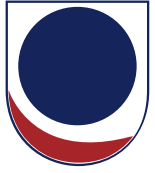




# Introdução a Séries Temporais

Prof. Eraylson Galdino



# Agenda

---

- O que são séries temporais
- Análise de séries temporais:
  - Objetivo
  - Aplicações
- Modelos de Séries Temporais
- Componentes
- Função de autocorrelação
- Correlograma
- Processo estacionário



# Séries Temporais: Definição

---



George Box & Gwilym Jenkins

*Uma série temporal é um conjunto de **observações ordenadas**,  $x_t$ , cada uma observada em um instante de tempo.*



Peter J Brockwell

*Uma série temporal é um conjunto de observações,  $x_t$ , cada uma **registrada num tempo específico  $t$ .***



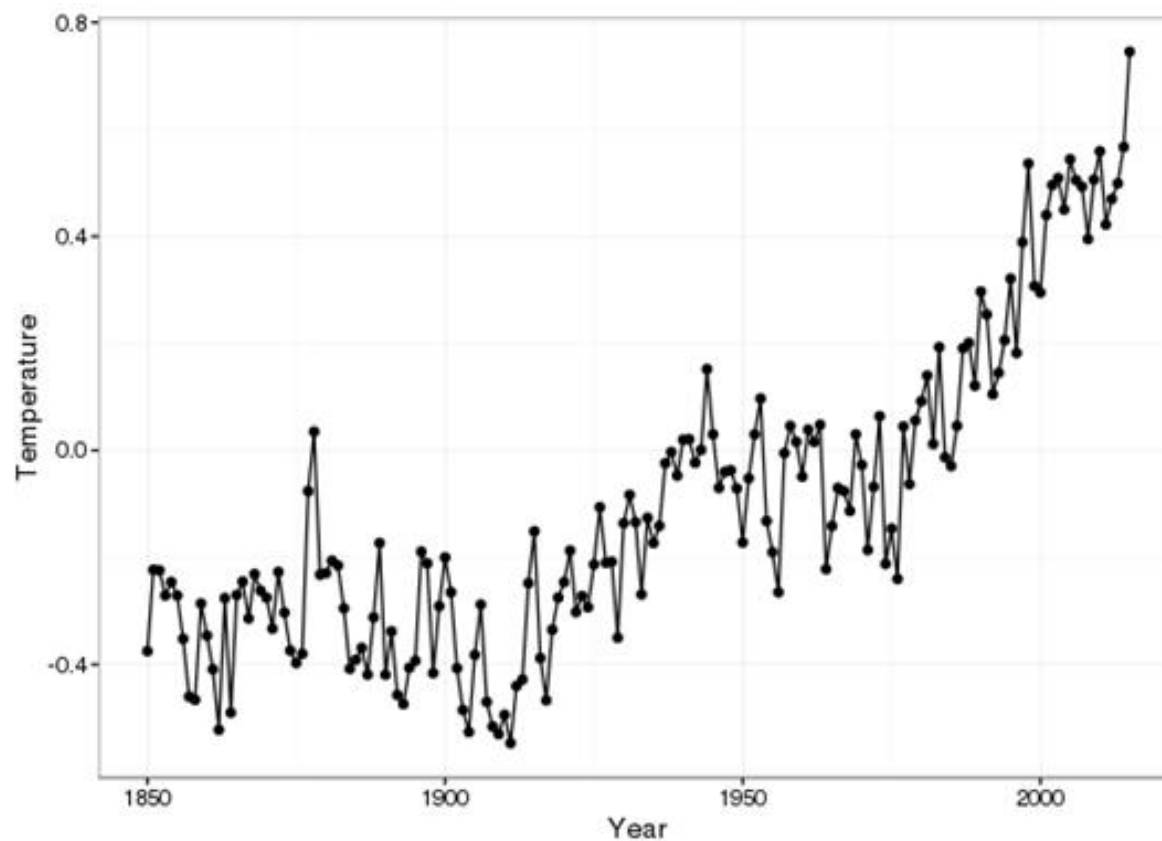
# Séries Temporais: Definição

---

- **Séries temporais discretas:** são séries em que o intervalo de observações ( $t$ ) pertence a um conjunto discreto. Ou seja, as observações são feitas em intervalos de tempo fixos;
- **Séries temporais contínuas:** são séries em que as observações são obtidas continuamente através de algum intervalo no tempo, por exemplo, quando  $T = [0, 1]$ ;



# Séries Temporais: exemplo

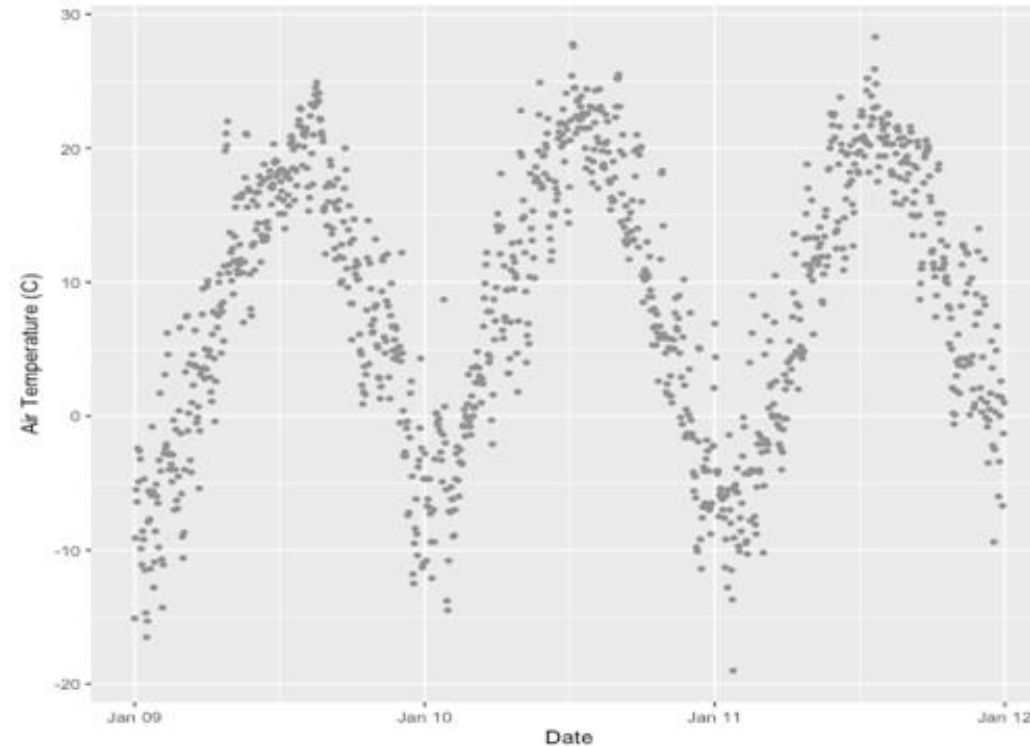


Exemplo de uma série temporal **discreta**



# Séries Temporais: exemplo

---



Exemplo de uma série temporal **contínua**



# Análise de Séries Temporais

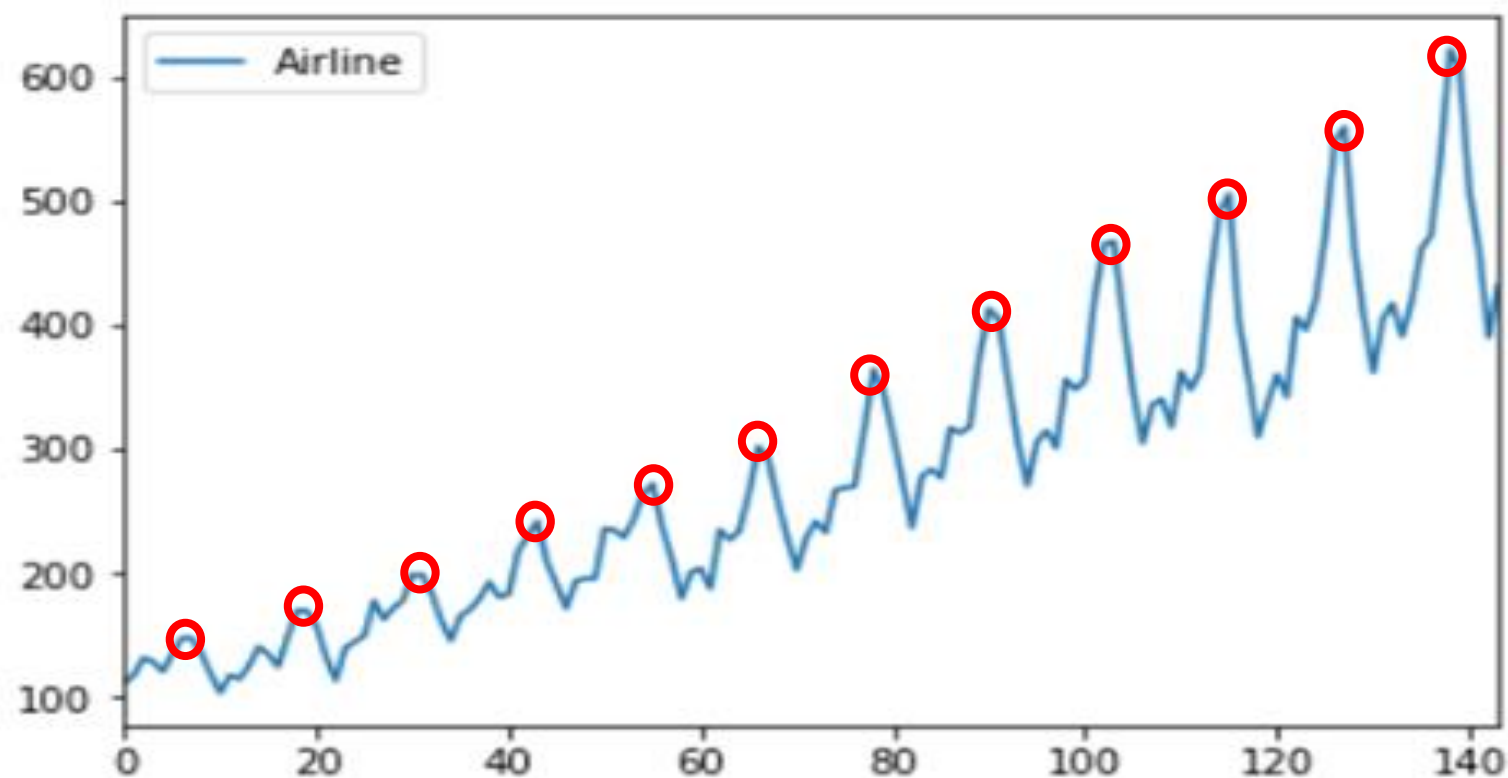
---

- Objetivos:
  - **Compreender** as **características** de tal fenômeno temporal;
  - **Selecionar** e **estimar** um modelo estocástico que possivelmente tenha gerado o conjunto de dados;
- Aplicações:
  - Entendimento de um comportamento temporal;
  - Classificar um comportamento temporal;
  - Detecção de anomalias;
  - Previsão;



# Análise de Séries Temporais

- Entendimento de um comportamento temporal

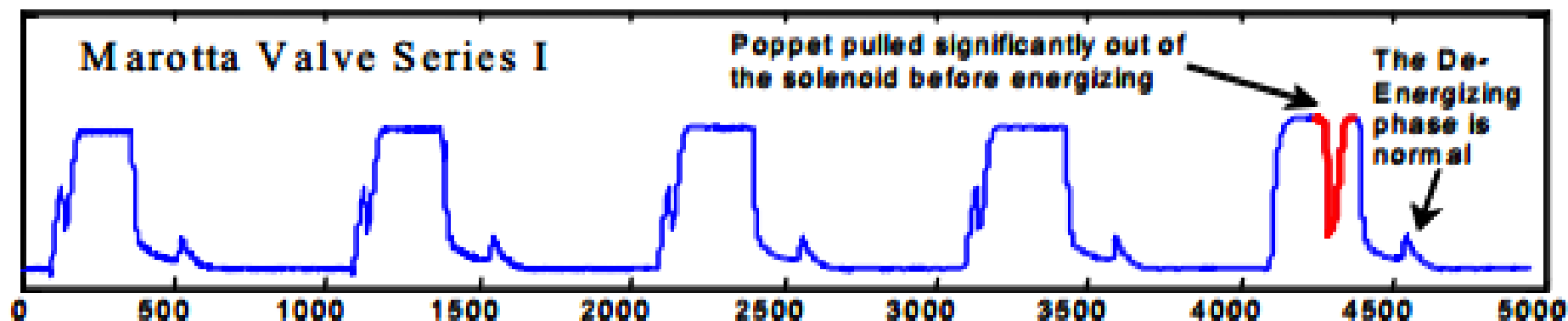






# Análise de Séries Temporais

- Detecção de anomalias

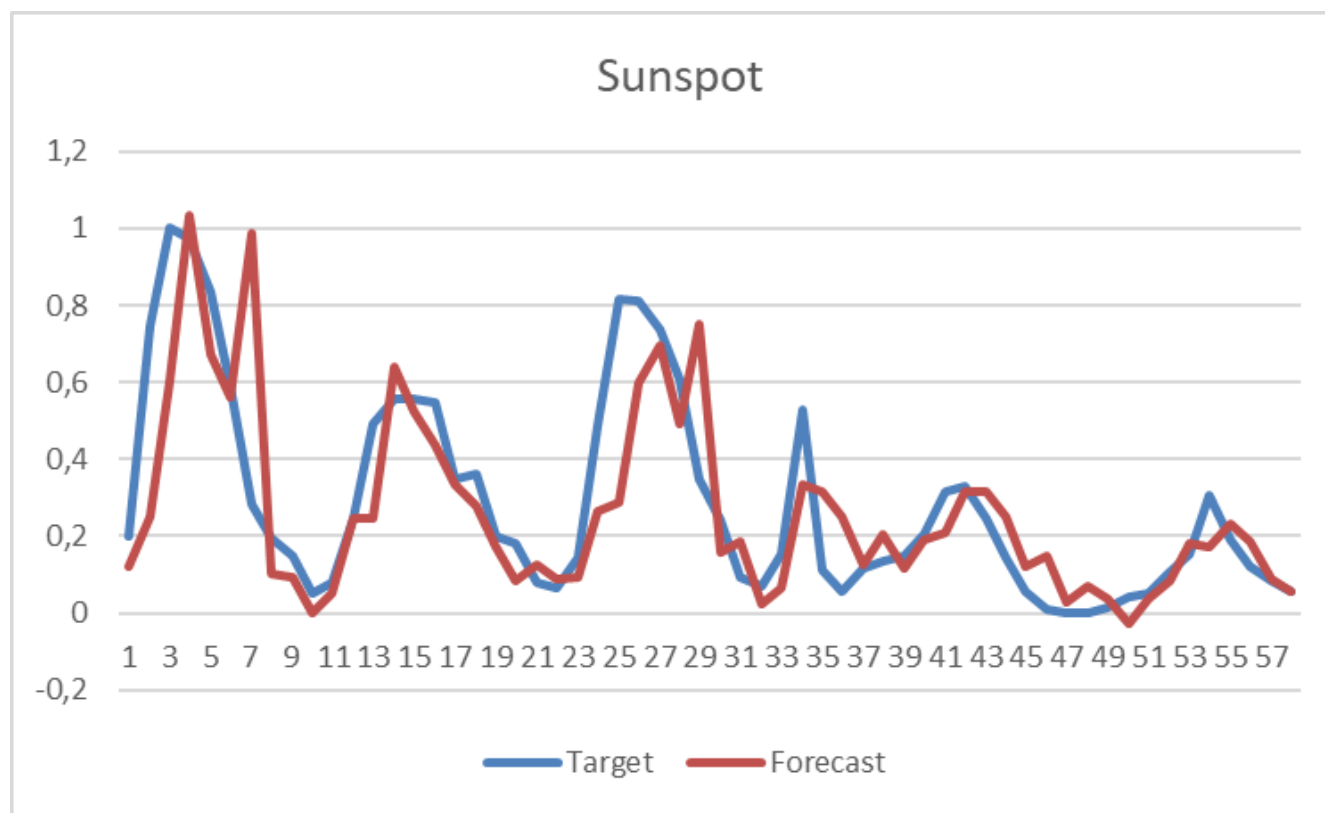


**Figure 6: An example of an annotated Marotta Valve time series. The discord discovered (highlighted in bold) exactly corresponds with the expert's annotation**



# Análise de Séries Temporais

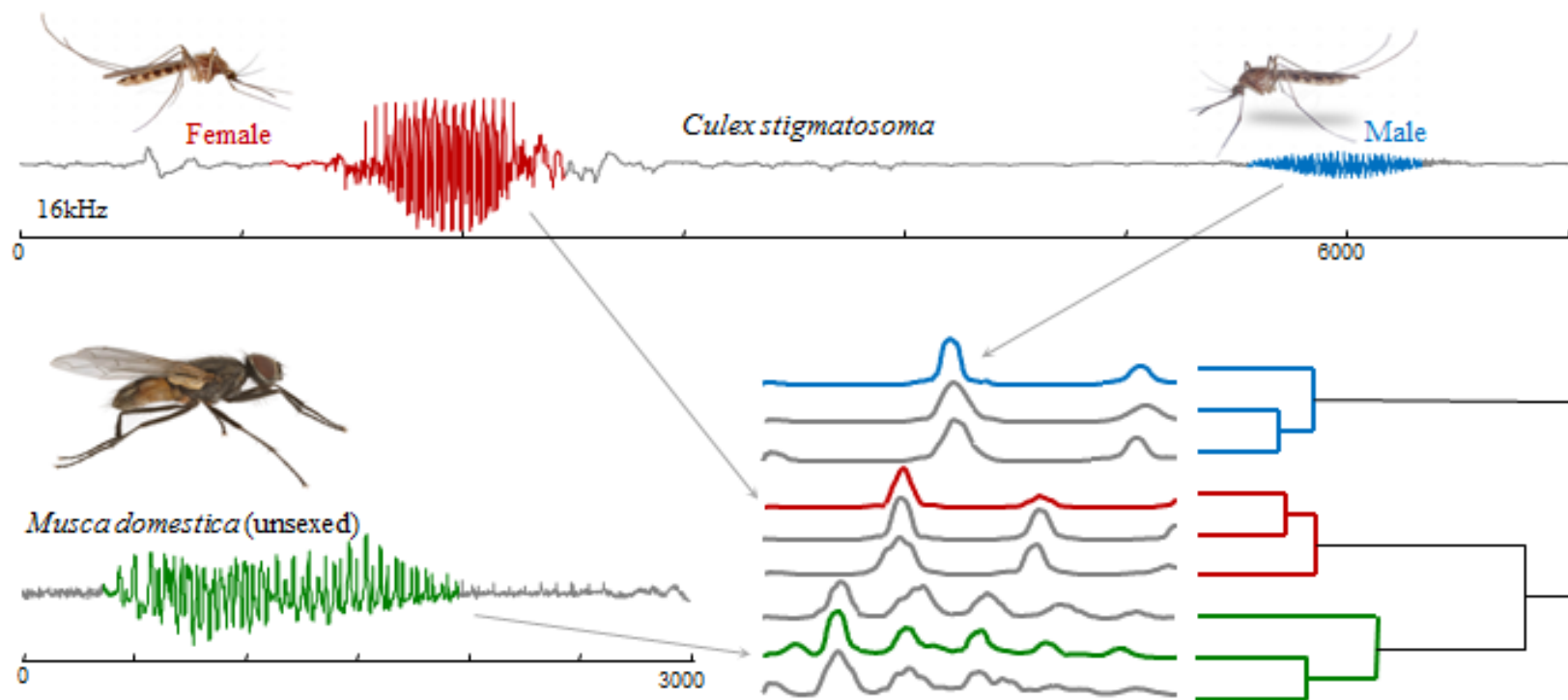
- Previsão

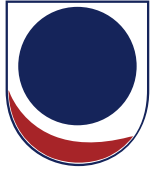




# Análise de Séries Temporais

- Classificação





# Análise de Séries Temporais

---

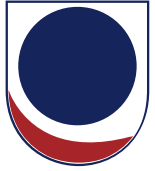
- Campos de Aplicação:
  - Mercado Financeiro;
  - Meteorologia;
  - Medicina;
  - Controle de Qualidade e Processos;
  - Governo;
  - Epidemiologia;
  - Transporte;



# Análise de Séries Temporais

---

- No geral, análise de Séries Temporais consiste em:
  - Analisar um conjunto de dados (ao longo do tempo);
  - Selecionar e estimar um modelo matemático que possivelmente tenha gerado o conjunto de dados;



# Modelos de Séries Temporais

---



George Box

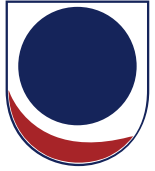
*Todos os modelos estão errados, mas  
alguns são úteis.*



# Modelos Simples

---

- Selecionar um modelo probabilístico adequado para os dados é uma das partes mais importantes da análise de séries temporais;
- Geralmente, é suposto que cada observação  $x$  seja um valor resultante de uma determinada variável aleatória  $X$ ;
- Um modelo de série temporal para um determinado dado observado  $x(t)$  é a especificação da composição de distribuições da sequência da variável aleatória  $X(t)$  em que  $x(t)$  é uma observação;



# Modelos Simples

---

- Modelos com média zero:
  - Ruído i.i.d. (independente e identicamente distribuído);
  - Processo binário;
  - Random Walk (Passeio Aleatório);
- Modelos com tendência e sazonalidade

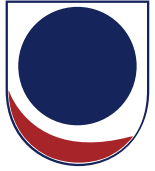




# Modelos Simples

---

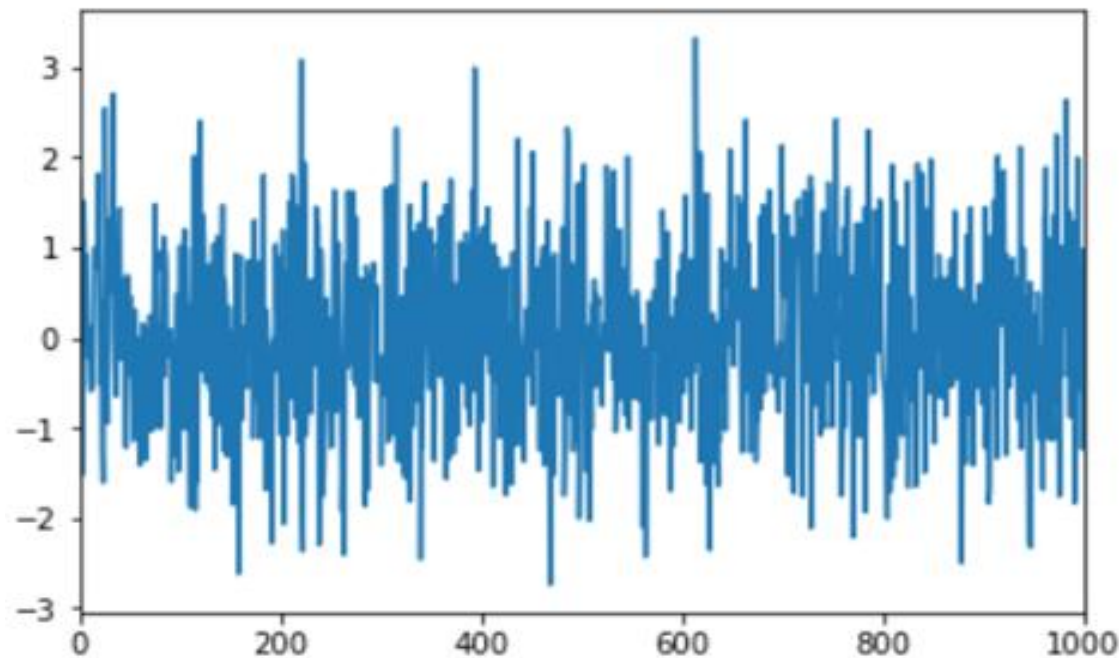
- Modelo com média zero: Ruído i.i.d
  - Modelo mais simples;
  - Sem os componentes de tendência e sazonalidade;
  - As observações são resultados de variáveis aleatórias i.i.d. com média zero;
  - Não existe dependência entre as observações;
  - Através do valor  $x(t)$  não é possível prever o valor de  $x(t+h)$ ;
  - Apesar de não ser interessante para predição, é um **modelo importante para construção e entendimento** de modelos mais complexos;



# Modelos Simples

---

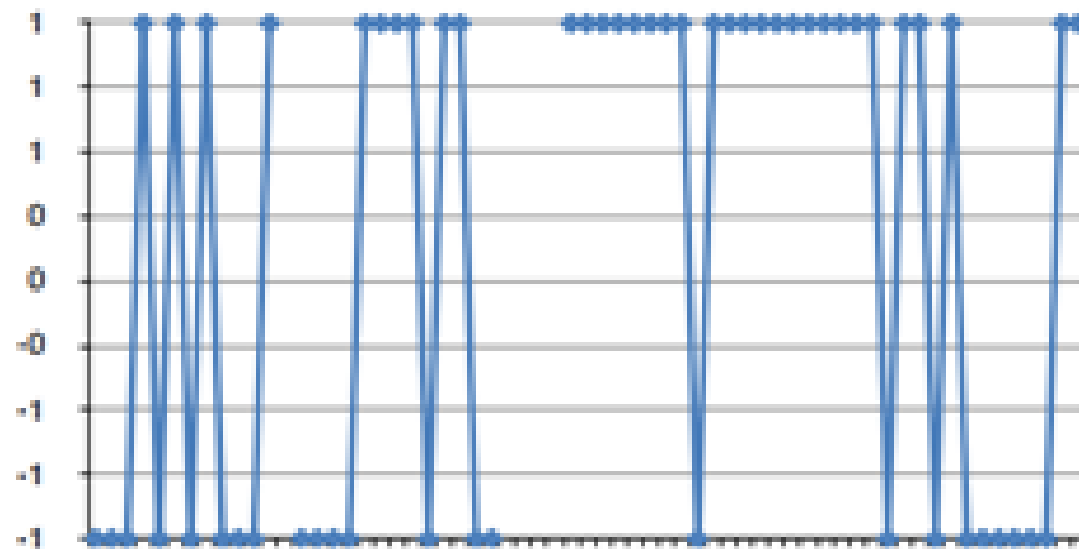
- Um ruído i.i.d com distribuição normal com média zero e variância  $\sigma^2$  é também chamado de ruído branco gaussiano;





# Modelos Simples

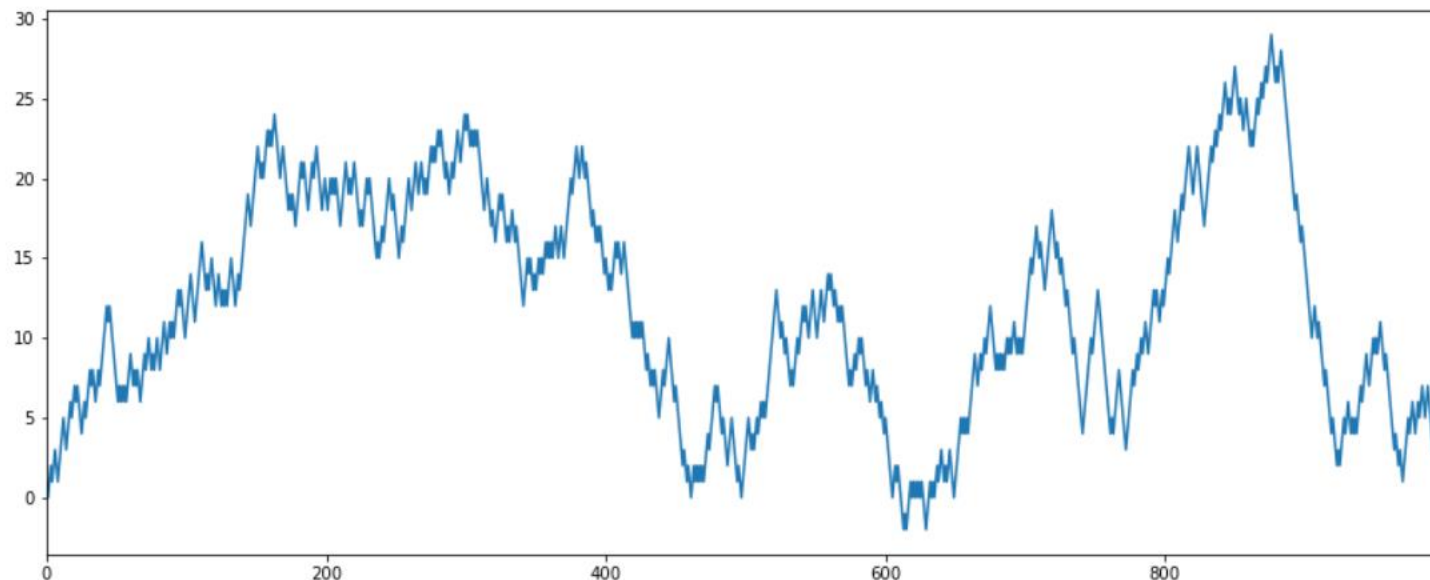
- Processo binário:  
As observações só podem assumir dois valores possíveis;





# Modelos Simples

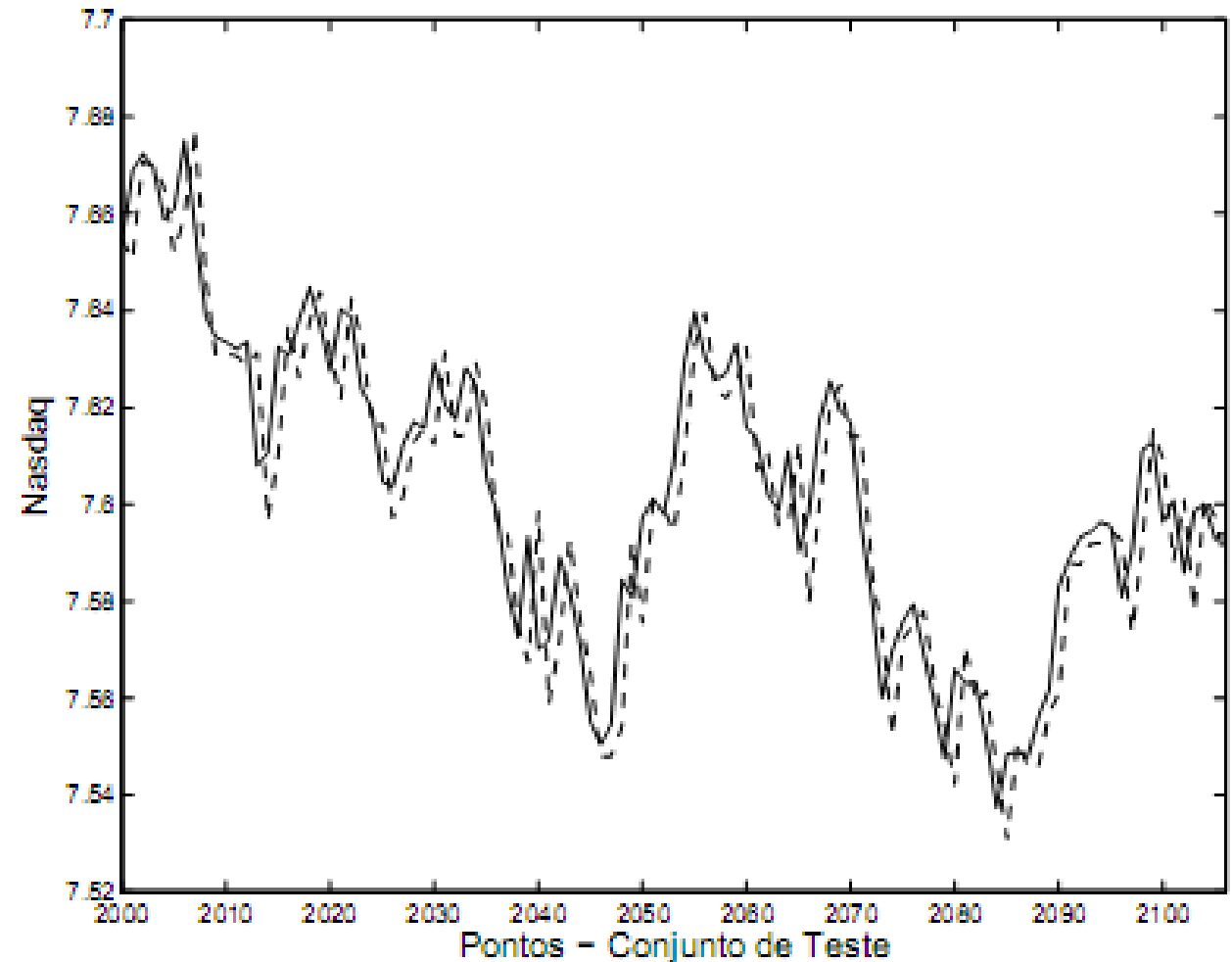
- Random Walk:
  - Ideia remete a “caminhada do bêbado”. Tomada de vários passos consecutivos, cada qual em uma direção aleatória;
  - É obtido através da soma cumulativa de variáveis i.i.d aleatórias com o valor atual;

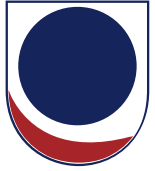




# Modelos Simples

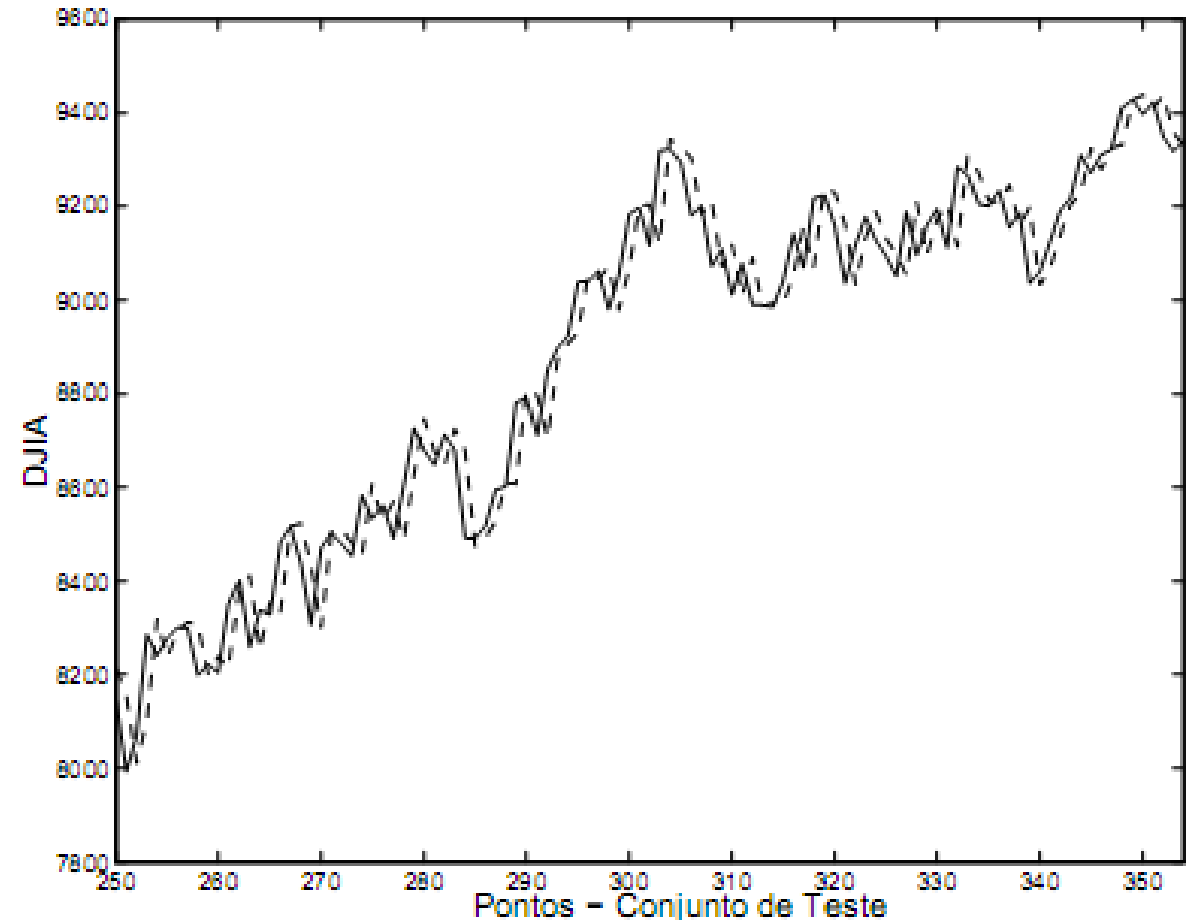
- Random Walk:
  - Problema de previsão





# Modelos Simples

- Random Walk:
  - Problema de previsão

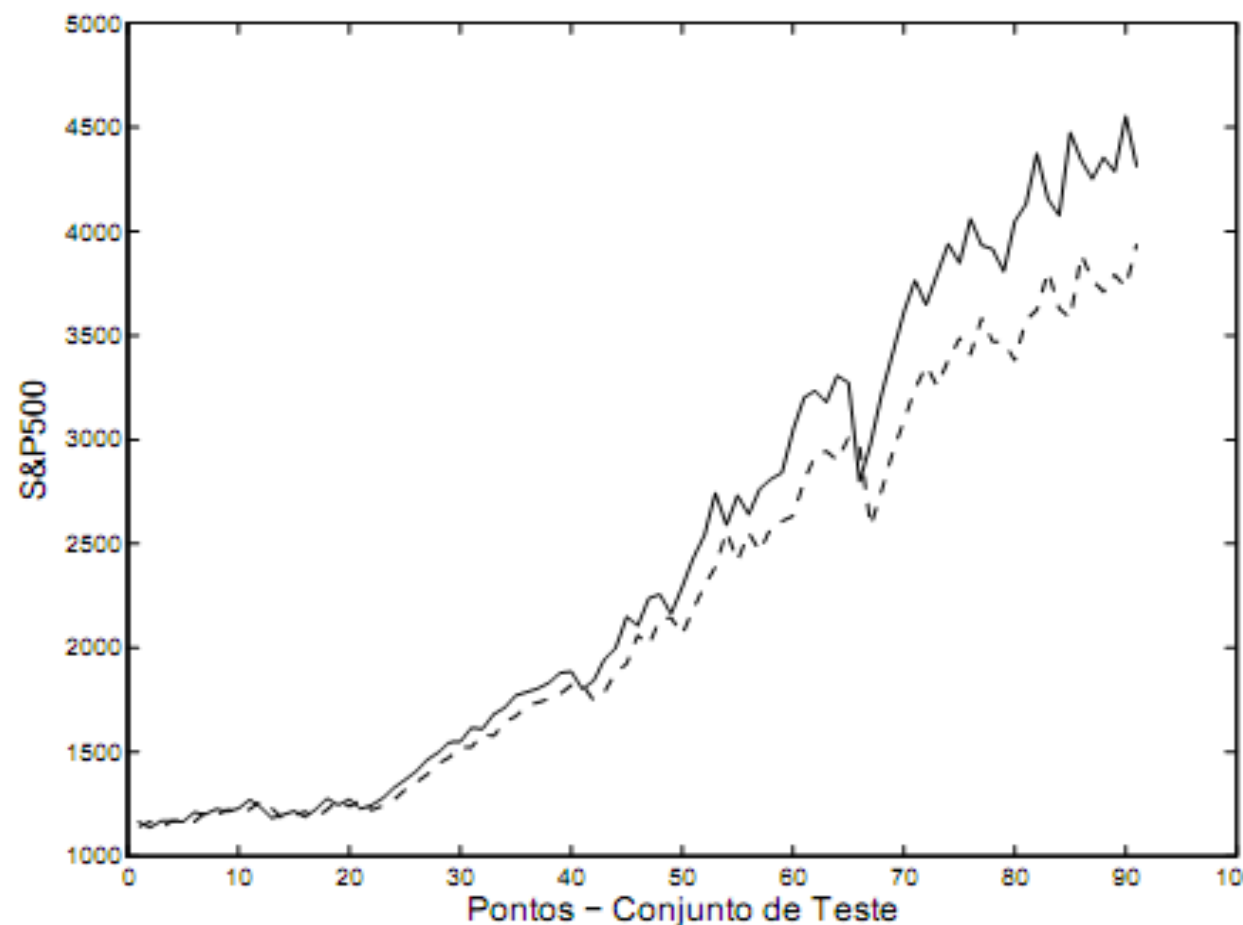


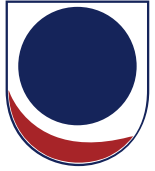


# Modelos Simples

- Random Walk:
  - Problema de previsão

Real	Previsão
0,685264	0,69427
0,690659	0,688232
0,681649	0,69365
0,685226	0,684601
0,691443	0,688193
0,700727	0,694438
0,707789	0,703763





# Modelos com tendência e sazonalidade

---

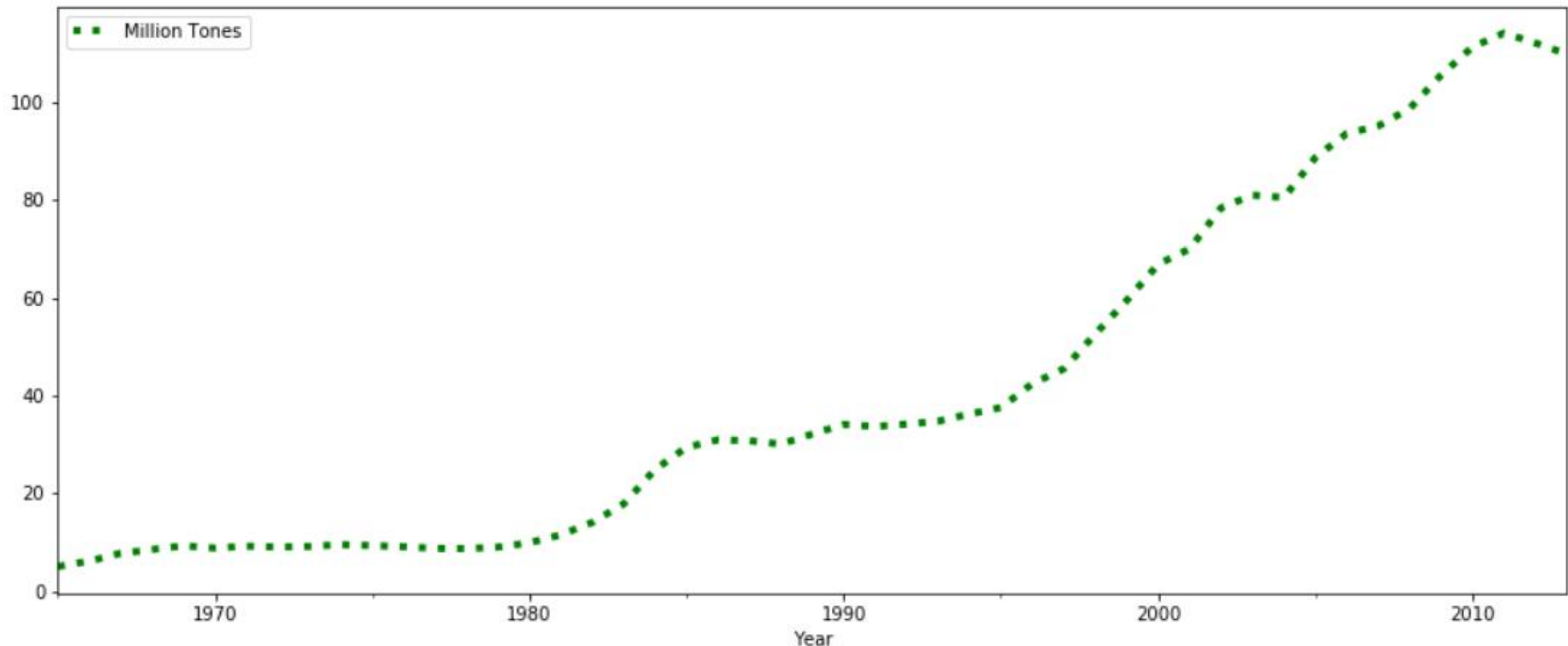
- Algumas séries é perceptível que **não podem ser modeladas** por modelos simples de média zero;
- São geradas com componentes de tendência e sazonalidade;
- **Tendência**: mudança sistemática na série temporal que não aparenta ser periódico;
- **Sazonalidade**: comportamento que se repete durante um período de tempo;

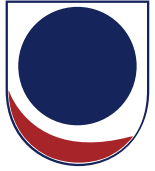




# Modelos com tendência e sazonalidade

- Série com tendência

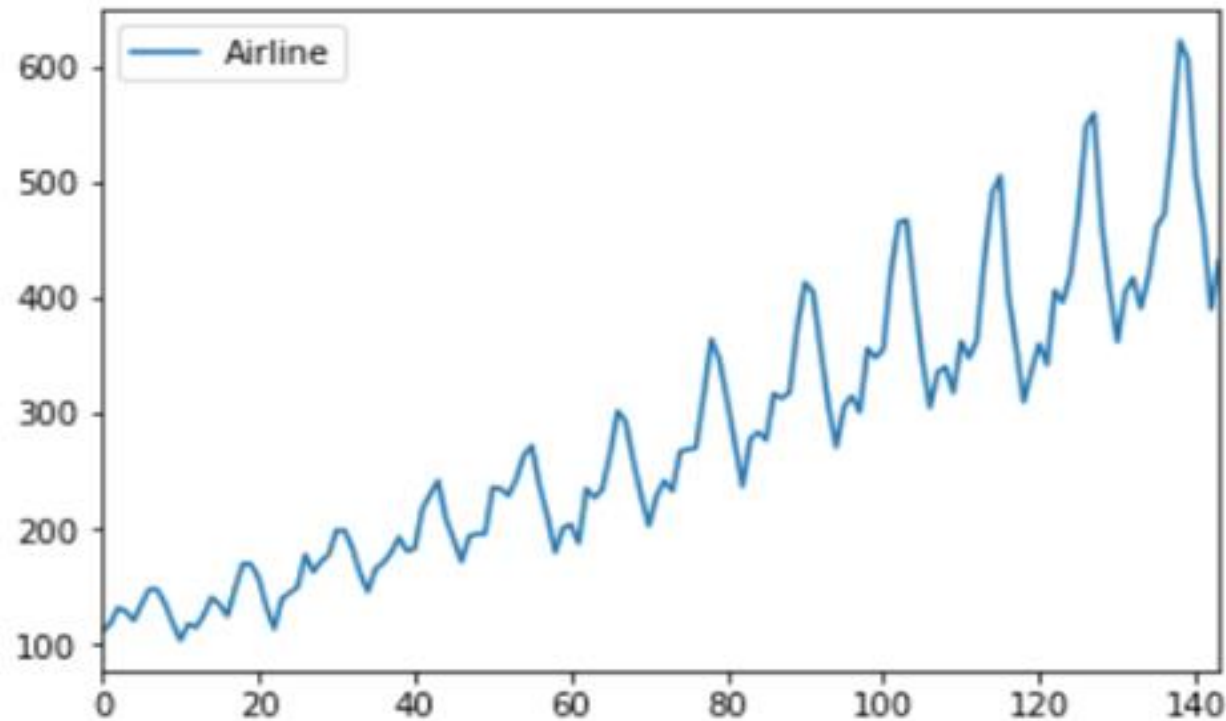


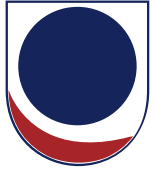


# Modelos com tendência e sazonalidade

---

- Série com sazonalidade

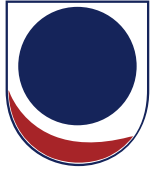




# Modelos com tendência e sazonalidade

---

- Modelo Aditivo:
  - $Z(t) = \text{Tendência} + \text{Sazonalidade} + \text{Ruído}$
  - As mudanças são consistentes
  - Uma tendência linear é uma linha reta
  - Uma sazonalidade linear tem a mesma frequência (comprimento dos ciclos) e amplitude (altura dos ciclos)



# Modelos com tendência e sazonalidade

---

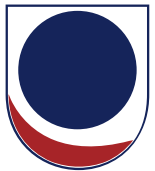
- Modelo Multiplicativo:
  - $Z(t) = \text{Tendência} * \text{Sazonalidade} * \text{Ruído}$
  - É não linear (quadrático, exponencial, etc. )
  - Mudanças ao longo do tempo aumentam ou diminuem
  - Uma tendência não linear é uma curva
  - Uma sazonalidade não-linear tem frequência e amplitude variável



# Componentes de séries temporais

---

- **Abstração** para entender melhor a série temporal;
- Não necessariamente para previsão:
  - Informações que podem ser aplicadas **na preparação de dados** e **na seleção dos modelos**;
  - **Remoção** de alguns desses componentes pode **aumentar** o desempenho do **modelo** de Machine Learning;
- Problemas do mundo real são caóticos e ruidosos:
  - Componentes aditivos e multiplicativos na mesma série;
  - Tendência de crescimento e decrescimento;
  - Sazonalidade variável;
  - Mudança de comportamento;



# Função de Autocorrelação (FAC)

---

- Seja  $X(t)$  uma série estacionária, a Função de Autocovariância do lag  $h$  é:

$$\gamma_x = Cov(X_{t+h}, X_t)$$

- A Função de Autocorrelação de  $X(t)$  do lag  $h$  é definida como:

$$\rho_x(h) \equiv \frac{\gamma_x(h)}{\gamma_x(0)} = Cor(X_{t+h}, X_t)$$

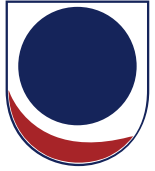
- A autocorrelação é a correlação entre uma série e ela mesma defasada;



# Função de Autocorrelação (FAC)

---

- Utilizada para analisar o grau de dependência temporal na série;
- A relação entre uma observação atual e observações é chamado de lags;
- Auxilia na seleção de possíveis modelos de séries temporais estacionárias;
- Através do coeficiente de correlação de Pearson:
  - -1 correlação negativa;
  - 1 correlação positiva;
  - 0 nenhuma correlação;



# Função de Autocorrelação Parcial

---

- As autocorrelações para intervalos sucessivos são formalmente dependentes;
- A F.A.C parcial é utilizada para obter uma informação sobre autocorrelação na série sem esta influência em cascata;
- A autocorrelação parcial de atraso  $k$  corresponde a autocorrelação entre  $x_t$  e  $x_{(t-k)}$  que não é explicada pelos atrasos de 1 a  $k$ ;

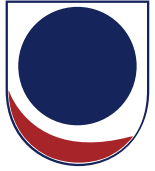




# Correlação x Causalidade

---

- Correlação:
  - Uma variável A tem um comportamento semelhante à B;
- Causalidade:
  - Uma variável B influencia o comportamento da variável B;
- <http://www.tylervigen.com/spurious-correlations>



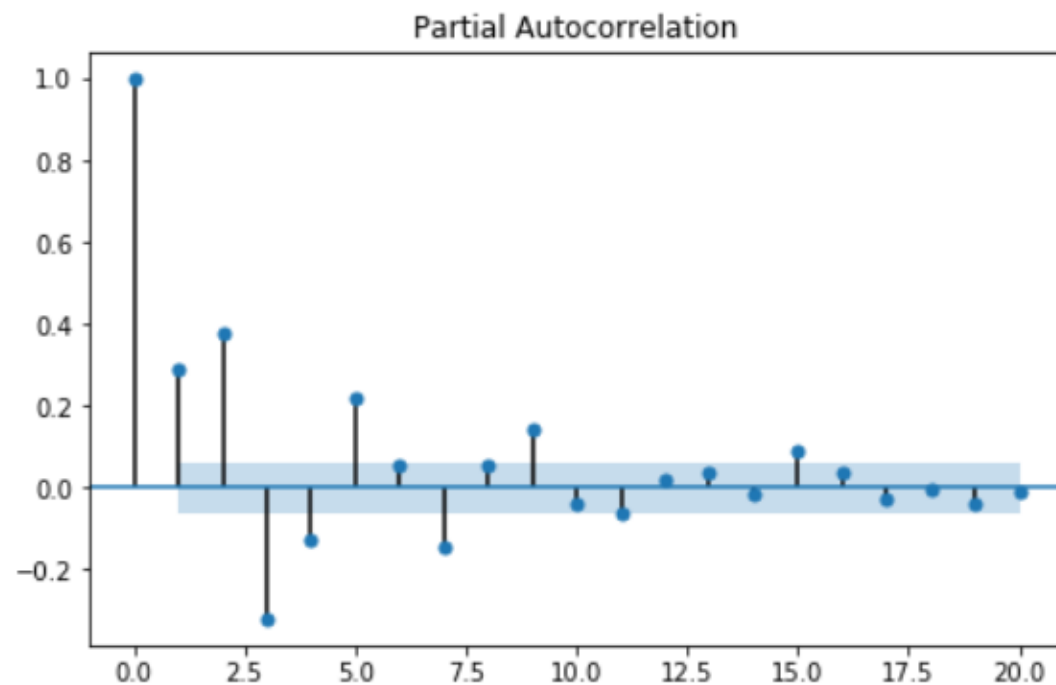
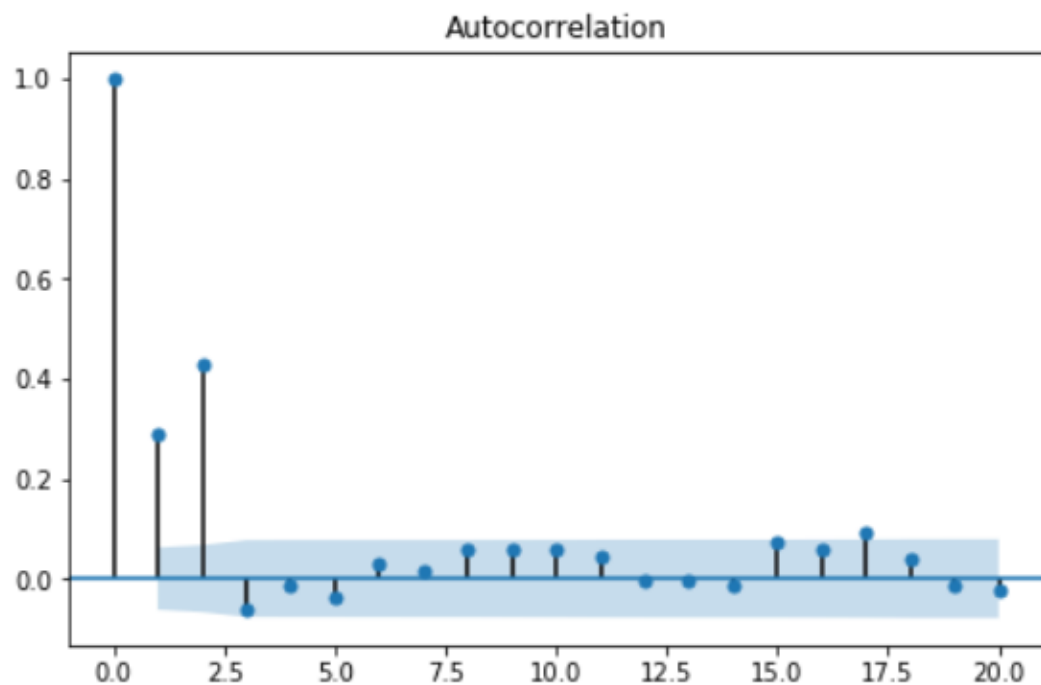
# Correlograma

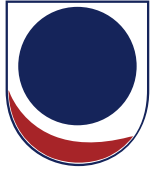
---

- Forma gráfica de analisar a autocorrelação;
- O correlograma traça as autocorrelações em diversas defasagens;
- Através da análise utilizando o correlograma é possível entender se a série é aleatória ou possui alguma tendência ou sazonalidade;
- Frequentemente utilizado para analisar os resíduos de um modelo;



# Correlograma





# Correlograma

---

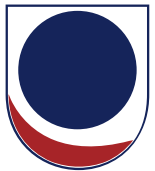
- Para inferir se em um dado lag existe ou não correlação é utilizado intervalos de confiança;
- Para um intervalo de confiança de 95% os limites são  $\pm 1.96/\sqrt{n}$
- Os lag com coeficiente fora do intervalo de confiança são considerados significantes.
- O coeficiente com valor 1 representa correlação máxima positiva;
- O coeficiente com valor -1 representa correlação máxima negativa;
- O coeficiente com valor 0 representa que não existe correlação;



# Correlograma

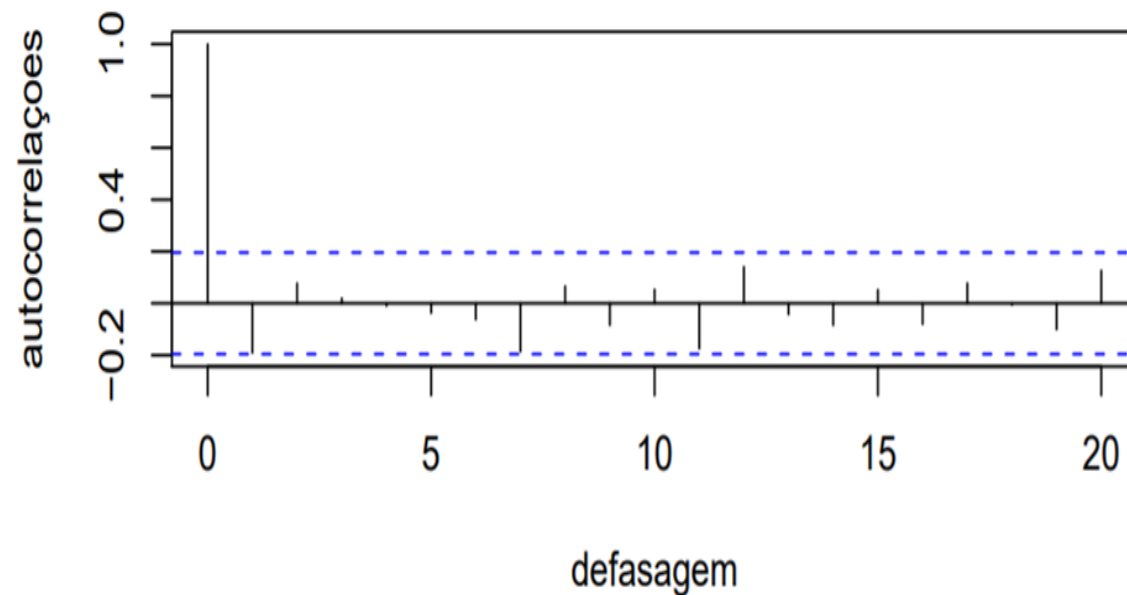
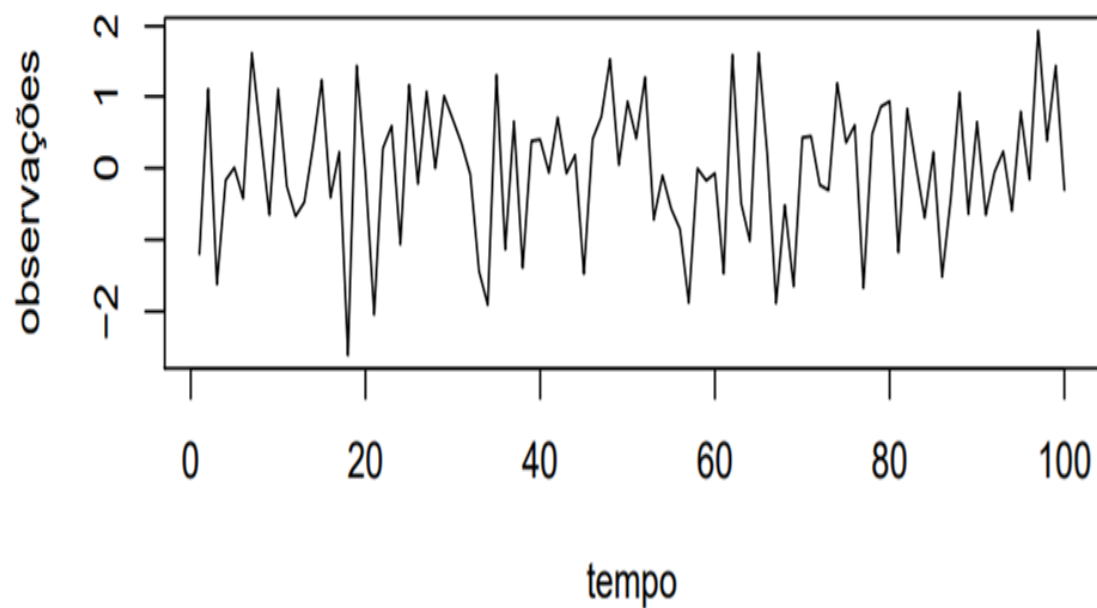
---

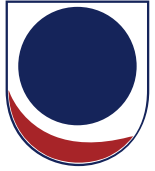
- Série aleatória:
  - Através do correlograma é possível analisar se uma série é aleatória ou não;
  - Em uma série completamente aleatória os lags são não correlacionados, ou seja, espera-se que o coeficiente de autocorrelação amostral  $k$  seja próximo à zero;



# Correlograma

- Série aleatória:





# Correlograma

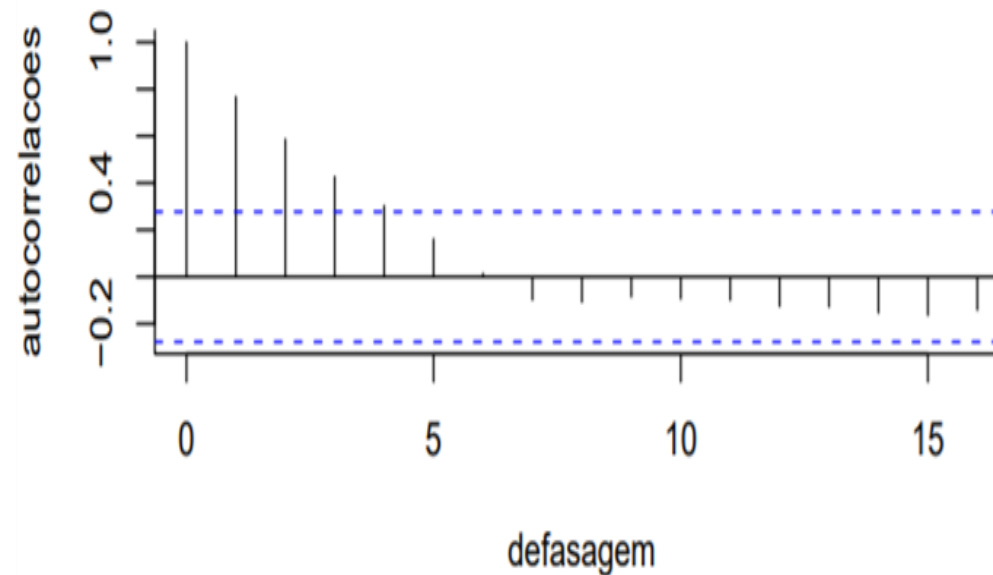
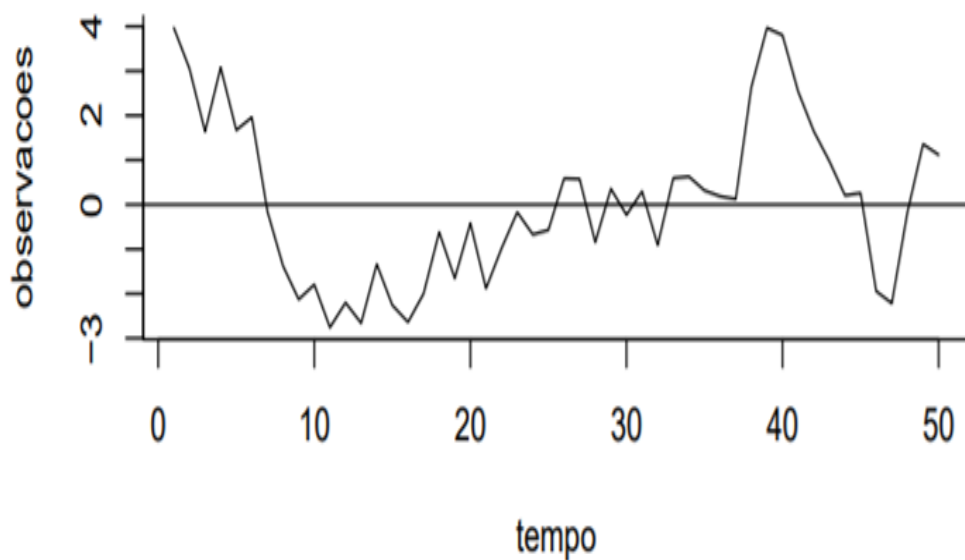
---

- Correlação de curto prazo:
  - É dita de **curto prazo quando** uma **observação acima da média** tende a ser **seguida** por uma ou **mais observações acima** da média. O mesmo ocorre para observações abaixo da média.
  - A partir de uma certa defasagem  $k$  os valores do coef de  $k$  tendem a ser aproximadamente zero.



# Correlograma

- Correlação de curto prazo:







# Correlograma

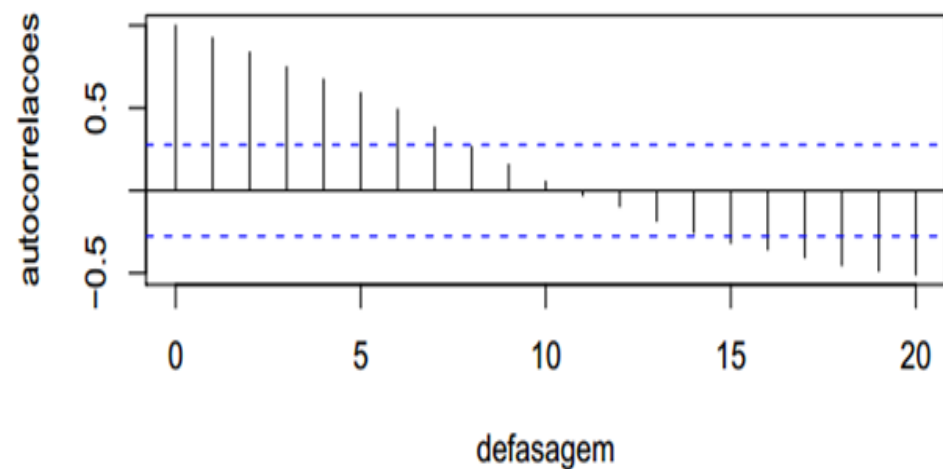
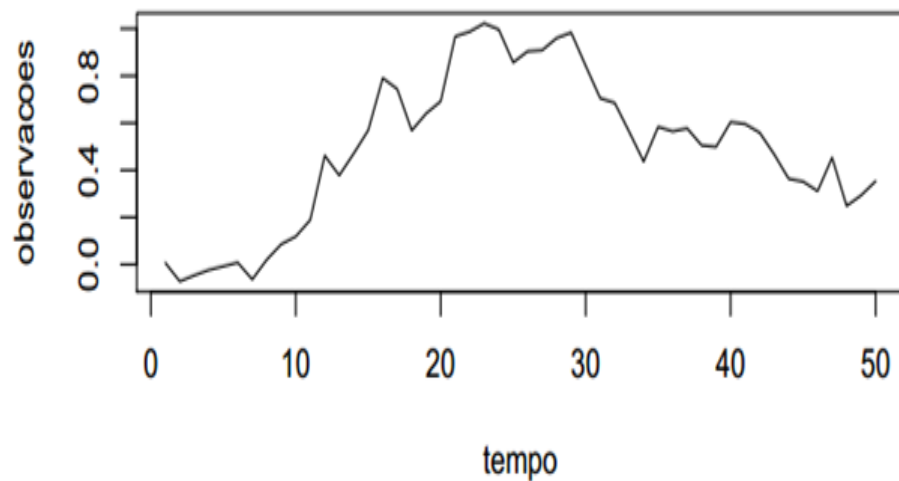
---

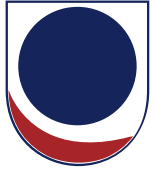
- Séries não estacionária:
  - Para uma série com tendência **os valores do coeficiente  $r$  não decaem para zero** a não ser em defasagens grandes. Isso ocorre **pois uma observação de um lado da média tende a ser seguida por um grande número de observações da mesma média** (mesmo lado) por conta da tendência;
  - Nesse caso, pouca ou nenhuma informação pode ser obtida do correlograma pois a tendência dominará outras características;



# Correlograma

- Séries não estacionária:





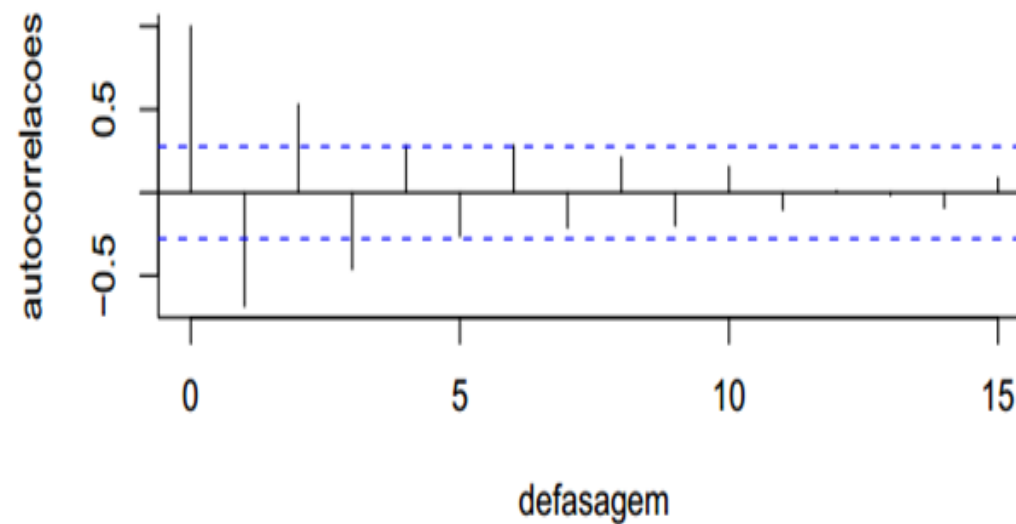
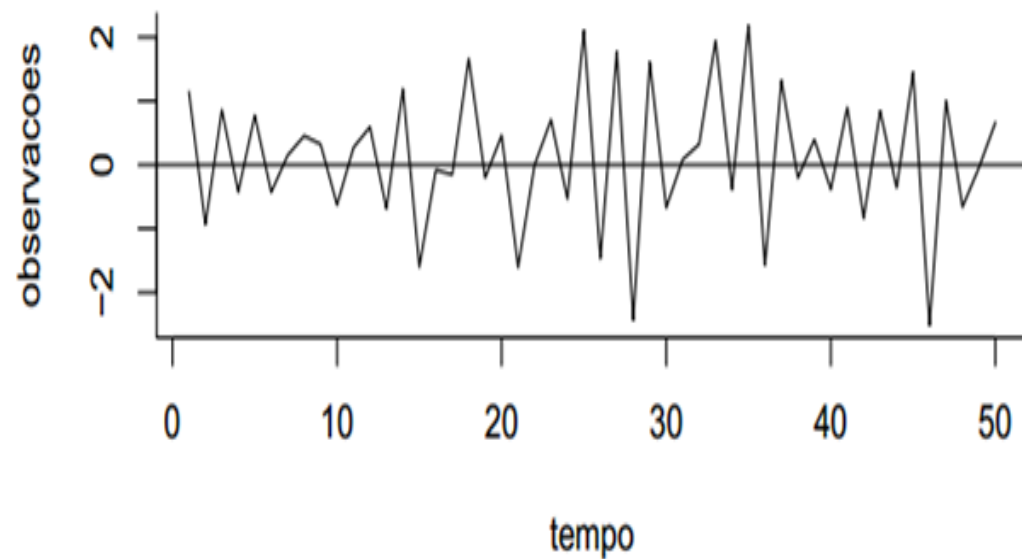
# Correlograma

---

- Correlação negativa:
  - Quando os valores das observações tendem a se alternar acima e abaixo de uma média, o coeficiente de correlação também tende a se alternar;
  - O valor de  $r_1$  será negativo enquanto o valor de  $r_2$  será positivo já que as observações defasadas de 2 períodos (lags) tendem a estar do mesmo lado da média;



# Correlograma

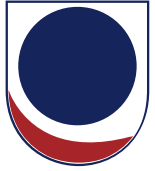




# Abordagem Geral

---

- A abordagem consiste em plotar a série e analisar alguns aspectos:
  - Tendência
  - Sazonalidade
  - Alterações acentuadas no comportamento
  - Observações discrepantes com os dados
- Remover tendências e componentes sazonais para obter resíduos estacionários;
- Escolher um modelo para ajustar aos resíduos. Utilizando várias estatísticas amostrais como função de autocorrelação
- Alcançar a previsão original da série através da previsão dos resíduos junto com os valores estimados da tendência e sazonalidade;



# Processo Estacionário

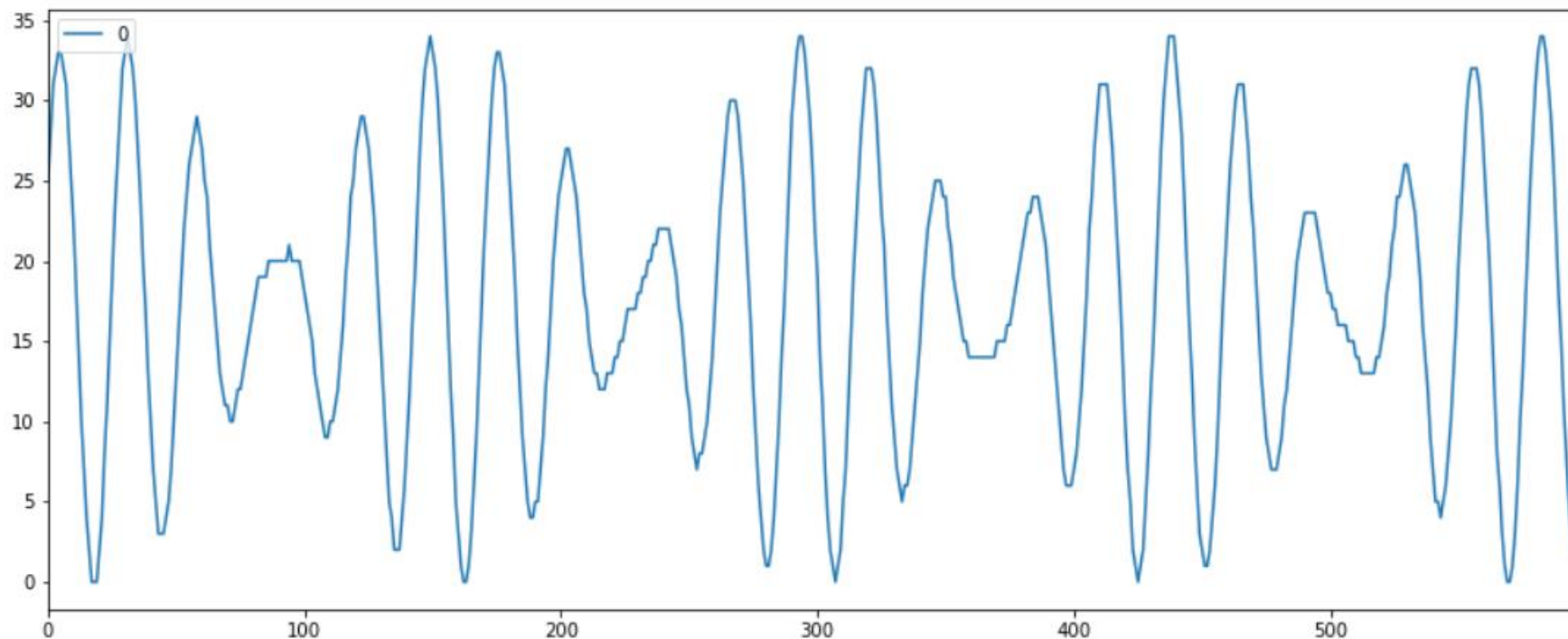
---

- É um processo que se mantém em equilíbrio estatístico com propriedades probabilísticas que não se alteram no tempo;
- O processo é considerado fracamente estacionário se:
  - As características de  $x_t$  são iguais para todo  $t$ ;
  - Média e variância constantes;



# Processo Estacionário

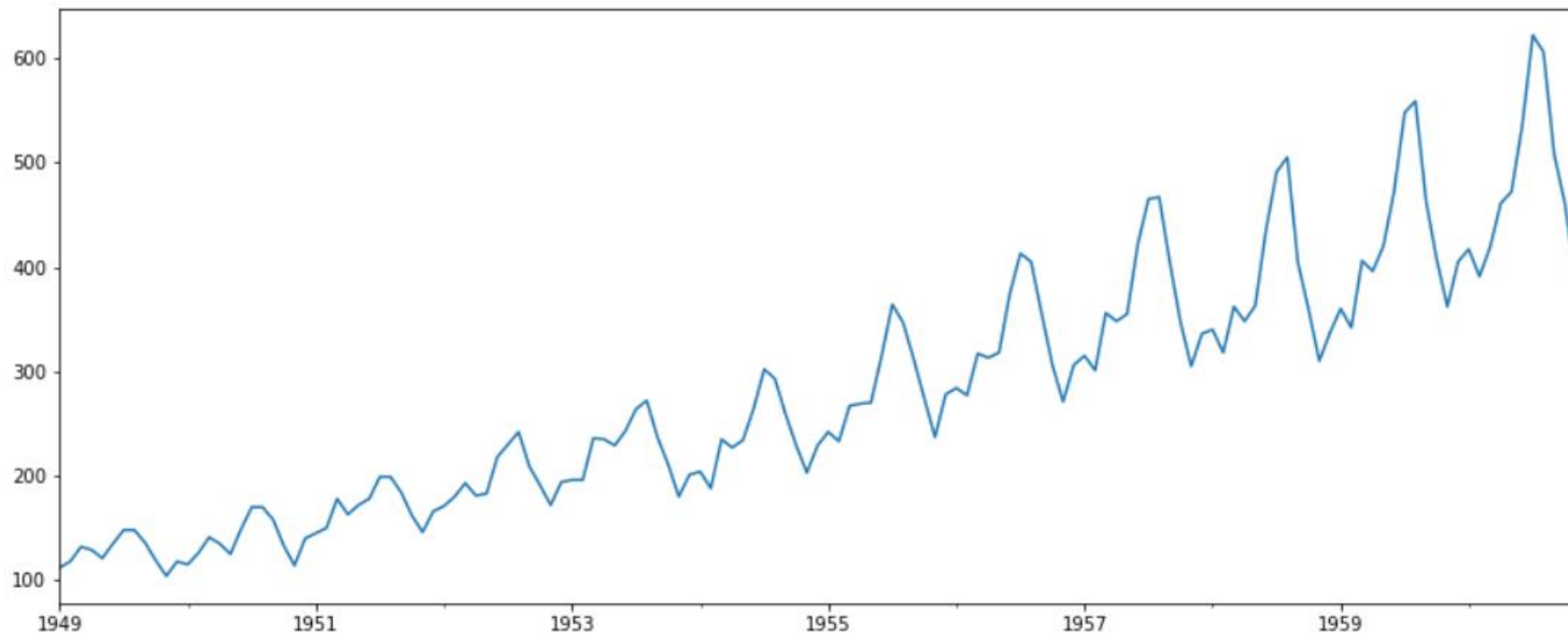
---





# Processo Estacionário

---







# Processo Estacionário

---

- Como testar se uma série é estacionária
  - Testes estatísticos:
    - Teste de Dickey-Fuller ( $H_0$ : série não-estacionária)
    - Teste de KPSS ( $H_0$ : série estacionária)
  - Gráficos:
    - Média e desvio constante
    - Histograma



# Hora de programar

---

