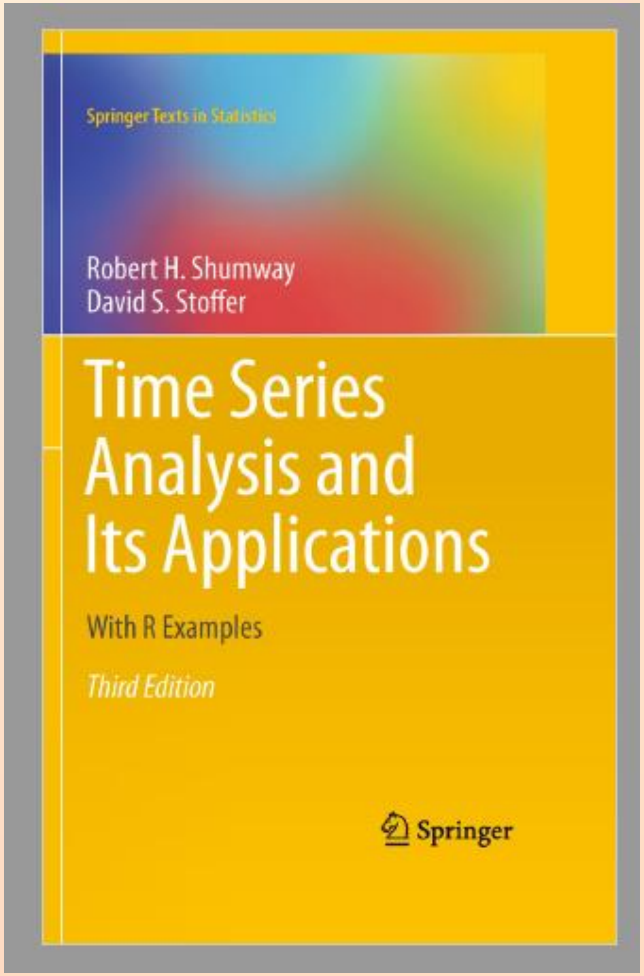
# Análise de Séries Temporais

# Análise de Séries Temporais



# Definições

- Uma série temporal é qualquer conjunto de observações ordenadas no tempo.

Por exemplo:

- Valores diários de poluição na cidade de São Paulo;
- Valores mansais de temperatura na cidade de São Paulo;
- Precipitação atmosférica anual na cidade de Fortaleza.

# Objetivo da análise de séries temporais

- Investigar o mecanismo gerador da série temporal;
- Fazer previsões de valores futuros da série;
- Descrever apenas o comportamento da série: existência de tendência, ciclos e variações sazonais;
- Procurar periodicidades relevantes nos dados.

# Tipos de Séries Temporais

Uma série temporal pode ser

. Discreta: X(t), t=1,2 , ..., n

- valores semanais do número de casos de Aids em São Paulo;
- taxa de mortalidade(mensais, anuais);
- gastos com a saúde (mensais, anuais)

. Contínua: X(t),

- valores do eletrocardiograma;
- medições de temperatura e umidade.

# Ferramentas

- **Descrever o comportamento da série: gráficos e testes para avaliar tendências, ciclos, variações sazonais;**
- **Inferências estatísticas;**
- **Modelagem do fenômeno estudado;**
- **Previsões.**

# Séries Temporais  - Caracteristicas

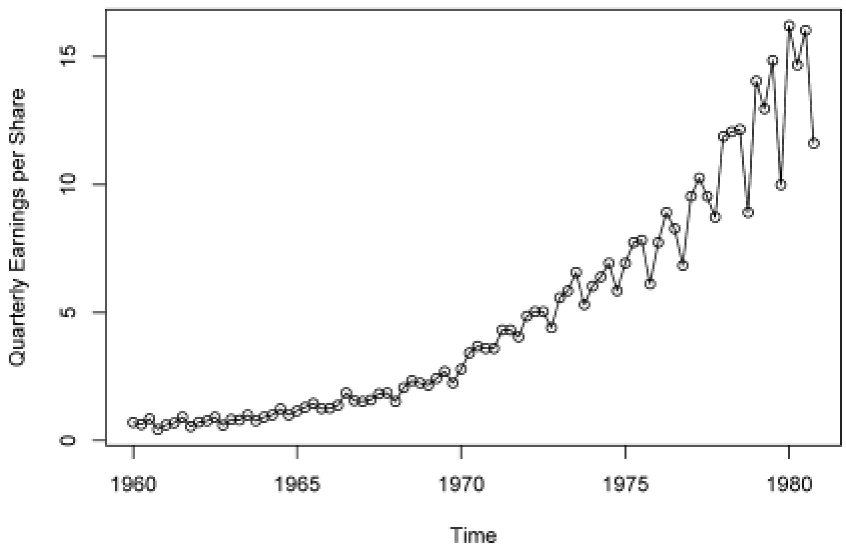



**Fig. 1.1.** Johnson & Johnson quarterly earnings per share, 84 quarters, 1960-I to 1980-IV.

**Example 1.1 Johnson & Johnson Quarterly Earnings**

Figure 1.1 shows quarterly earnings per share for the U.S. company Johnson & Johnson, furnished by Professor Paul Griffin (personal communication) of the Graduate School of Management, University of California, Davis. There are 84 quarters (21 years) measured from the first quarter of 1960 to the last quarter of 1980. Modeling such series begins by observing the primary patterns in the time history. In this case, note the gradually increasing underlying trend and the rather regular variation superimposed on the trend that seems to repeat over quarters. Methods for analyzing data such as these are explored in Chapter 2 (see Problem 2.1) using regression techniques and in Chapter 6, §6.5, using structural equation modeling.

# Séries Temporais - Caracteristicas

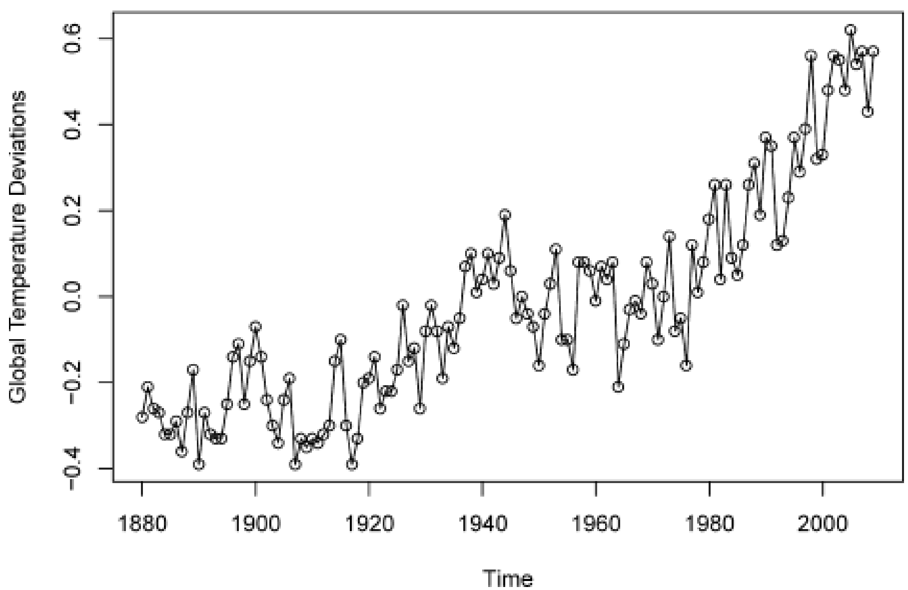



**Fig. 1.2.** Yearly average global temperature deviations (1880–2009) in degrees centigrade.

**Example 1.2 Global Warming**

Consider the global temperature series record shown in Figure 1.2. The data are the global mean land–ocean temperature index from 1880 to 2009, with the base period 1951-1980. In particular, the data are deviations, measured in degrees centigrade, from the 1951-1980 average, and are an update of Hansen et al. (2006). We note an apparent upward trend in the series during the latter part of the twentieth century that has been used as an argument for the global warming hypothesis. Note also the leveling off at about 1935 and then another rather sharp upward trend at about 1970. The question of interest for global warming proponents and opponents is whether the overall trend is natural or whether it is caused by some human-induced interface.

# Séries Temporais - Caracteristicas

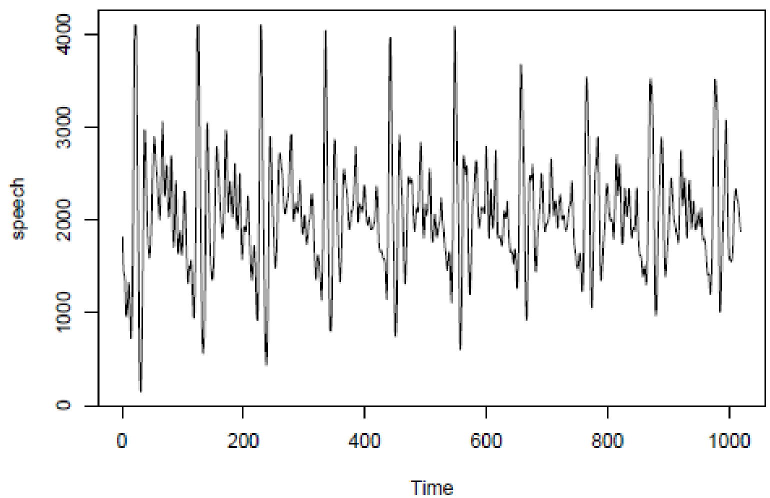



**Fig. 1.3.** Speech recording of the syllable $aaa \cdots hhh$ sampled at 10,000 points per second with $n = 1020$ points.

**Example 1.3 Speech Data**

More involved questions develop in applications to the physical sciences. Figure 1.3 shows a small .1 second (1000 point) sample of recorded speech for the phrase $aaa \cdots hhh$, and we note the repetitive nature of the signal and the rather regular periodicities. One current problem of great interest is computer recognition of speech, which would require converting this particular signal into the recorded phrase $aaa \cdots hhh$. Spectral analysis can be used in this context to produce a signature of this phrase that can be compared with signatures of various library syllables to look for a match.

# Séries Temporais - Caracteristicas

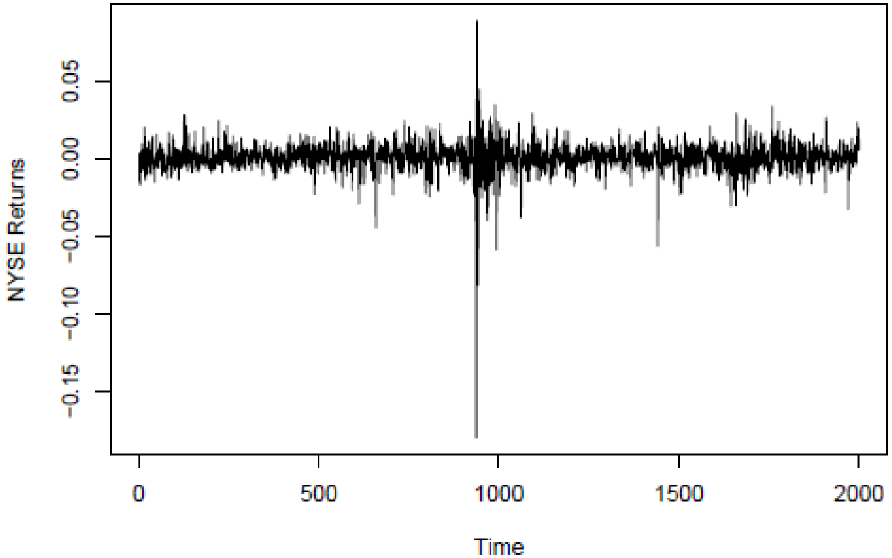



**Fig. 1.4.** Returns of the NYSE. The data are daily value weighted market returns from February 2, 1984 to December 31, 1991 (2000 trading days). The crash of October 19, 1987 occurs at $t = 938$.

**Example 1.4 New York Stock Exchange**

As an example of financial time series data, Figure 1.4 shows the daily returns (or percent change) of the New York Stock Exchange (NYSE) from February 2, 1984 to December 31, 1991. It is easy to spot the crash of October 19, 1987 in the figure. The data shown in Figure 1.4 are typical of return data. The mean of the series appears to be stable with an average return of approximately zero, however, the volatility (or variability) of data changes over time. In fact, the data show volatility clustering; that is, highly volatile periods tend to be clustered together. A problem in the analysis of these type of financial data is to forecast the volatility of future returns.

# Séries Temporais - Caracteristicas

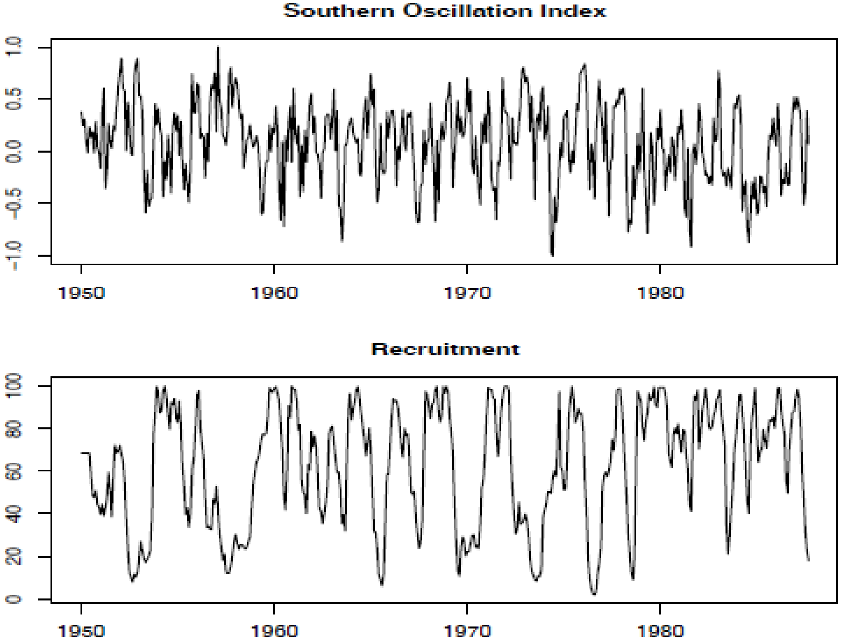



**Fig. 1.5.** Monthly SOI and Recruitment (estimated new fish), 1950-1987.

**Example 1.5 El Niño and Fish Population**

We may also be interested in analyzing several time series at once. Figure 1.5 shows monthly values of an environmental series called the Southern Oscillation Index (SOI) and associated Recruitment (number of new fish) furnished by Dr. Roy Mendelssohn of the Pacific Environmental Fisheries Group (personal communication). Both series are for a period of 453 months ranging over the years 1950–1987. The SOI measures changes in air pressure, related to sea surface temperatures in the central Pacific Ocean. The central Pacific warms every three to seven years due to the El Niño effect, which has been blamed, in particular, for the 1997 floods in the midwestern portions of the United States. Both series in Figure 1.5 tend to exhibit repetitive behavior, with regularly repeating cycles that are easily visible. This periodic behavior is of interest because underlying processes of interest may be regular and the rate or frequency of oscillation characterizing the behavior of the underlying series would help to identify them. One can also remark that the cycles of the SOI are repeating at a faster rate than those of the Recruitment series. The Recruitment series also shows several kinds of oscillations, a faster frequency that seems to repeat about every 12 months and a slower frequency that seems to repeat about every 50 months. The study of the kinds of cycles and their strengths is the subject of Chapter 4. The two series also tend to be somewhat related; it is easy to imagine that somehow the fish population is dependent on the SOI. Perhaps even a lagged relation exists, with the SOI signaling changes in the fish population. This possibility

# Séries Temporais - Caracteristicas

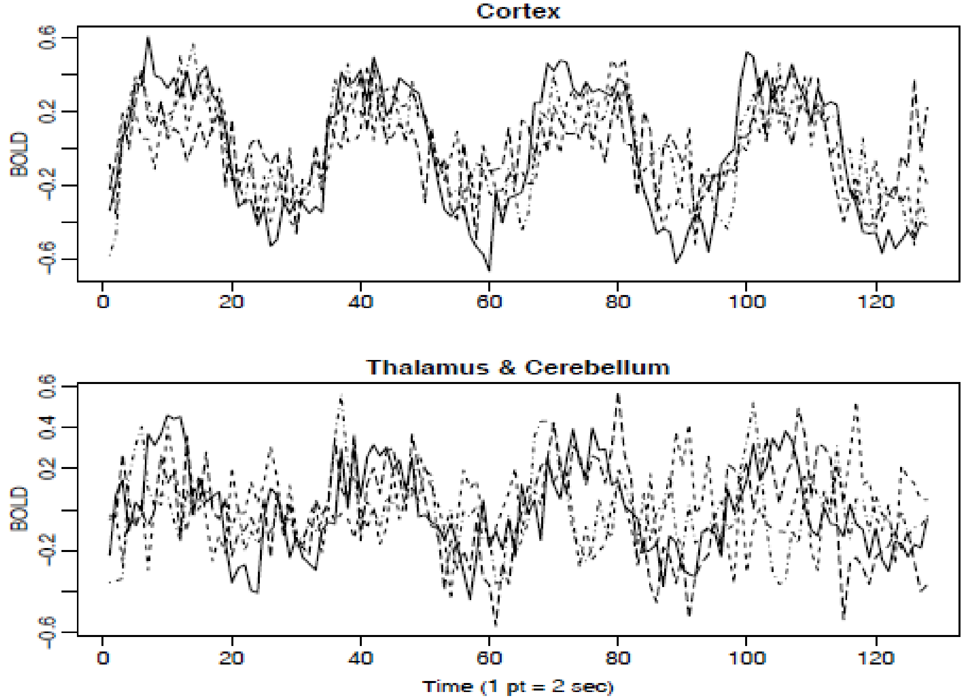



**Fig. 1.6.** fMRI data from various locations in the cortex, thalamus, and cerebellum; $n = 128$ points, one observation taken every 2 seconds.

**Example 1.6 fMRI Imaging**

A fundamental problem in classical statistics occurs when we are given a collection of independent series or vectors of series, generated under varying experimental conditions or treatment configurations. Such a set of series is shown in Figure 1.6, where we observe data collected from various locations in the brain via functional magnetic resonance imaging (fMRI). In this example, five subjects were given periodic brushing on the hand. The stimulus was applied for 32 seconds and then stopped for 32 seconds; thus, the signal period is 64 seconds. The sampling rate was one observation every 2 seconds for 256 seconds ($n = 128$). For this example, we averaged the results over subjects (these were evoked responses, and all subjects were in phase). The series shown in Figure 1.6 are consecutive measures of blood oxygenation-level dependent (BOLD) signal intensity, which measures areas of activation in the brain. Notice that the periodicities appear strongly in the motor cortex series and less strongly in the thalamus and cerebellum. The fact that one has series from different areas of the brain suggests testing whether the areas are responding differently to the brush stimulus. Analysis of variance

# Séries Temporais  - Caracteristicas

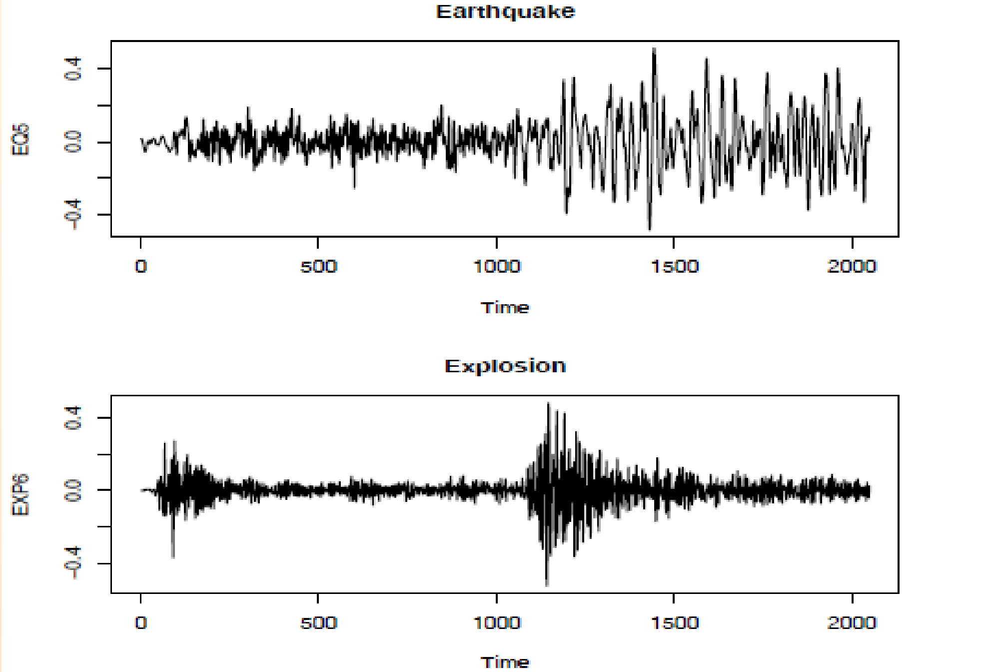
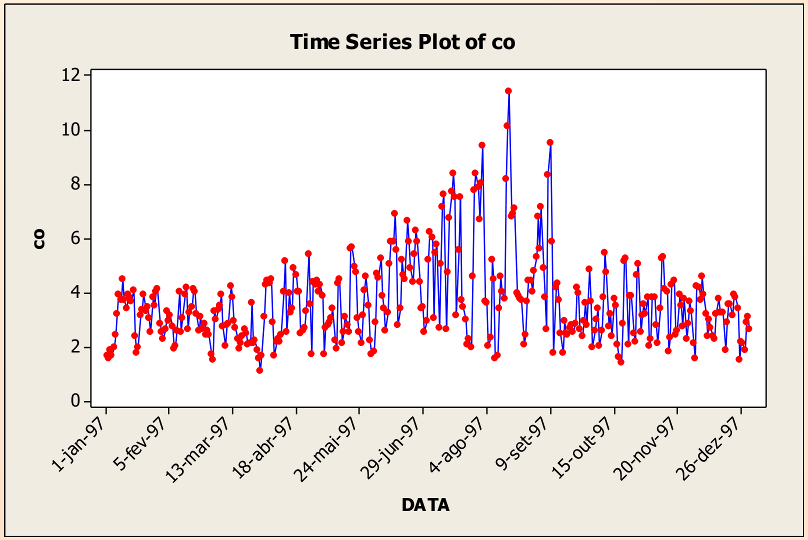
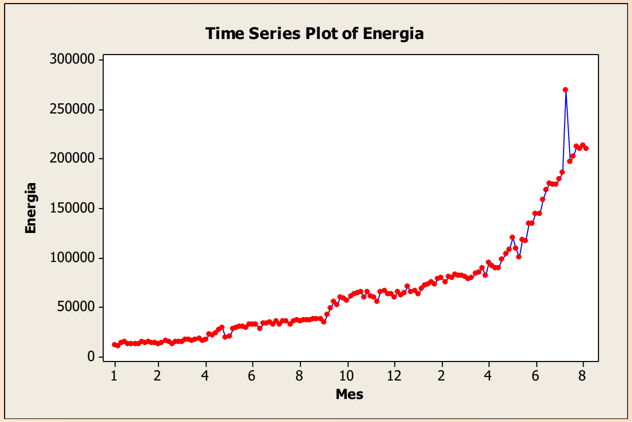


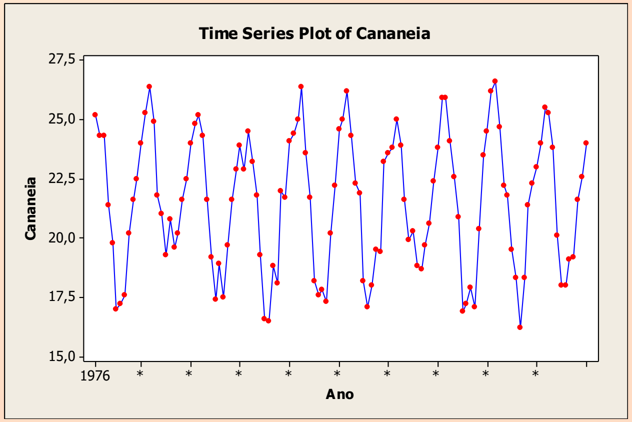

**Fig. 1.7.** Arrival phases from an earthquake (top) and explosion (bottom) at 40 points per second.

**Example 1.7 Earthquakes and Explosions**

As a final example, the series in Figure 1.7 represent two phases or arrivals along the surface, denoted by P ($t = 1, \ldots, 1024$) and S ($t = 1025, \ldots, 2048$), at a seismic recording station. The recording instruments in Scandinavia are observing earthquakes and mining explosions with one of each shown in Figure 1.7. The general problem of interest is in distinguishing or discriminating between waveforms generated by earthquakes and those generated by explosions. Features that may be important are the rough amplitude ratios of the first phase P to the second phase S, which tend to be smaller for earthquakes than for explosions. In the case of the two events in Figure 1.7, the


Time Series Plot of co
co
DATA
0
2
4
6
8
10
12
1-jan-97
5-fev-97
13-mar-97
18-abr-97
24-mai-97
29-jun-97
4-ago-97
9-set-97
15-out-97
20-nov-97
26-dez-97


Time Series Plot of Energia
Energia
Mes
0
50000
100000
150000
200000
250000
300000
1
2
4
6
8
10
12
2
4
6
8


Time Series Plot of Cananeia
Cananeia
Ano
15,0
17,5
20,0
22,5
25,0
27,5
1976

# Por que fazer análise da série temporal(histórica)

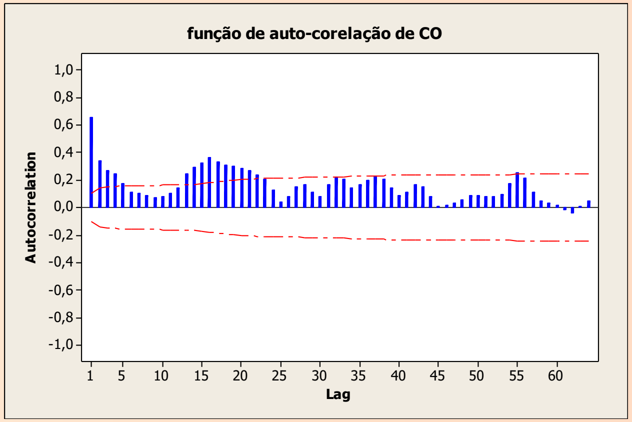
- Deseja-se modelar o fenômeno estudado para, a partir daí, descrever o comportamento da série, fazer estimativas e, por último, avaliar quais os fatores que influenciaram no comportamento da mesma, tentando definir relações de causa e efeito entre 2 ou mais séries.
- Para tanto, há diversas técnicas estatísticas disponíveis que dependem do modelo definido para a série, bem como do tipo de série analisada e do objetivo do trabalho.

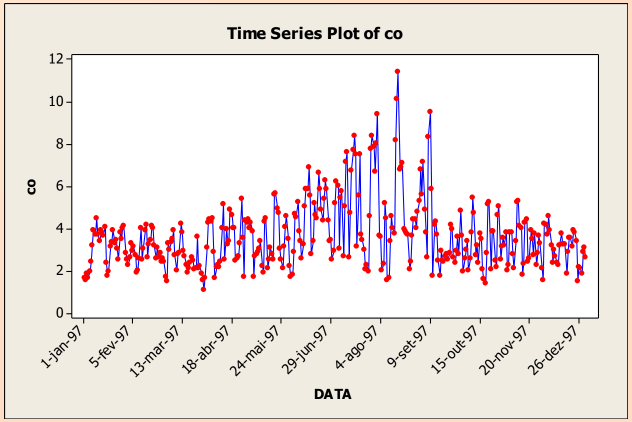
# Processos estacionários

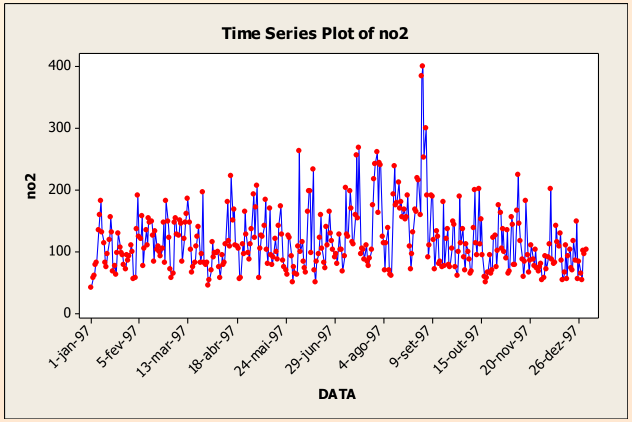
- É importante definir se a série é estacionária ou não para, a partir daí, estabelecer a estrutura do modelo probabilístico que estimará a mesma.
- Uma série é considerada estacionária quando suas observações ocorrem, aleatoriamente, ao redor de uma média constante e a correlação entre dois pontos dependem somente da defasagem entre eles.

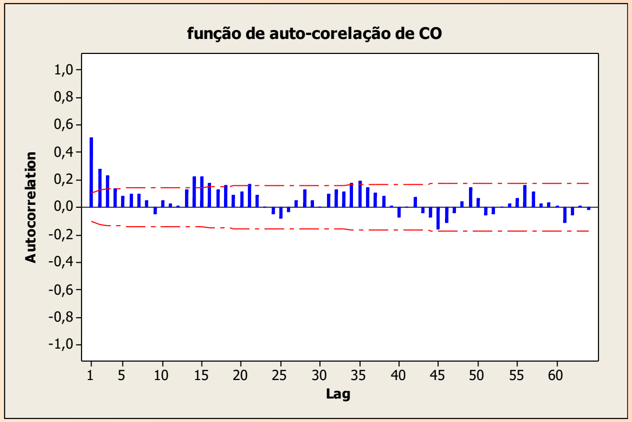
# Função de auto-correlação

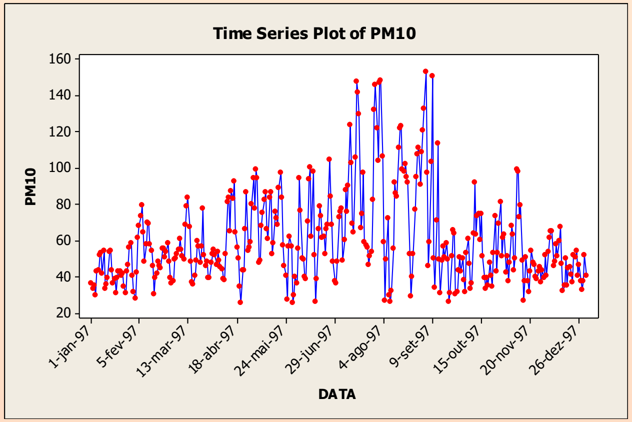
- O coeficiente de correlação entre as observações X(t) e x(t+j).

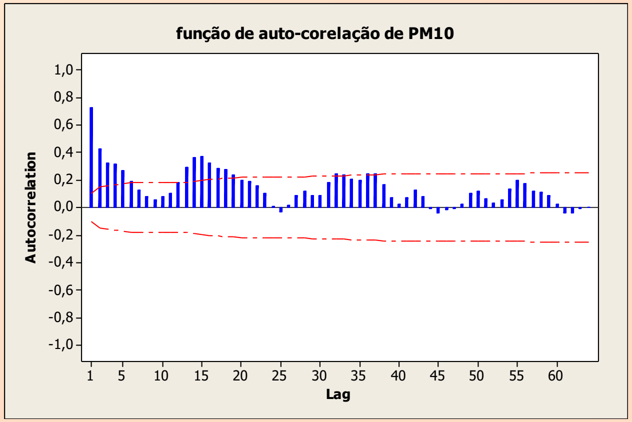



Time Series Plot of no2
no2
400
300
200
100
0
1-jan-97
5-fev-97
13-mar-97
18-abr-97
24-mai-97
29-jun-97
4-ago-97
9-set-97
15-out-97
20-nov-97
26-dez-97
DATA


função de auto-corelação de CO
Autocorrelation
1,0
0,8
0,6
0,4
0,2
0,0
-0,2
-0,4
-0,6
-0,8
-1,0
1
5
10
15
20
25
30
35
40
45
50
55
60
Lag


Time Series Plot of PM10
PM10
160
140
120
100
80
60
40
20
1-jan-97
5-fev-97
13-mar-97
18-abr-97
24-mai-97
29-jun-97
4-ago-97
9-set-97
15-out-97
20-nov-97
26-dez-97
DATA


função de auto-corelação de PM10
Autocorrelation
1,0
0,8
0,6
0,4
0,2
0,0
-0,2
-0,4
-0,6
-0,8
-1,0
1
5
10
15
20
25
30
35
40
45
50
55
60
Lag

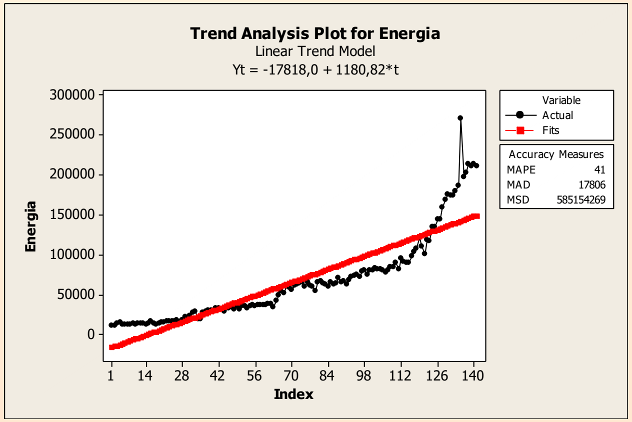
# Componentes de uma série temporal

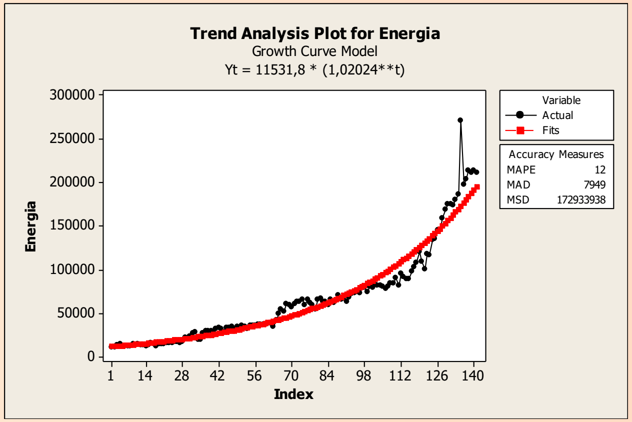
- Tendência: T(t);
- Sazonalidade: S(t)
- Ruído branco: a(t):

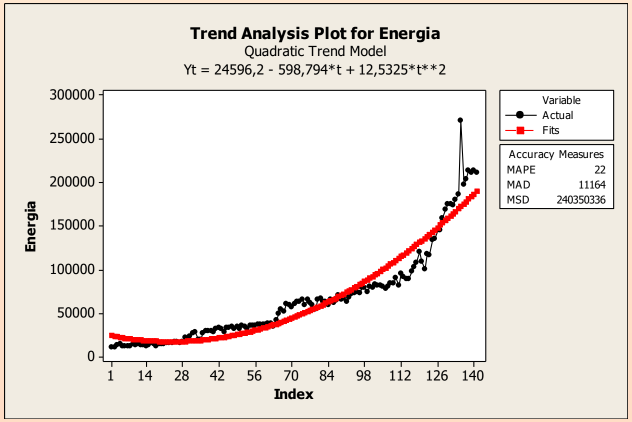
$$X(t) = T(t) + S(t) + a(T)$$

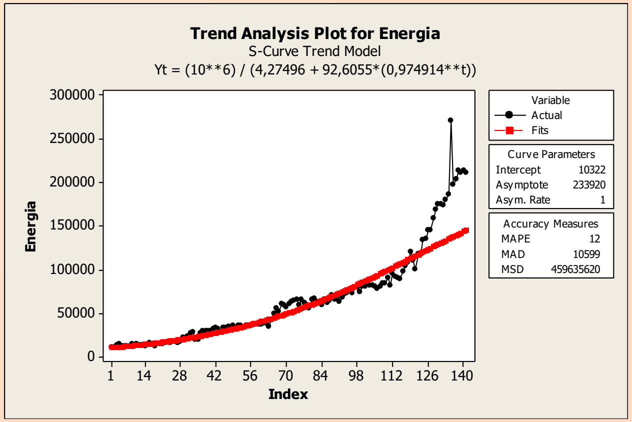
# Tendência

- Ajustar uma função polinomial do tempo(estimar um modelo de regressão)
- Suavizar os valores da série ao redor de um ponto, para estimar a tendência naquele ponto;
- Suavizar os valores da série através de sucessivos ajustes de retas de mínimos quadrados ponderados;
- Tomar diferença para eliminar a tendência.


Trend Analysis Plot for Energia
Linear Trend Model
Yt = -17818,0 + 1180,82*t
Energia
Index
Variable
Actual
Fits
Accuracy Measures
MAPE 41
MAD 17806
MSD 585154269

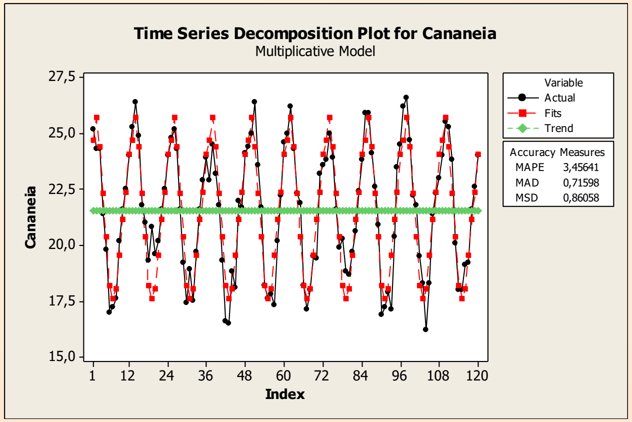

Trend Analysis Plot for Energia
Growth Curve Model
Yt = 11531,8 * (1,02024**t)
Energia
Index
Variable
Actual
Fits
Accuracy Measures
MAPE 12
MAD 7949
MSD 172933938

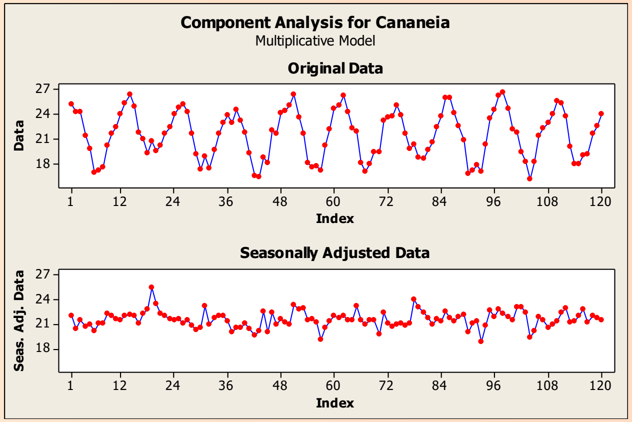

Trend Analysis Plot for Energia
Quadratic Trend Model
Yt = 24596,2 - 598,794*t + 12,5325*t**2
Energia
Index
Variable
Actual
Fits
Accuracy Measures
MAPE 22
MAD 11164
MSD 240350336

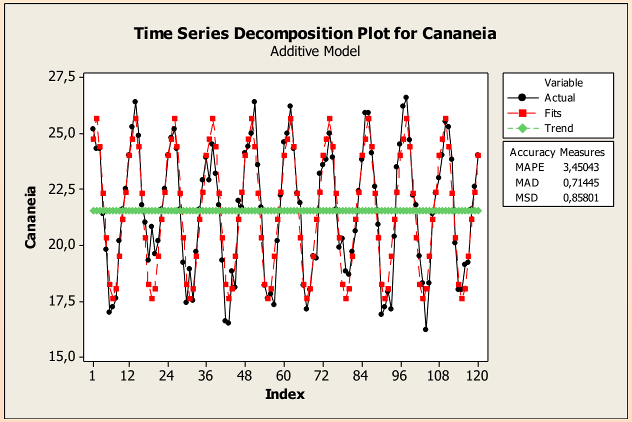

Trend Analysis Plot for Energia
S-Curve Trend Model
Yt = (10**6) / (4,27496 + 92,6055*(0,974914**t))
Energia
Index
Variable
Actual
Fits
Curve Parameters
Intercept 10322
Asymptote 233920
Asym. Rate 1
Accuracy Measures
MAPE 12
MAD 10599
MSD 459635620

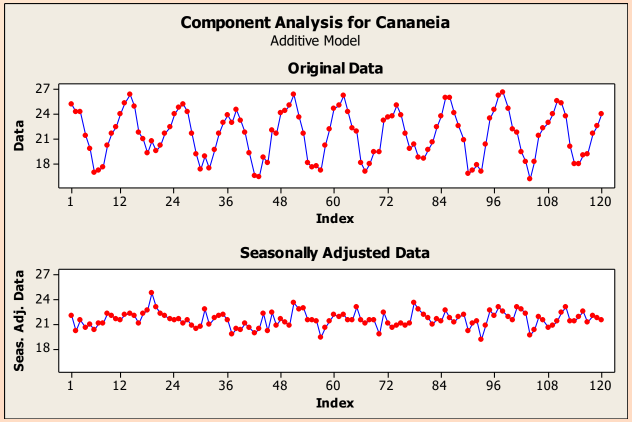
# Sazonalidade

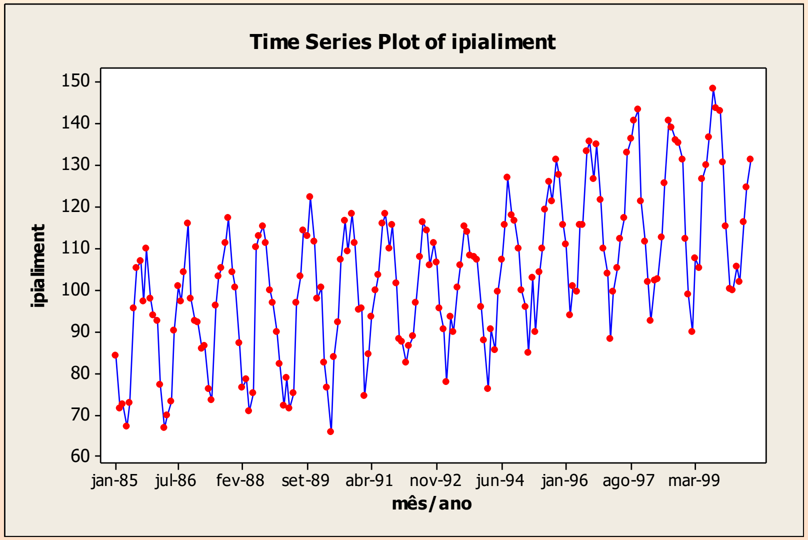
- Defini-se um fenômeno sazonal aquele que ocorre regularmente em períodos fixos de tempo.
- Sazonalidade determinística – método de regressão que incorporem funções do tipo seno ou cosseno à variável tempo.
- Sazonalidade estocástica: método de médias móveis.


Time Series Decomposition Plot for Cananeia
Multiplicative Model
Cananeia
Index
Variable
Actual
Fits
Trend
Accuracy Measures
MAPE 3,45641
MAD 0,71598
MSD 0,86058

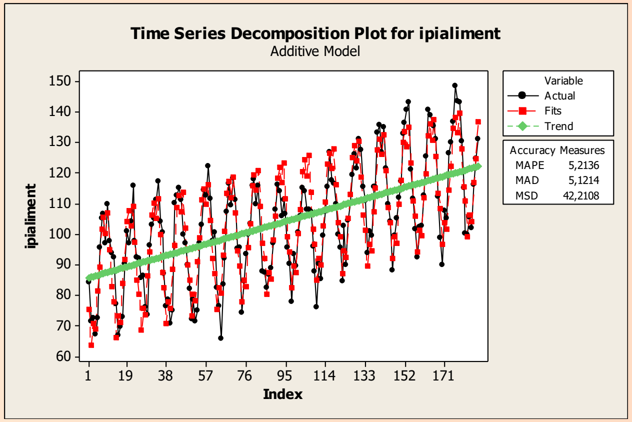

Component Analysis for Cananeia
Multiplicative Model
Original Data
Data
Index
Seasonally Adjusted Data
Seas. Adj. Data
Index


Time Series Decomposition Plot for Cananeia
Additive Model
Cananeia
Index
Variable
Actual
Fits
Trend
Accuracy Measures
MAPE 3,45043
MAD 0,71445
MSD 0,85801

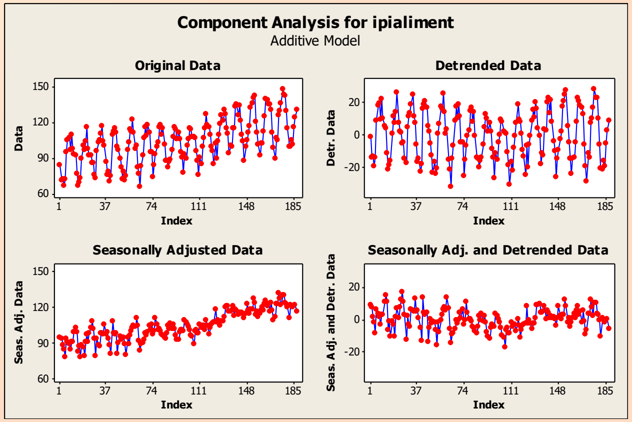

Component Analysis for Cananeia
Additive Model
Original Data
Data
Index
Seasonally Adjusted Data
Seas. Adj. Data
Index



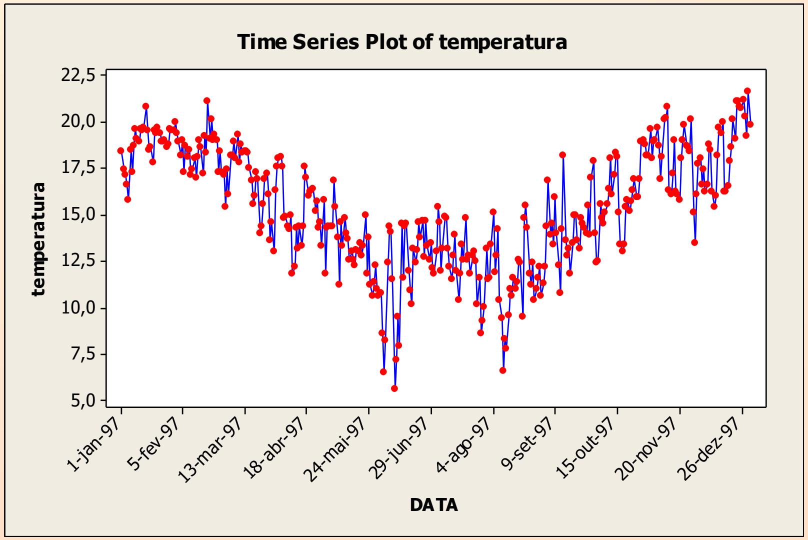
Time Series Plot of ipialiment
ipialiment
mês/ano
jan-85
jul-86
fev-88
set-89
abr-91
nov-92
jun-94
jan-96
ago-97
mar-99

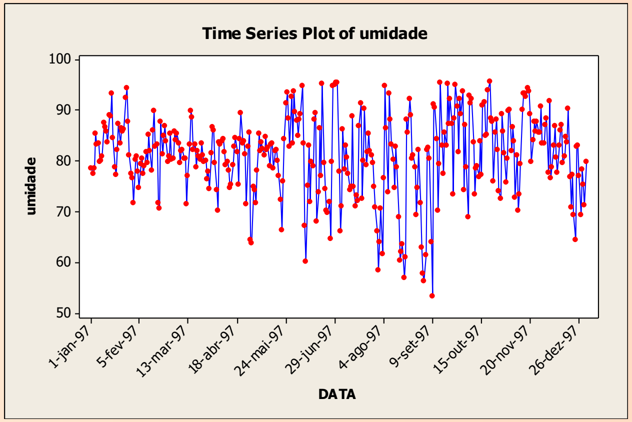



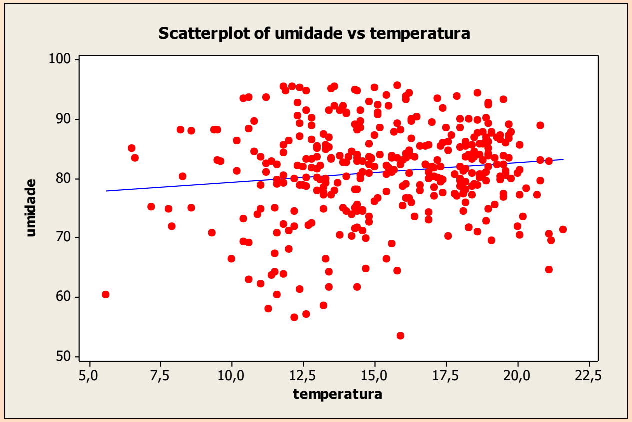
Time Series Decomposition Plot for ipialiment
Additive Model
ipialiment
Index
Variable
Actual
Fits
Trend
Accuracy Measures
MAPE 5,2136
MAD 5,1214
MSD 42,2108

Component Analysis for ipialiment
Additive Model
Original Data
Detrended Data
Seasonally Adjusted Data
Seasonally Adj. and Detrended Data
Data
Detr. Data
Seas. Adj. Data
Seas. Adj. and Detr. Data
Index


Time Series Plot of temperatura
temperatura
22,5
20,0
17,5
15,0
12,5
10,0
7,5
5,0
1-jan-97
5-fev-97
13-mar-97
18-abr-97
24-mai-97
29-jun-97
4-ago-97
9-set-97
15-out-97
20-nov-97
26-dez-97
DATA


Time Series Plot of umidade
umidade
100
90
80
70
60
50
1-jan-97
5-fev-97
13-mar-97
18-abr-97
24-mai-97
29-jun-97
4-ago-97
9-set-97
15-out-97
20-nov-97
26-dez-97
DATA


Scatterplot of umidade vs temperatura
umidade
100
90
80
70
60
50
5,0
7,5
10,0
12,5
15,0
17,5
20,0
22,5
temperatura

# Modelos Estatísticos de Séries Temporais

**Example 1.8 White Noise**

A simple kind of generated series might be a collection of uncorrelated random variables, $w_t$, with mean 0 and finite variance $\sigma_w^2$. The time series generated from uncorrelated variables is used as a model for noise in engineering applications, where it is called *white noise*; we shall sometimes denote this process as $w_t \sim wn(0, \sigma_w^2)$. The designation white originates from the analogy with white light and indicates that all possible periodic oscillations are present with equal strength.

**Example 1.9 Moving Averages**

We might replace the white noise series $w_t$ by a moving average that smooths the series. For example, consider replacing $w_t$ in Example 1.8 by an average of its current value and its immediate neighbors in the past and future. That is, let

$$v_t = \tfrac{1}{3}(w_{t-1} + w_t + w_{t+1}), \tag{1.1}$$

which leads to the series shown in the lower panel of Figure 1.8. Inspecting the series shows a smoother version of the first series, reflecting the fact that the slower oscillations are more apparent and some of the faster oscillations are taken out. We begin to notice a similarity to the SOI in Figure 1.5, or perhaps, to some of the fMRI series in Figure 1.6.

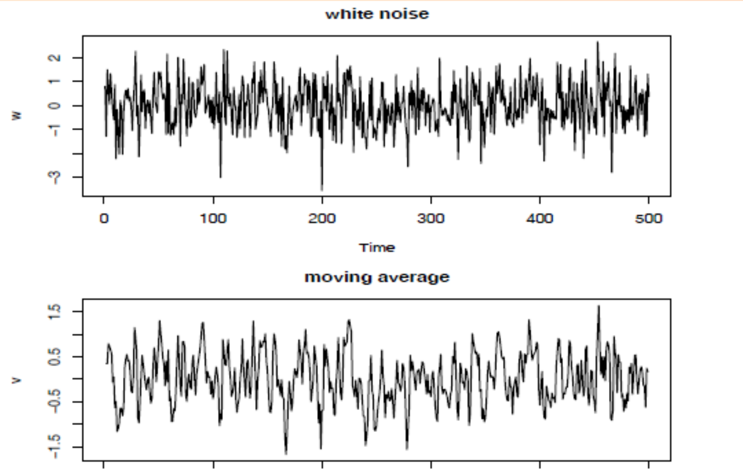



**Fig. 1.8.** Gaussian white noise series (top) and three-point moving average of the Gaussian white noise series (bottom).

# Modelos Estatísticos de Séries Temporais

**Example 1.10 Autoregressions**
Suppose we consider the white noise series $w_t$ of Example 1.8 as input and calculate the output using the second-order equation

$$x_t = x_{t-1} - .9x_{t-2} + w_t \qquad (1.2)$$

successively for $t = 1, 2, \ldots, 500$. Equation (1.2) represents a regression or prediction of the current value $x_t$ of a time series as a function of the past two values of the series, and, hence, the term autoregression is suggested

**Example 1.11 Random Walk with Drift**
A model for analyzing trend such as seen in the global temperature data in Figure 1.2, is the random walk with drift model given by

$$x_t = \delta + x_{t-1} + w_t \qquad (1.3)$$

for $t = 1, 2, \ldots$, with initial condition $x_0 = 0$, and where $w_t$ is white noise. The constant $\delta$ is called the drift, and when $\delta = 0$, (1.3) is called simply a random walk. The term random walk comes from the fact that, when $\delta = 0$, the value of the time series at time $t$ is the value of the series at time $t-1$ plus a completely random movement determined by $w_t$. Note that we may rewrite (1.3) as a cumulative sum of white noise variates. That is,

$$x_t = \delta\, t + \sum_{j=1}^{t} w_j \qquad (1.4)$$

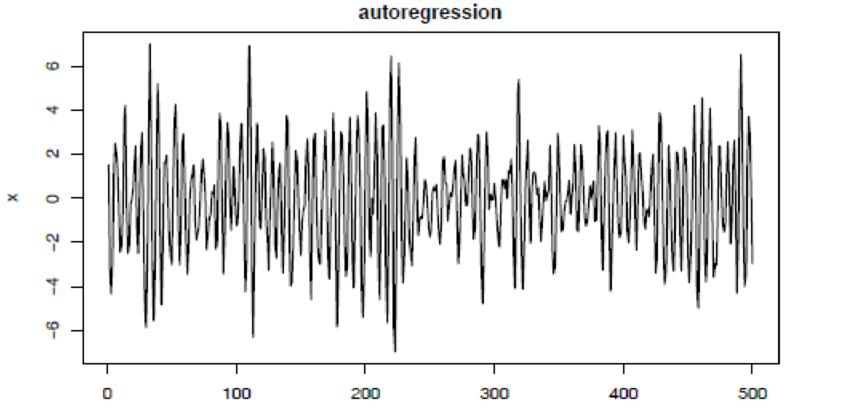


**Fig. 1.9.** Autoregressive series generated from model (1.2).

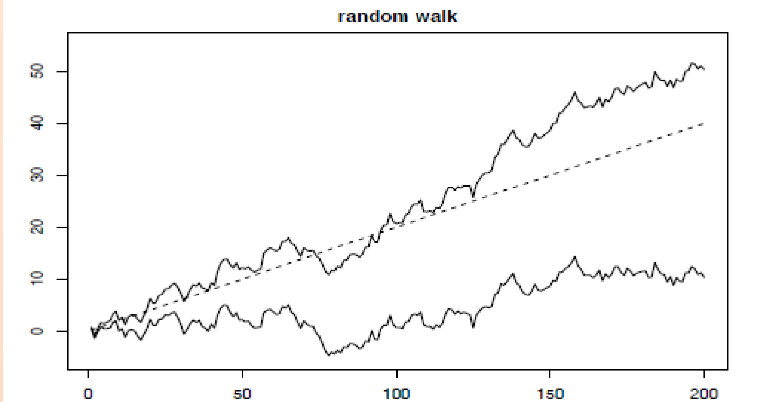


**Fig. 1.10.** Random walk, $\sigma_w = 1$, with drift $\delta = .2$ (upper jagged line), without drift, $\delta = 0$ (lower jagged line), and a straight line with slope .2 (dashed line).

# Modelos Estatísticos de Séries Temporais

**Example 1.12 Signal in Noise**

Many realistic models for generating time series assume an underlying signal with some consistent periodic variation, contaminated by adding a random noise. For example, it is easy to detect the regular cycle fMRI series displayed on the top of Figure 1.6. Consider the model

$$x_t = 2\cos(2\pi t/50 + .6\pi) + w_t \tag{1.5}$$

for $t = 1, 2, \ldots, 500$, where the first term is regarded as the signal, shown in the upper panel of Figure 1.11. We note that a sinusoidal waveform can be written as

$$A\cos(2\pi\omega t + \phi), \tag{1.6}$$

where $A$ is the amplitude, $\omega$ is the frequency of oscillation, and $\phi$ is a phase shift. In (1.5), $A = 2$, $\omega = 1/50$ (one cycle every 50 time points), and $\phi = .6\pi$.

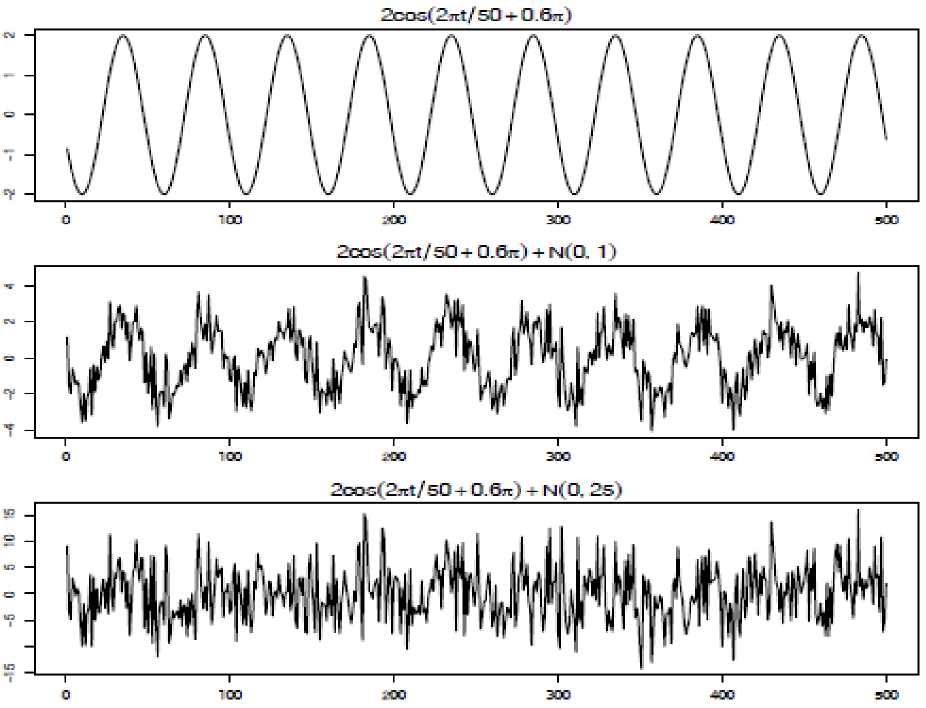



**Fig. 1.11.** Cosine wave with period 50 points (top panel) compared with the cosine wave contaminated with additive white Gaussian noise, $\sigma_w = 1$ (middle panel) and $\sigma_w = 5$ (bottom panel); see (1.5).

# Bibliografia


- Shumway, R.H. and Stoffer, D.S. (2010). Time Series Analysis and Its Applications with R Examples. Springer. 3rd Edtion.
- Morettin, P.A e Toloi, C.M.C. (2004). Análise de Séries Temporais. Edgard Blucher.