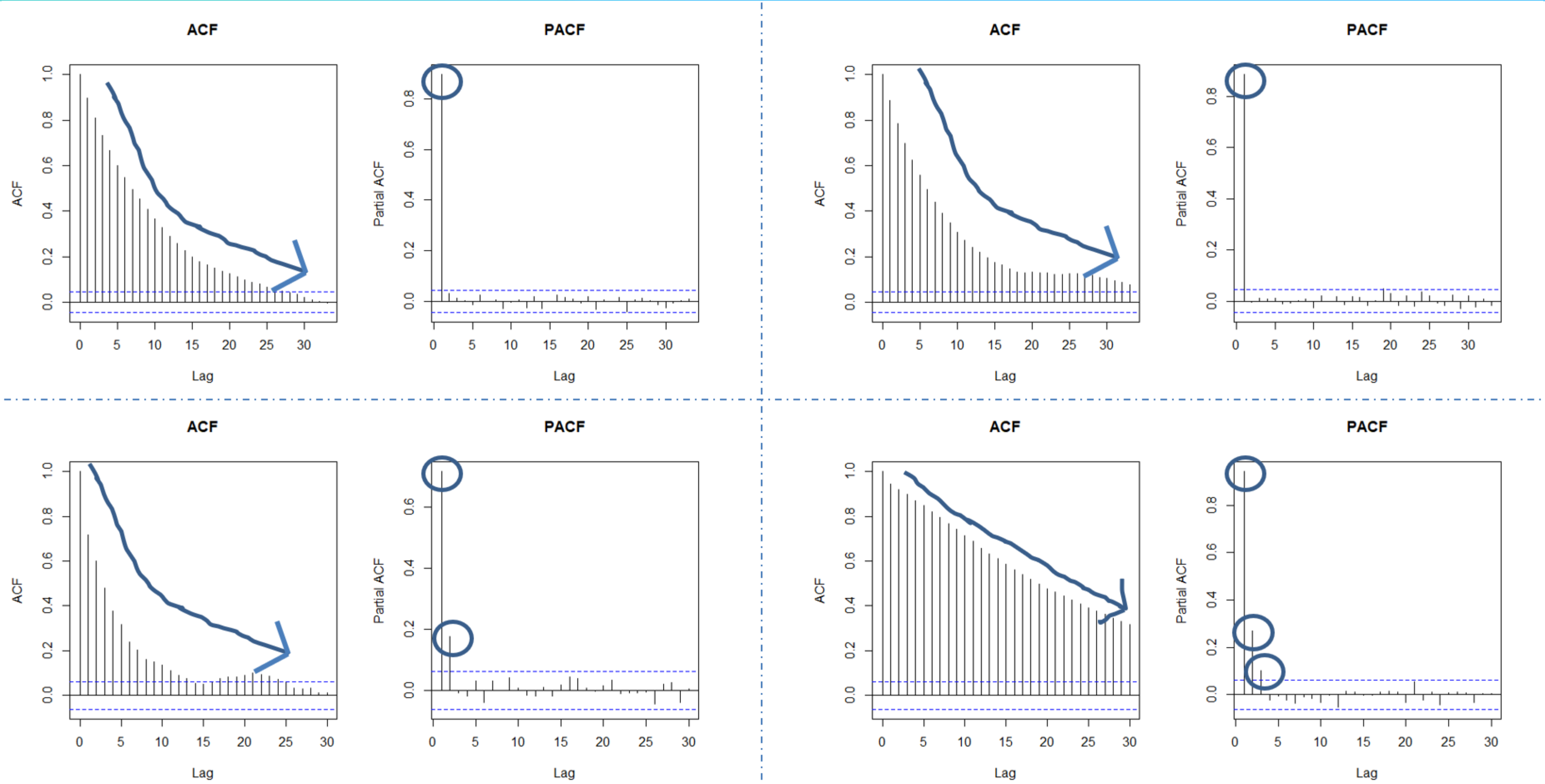
PROJETO ANALYTICS

Conteúdo da Aula

1. Introdução
2. Modelo Médias-móveis: Ordens 1, 2, q
    i. Identificação
    ii. Estimação
3. Exercícios para Fixação

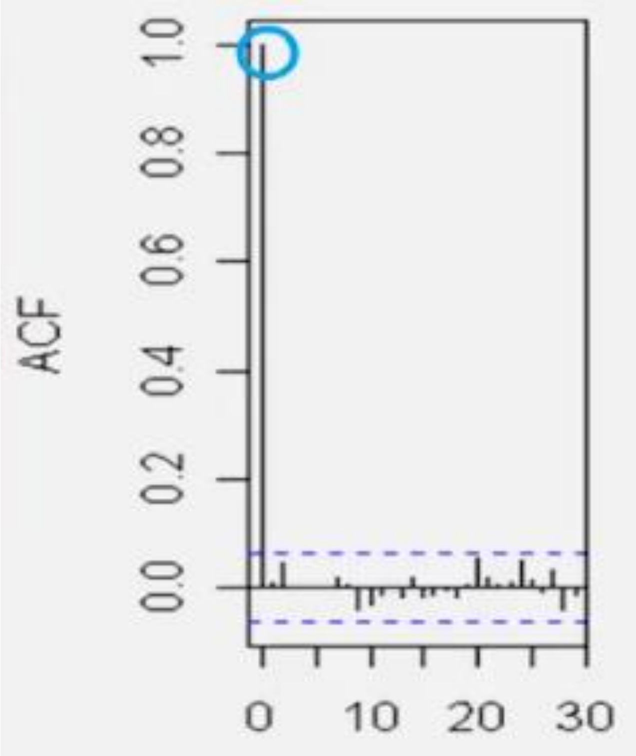
1. Introdução

Vimos algo em comum em todas as séries até agora: existem várias autocorrelações diferentes de zero.



Em alguns casos, não existem várias autocorrelações diferentes de zero.

Nesses casos, não é possível ajustar um modelo autorregressivo(AR).



2. Modelo de Médias Móveis

Modelo Médias Móveis

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.

Erro é o valor observado da série temporal menos o valor ajustado pelo modelo.

Quando uma empresa projeta uma venda em 10 milhões de reais e a venda real é de 8 milhões de reais, o erro do modelo é de -2 milhões de reais.

Exemplo: MA(1)

Uma série temporal segue um processo de médias móveis de ordem 1 ou MA(1) quando o valor da série no tempo t depende somente dos valores dos erros no modelo em t-1.

$$Y_t = 500 + 0{,}12 * \varepsilon_{t-1}$$

Yt: valor previsto da série no instante t

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 março sabendo que o erro do modelo em fevereiro foi de 50 unidades.

Exemplo: MA(1)

Uma série temporal segue um processo de médias móveis de ordem 1 ou MA(1) quando o valor da série no tempo t depende somente dos valores dos erros no modelo em t-1.

$$Y_t = 500 + 0,12 * \varepsilon_{t-1}$$

Yt: valor previsto da série no instante t

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 março sabendo que o erro do modelo em fevereiro foi de 50 unidades.

$$Y_t = 500 + 0,12 * \varepsilon_{t-1}$$

$$Y_t = 500 + 0,12 * 50$$

$$Y_t = 506$$

Exercício: MA(1)

Uma série temporal segue um processo de médias móveis de ordem 1 ou MA(1) quando o valor da série no tempo t depende somente dos valores dos erros no modelo em t-1.

$$Y_t = 600 + 0{,}45 * \varepsilon_{t-1}$$

Yt: valor previsto da série no instante t

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

Seja Y, a variável vendas mensais de uma empresa de esportes, a diretora da área gostaria de saber qual será a venda de outubro sabendo que o erro do modelo em setembro foi de 30 unidades.

Exemplo: MA(2)

Uma série temporal segue um processo de médias móveis de ordem 2 ou MA(2) quando o valor da série no tempo t depende somente dos valores dos erros no modelo em t-1 e t-2.

$$Y_t = 300 + 0{,}12 * \varepsilon_{t-1} + 0{,}08 * \varepsilon_{t-2}$$

Yt: valor previsto da série no instante t

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

ε_{t-2}: valor do erro do modelo no instante t-2

Seja Y_t a variável vendas mensais de uma empresa de eletrodomésticos, o coordenador da área gostaria de saber qual será a venda de julho sabendo que o erro do modelo em junho foi de 50 unidades e o erro em maio foi de 30 unidades.

Exemplo: MA(2)

Uma série temporal segue um processo de médias móveis de ordem 2 ou MA(2) quando o valor da série no tempo t depende somente dos valores dos erros no modelo em t-1 e t-2.

$$Y_t = 300 + 0{,}12 * \varepsilon_{t-1} + 0{,}08 * \varepsilon_{t-2}$$

Yt: valor previsto da série no instante t

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

ε_{t-2}: valor do erro do modelo no instante t-2

Seja Y_t a variável vendas mensais de uma empresa de eletrodomésticos, o coordenador da área gostaria de saber qual será a venda de julho sabendo que o erro do modelo em junho foi de 50 unidades e o erro em maio foi de 30 unidades.

$$Y_t = 300 + 0{,}12 * \varepsilon_{t-1} + 0{,}08 * \varepsilon_{t-2}$$

$$Y_t = 300 + 0{,}12 * 50 + 0{,}08 * 30$$

$$Y_t = 308$$

Exercício: MA(2)

Uma série temporal segue um processo de médias móveis de ordem 2 ou MA(2) quando o valor da série no tempo t depende somente dos valores dos erros no modelo em t-1 e t-2.

$$Y_t = 100 + 0{,}22 * \varepsilon_{t-1} + 0{,}15 * \varepsilon_{t-2}$$

Yt: valor previsto da série no instante t

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

ε_{t-2}: valor do erro do modelo no instante t-2

Seja Y, a variável número de festas mensais de um buffet, a coordenadora da área gostaria de saber qual será a venda de julho sabendo que o erro do modelo em junho foi de 10 unidades e o erro em maio foi de 18 unidades.

Exercício: MA(2)

Uma série temporal segue um processo de médias móveis de ordem 2 ou MA(2) quando o valor da série no tempo t depende somente dos valores dos erros no modelo em t-1 e t-2.

$$Y_t = 100 + 0,22 * \varepsilon_{t-1} + 0,15 * \varepsilon_{t-2}$$

Yt: valor previsto da série no instante t

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

ε_{t-2}: valor do erro do modelo no instante t-2

Seja Y_t a variável número de festas mensais de um buffet, a coordenadora da área gostaria de saber qual será a venda de julho sabendo que o erro do modelo em junho foi de 10 unidades e o erro em maio foi de 18 unidades.

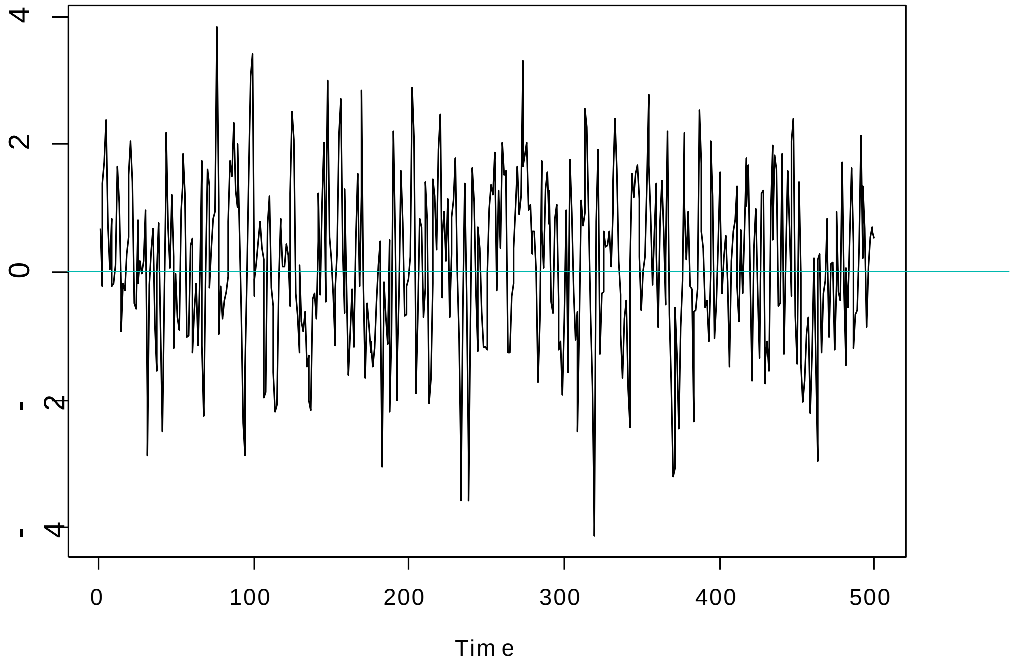
$$Y_t = 100 + 0,22 * \varepsilon_{t-1} + 0,15 * \varepsilon_{t-2}$$

$$Y_t = 100 + 0,22 * 10 + 0,15 * 18$$

$$Y_t = 105$$

Séries Estacionárias

Uma série temporal é estacionária quando oscila ao redor de uma **média constante,** refletindo um comportamento estável e aleatório ao longo do tempo.



A média da série é constante

Para o ajuste de uma série temporal por meio da metodologia de Box e Jenkins é necessário verificar se a série é estacionária, pois a metodologia de Box e Jenkins é aplicada quando a série é estacionária.

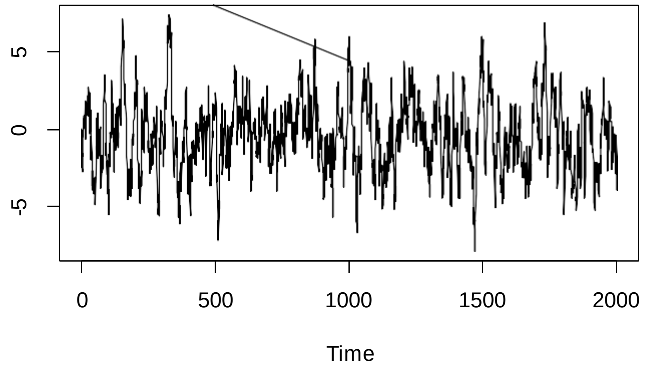
Apesar de ser possível visualizar graficamente a estacionariedade da série, pode-se utilizar o **teste de estacionariedade de Dickey-Fuller Aumentado** (Dickey & Fuller, 1979) , com as seguintes hipóteses:

H_0: a série não é estacionária

H_1: a série é estacionária

Regra de decisão: Quando o nível descritivo é < 0,10 rejeitamos H_0, ou seja, há evidência de que a série é estacionária.

Esta série temporal parece estacionária, pois oscila em torno de uma média constante.



Metodologia de Box e Jenkins

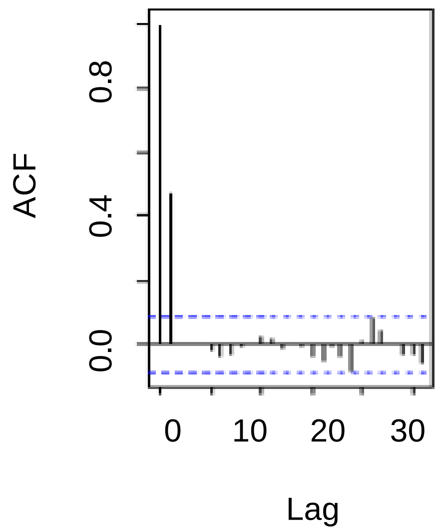
Uma vez confirmado pelo teste de Dickey-Fuller Aumentado que a série é estacionária, pode-se utilizar a metodologia de Box e Jenkins, que consiste na realização das seguintesfases:

i. Identificação
ii. Estimação
iii. Previsão

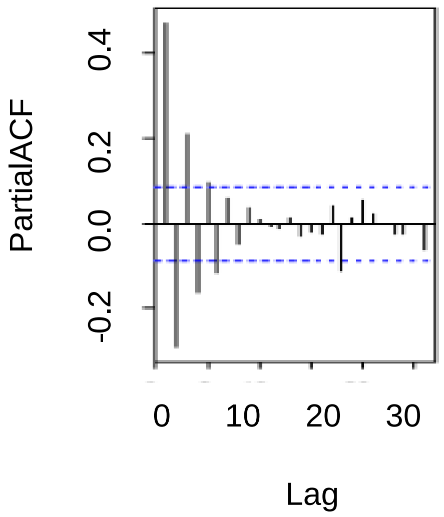
Para identificar se a série temporal segue um modelo de Médias Móveis, deve-se analisar as funções de autocorrelação (ACF – *autocorrelation function*) e autocorrelação parcial (PACF – *partial autocorrelation function*).

Esta é uma análise exploratória da série de dados para investigar as ordens a serem sugeridas para o modelo MA de maneira a se obter uma estrutura de modelo parcimoniosa.

Função Autocorrelação



Função de Autocorrelação Parcial



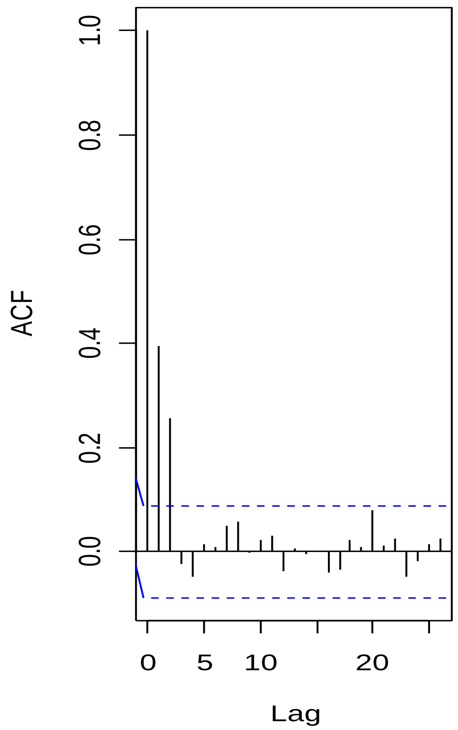
Para o modelo ser Médias Móveis, pela análise da PACF, várias autocorrelações parciais devem apresentar valores diferentes de zero.

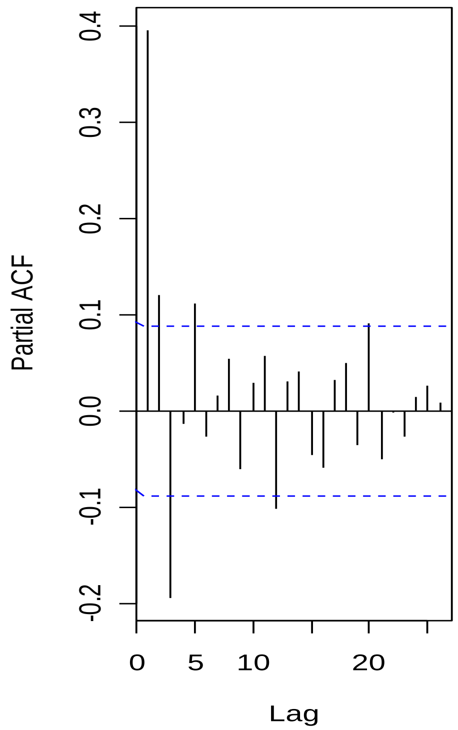
Note que a **ACF na *lag* zero será igual a um**, pois é a correlação da série no mesmo instante t contra ela mesma, ou seja, sem *lag* de defasagem.

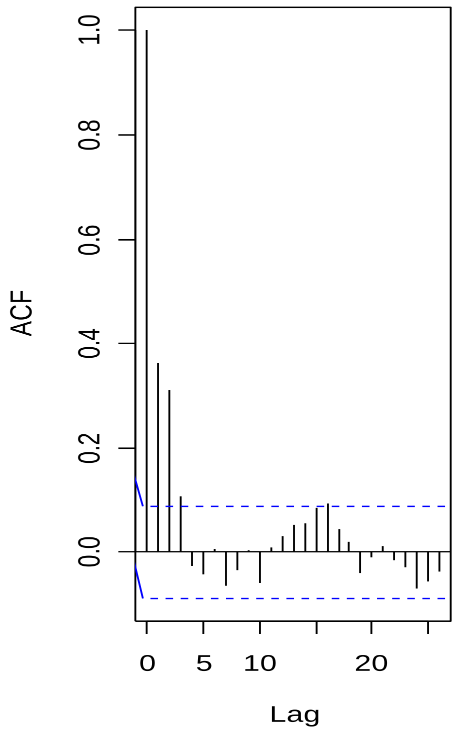
Além disso, pela análise da ACF, devemos ter algumas autocorrelações diferentes de zero. Como apenas a autocorrelação de ordem 1 é diferente de zero, deve-se ajustar um modelo de médias móveis de ordem 1, MA(1).

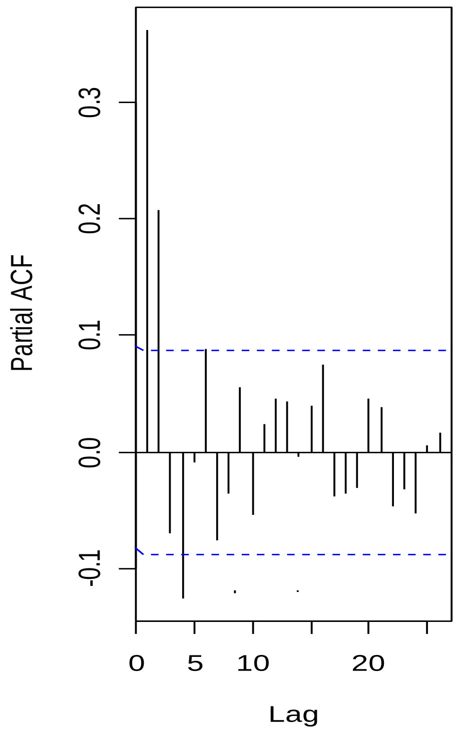
Gráfico	AR	MA
ACF	Várias autocorrelações diferentes de 0	Ordem q
PACF	Ordem p	Várias autocorrelações diferentes de 0

Exercício: Identifique pelas ACF e ACFP se seriam modelos MA e qual sua ordem

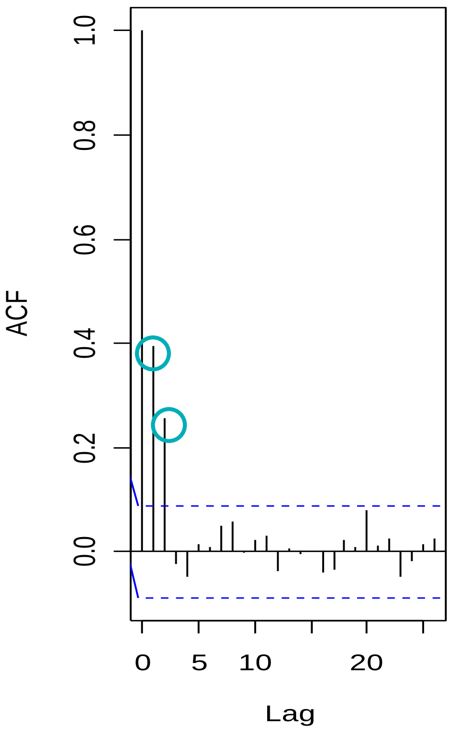


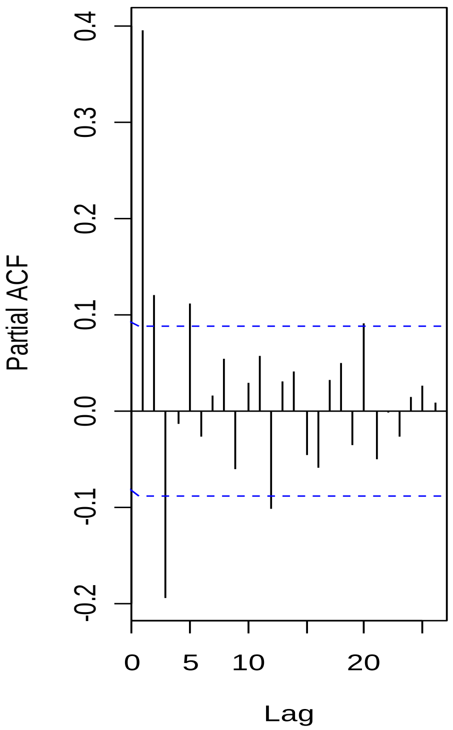


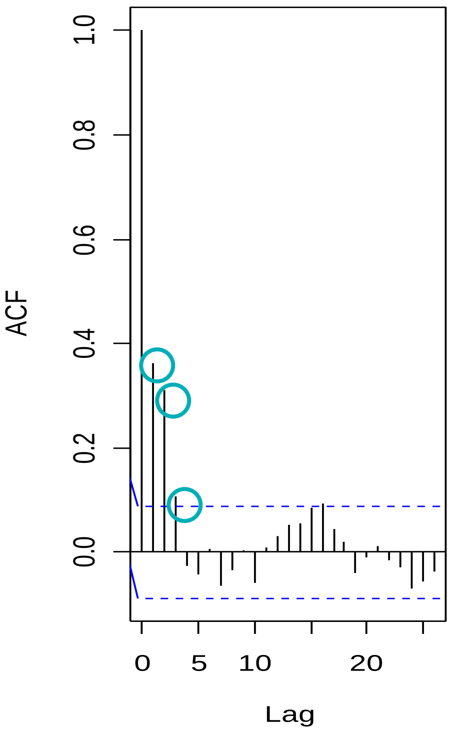


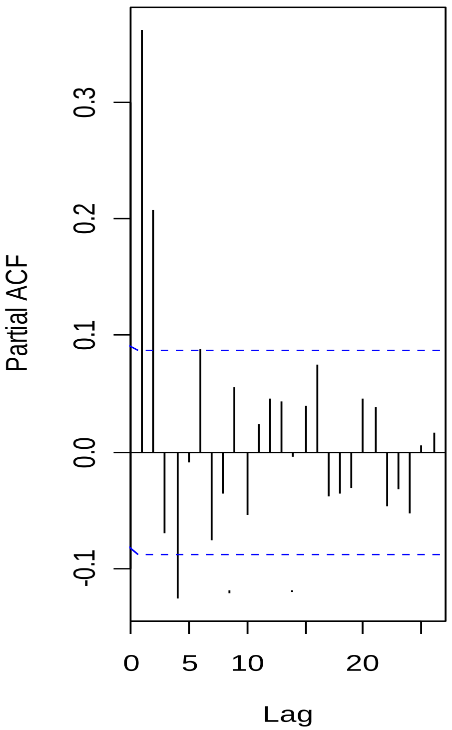


Exercício: Identifique pelas ACF e ACFP se seriam modelos MA e qual sua ordem









Há evidências de que a série temporal segue um modelo MA(2), pois existem várias autocorrelações parciais diferentes de zero e apenas as autocorrelações 1 e 2 são diferentes de zero.

Há evidências de que a série temporal segue um modelo MA(3), pois existem várias autocorrelações parciais diferentes de zero e apenas as autocorrelações 1, 2 e 3 são diferentes de zero.

2.2 Fase de Estimação

2.ii. ESTIMAÇÃO | METODOLOGIA DE BOX EJENKINS

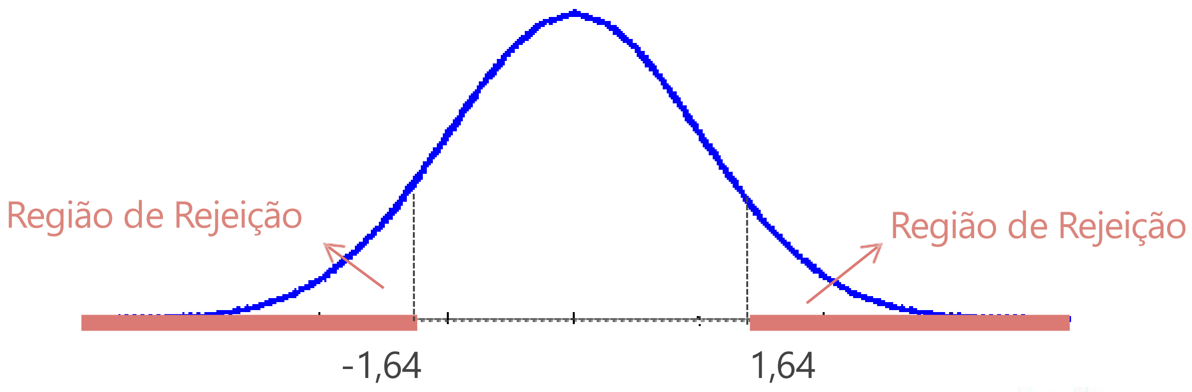
Uma vez identificada a ordem do modelo MA a ser testada, o próximo passo é estimar os parâmetros do modelo:

$$Y_t = \theta_0 + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$$

Teste de hipótese para cada um dos parâmetros i=0,1, ..., q:

$$H_0: \theta_i = 0$$
$$H_1: \theta_i \neq 0$$

Considerando 90 % de confiança da Distribuição Normal:

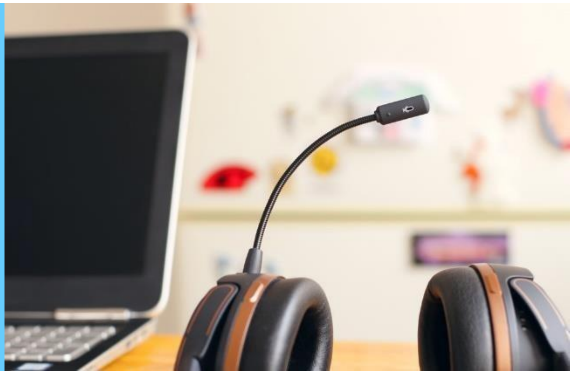


Quando o p-valor do teste estiver abaixo do nível descritivo adotado de 10% ou quando a estatística do teste estiver na região de rejeição, deve-se rejeitar H_0.

Case *Serie 14*: Realize as fase de identificação e estimação da série

2.ii. ESTIMAÇÃO | METODOLOGIA DE BOX EJENKINS

Ao observar a queda no número de vendas de um produto, um gestor gostaria de saber qual será a venda estimada para o próximo mês, uma vez que, ele gostaria de saber se o número de itens no estoque será suficiente para atender a demanda ou se vale a pena fazer um novo pedido.



(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 MA.

(e) Ajuste um modelo Médias Móveis MA 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

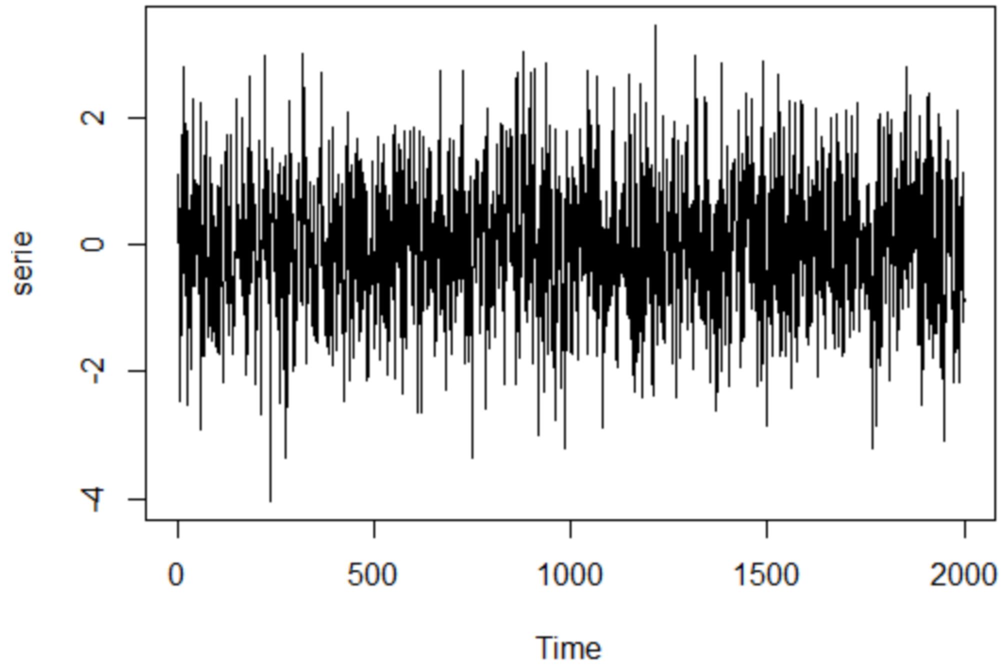
Vamos fazer juntos?

RStudio

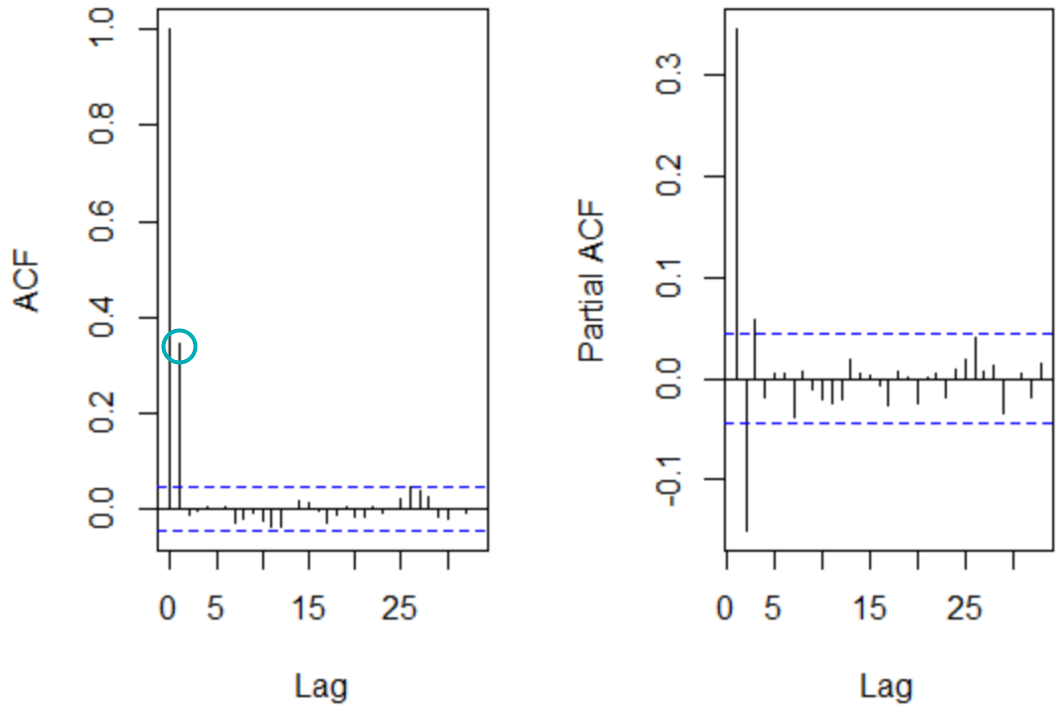
Case *Serie 14* : Análise Exploratória

2.ii. ESTIMAÇÃO | METODOLOGIA DE BOX EJENKINS

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-4.034	-0.736	-0.034	-0.020	0.689	3.457



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



Adiciona o intercepto

```
modelo <- arima(serie, order = c( 0 ,0,1), fixed = c(NA,NA),method=c("ML"))
```

θ_1

Estima os parâmetros pelo método de máxima verossimilhança

A base de dados está armazenada na variável série

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

```
z test of coefficients:

           Estimate Std. Error z value              Pr(>|z|)
ma1        0.408268   0.020300  20.112 <0.0000000000000002 ***
intercept -0.020364   0.031332  -0.650               0.5157
---
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 14* : Ajuste do modelo no R

Retira o intercepto

```
modelo <- arima(serie, order = c( 0 ,0,1), fixed = c(NA,0),method=c("ML"))
```

θ_1

Estima os parâmetros pelo método de máxima verossimilhança

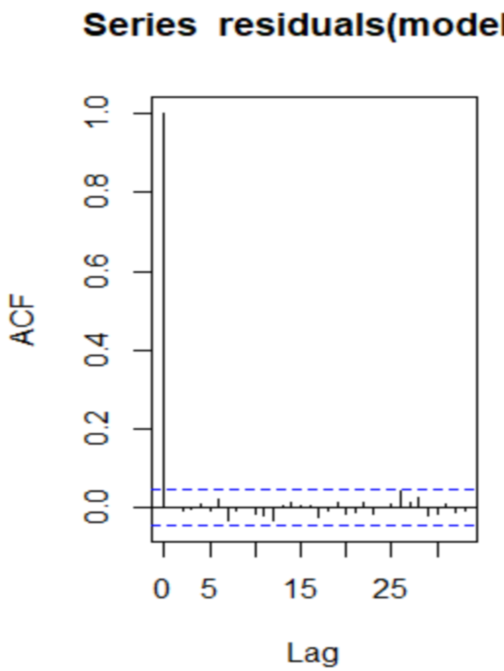
A base de dados está armazenada na variável série

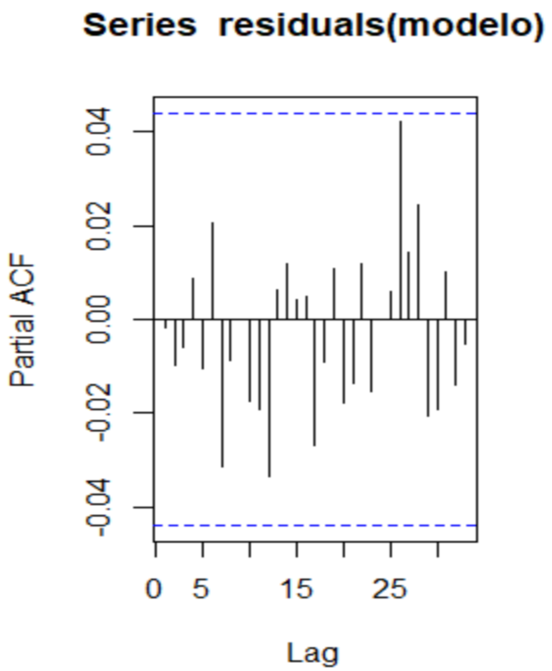
O valor 1 sinaliza que será ajustado um modelo MA(1)

Case *Serie 14* : Ajuste do modelo no R


```
z test of coefficients:

    Estimate Std. Error z value                  Pr(>|z|)
ma1 0.408389   0.020295  20.122 < 0.00000000000000022 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```





Equação do Modelo:

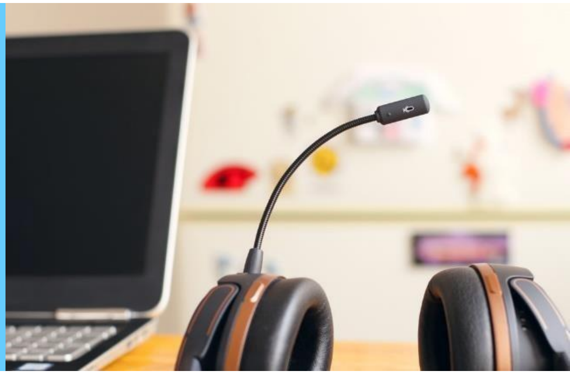
$$\hat{Y}_t = 0{,}4084 * \varepsilon_{t-1}$$

3. Exercícios para fixação

Exercícios para Fixação

Bases de dados:

- Série 13
- Série 22
- Série 33



(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 MA.

(e) Ajuste um modelo Médias Móveis MA 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?

Referências

LIVROS-TEXTO | SÉRIES TEMPORAIS

- 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 Blucher.
- Morettin, P.A. (2008). *Econometria financeira: um curso em séries temporais financeiras*. São Paulo: Edgard Blücher.