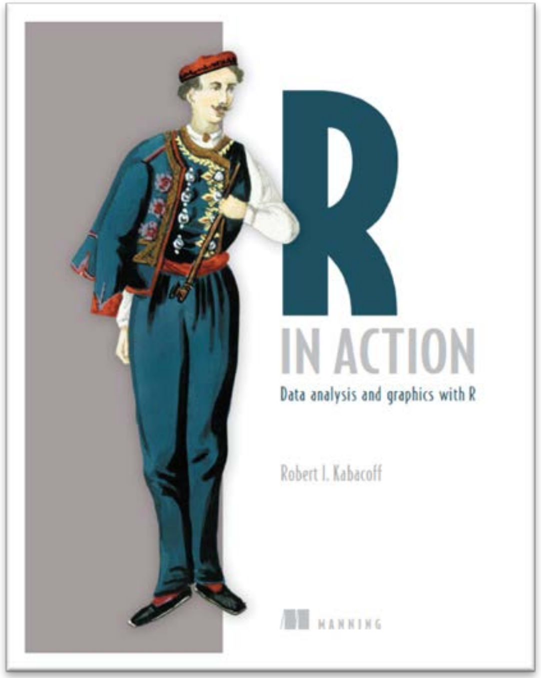
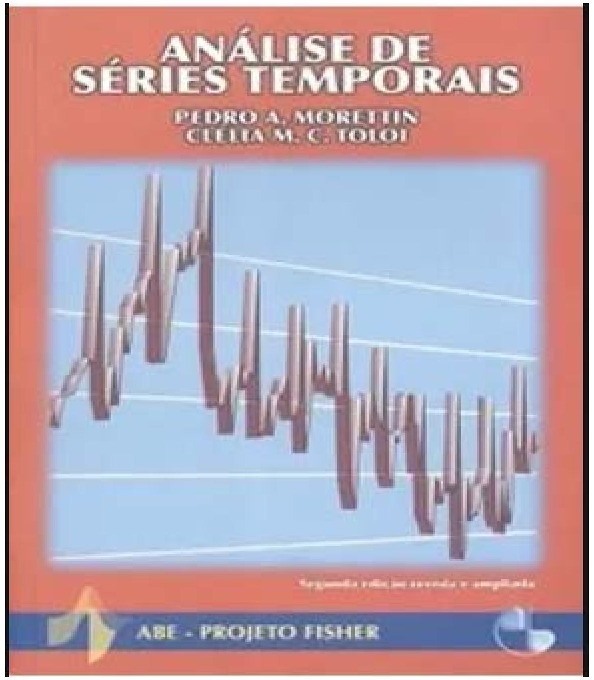
Uma Introdução à Análise de Séries Temporais com R

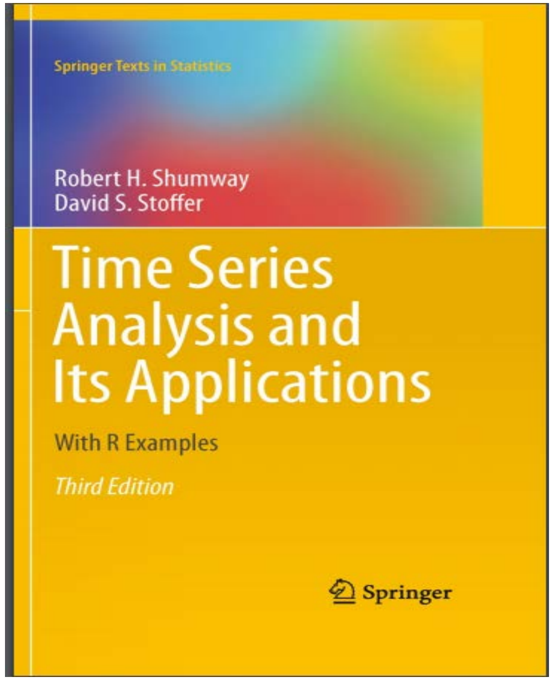
Dr. Everaldo Freitas Guedes

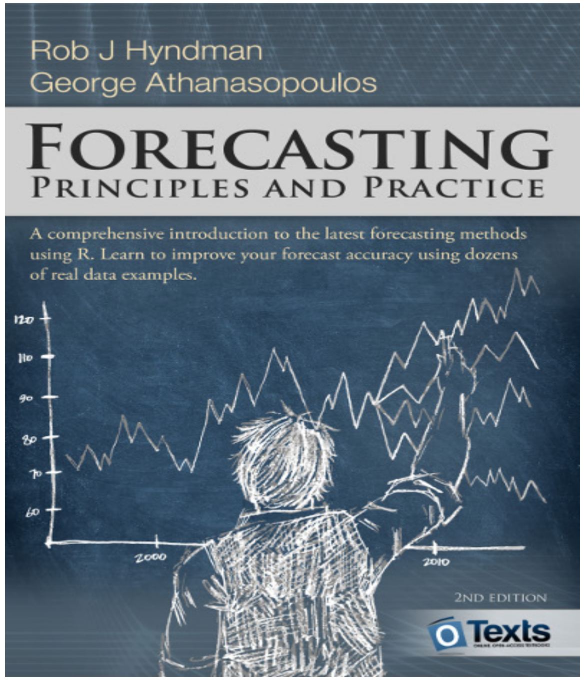
Salvador, 20 de novembro de 2019

Referências









Objetivos

- ✓ Introduzir os conceitos da análise de séries temporais;
- ✓ Apresentar de modelos clássicos de predição de séries temporais;
- ✓ Utilizar do software R para o ajuste e simulação de alguns modelos de séries temporais.

1. Introdução

1.1 O que é análise de séries temporais?

A análise de séries temporais é um ramo da ciência estatística dedicada ao tratamento analítico de uma séries de instâncias ou observações dependentes.

1. Introdução

1.2 Noções

Uma série temporal é definida como qualquer conjunto de observações ordenadas no tempo. Entretanto, pode ser substituída por uma outra, a exemplo de espaço e profundidade.

Matematicamente, uma série temporal é definida pelos valores $Y_1, \ldots, Y_N$ de uma variável Y, nos tempos $t_1, \ldots, t_N$. Portanto, Y é uma função simbolizada $Y = f(t)$.

1. Introdução

1.3 Classificação

- ✓ **Discreta**: quando *T* (conjunto de índices) é um conjunto finito de pontos, $t = 1, 2, \ldots, n$.
- ✓ **Contínua**: quando *T* é um intervalo finito $T = [t\colon t_1 < t < t_2]$. Por exemplo, o registro da maré no Porto de Salvador durante 1 ano $T = [0, 23]$ se a unidade de tempo é a hora.

1. Introdução

1.3 Classificação

- ✓ **Determinística**: quando uma função matemática pode ser utilizada para estabelecer exatamente os valores futuros da série.
- ✓ **Estocástica:** quando os valores futuros da série somente podem ser estabelecidos em termos probabilísticos, pois o modelo compõe-se também de um termo aleatório.

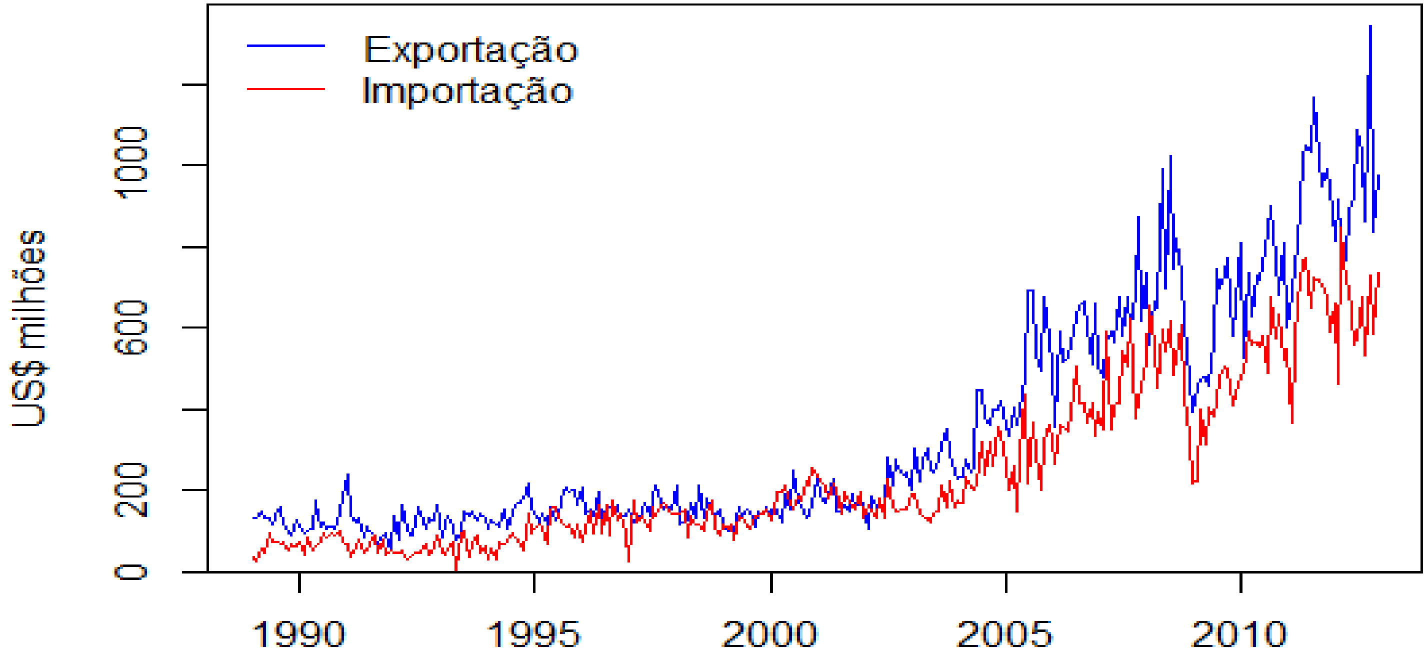
1. Introdução

1.4 Aplicações de séries temporais

Economia -> preços diários de ações; taxa mensal de desemprego.
Medicina -> eletrocardiograma, eletroencefalograma.
Epidemiologia -> número mensal de novos casos de dengue.
Meteorologia -> precipitação pluviométrica.
Climatologia -> humidade relativa do ar, temperatura.

1. Introdução

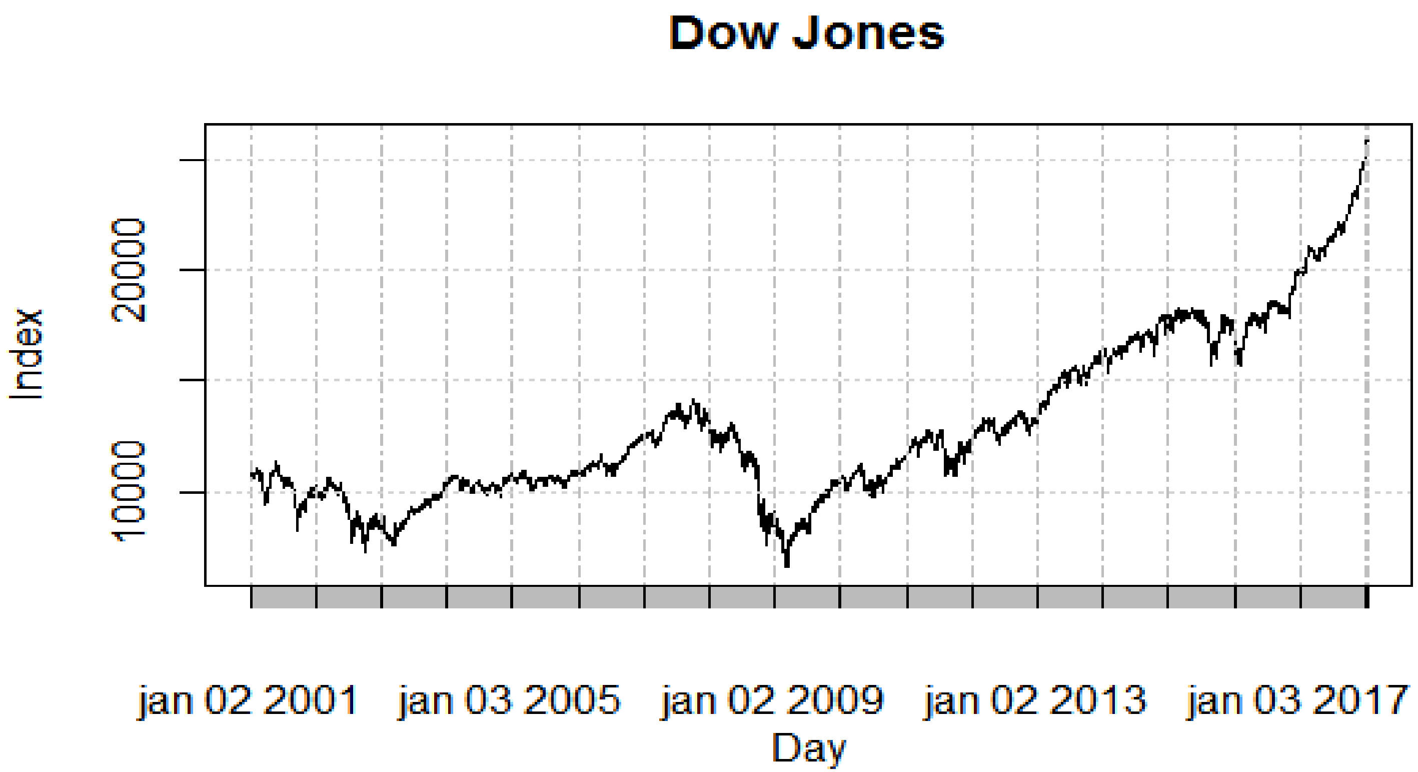
Exemplos:



Bahia: Evolução mensal das exportações e importações (Jan/1989 a Dez/2013).
Fonte: MDIC. Nota: Elaborado pelo autor.

1. Introdução

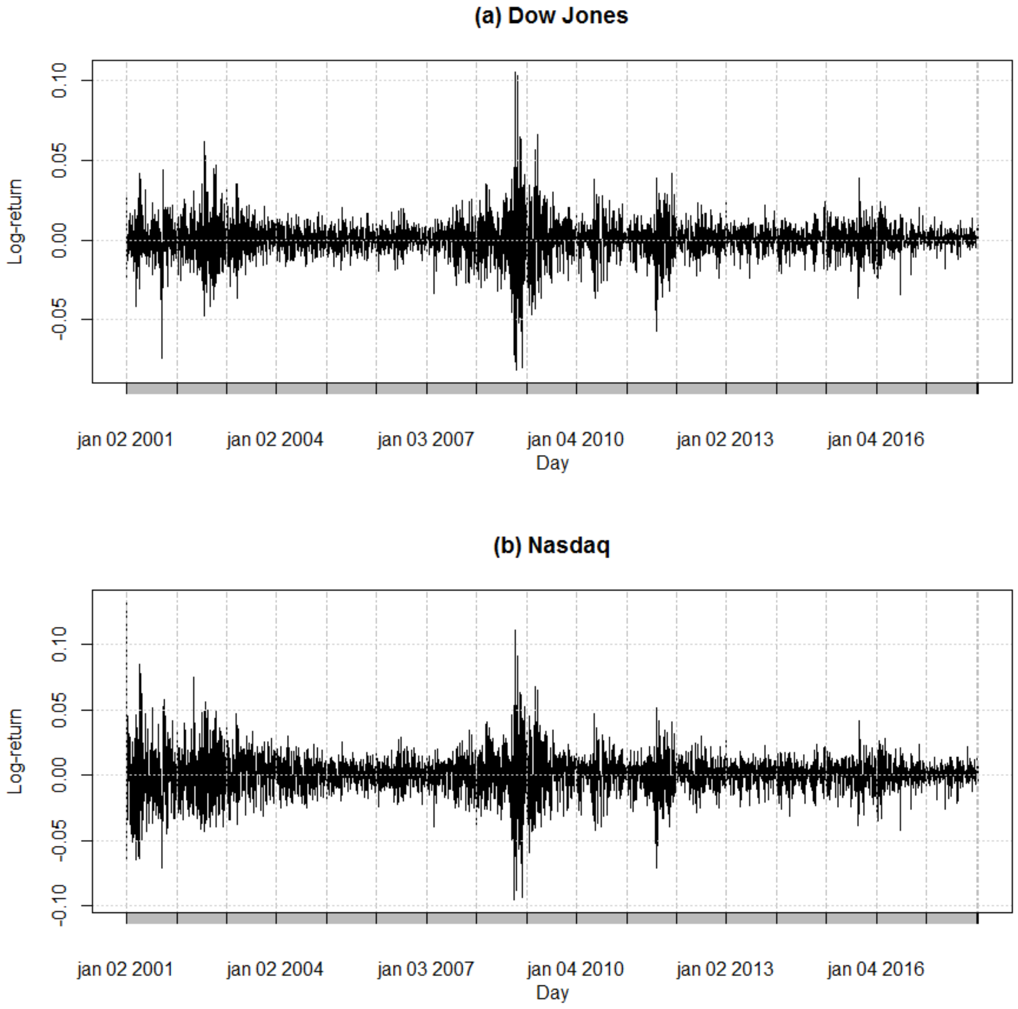
Exemplos:



Índice de Fechamento do Dow Jones.
Nota: Elaborado pelo autor.

1. Introdução

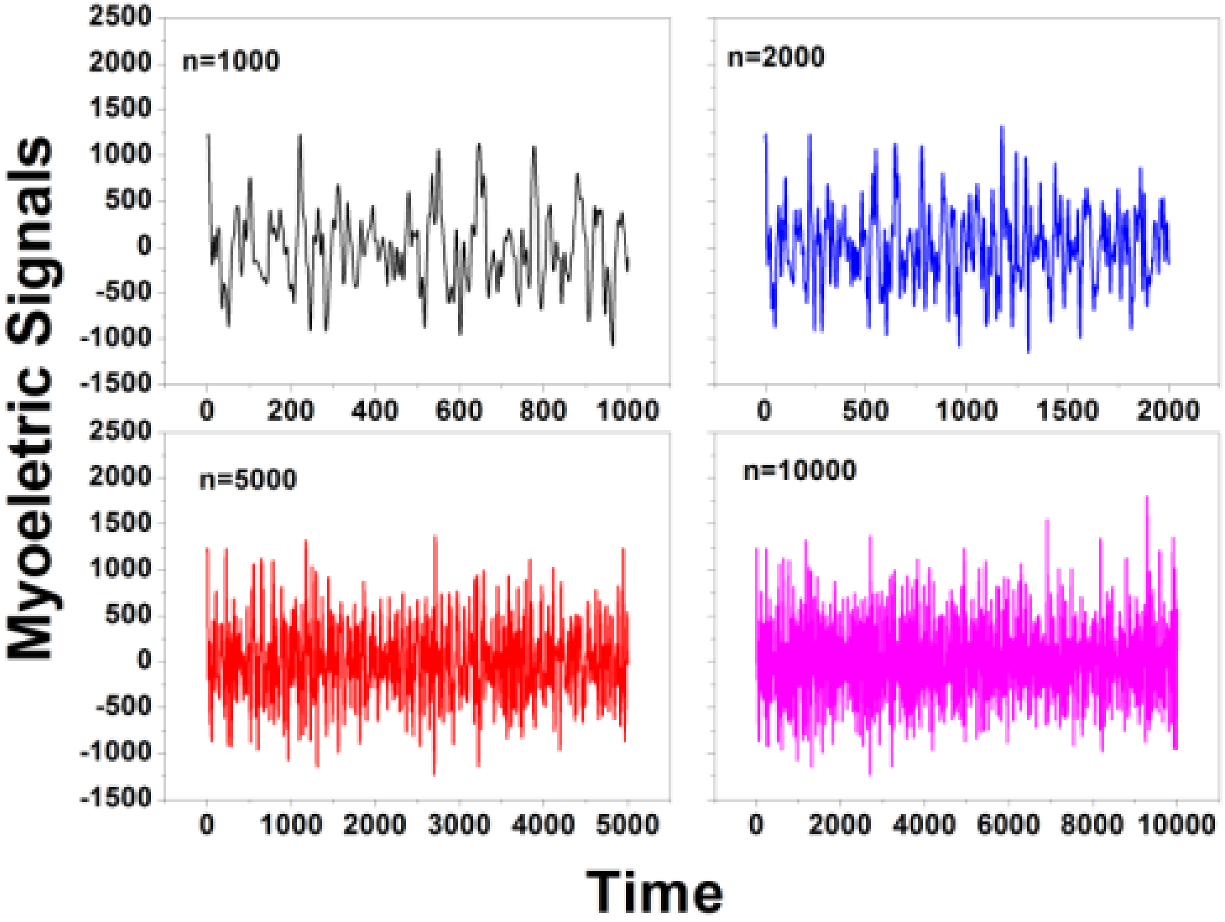
Exemplos:



Retorno dos mercados Dow Jones e Nasdaq.
Nota: Elaborado pelo autor.

1. Introdução

Exemplos:



Sinais Mioelétricos do movimento de abrir e fechar a mão.
Nota: Elaborado pelo autor.

1. Introdução

1.5 Objetivos

- Em linhas gerais, pode-se dizer que o objetivo global do estudo de séries temporais é sumarizar as propriedades da série e caracterizar seu comportamento, identificando ou sugerindo um modelo adequado.

1. Introdução

1.5 Objetivos

- ✓ **Explicação**: usar a variação de uma série para explicar a variação de outra série.
- ✓ **Predição**: predizer valores futuros com base em valores passados. Esta poderá ser de curto prazo (1 a 3 meses), médio prazo (3 a 24 meses) ou longo prazo (acima de 24 meses).

1. Introdução

1.5 Objetivos

- ✓ **Controle de processos**: controle estatístico de qualidade. Controlar a qualidade de uma série temporal é importante por permitir ajustar o modelo à série de dados, possibilitando tomar medidas corretivas nas séries para evitar que a qualidade se afaste de um nível estabelecido.
- ✓ **Descrição**: propriedades da série, a exemplo de tendência, sazonalidade, *outliers*, alterações estruturais e etc.

1. Introdução

1.6 Componentes

- ✓ Os movimentos característicos das séries de tempo podem ser classificados em três tipos principais, frequentemente denominados componentes.

$$Y_t = T_t + S_t + A_t \ ou \ Y_t = T_t * S_t * A_t$$

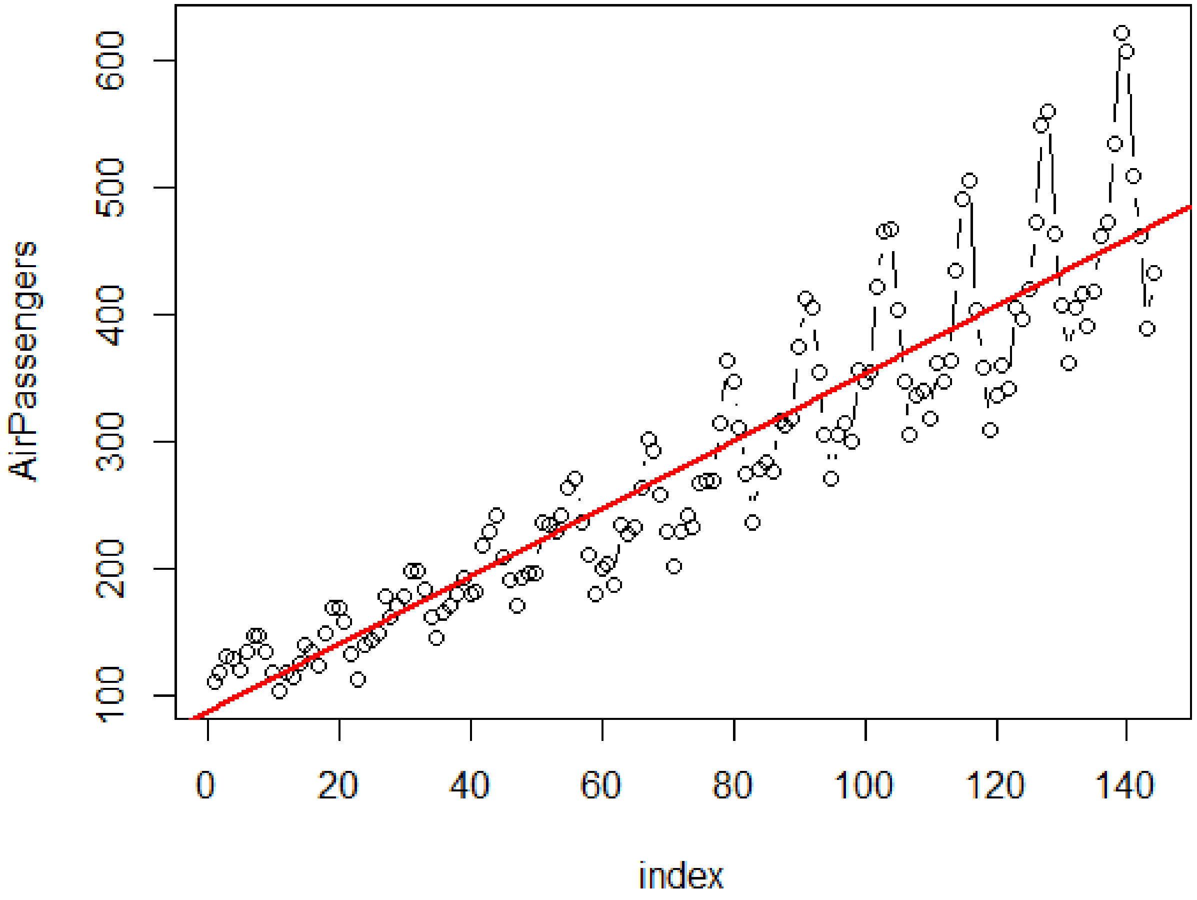
1. Introdução

1.6 Componentes

- ✓ **Tendência**: Representa a componente macro de uma série; é a indicadora da direção global dos dados (ou o movimento geral da variável). Esse movimento pode ser de crescimento/decréscimo linear ou não-linear.

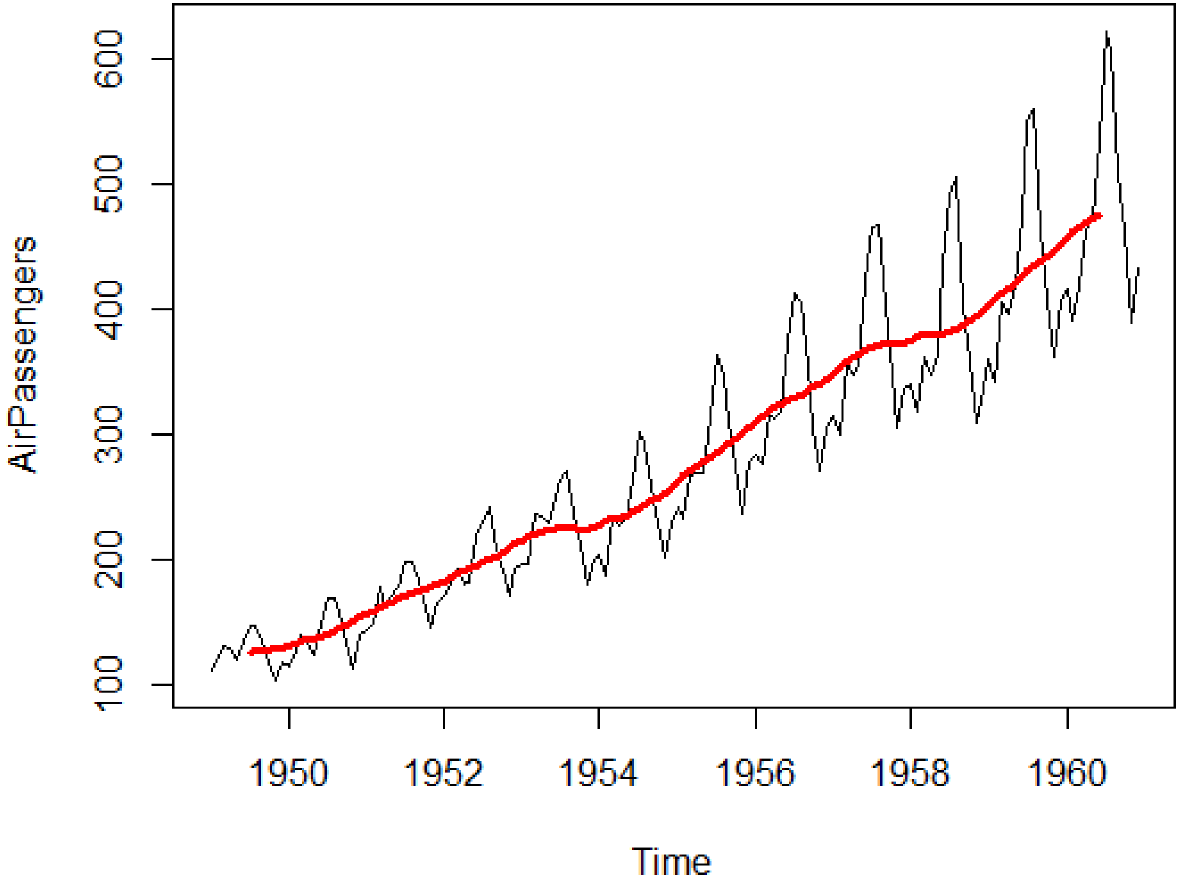
1. Introdução

Exemplo:



1. Introdução

Exemplo:



1. Introdução

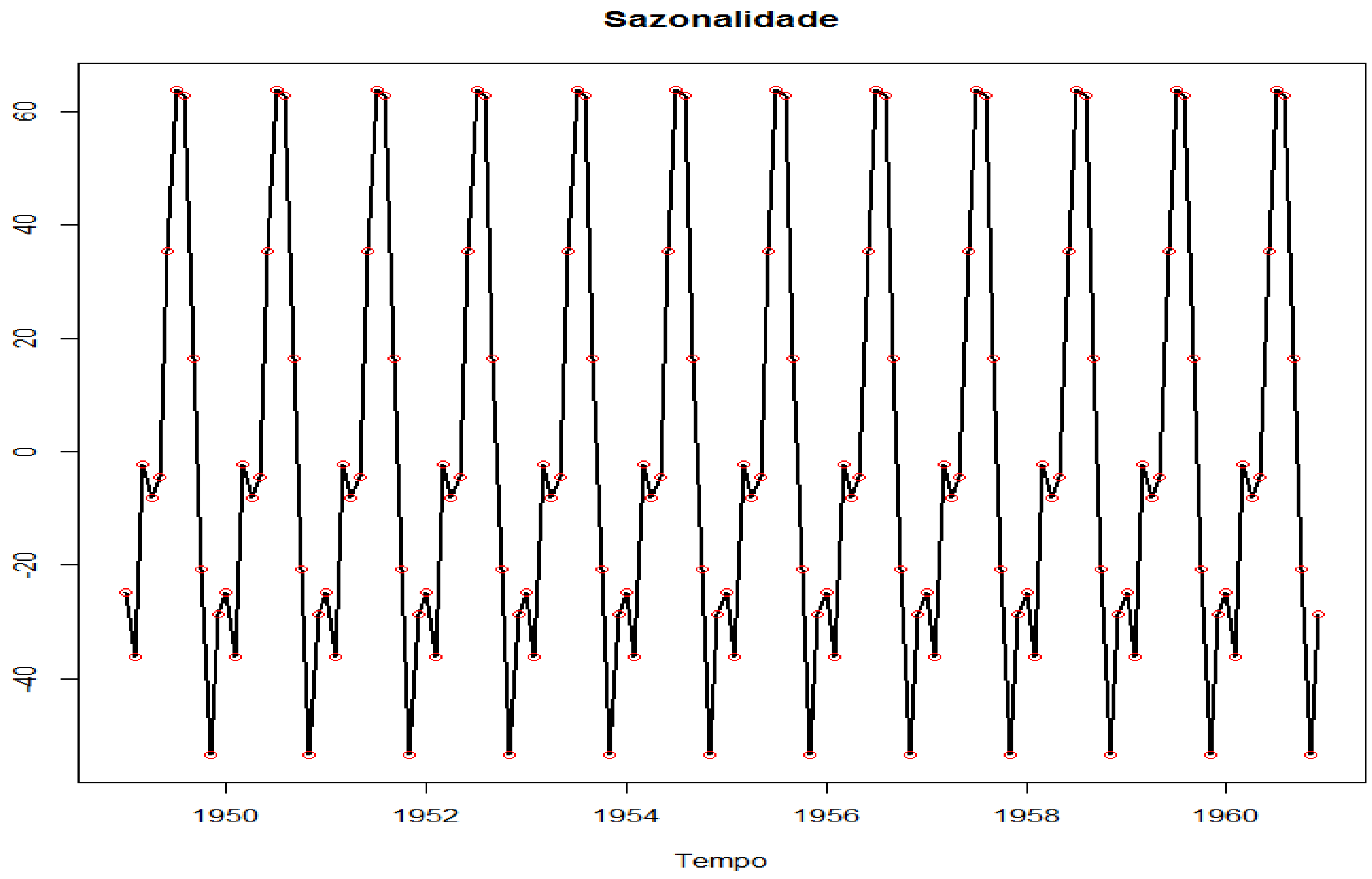
1.6 Componentes

- **Sazonal**: Representa padrões idênticos, ou quase, que uma série temporal parece obedecer durante uma determinada época do ano. Esse movimento refere-se aos ciclos de curto prazo em torno da tendência.

- Ex.: Refere-se a eventos ligados às estações do ano, vinculados ao calendário e repetidos a cada doze meses, relacionado às causas climáticas, ciclos vegetativos, usos e costumes, festas sociais e religiosas.

1. Introdução

Exemplo:



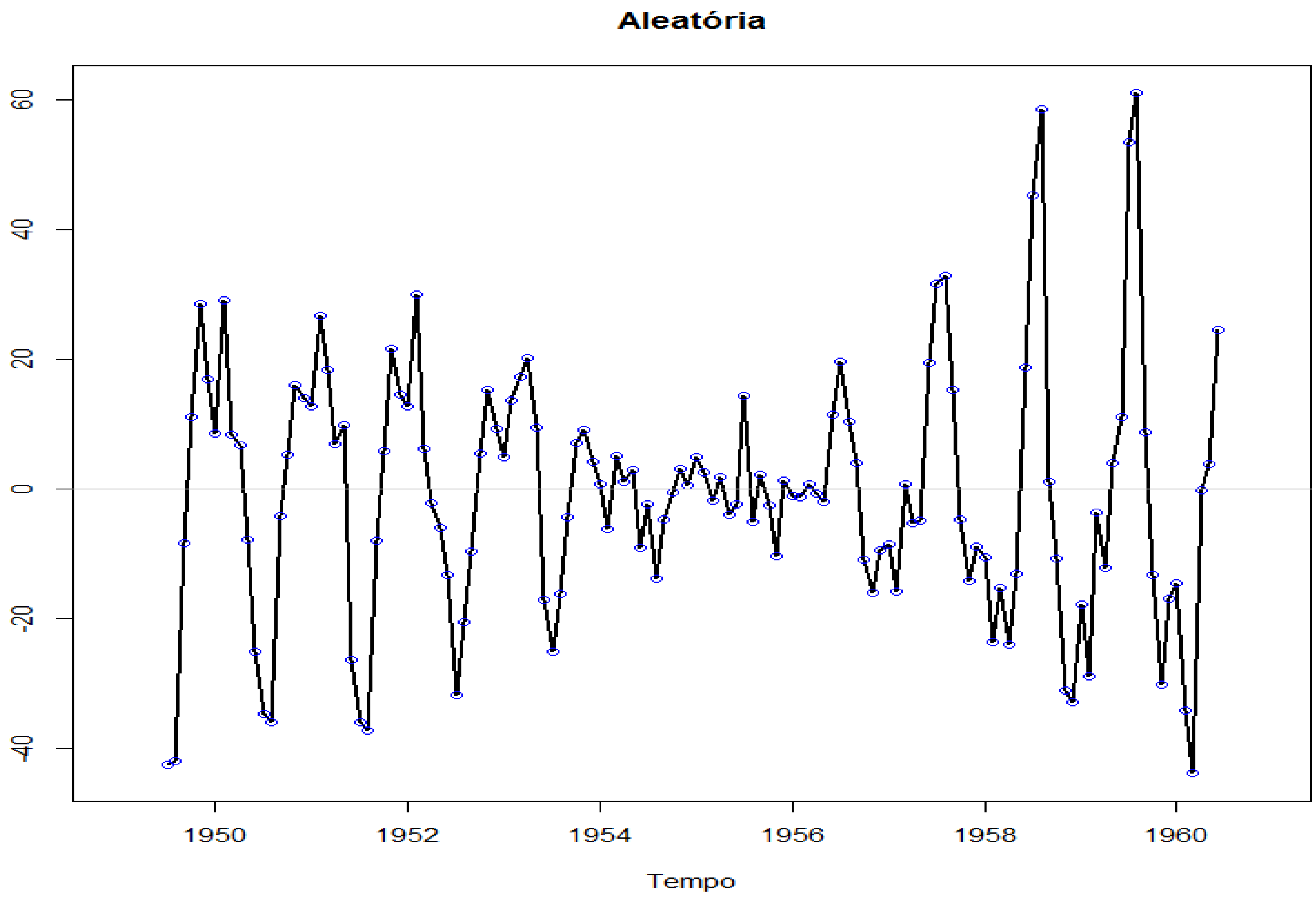
1. Introdução

1.6 Componentes

- ✓ **Aleatória**: Refere-se aos deslocamentos esporádicos das séries temporais, provocados por eventos casuais. A importância dessa componente vincula-se ao poder de alterar tanto a direção da tendência quanto a amplitude dos ciclos existentes.

1. Introdução

Exemplo:



1. Introdução

1.6 Componentes

- ✓ A análise de séries temporais consiste em uma modelagem dessas componentes. Nesse sentido, a maneira clássica de decomposição baseia-se nos seguintes modelos:

1. Introdução

Exercício 1

1) Realizar a estimação da tendência através das técnicas de regressão linear simples e de média móvel das seguintes séries:

    a) AirPassengers
    b) co2
    c) Nile
    d) nottem
    e) ausbeer #library(fpp)
    f) visitors #library(fpp)
    g) JohnsonJohnson

1.7 Decomposição de séries temporais

Muitas das propriedades observadas em uma série temporal Y_t podem ser captadas assumindo-se as seguintes formas de decomposição:

$$Y_t = T_t + S_t + A_t \ ou \ Y_t = T_t * S_t * A_t$$

Tipos:

- ✓ Clássica;
- ✓ STL (Seasonal Trend Decomposition Loess)

1. Introdução

Exercício 2

1) Realizar a decomposição clássica das seguintes séries:

   a) AirPassengers
   b) co2
   c) Nile
   d) nottem
   e) ausbeer #library(fpp)
   f) visitors #library(fpp)
   g) JohnsonJohnson

O que é possível concluir?

Processos estocásticos

2. Processos estocásticos

2.1 Conceitos básicos

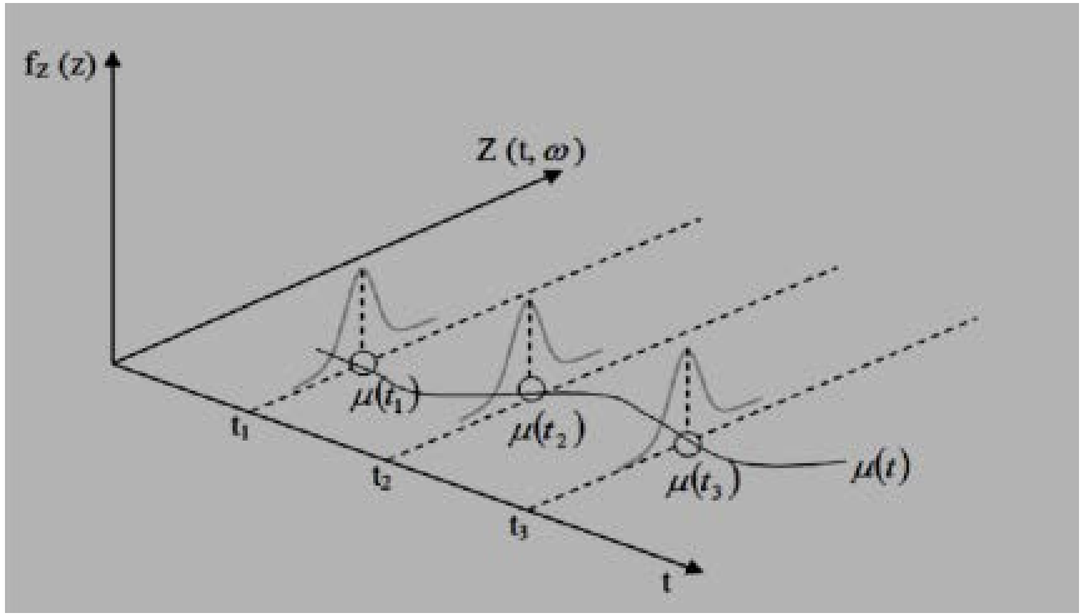
Definições:

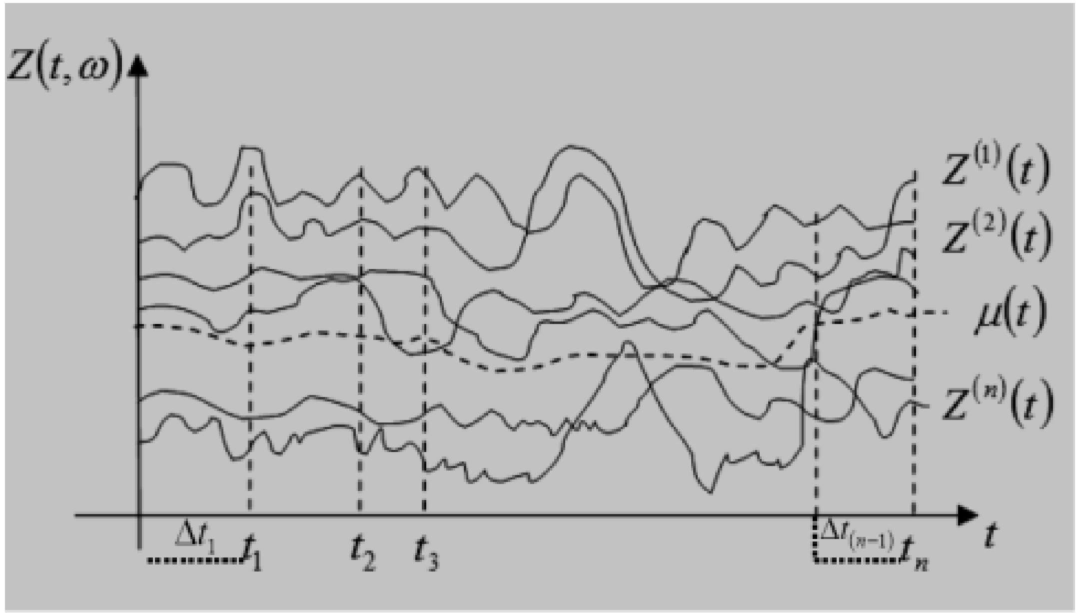
“é uma família de variáveis aleatórias representando a evolução de um sistema de valores com o tempo”.

“é o conjunto de todas as possíveis trajetórias de um fenômeno ou uma coleção de realizações do processo físico”.

2. Processos estocásticos

Ilustração:





2. Processos estocásticos

2.1 Conceitos básicos

Usualmente, descreve-se um processo estocástico através de suas funções média, autocovariância e variância.

$$E[Y(t)] = \mu_t$$

$$\gamma_{(t1,t2)} = E[(\mathrm{Y}(t_1) - \mu_{t1})][(\mathrm{Y}(t_2) - \mu_{t2})]$$

$$\sigma^2 = E[Y^2(t) - \mu^2{}_t]$$

2. Processos estocásticos

2.2 Estacionários

- ✓ Quando é invariante no tempo (estrita ou fraca).
- ✓ Um P.E. é estritamente estacionário se a média e a variância forem constantes ao longo do tempo.
- ✓ Um P.E. será considerado fracamente estacionário se e somente se

$$\mathrm{E}\{\mathrm{X}(\mathrm{t})\} = \mu(\mathrm{t}) = \mu, \text{ constante, para todo t} \in \mathrm{T};$$

$$\mathrm{E}\{X^2(\mathrm{t})\} < \infty \text{ , para todo t} \in \mathrm{T};$$

$$\gamma(t_1, t_2) = \mathrm{CovX}(t_1), \mathrm{X}(t_2) \text{ é uma função de } |t_1 - t_2|.$$

2. Processos estocásticos

2.2 Estacionários

- ✓ Quando é invariante no tempo. (estrita ou fraca)
- ✓ Um P.E. é estritamente estacionário se as distribuições unidimensionais são invariantes no tempo, ou seja, se a média e a variância forem constantes.
- ✓ Um P.E. será considerado fracamente estacionário se e somente se

$$\mathrm{E}\{\mathrm{X}(\mathrm{t})\} = \mu(\mathrm{t}) = \mu, \text{ constante, para todo t} \in \mathrm{T};$$

$$\mathrm{E}\{X^2(\mathrm{t})\} < \infty \text{ , para todo t} \in \mathrm{T};$$

$$\gamma(t_1, t_2) = \mathrm{Cov}\mathrm{X}(t_1), \mathrm{X}(t_2) \text{ é uma função de } |t_1 - t_2|.$$

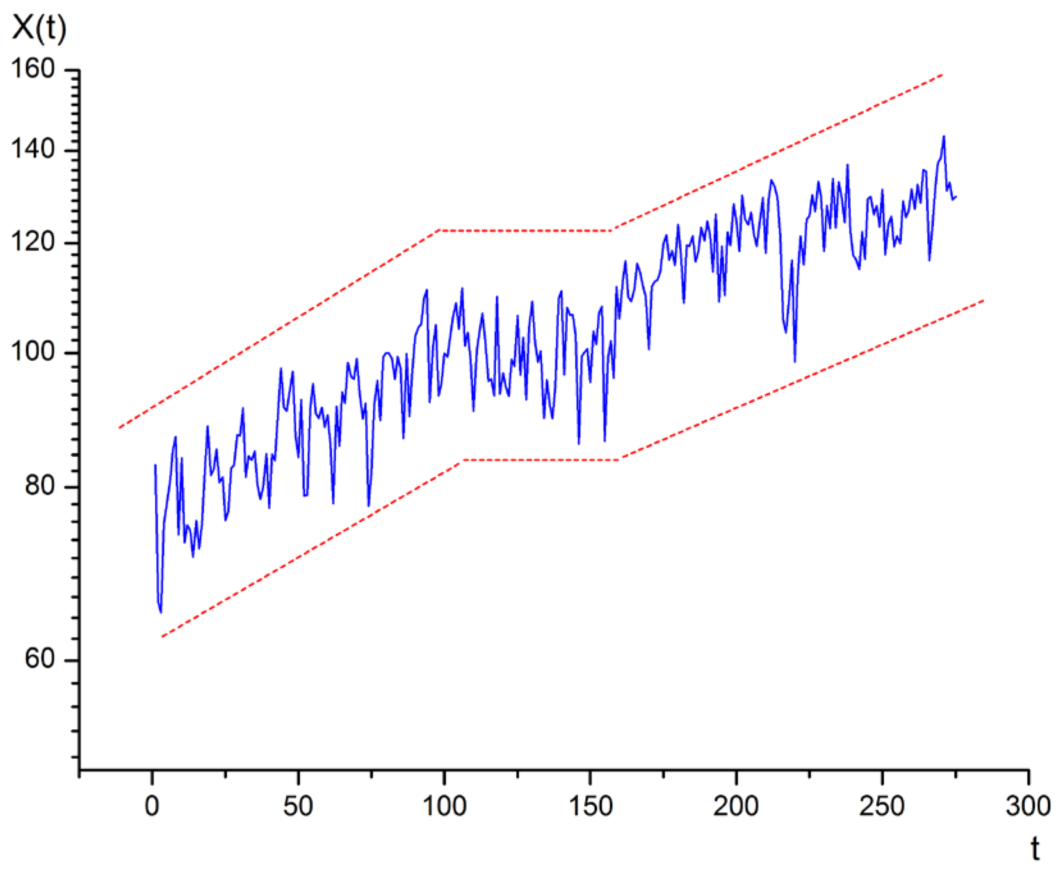
2. Processos estocásticos

2.2 Estacionários

Obs.: uma série poderá ser estacionaria durante períodos longos ou apenas em períodos breves, mudando de nível e/ou de inclinação.

2. Processos estocásticos

Ilustração:



2. Processos estocásticos

2.2 Estacionários

Obs.: Caso não seja, poderá se transformar em estacionária ao ser diferenciada em uma quantidade *n* finita de vezes.

$$\Delta X(t) = X(t) - X(t-1),$$

$$\Delta^2 X(t) = \Delta[\Delta X(t)] = \Delta[X(t) - X(t-1)],$$

$$\Delta^n X(t) = \Delta[\Delta^{n-1} X(t)].$$

2. Processos estocásticos

2.2 Estacionários

Obs.: No máximo duas diferenças são suficientes para tornar a série temporal estacionária. No entanto, quando as diferenças não são suficientes para alcançar tal suposição, costuma-se aplicar antes das diferenças, uma transformação não linear nos dados, como por exemplo, a transformação logarítmica.

2. Processos estocásticos

Exercício 3

1) Realizar a 1ª e 2ª diferença das séries abaixo e plotar as séries originais:

   a) AirPassengers
   b) co2
   c) Nile
   d) nottem
   e) ausbeer #library(fpp)
   f) visitors #library(fpp)
   g) JohnsonJohnson

2. Processos estocásticos

2.3 Função de autocorrelação

A FAC quantifica a interdependência de uma série temporal.

$$\hat{\rho_{(k)}} = \frac{\hat{\gamma_{(k)}}}{\hat{\gamma_{(0)}}} = \frac{\sum_{t=1}^{n-k}(Y_t - \bar{Y_t})(Y_{(t+k)} - \bar{Y_t})}{\sum_{t=1}^{n}(Y_t - \bar{Y_t})^2}$$

$\hat{\gamma_{(k)}}$ é a covariância na defasagem $k(k = 0, 1, 2, \ldots)$; $\hat{\gamma_{(0)}}$ é a variância amostral,

2. Processos estocásticos

2.3 Função de autocorrelação

Como a covariância e a variância apresentam as mesmas unidades de medida, os valores de $\rho_{(k)}$são adimensionais e variam entre -1 a 1, de modo que:

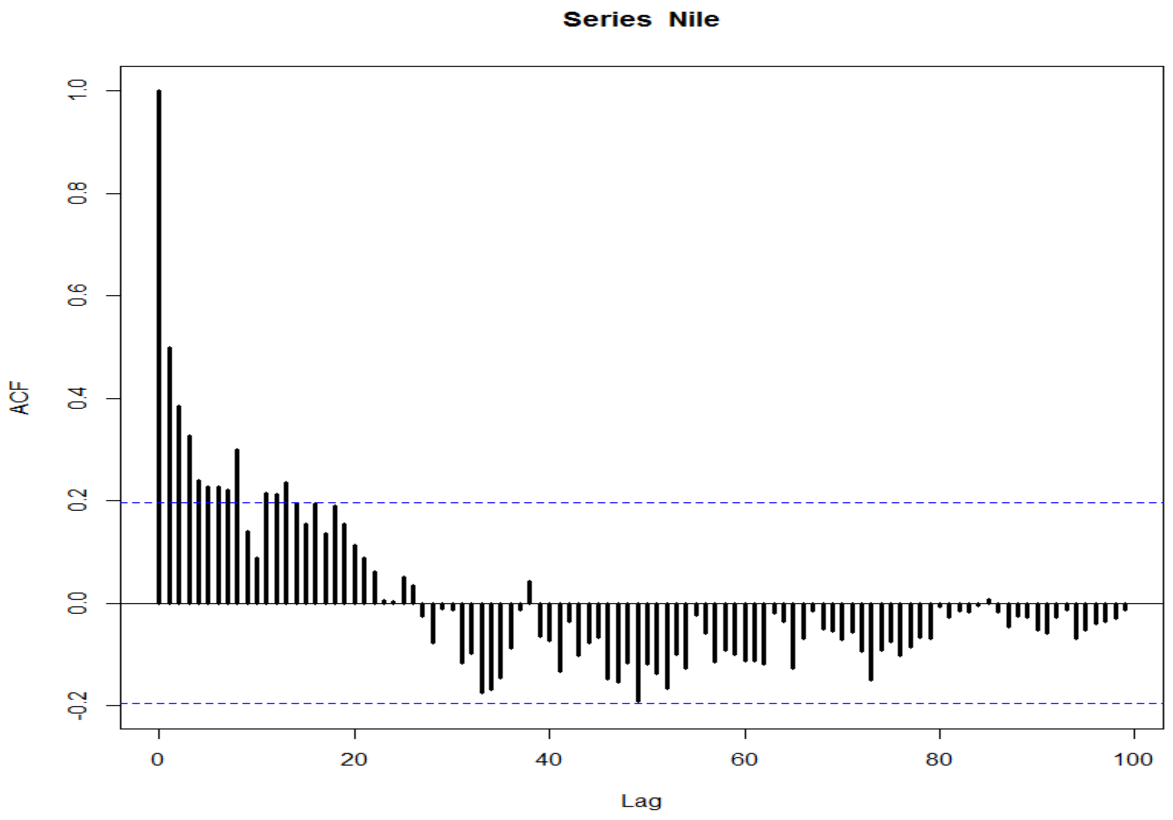
$0 < \hat{\rho_{(k)}} \leq 1$ a série possui autocorrelação positiva;

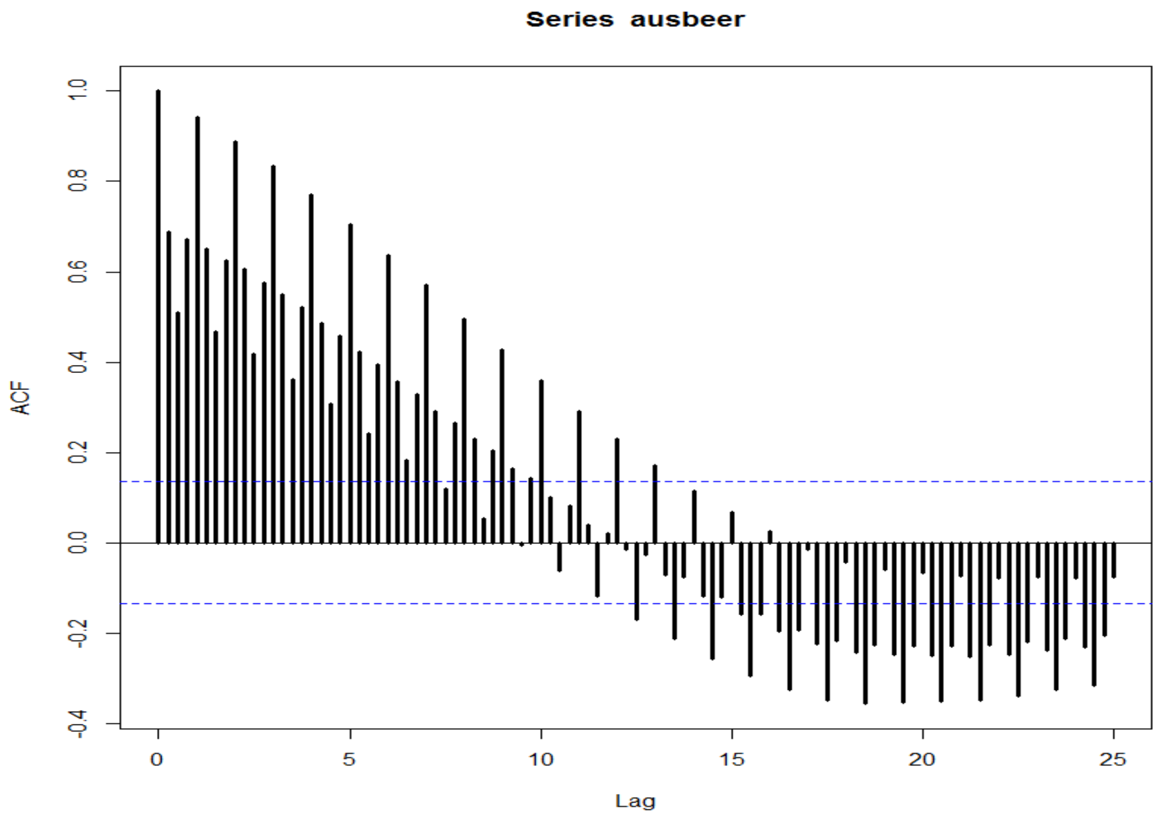
$\hat{\rho_{(k)}} = 0$ não existe autocorrelação na série; e,

-1 $\leq \hat{\rho_{(k)}} < 0$ a série possui autocorrelação negativa;

2. Processos estocásticos

Ilustração:





2. Processos estocásticos

2.3 Função de autocorrelação

A partir do correlograma é possível caracterizar uma séria temporal em:

a. Estacionária
b. Não estacionária;
c. Periódica;
d. Ruído branco;

2. Processos estocásticos

Exercício 4

1) Realizar a 1ª e 2ª diferença das séries abaixo. Plotar as séries originais e as funções de autocorrelação.

   a) AirPassengers
   b) co2
   c) Nile
   d) nottem
   e) ausbeer #library(fpp)
   f) visitors #library(fpp)
   g) JohnsonJohnson

O que é possível concluir?

2. Processos estocásticos

2.5 Tipos

um processo estocástico ε_t é especificado como **ruído branco** ou sequência aleatória se

$$\mathrm{E}[\varepsilon_t] = 0, \text{ constante, para todo t} \in \mathrm{T};$$

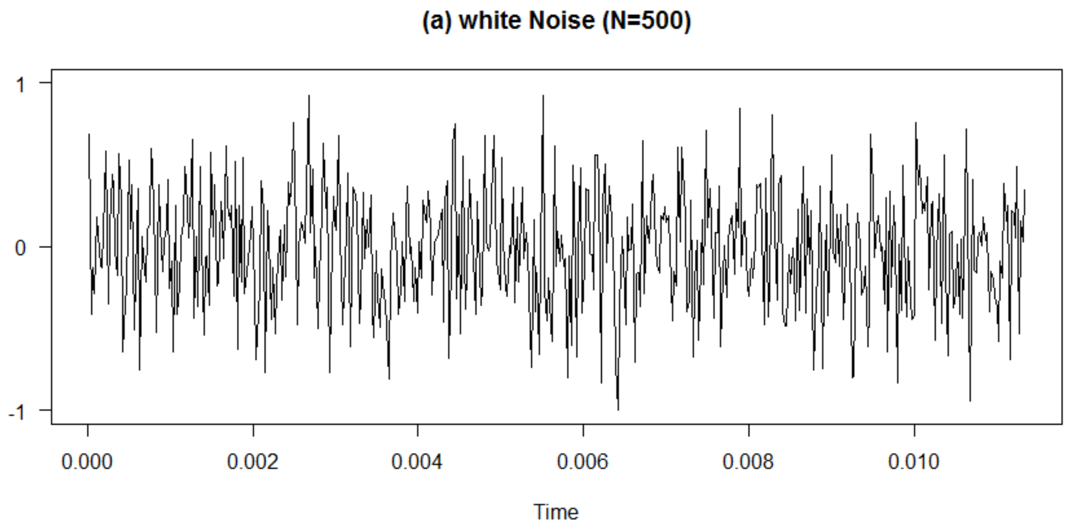
$$\mathrm{Var}[\varepsilon_t] = \sigma_\varepsilon^2, \text{ para todo t} \in \mathrm{T};$$

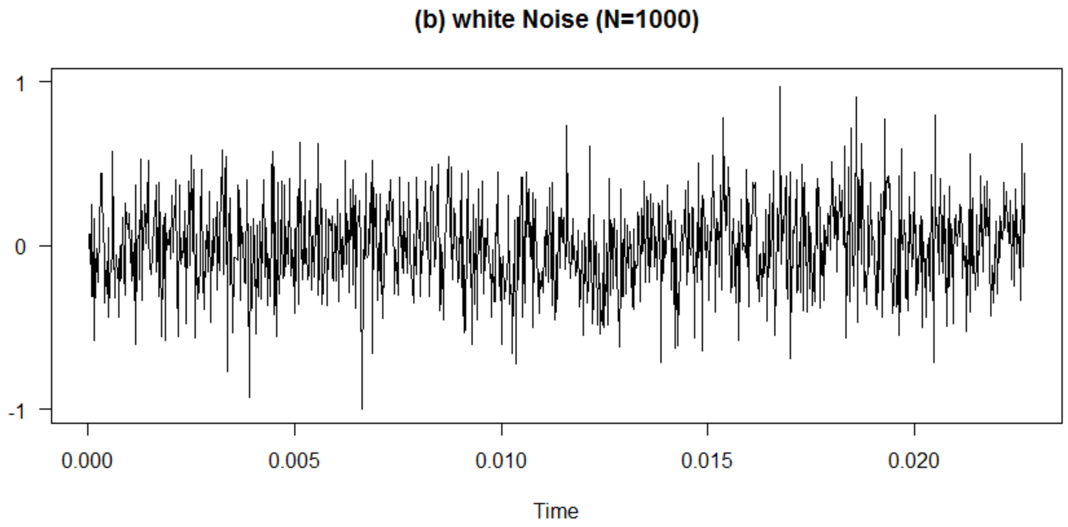
$$\gamma(\varepsilon_{t_1}, \varepsilon_{t_2}) = \mathrm{Cov}[\varepsilon_{t_1}, \varepsilon_{t_2}] = 0, \text{ para todo } t_1 \neq t_2.$$

Obs.: o RB é homogêneo, estacionário e sem dependência temporal. Além disso, possui distribuição normal com média zero, variância e covariâncias nulas.

2. Processos estocásticos

Ilustração:





2. Processos estocásticos

2.5 Tipos

Um processo estocástico é dito **passeio aleatório** se a primeira diferença deste resulta em um RB.

$$\mathrm{E}[Y_t] = \sum_{j=1}^{t} E[\varepsilon_j] = t\mu;$$

$$\mathrm{Var}[Y_t] = \sum_{j=1}^{t} Var[\varepsilon_j] = t\sigma_\varepsilon^2;$$

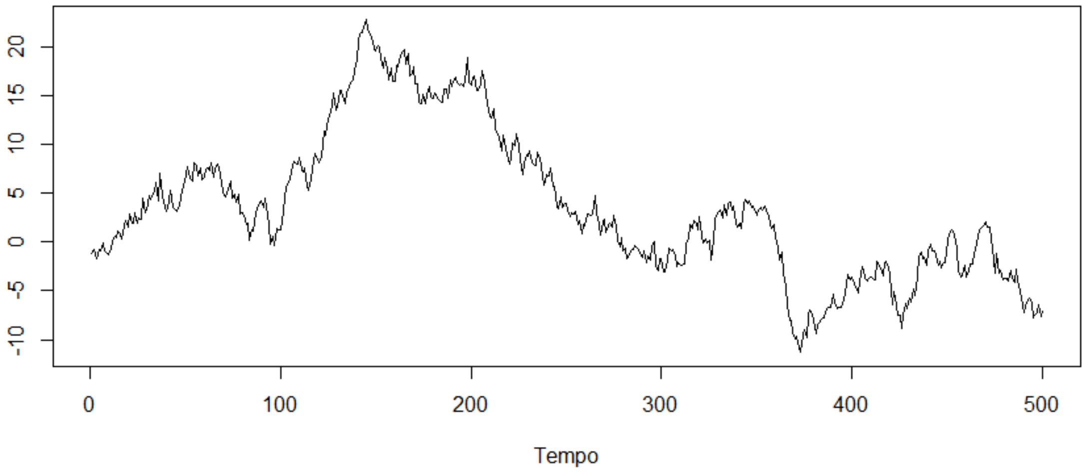
$$Cov(Y_t, Y_{t-k}) = Cov[\varepsilon_1 + \cdots + \varepsilon_{t-k} + \cdots + \varepsilon_t, \varepsilon_1 + \cdots + \varepsilon_{t-k}] = (t-k)\sigma_\varepsilon^2$$

Obs.: Apresenta tendência e, portanto, não é estacionário.

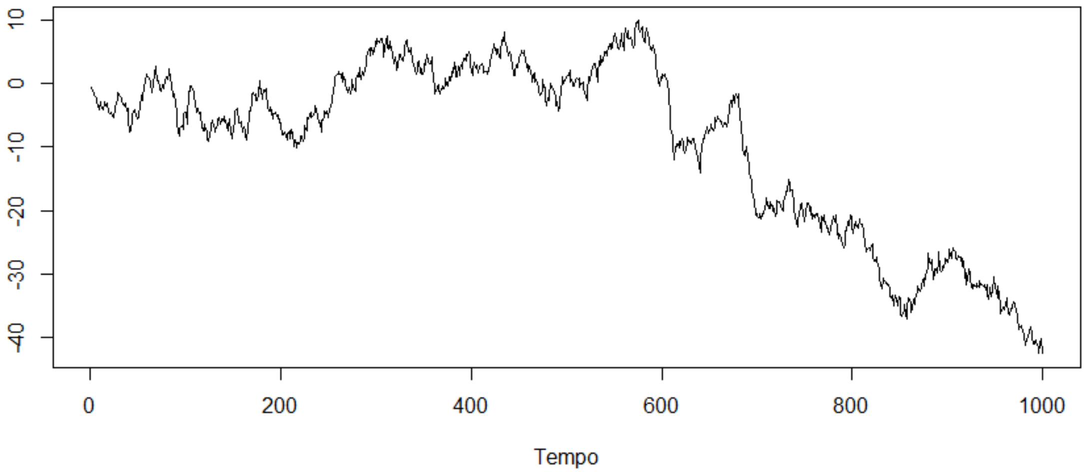
2. Processos estocásticos

Ilustração:

(a) Passeio Aleatório (N=500)



(b) Passeio Aleatório (N=1000)



2. Processos estocásticos

2.5 Tipos

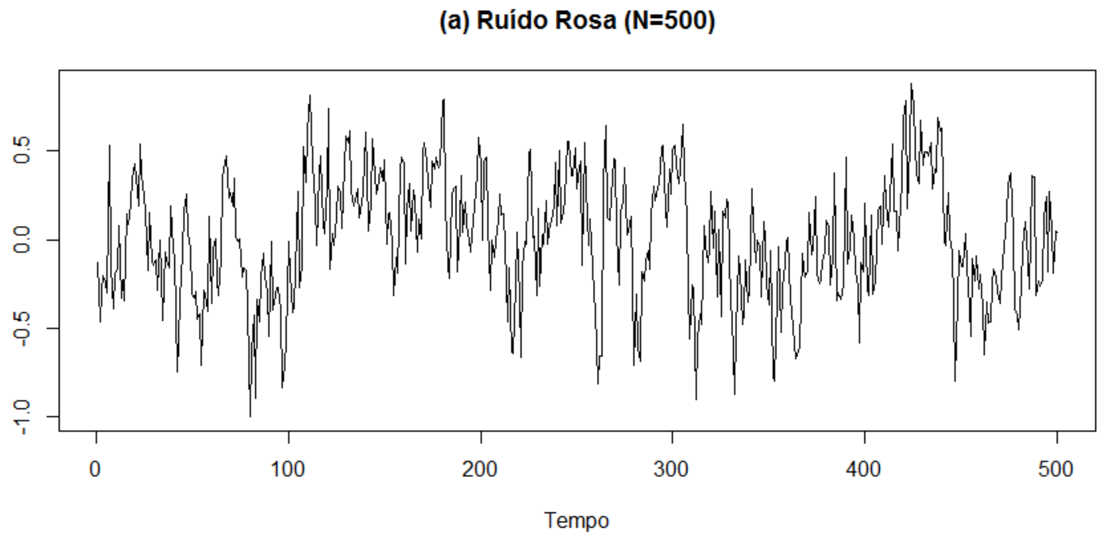
Ruído Rosa é um processo estocástico intermediário entre o ruído branco e passeio aleatório.

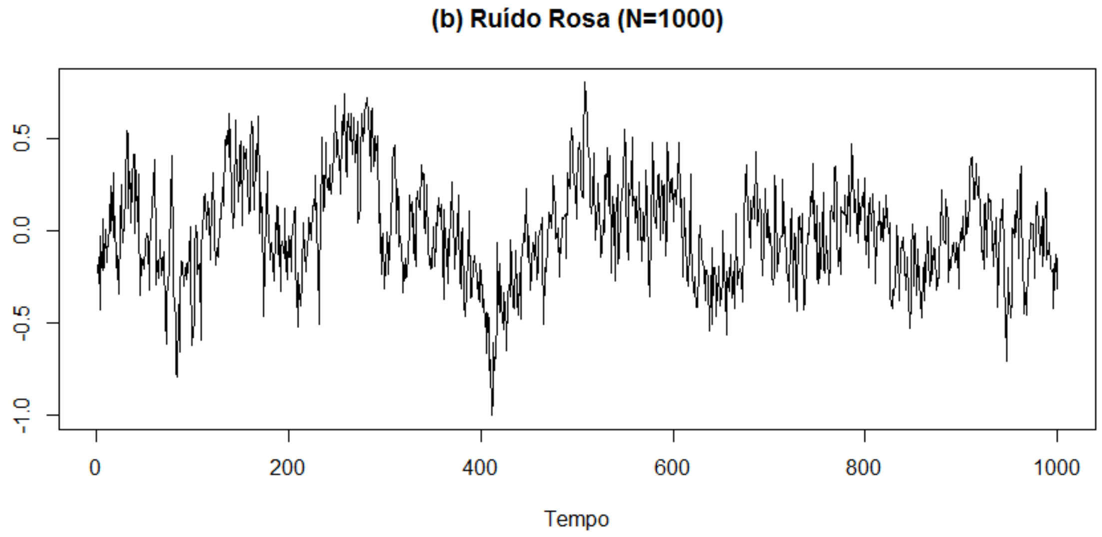
Obs.: A família de processos aleatórios estatisticamente auto-semelhantes ${}^{1}/_{f}$.

Dica: https://label2.tecnico.ulisboa.pt/vilela/Cursos/Beyond.pdf

2. Processos estocásticos

Ilustração:





2. Processos estocásticos

Exercício 5

1) Simular os processos estocásticos RB, PA e RR com os tamanhos 250, 500, 750, 1000.

2) Analisar a função de autocorrelação das séries simuladas. O que é possível concluir?

2. Processos estocásticos

2.5 Tipos

Um processo $\boldsymbol{AR\,(p)}$ é uma representação de um tipo de processo aleatório especifica que a variável de saída depende linearmente de seus próprios valores anteriores e de um termo estocástico.

$$\phi(\mathrm{B})X_t = c + \varepsilon_t$$

em que, $\phi(\mathrm{B}) = 1 - \phi_1\mathrm{B} - \phi_2\mathrm{B}^2 - \cdots - \phi_p\mathrm{B}^p$ representa o operador autorregressivo de ordem *p.*

2. Processos estocásticos

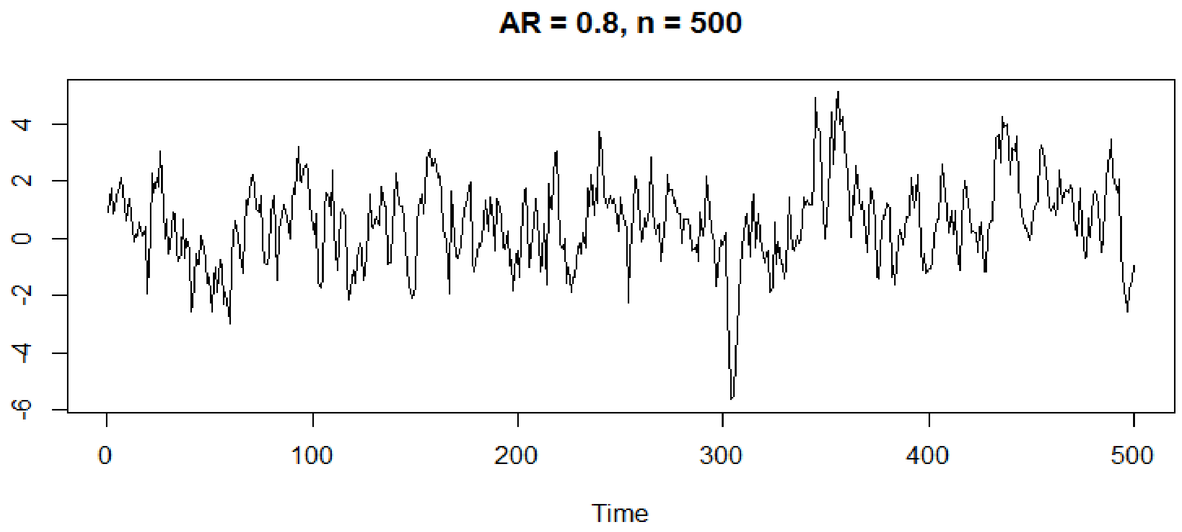
2.5 Tipos

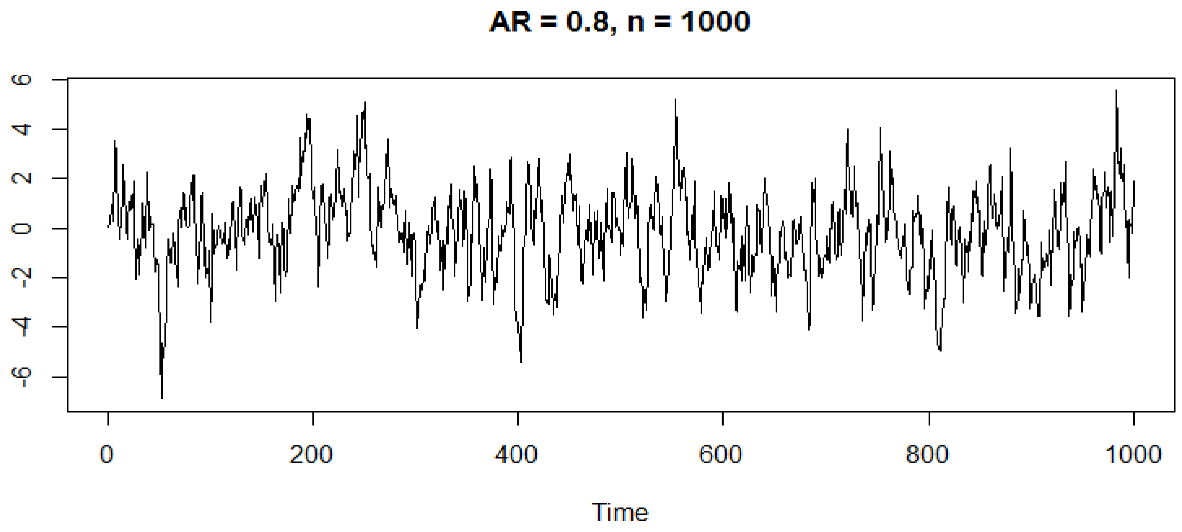
Características:

- ✓ A fac teórica de um processo autorregressivo decai segundo exponenciais e/ou senóides amortecidas;
- ✓ As autocorrelações parciais (facp) nos primeiros lags são estatisticamente diferentes de zero;
- ✓ O processo AR, geralmente, não é estacionário;

2. Processos estocásticos

Ilustração:





2. Processos estocásticos

Exercício 6

1) Simular os processos estocásticos AR com os tamanhos 250, 500, 750, 1000.

2) Analisar as funções de autocorrelação e autocorrelação parcial das séries simuladas. As séries são estacionárias?

2. Processos estocásticos

2.5 Tipos

Um processo $\boldsymbol{MA\,(q)}$ é um tipo de processo aleatório que especifica que a variável de saída depende linearmente de seus próprios valores anteriores e atuais mais um termo estocástico.

$$X_t = \mu + \theta(\mathrm{B})\varepsilon_t$$

em que, $\theta(\mathrm{B}) = 1 - \theta_1\mathrm{B} - \theta_2\mathrm{B}^2 - \cdots - \theta_q\mathrm{B}^q$ representa o operador de médias móveis de ordem *q*.

2. Processos estocásticos

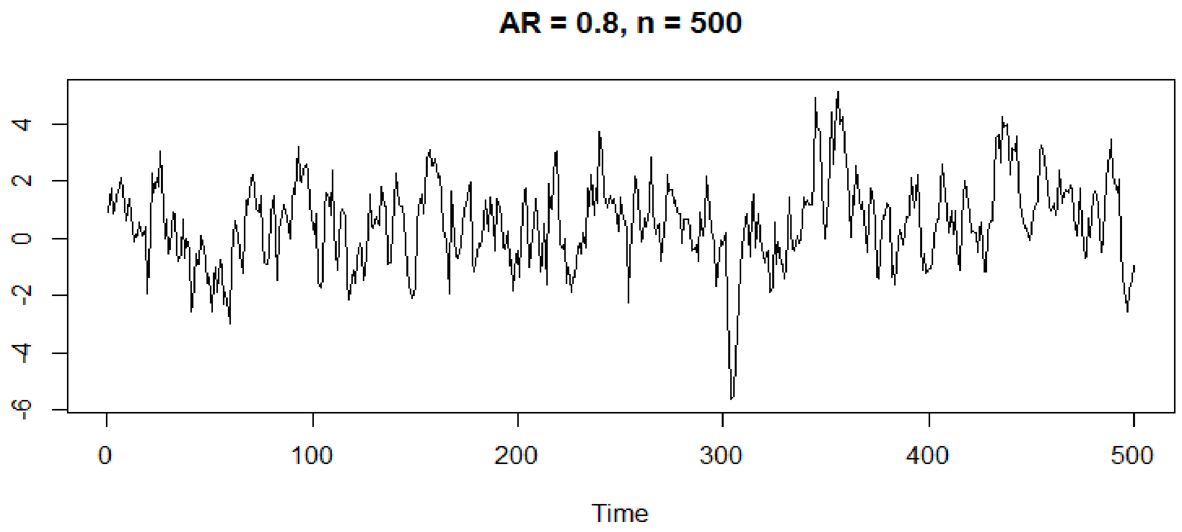
2.5 Tipos: MA

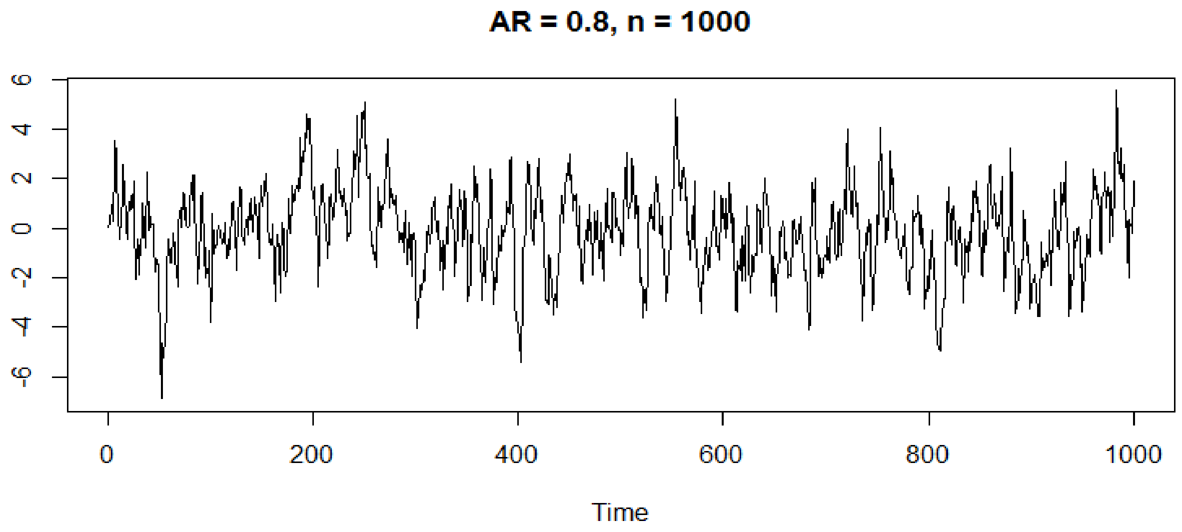
Características:

✓ O processo MA é estacionário;

2. Processos estocásticos

Ilustração:





2. Processos estocásticos

Exercício 7

1) Simular os processos estocásticos MA com os tamanhos 250, 500, 750, 1000.

2) Analisar as funções de autocorrelação e autocorrelação parcial das séries simuladas. As séries são estacionárias?

2. Processos estocásticos

2.5 Tipos

Um processo $\boldsymbol{ARMA\,(p,q)}$ é um processo misto.

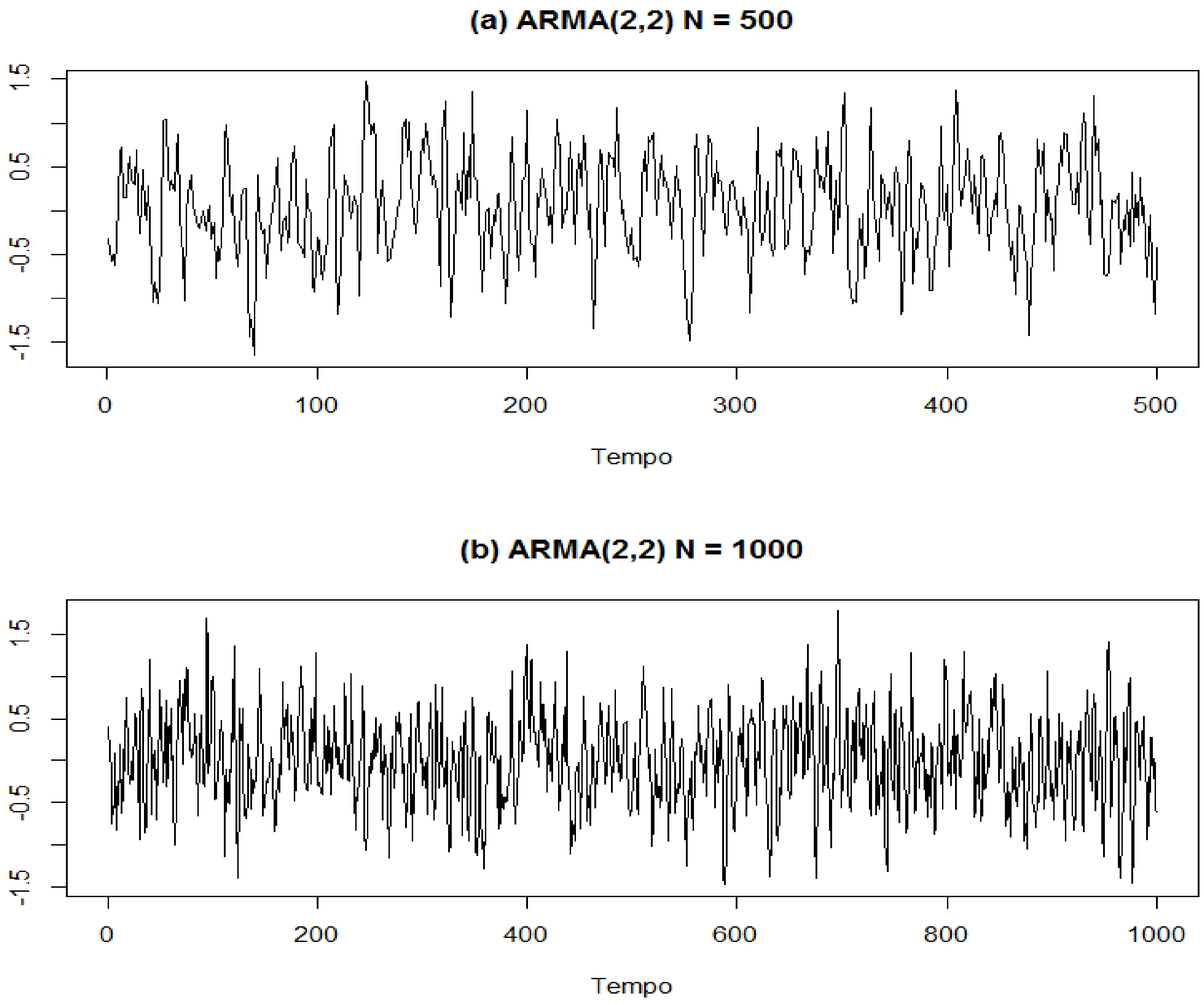
$$\phi(\mathrm{B})X_t = c + \theta(\mathrm{B})\varepsilon_t$$

em que,

- $\phi(\mathrm{B}) = 1 - \phi_1\mathrm{B} - \phi_2\mathrm{B}^2 - \cdots - \phi_p\mathrm{B}^p$ é o operador autorregressivo de ordem *p;*
- $\theta(\mathrm{B}) = 1 - \theta_1\mathrm{B} - \theta_2\mathrm{B}^2 - \cdots - \theta_q\mathrm{B}^q$ é o operador de médias móveis de ordem *q.*

2. Processos estocásticos

Ilustração:



2. Processos estocásticos

Exercício 8

1) Simular os processos estocásticos ARMA com os tamanhos 250, 500, 750, 1000.

2) Analisar as funções de autocorrelação e autocorrelação parcial das séries simuladas. As séries são estacionárias?

2. Processos estocásticos

2.5 Tipos

Um processo $\boldsymbol{ARIMA\ (p,d,q)}$ é um processo misto.

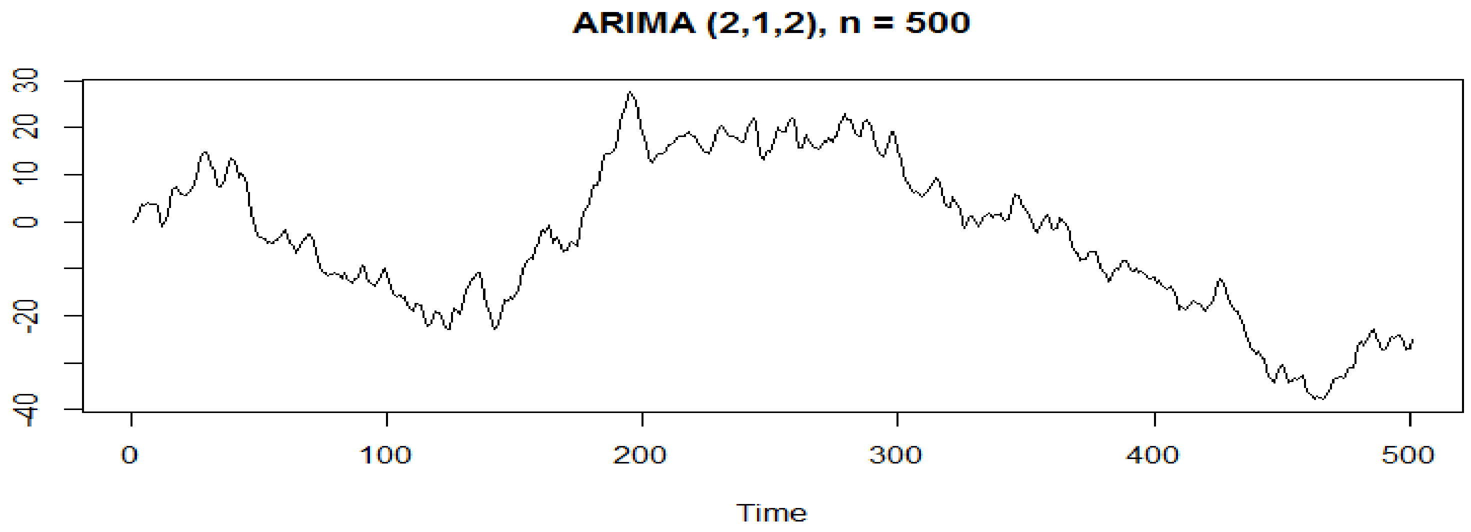
$$\phi(\mathrm{B})[(1-\mathrm{B})^{d}X_t - \mu] = c + \theta(\mathrm{B})\varepsilon_t$$

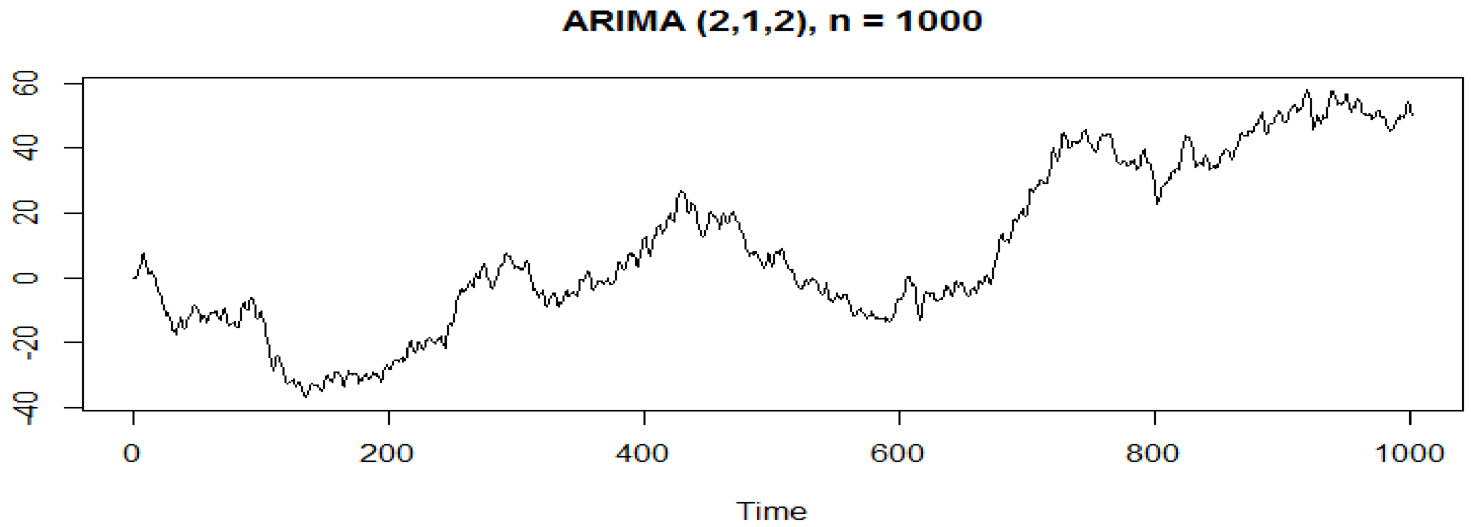
em que,

- $\phi(\mathrm{B}) = 1 - \phi_1\mathrm{B} - \phi_2\mathrm{B}^2 - \cdots - \phi_p\mathrm{B}^p$ é o operador autorregressivo de ordem *p;*
- μ é a média da série com *d* diferenças;
- $\theta(\mathrm{B}) = 1 - \theta_1\mathrm{B} - \theta_2\mathrm{B}^2 - \cdots - \theta_q\mathrm{B}^q$ é o operador de médias móveis de ordem *q.*

2. Processos estocásticos

Ilustração:





2. Processos estocásticos

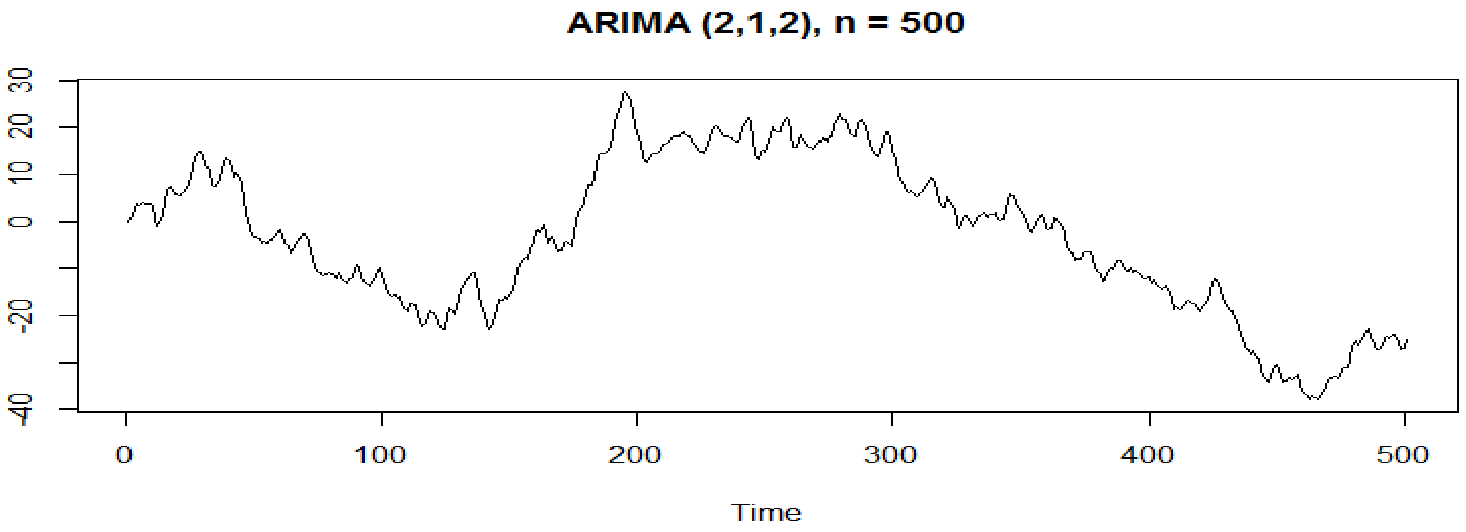
2.5 Tipos

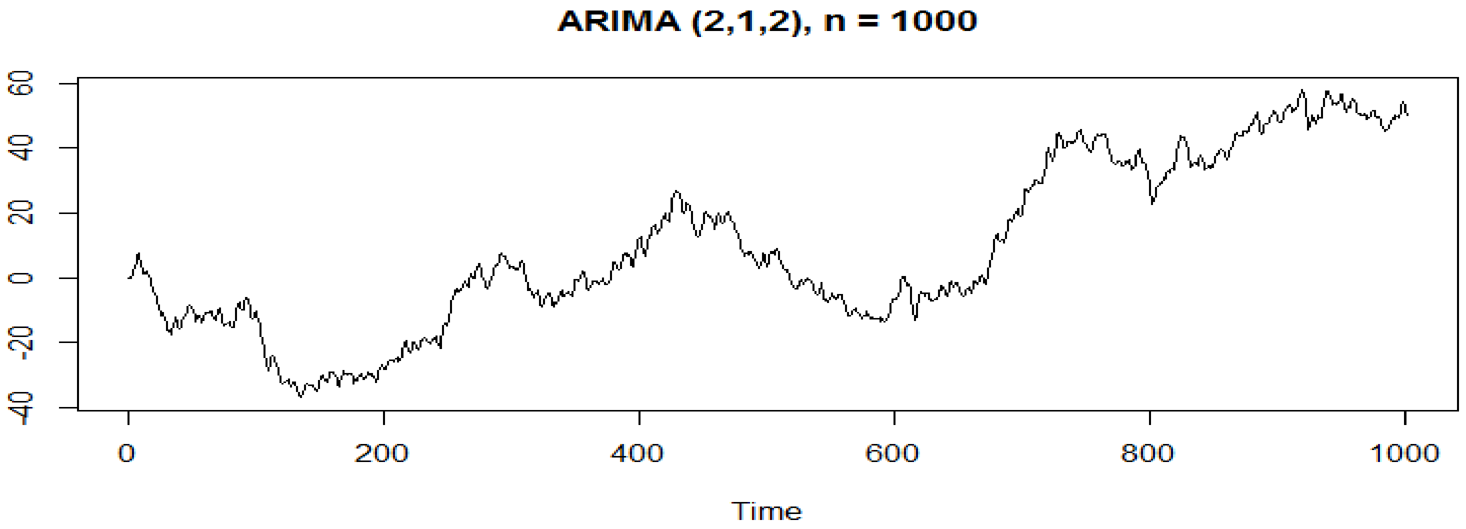
Um processo $SARIMA\,(p,d,q)x(P,D,Q)$ é um processo ARIMA misto com parte sazonal.

$$\phi(\mathrm{B}).\Phi_p(\mathrm{B}^s)\left[(1-\mathrm{B})^d(1-\mathrm{B})_s^D-\mu\right]X_t=\theta(\mathrm{B}).\Theta_Q.(\mathrm{B}^S).\xi_t$$

2. Processos estocásticos

Ilustração:





2. Processos estocásticos

Exercício 9

1) Simular os processos estocásticos ARIMA com os tamanhos 250, 500, 750, 1000.

2) Analisar as funções de autocorrelação e autocorrelação parcial das séries simuladas. As séries são estacionárias?

3) Repetir as simulações acima para os processos SARIMA.

Análise exploratória de Séries Temporais

3. Análise exploratória

3.1 – Imputação de valores ausentes

O objetivo aqui é preencher valores ausentes na série pela média ou pela mediana.

```
###código
library(imputeTS)
x <- ts(c(1, 2, 3, 4, 5, 6, 7, 8, NA, NA, 11, 12))
na.mean(x)
na.mean(x, option="median")
```

1. Introdução

Exercício 10

1) Na pasta “exercício 10”. Preencher, se possível, os valores da séries.

3. Análise exploratória

3.2 – Estatísticas descritivas

O objetivo aqui é caracterizar as séries temporais.

- Mínimo, máximo, amplitude;
- Média, variância, desvio padrão, coeficientes de variação, de assimetria e de curtose;
- Gráficos (boxplot, histograma);

Exercício 11

1) Determinar e analisar as estatísticas descritivas das séries abaixo. Interprete.

   a) AirPassengers
   b) co2
   c) Nile
   d) nottem
   e) ausbeer #library(fpp)
   f) visitors #library(fpp)
   g) JohnsonJohnson

3. Análise exploratória

3.3 – Verificar a hipótese de normalidade

- Teste Jarque-bera;
- Teste Shapiro-Wilk;

3. Análise exploratória

3.3.1 – Teste JB

O teste JB utiliza os coeficientes de assimetria e curtose para formar a seguintes estatística:

$$JB = n.\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right] \sim \chi^2_{(2)}$$

H_0: X_t é $normal$
H_1: X_t $não$ é $normal$

3. Análise exploratória

3.3.1 – Teste SW

O teste SW é o mais utilizado na identificação de normalidade. Tal teste baseia-se na estatística W denotada por

$$W = \frac{\left(\sum_{i=1}^{n} a_i y_i\right)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2}$$

sendo $a_1, a_2, \ldots, a_n$ constantes calculadas como a solução da equação

H_0: X_t é $normal$
H_1: X_t $não$ é $normal$

1. Introdução

Exercício 12

1) Verificar se as séries abaixo são possuem distribuição normal.

   a) AirPassengers
   b) co2
   c) Nile
   d) nottem
   e) ausbeer #library(fpp)
   f) visitors #library(fpp)
   g) JohnsonJohnson

3. Análise exploratória

3.4 – Verificar a hipótese de estacionariedade

- ✓ Teste Dickey-Fuller;
- ✓ Teste Dickey-Fuller Aumentado;

3. Análise exploratória

3.3.1 – Teste de Dickey-Fuller

O teste DF baseia-se numa regressão entre os valores de uma série temporal X_t e sua medição em um período imediatamente anterior $X_{(t-1)}$, ou seja, um modelo do tipo [AR(1)] representado por

$$\Delta X_t = \phi X_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim RB(0, \sigma_\varepsilon^2)$$

3.3.1 – Teste Dickey-Fuller

As hipóteses são:

i) $\Delta X_t = \pi X_{t-1} + \varepsilon_t$
Sem constante (α), sem tendência (α_t)

ii) $\Delta X_t = \alpha + \pi X_{t-1} + \varepsilon_t$
Com constante (α), sem tendência (α_t)

iii) $\Delta X_t = \alpha + \alpha_t + \pi X_{t-1} + \varepsilon_t$
Com constante (α) e tendência (α_t)

H_0: X_t *não é estacionária*
H_1: X_t *é estacionária*

3. Análise exploratória

3.3.2 – Teste Dickey-Fuller Aumentado

O teste ADF introduz um operador de defasagens no teste DF, para resolver o problema da autocorrelação serial tal como:

$$\Delta X_t = \phi X_{t-1} + \varepsilon_t$$

$$\Delta X_t = \alpha X_{(t-1)} + \sum_{k=1}^{\cdot\cdot} \beta_k \Delta X_{(t-k)} + \varepsilon_t$$

$$\varepsilon_t \sim RB(0, \sigma_\varepsilon^2)$$

3. Análise exploratória

3.3.2 – Teste Dickey-Fuller Aumentado

As hipóteses são:

i) $\Delta X_t = \pi X_{(t-1)} + \sum_{k=1}^{n} \beta_k \Delta X_{(t-k)} + \varepsilon_t$
Sem constante (α), sem tendência (α_t)

ii) $\Delta X_t = \alpha + \pi X_{(t-1)} + \sum_{k=1}^{n} \beta_k \Delta X_{(t-k)} + \varepsilon_t$
Com constante (α), sem tendência (α_t)

iii) $\Delta X_t = \alpha + \alpha_t + \pi X_{(t-1)} + \sum_{k=1}^{n} \beta_k \Delta X_{(t-k)} + \varepsilon_t$
Com constante (α) e tendência (α_t)

H_0: X_t *não* é *estacionária*
H_1: X_t é *estacionária*

Exercício 13

1) Executar o teste ADF nas séries listadas abaixo. O que é possível concluir?

   a) AirPassengers
   b) co2
   c) Nile
   d) nottem
   e) ausbeer #library(fpp)
   f) visitors #library(fpp)
   g) JohnsonJohnson

1. Introdução

Exercício 13

2) Executar o teste ADF na duas primeiras diferenças das séries abaixo. O que é possível concluir?

   a) AirPassengers
   b) co2
   c) Nile
   d) nottem
   e) ausbeer #library(fpp)
   f) visitors #library(fpp)
   g) JohnsonJohnson

3. Análise exploratória

3.3.3 – Teste para memória

- Método DFA;

3. Análise exploratória

3.3.3 – DFA

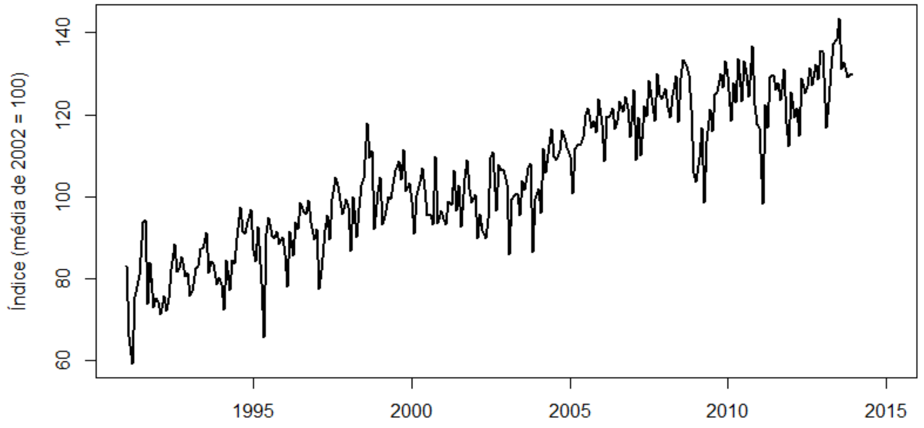
O método DFA é um método computacional de análise de escala utilizado para estimar expoentes que caracterizam o decaimento lento da função de autocorrelação e possibilita a identificação de auto-afinidade em séries temporais não estacionárias.

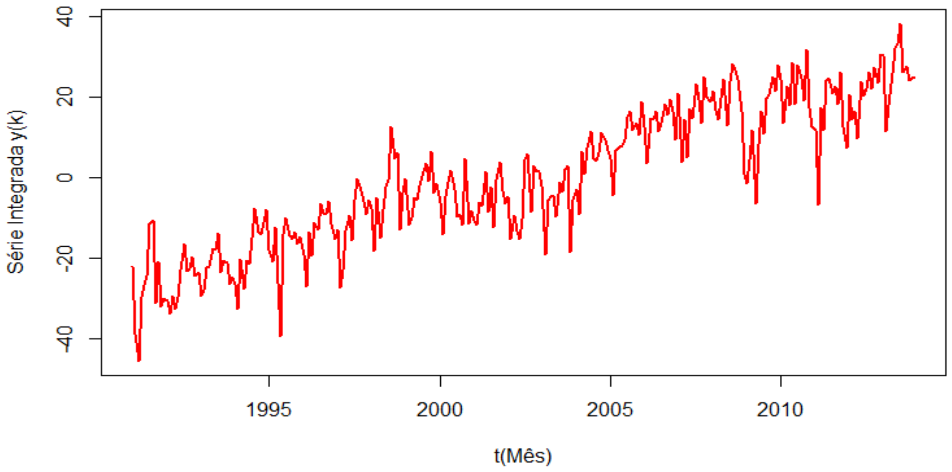
3. Análise exploratória

Passos do DFA:

1º – Obter a série integrada $y_{(k)}$

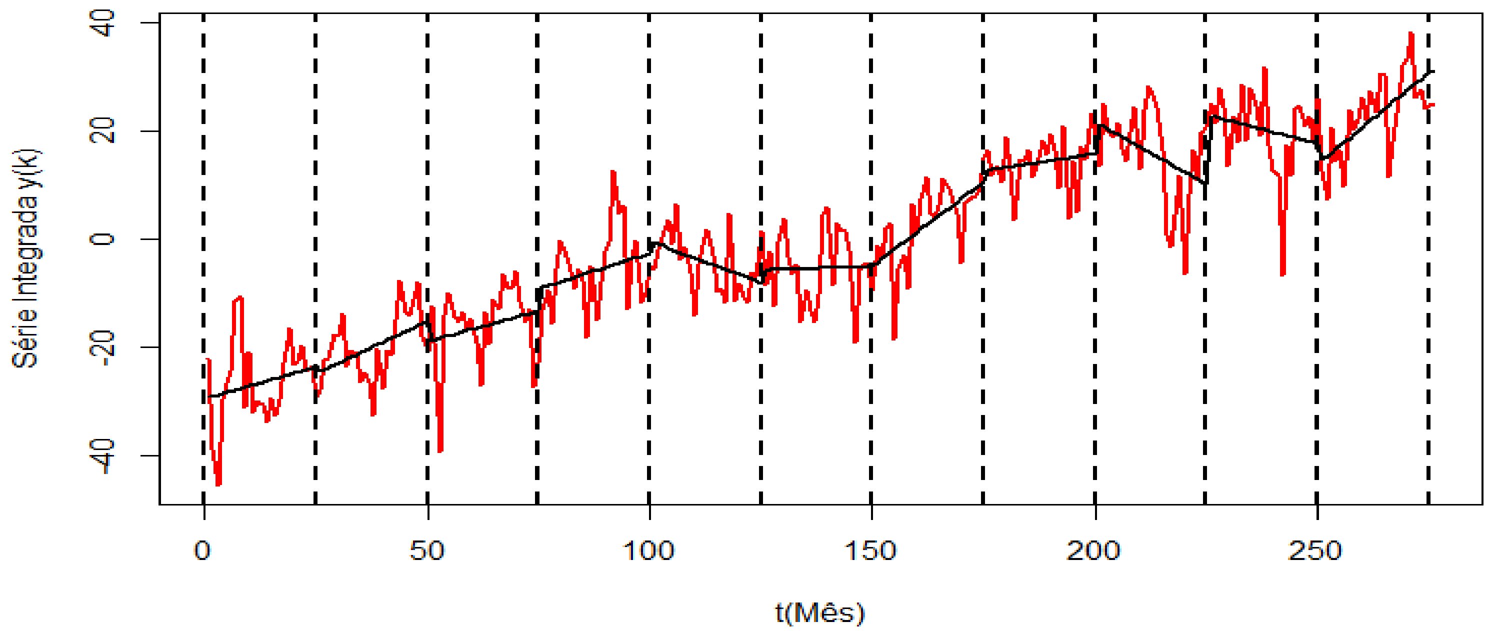
$$y(k) = \sum_{i=1}^{n} (X_i - \langle X \rangle),$$





3. Análise exploratória

2º – A série integrada $y_{(k)}$ é dividida em intervalos (*box*) de tamanhos iguais *n*, que não se interceptam. Em seguida, ajusta-se um polinômio de grau ≥ 1 a $y_{(k)}$ para cada *box* de tamanho *n*. O polinômio é do tipo $(y_n(k) = a + bx)$.



3. Análise exploratória

3º – A série integrada $y_{(k)}$ é subtraída de $y_n(k)$ em cada intervalo de tamanho *n*. Assim, é calculada a raiz quadrática média $F_{(n)}$ para cada amplitude de tamanho *n*, isto é,

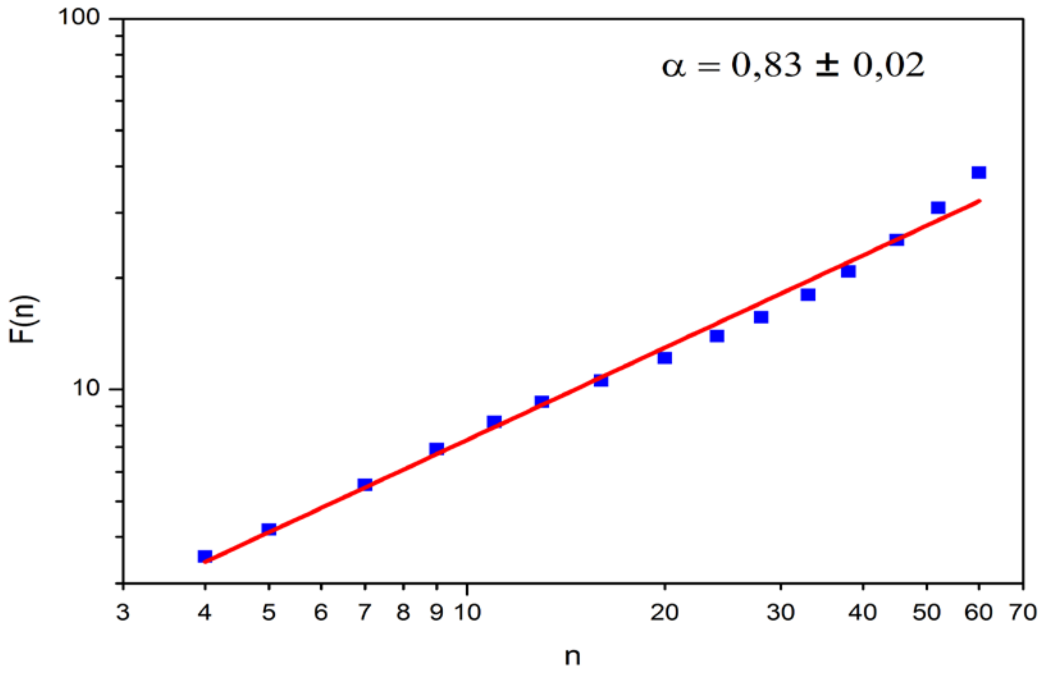
$$F(n) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}[y(k) - y_n(k)]^2},$$

Nota: O cálculo descrito anteriormente deve ser repetido sistematicamente para diferentes *boxes* de tamanho *n* $(4 \leq n \leq {}^{n}/_{4})$.

3. Análise exploratória

3º – Por último, determina-se o comportamento de escala da flutuação $F_{(n)}$ através da relação linear entre *log* $F_{(n)}$ em função de *log(n)*.

$$F_{(n)} \sim n^{\alpha}$$



3. Análise exploratória

Tabela 1.3: Tipos de comportamento considerados na análise do DFA

DFA	Tipos de comportamentos
$\alpha_{DFA} < 0{,}5$	Antipersistente
$\alpha_{DFA} \simeq 0{,}5$	Não correlacionado (Ruído Branco)
$\alpha_{DFA} > 0{,}5$	Persistente
$\alpha_{DFA} \simeq 1{,}0$	Ruído $1/f$
$\alpha_{DFA} > 1{,}0$	Não estacionário (Passeio Aleatório)
$\alpha_{DFA} \simeq 1{,}5$	Ruído Browniano

3. Análise exploratória

Exercício 14

1) Testar a memória das séries abaixo. O que é possível concluir?

   a) AirPassengers
   b) co2
   c) Nile
   d) nottem
   e) ausbeer #library(fpp)
   f) visitors #library(fpp)
   g) JohnsonJohnson

3. Análise exploratória

Exercício 14

2) Simule séries do tipo RB, RR, PA, AR, MA, ARMA, ARIMA com tamanho 1000. Para cada série simulada pede-se

a. Plote a função de autocorrelação.

b. Aplique o DFA e classifique as séries com base no α.

Métodos Preditivos

4. Métodos Preditivos

O que é previsão?

Etimologicamente (*prae e videre*), a palavra previsão sugere que se quer ver uma coisa antes que ela exista.

Alguns autores sugerem a palavra predição, para indicar algo que deverá existir no futuro. Outros autores, preferem o termo projeção.

Obs.: Salienta-se que a projeção é apenas um meio de fornecer informações para uma tomada de decisão.

4. Métodos Preditivos

Abordagens

- ✓ Econométrico
- ✓ De Séries Temporais

4. Métodos Preditivos

Métodos de Séries Temporais

- ✓ Média;
- ✓ Naive;
- ✓ SES;
- ✓ SEH;
- ✓ SEHW;
- ✓ ETS
- ✓ TBATS
- ✓ ARIMA;
- ✓ SSA;

4. Métodos Preditivos

4.1 – Indicadores de acurácia

- ✓ É difícil encontrar modelos que vinculem plenamente o mundo real aos dados observados.
- ✓ Vários indicadores de performance foram desenvolvidos para avaliar o desempenho preditivo (acurácia) de modelos.

4. Métodos Preditivos

4.1 – Indicadores de acurácia

$$ME = \frac{\sum_{t=1}^{n}(X_t - \hat{X}_t)}{n}$$

$$MAE = \frac{\sum_{t=1}^{n}|X_t - \hat{X}_t|}{n}$$

$$MSE = \frac{\sum_{t=1}^{n}(X_t - \hat{X}_t)^2}{n}$$

$$MAPE = \left[\frac{1}{n}\sum_{t=1}^{n}\frac{|X_t - \hat{X}_t|}{X_t}\right] * 100$$

4. Métodos Preditivos

4.2 – Média

As previsões de todos os valores futuros são iguais à média dos dados históricos.

$$\hat{Y}_{T+h} = (Y_1, \ldots, Y_N)/T$$

Ex.:
```
plot(AirPassengers)
library(forecast)
model <- meanf(AirPassengers)
lines(model$mean, col="red")
summary(model)
```

1. Métodos Preditivos

Exercício 15

1) Realize a modelagem das séries abaixo com o método de médias e estime a valor futuro considerando-se $h = 5$. Os valores são iguais? Por que?

   a) co2
   b) Nile
   c) ausbeer #library(fpp)
   d) visitors #library(fpp)

4. Métodos Preditivos

4.3 – Naive

As previsões são definidas como o valor da última observação Y_t.

- ✓ Simples
- ✓ Drift
- ✓ Sazonal

4. Métodos Preditivos

4.2.1 – Naive Simples

As previsões são definidas como o valor da última observação Y_t.

$$\hat{Y}_{T+h|T} = Y_T$$

Ex.:
```
plot(AirPassengers)
library(forecast)
model1 <- naive(AirPassengers, h = 3)
summary(model1)
plot(model1)
```

4. Métodos Preditivos

4.2.2 – Naive drift

Uma variação do método Naive que permite que as previsões aumentem ou diminuam com o tempo Y_t.

$$\hat{Y}_{T+h|T} = Y_T + \frac{h}{T-1}(Y_T - Y_1)$$

Ex.:

```
plot(AirPassengers)
library(forecast)
model2 <- rwf(AirPassengers, h = 3, drift=TRUE)
summary(model2)
plot(model2)
```

4. Métodos Preditivos

4.2.3 – Naive seasonal

É útil para dados altamente sazonais. Neste caso, cada previsão é igual ao último valor observado da mesma estação do ano.

$$\hat{Y}_{T+h|T} = Y_{T+h-km}, \quad k = \left[\frac{h-1}{m}\right] + 1$$

Ex.:
```
plot(AirPassengers)
library(forecast)
model3 <- snaive(AirPassengers, h = 3)
summary(model3)     plot(model3)
```

1. Métodos Preditivos

Exercício 16

1) Realize a modelagem das séries abaixo usando os métodos Naive simples, Naive drift e Naive Seasonal. Qual deles apresenta a melhor acurácia?

   a) co2
   b) Nile
   c) ausbeer #library(fpp)
   d) visitors #library(fpp)

4. Métodos Preditivos

4.4 – SES

É adequado para prever dados sem tendência ou padrão sazonal.

$$\hat{Y}_{t+1} = \alpha Y_t + (1-\alpha)\hat{Y}_t \qquad 0 \leq \alpha \leq 1 \qquad \hat{Y}_0 = Y_t$$

$$\boxed{\hat{Y}_{T+h|T} = N_t}$$

Obs.: Explicar Utilizando o excel

1. Métodos Preditivos

Exercício 17

1) Realize a modelagem das séries abaixo usando o método SES. Qual foi o erro obtido?

    a) Nile
    b) nottem

4. Métodos Preditivos

4.5 – SEH

É adequado para prever dados com tendência.

$$N_{t+1} = \alpha R_t + (1-\alpha)(N_{t-1} + T_{t-1}) \qquad 0 \leq \alpha \leq 1$$

$$T_{t+1} = \beta(N_t - N_{t-1}) + (1-\beta)T_{t-1} \qquad 0 \leq \beta \leq 1$$

$$N_0 = Y_t$$

$$T_0 = Y_{t+1} - Y_t$$

$$\boxed{\widehat{Y}_{T+h|T} = N_t + hT_t}$$

Obs.: Explicar Utilizando o excel

1. Métodos Preditivos

Exercício 18

1) Realize a modelagem das séries abaixo usando o método SEH. Qual foi o erro obtido?

   a) Co2
   b) AirPassengers
   c) JohnsonJohnson

4. Métodos Preditivos

4.7 – SEHW

É adequado para prever dados com tendência e sazonalidade.

Aditivo	Multiplicativo
$N_t = \alpha(Y_t - S_{t-m}) + (1-\alpha)(N_{t-1} + T_{t-1})$	$N_t = \alpha(\frac{Y_t}{S_{t-m}}) + (1-\alpha)(N_{t-1} + T_{t-1})$
$T_t = \beta(N_t - N_{t-1}) + (1-\beta)T_{t-1}$	$T_t = \beta(N_t - N_{t-1}) + (1-\beta)T_{t-1}$
$S_t = \gamma(Y_t - N_{t-1} - T_{t-1}) + (1-\gamma)S_{t-m}$	$S_t = \gamma(\frac{Y_t}{N_{t-1} + T_{t-1}}) + (1-\gamma)S_{t-m}$
$\hat{Y}_{T+h\|T} = N_t + hT_t + S_{t-m}$	$\hat{Y}_{T+h\|T} = (N_t + hT_t)\, S_{t-m}$

$0 \le \alpha \le 1$

$0 \le \beta \le 1$

$0 \le \gamma \le 1$

1. Métodos Preditivos

Exercício 19

1) Realize a modelagem das séries abaixo usando o método SEHW. Qual foi o erro obtido?

    a) Co2
    b) AirPassengers
    c) JohnsonJohnson

Dica: Verifiquem se a sazonalidade é aditiva ou multiplicativa antes de modelar.

4. Métodos Preditivos

4.6 – ARIMA

Um processo ***ARIMA*** $(\boldsymbol{p}, \boldsymbol{d}, \boldsymbol{q})$ é um processo misto. É adequado para prever séries com tendência.

$$\phi(\mathrm{B})[(1-\mathrm{B})^{d} X_t - \mu] = c + \theta(\mathrm{B})\varepsilon_t$$

em que,

- $\phi(\mathrm{B}) = 1 - \phi_1\mathrm{B} - \phi_2\mathrm{B}^2 - \cdots - \phi_p\mathrm{B}^p$ é o operador autorregressivo de ordem *p;*
- μ é a média da série com *d* diferenças;
- $\theta(\mathrm{B}) = 1 - \theta_1\mathrm{B} - \theta_2\mathrm{B}^2 - \cdots - \theta_q\mathrm{B}^q$ é o operador de médias móveis de ordem *q.*

4. Métodos Preditivos

4.6 – SARIMA

Um processo $SARIMA\,(p,d,q)x(P,D,Q)$ é um processo misto. É adequado para prever séries com tendência e sazonalidade.

$$\phi(B).\Phi_p(B^s)\left[(1-B)^d(1-B)_s^D-\mu\right]X_t=\theta(B).\Theta_Q.(B^S).\xi_t$$

4.6 – Formas

a) **Forma de Equação de Diferenças**

Representado em termos de valores prévios de Y_t e do valor atual. É bastante utilizado para calcular previsões.

$$Y_t = \xi_1 Y_{t-1} + \cdots + \xi_{p+d} Y_{t-p-d} + u_t - \theta_1 u_{t-1} - \cdots - \theta_q u_{t-q}$$

b) **Forma de Choques Aleatórios**

Representado em termos do valor atual e prévio de u_t. É uma forma conveniente para se calcular a variância dos erros de previsão.

$$Y_t = u_t + \psi_1 u_{t-1} + \psi_2 u_{t-2} + \cdots = \psi(B) u_t$$

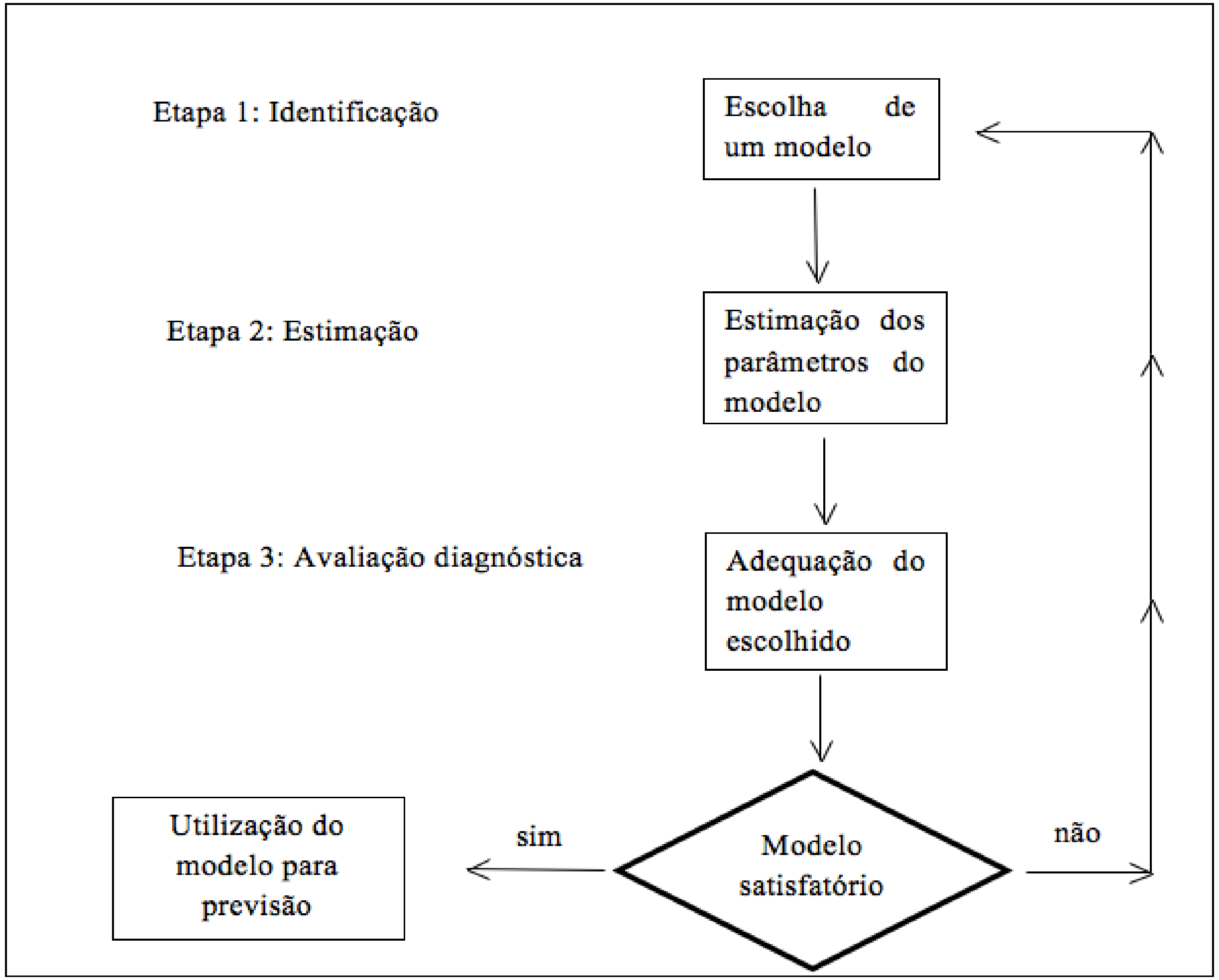
c) **Forma Invertida**

Representada em termos dos valores prévios de Y_t e do valor atual de u_t.

$$\pi(B) Y_t = Y_t - \pi_1 Y_{t-1} - \pi_2 Y_{t-2} - \cdots = u_t$$

4. Métodos Preditivos

4.6 – ARIMA/SARIMA



1 - Uma classe geral de modelos é considerada para a análise. A estimação é realizada considerando-se a ferramentas: FAC, FACP e critérios de informação (AIC, BIC e etc).

2 – Os parâmetros do modelo identificado são estimados com base nos dados (máxima verossimilhança ou mínimos quadrados).

3 – Avaliação residual. Espera-se que os resíduos não apresentem correlação serial (RB).

4. Métodos Preditivos

4.6 – ARIMA/SARIMA

Tabela 2.1 - Comportamento teórico da AC e PACF para alguns modelos

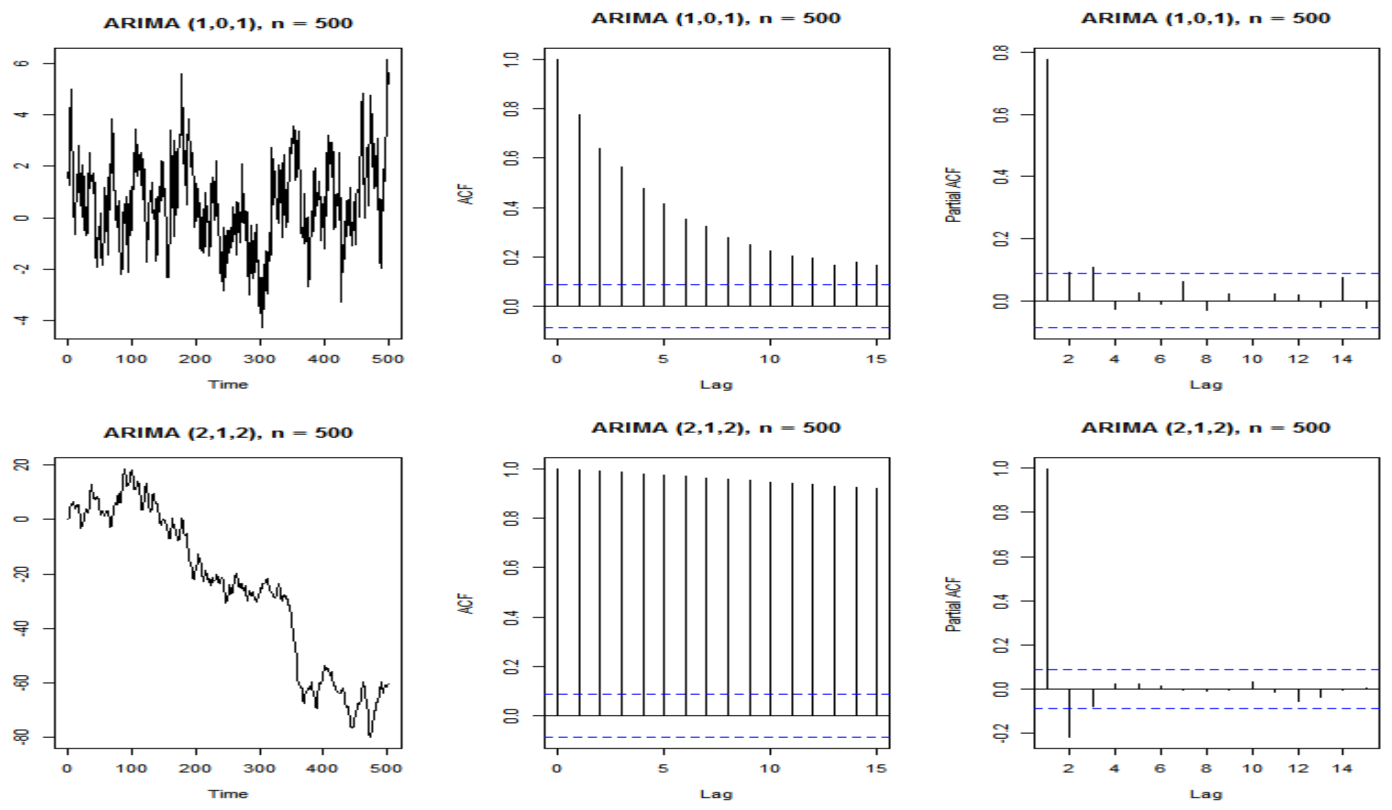
Modelo	ACF	PACF
MA(1)	1 pico no *lag* 1	Decrescimento exponencial
AR(1)	Decrescimento exponencial	1 pico no *lag* 1
MA(2)	1 pico no *lag* 1 e 1 pico no *lag* 2	Mistura de exponenciais ou ondas senóides amortecidas
AR(2)	Mistura de exponenciais ou ondas senóides amortecidas	1 pico no *lag* 1 e 1 pico no *lag* 2

$$AIC = -2ln(L) + 2(p+q)$$

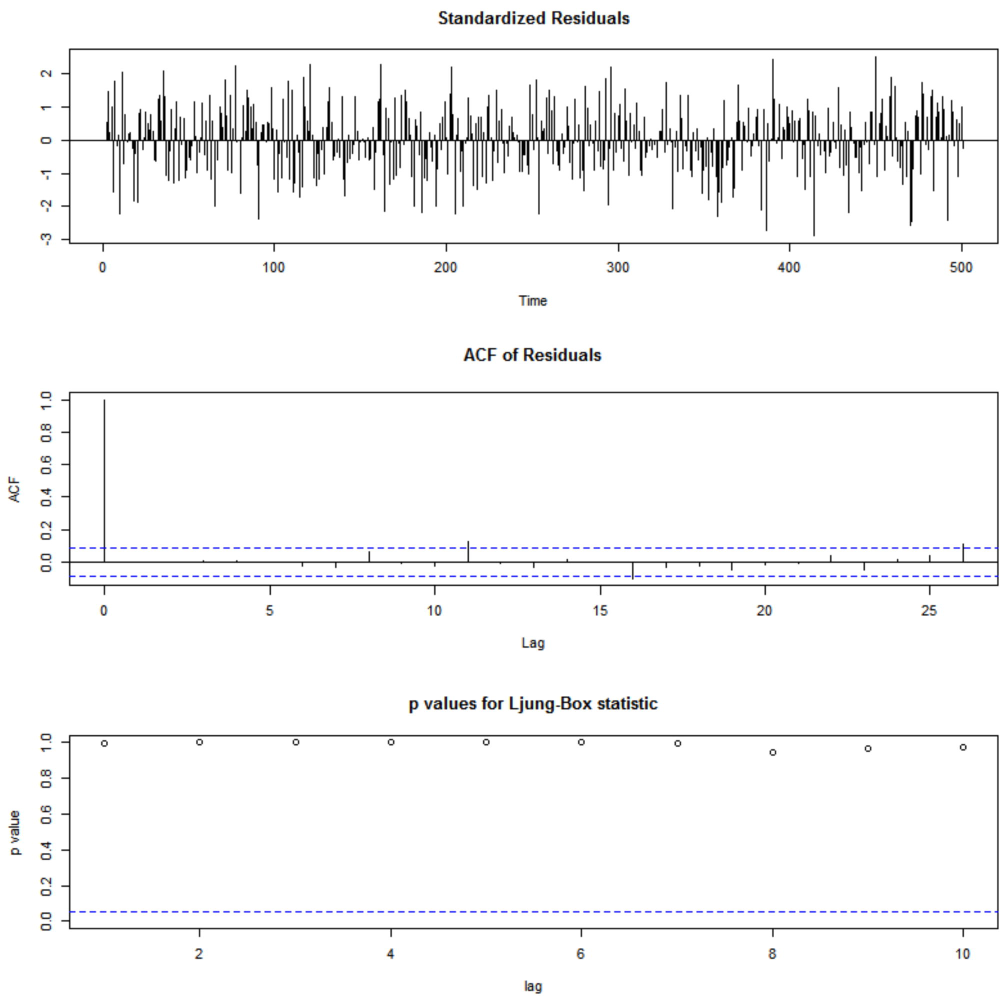
$$BIC = -2ln(L) + (p+q)ln(N)$$

L a verossimilhança maximizada.

```
library(forecast)
model <- auto.arima(AirPassengers)
summary(model)
```



4.6 – ARIMA/SARIMA




```
tsdiag(model)
```

$$H_0\colon\ \hat{\rho(1)} = \hat{\rho(2)} = \cdots \hat{\rho(k)} = 0$$

$$\hat{\rho(1)} \neq \hat{\rho(2)} \neq \cdots \hat{\rho(k)} \neq 0.$$

1. Métodos Preditivos

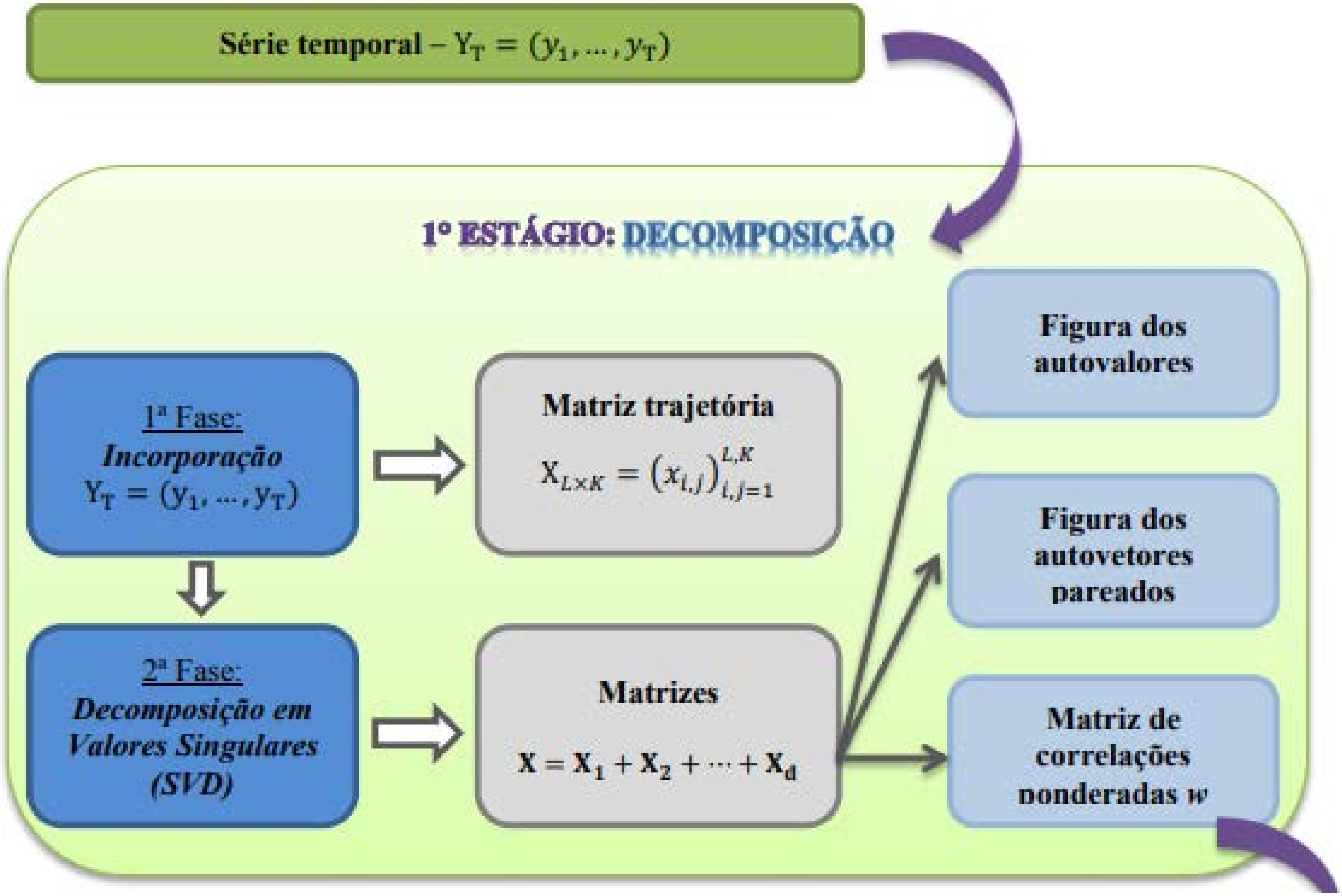
Exercício 20

1) Realize a modelagem das séries abaixo usando o método ARIMA/SARIMA.

    a) Co2
    b) AirPassengers
    c) JohnsonJohnson
    d) nottem

4. Métodos Preditivos

4.7 – Singular Spectrum Analysis (SSA)

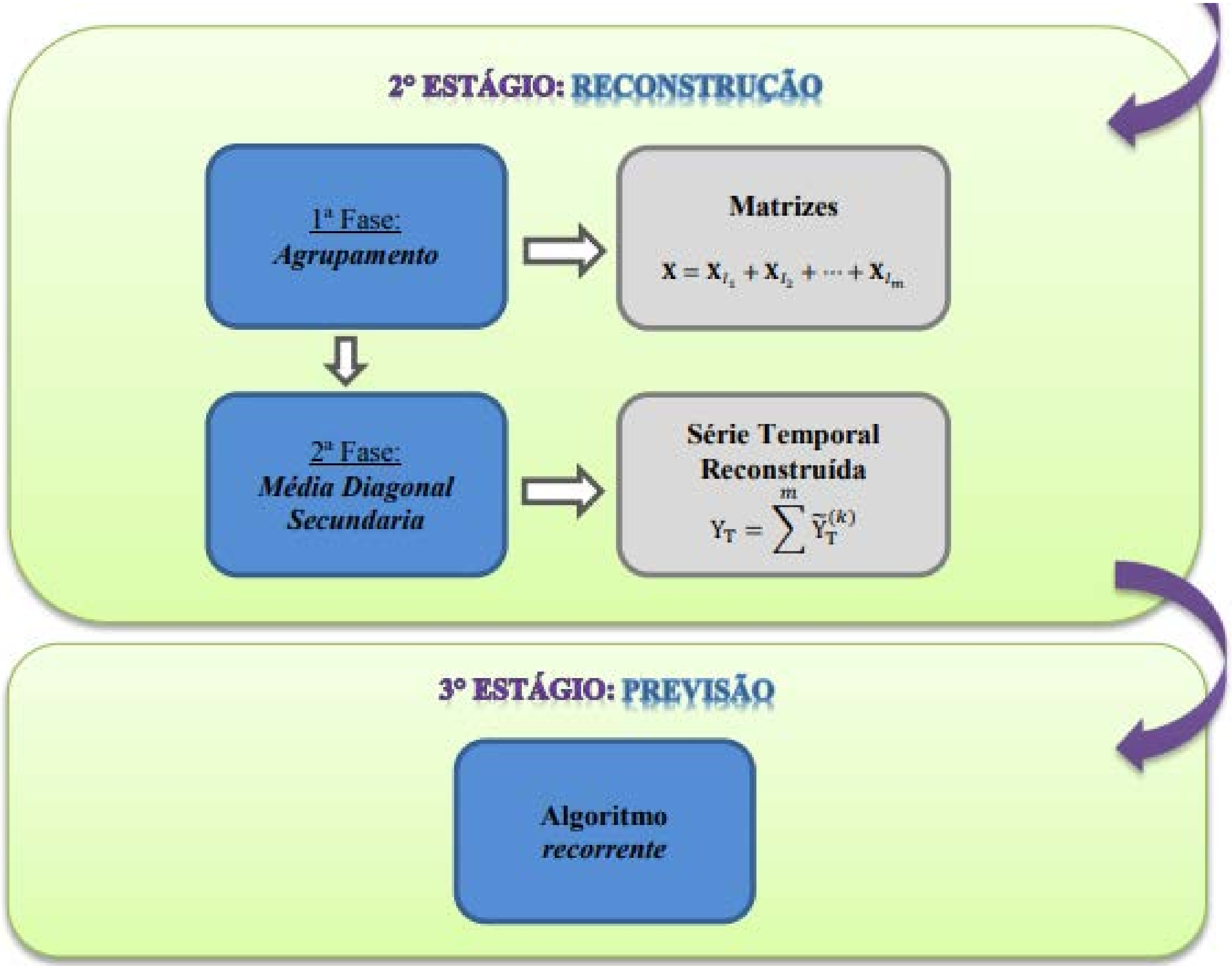


$$\begin{bmatrix} Y_1 & Y_2 & \cdots & Y_k \\ Y_2 & Y_3 & \cdots & Y_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ Y_L & Y_{L+1} & \cdots & Y_N \end{bmatrix}$$

$$K = N - L + 1$$

4. Métodos Preditivos

4.7 – Singular Spectrum Analysis (SSA)



Obs.: Demostrar SSA no R

1. Métodos Preditivos

Exercício 21

1) Aplique o método SSA nas séries abaixo.

   a) Co2
   b) JohnsonJohnson
   d) nottem

4. Métodos Preditivos

4.8 – Tópico especial

Importante! Antes de aplicar um método preditivo é necessário realizar o *data partition.* Pode ser:

- *train set, predict set* [70%, 30%];
- [80%, 20%];
- [90%, 10%];

Obs.: Rodar o script nº21.

Obrigado!

Everaldo Guedes
efgestatistico@gmail.com