Análise e Previsão de Séries Temporais

Aula – 2: Introdução às séries temporais

Eraylson Galdino

egs@cin.ufpe.br

Agenda

- Resumo da Aula anterior;
- Estimação e eliminação dos componentes sazonais e de tendência;
- Estimativa e Eliminação da Tendência, na ausência da sazonalidade;
- Estimação e Eliminação de ambos: Tendência e Sazonalidade;
- Avaliando e Estimando a sequência de ruído;

Resumo da aula anterior

- Conceitos:
  - Séries Temporais;
  - Modelos simples com média zero
    - Ruído I.I.D
    - Processo Binário
    - Random Walk
  - Modelos com tendência e sazonalidade
    - Tendência
    - Sazonalidade
  - Modelos Estacionários
    - Estacionariedade
    - Função de Autocorrelação
    - Correlograma

Estimação e eliminação dos componentes sazonais e de tendência

- A série temporal deve ser analisada e pode ser transformada num modelo mais compatível para ser trabalhado;
- Modelo de decomposição clássico:

$$X_t = m_t + s_t + Y_t$$

  - $\boldsymbol{m_t}$ é o componente de tendência
  - $\boldsymbol{s_t}$ é o componente sazonal
  - $\boldsymbol{Y_t}$ é a variável aleatória estacionária

- O objetivo é estimar componentes determinísticos como $\boldsymbol{m_t}$ e $\boldsymbol{s_t}$ e tornar a variável aleatória estacionária;

Estimativa e Eliminação da Tendência, na ausência da sazonalidade

- Sem sazonalidade o modelo é descrito como:

$$X_t = m_t + Y_t,\ t = 1, \ldots, n$$

$$EY_t = 0$$

- Métodos para estimação de tendência:
  - Suavização com o filtro de médias móveis;
  - Suavização exponencial;
  - Suavização pela eliminação de componentes de alta frequência;
  - Ajuste Polinomial
- Eliminação de tendência:
  - Diferenciação;

Estimação de tendência

- Suavização com filtro de médias móveis:
  - Médias Móveis: média dos **n** valores anteriores à observação atual $\mathbf{x}_t$

$$\hat{m}_t = \frac{1}{k}\sum_{n=0}^{k-1} X_{t-n}$$

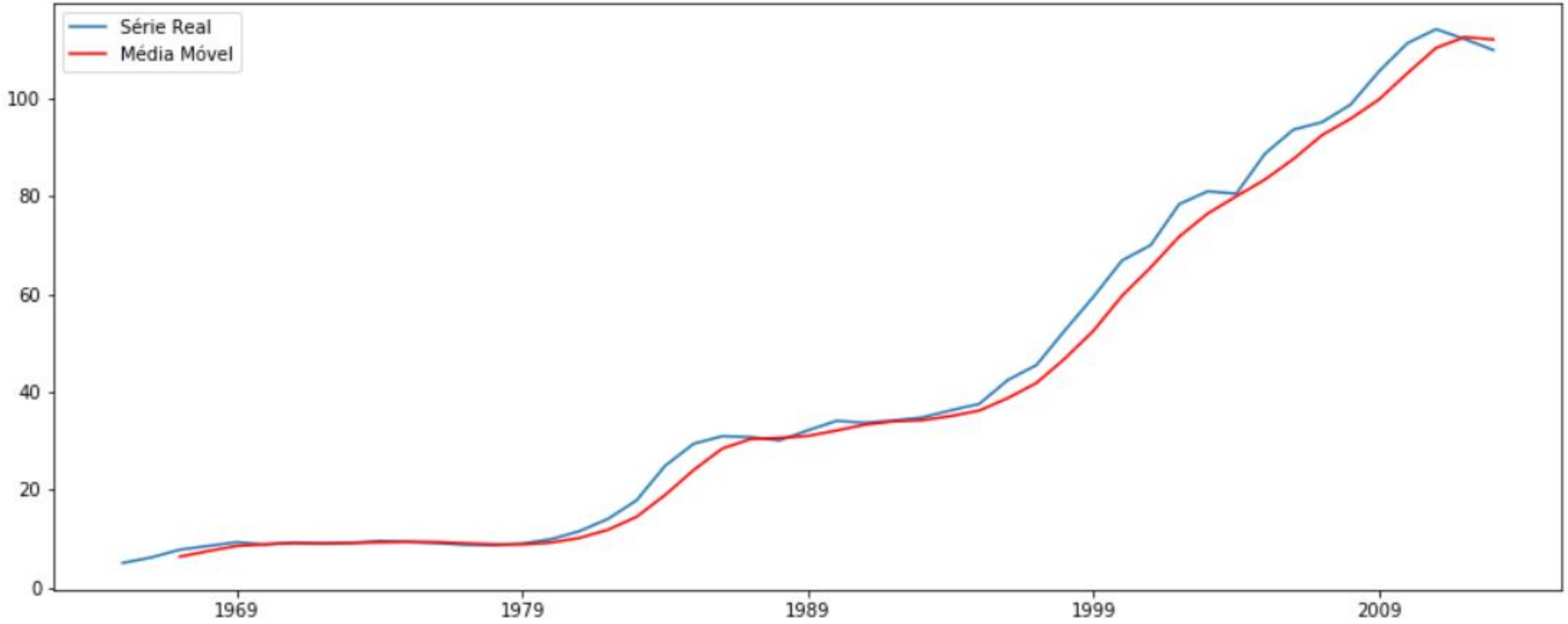
  - Funciona como um filtro removendo apenas as flutuações (alta frequência);
- Suavização Exponencial:
  - O componente de tendência é definido pelas médias móveis, sendo α um coeficiente entre [0, 1] que pondera a contribuição dos valores passados:

$$\hat{m}_t = \alpha X_t + (1-\alpha)\hat{m}_{t-1}, \quad t = 2, \ldots, n,$$

  - O X_t é o valor real da série
  - O $\hat{m}_t$ é a previsão através da média móvel

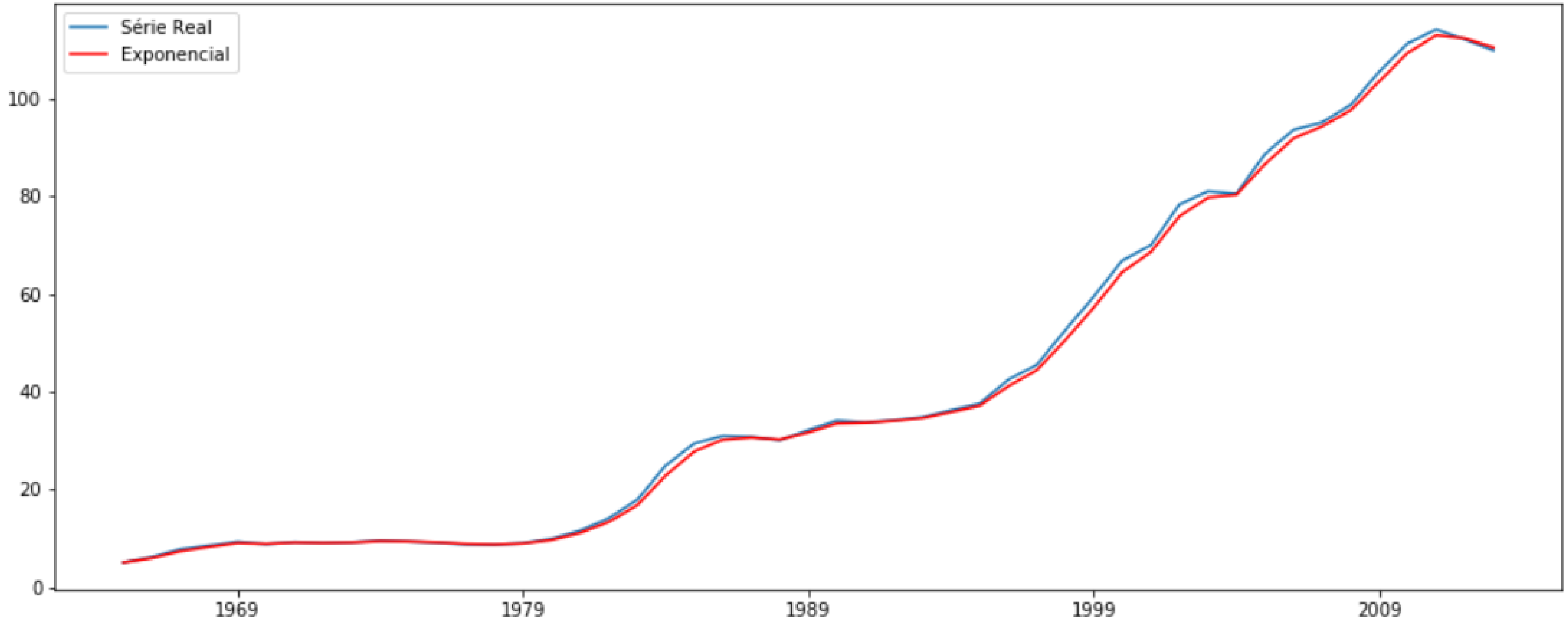
Estimação de tendência

- Utilizando Média Móvel



Estimação de tendência

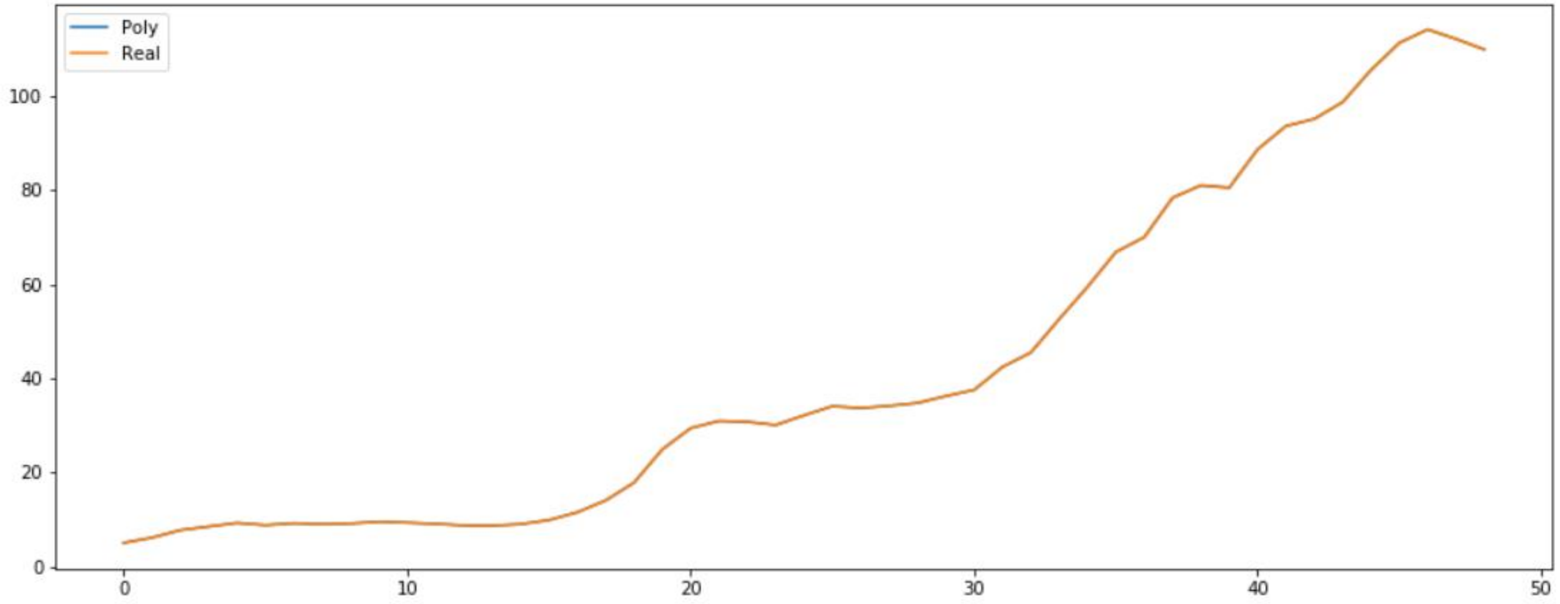
- Utilizando Suavização Exponencial
  - α = 0.5



Estimação de tendência

- Ajuste polinomial:
  - Utilizado para minimizar soma dos quadrados: $\sum_{t=1}^{n}(x_t - m_t)^2$
  - Exemplo:

$$m_t = a_0 + a_1 t + a_2 t^2$$



Eliminação da Tendência

- Diferenciação:
  - Utilizado o operador de diferenciação para obter a série diferenciada em lag-1:

$$\nabla X_t = X_t - X_{t-1} = (1 - B) X_t,$$

  - **B** é o operador de deslocamento para trás:

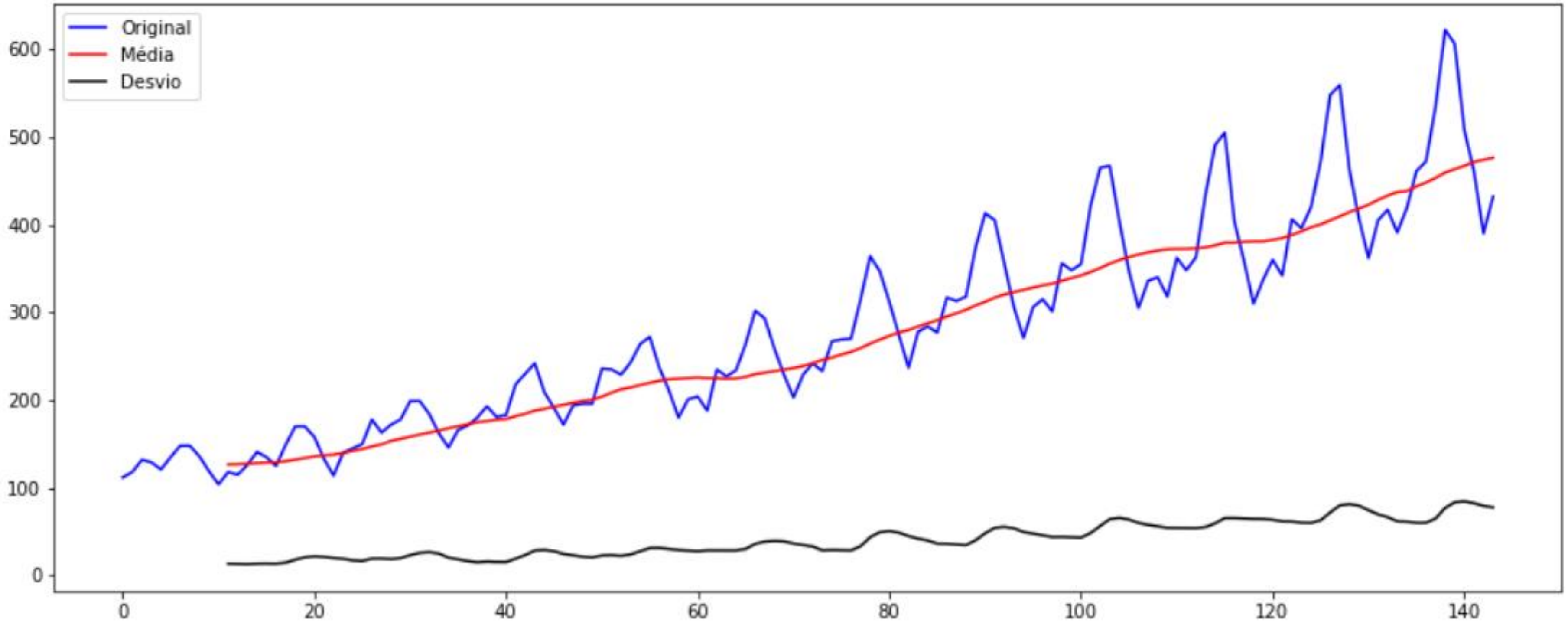
$$BX_t = X_{t-1}$$

$$BX_t = X_{t-1} \text{ e } \nabla^j (X_t) = \nabla(\nabla^{j-1}(X_t)), j \geq 1 \text{ com } \nabla^0 (X_t) = X_t$$

  - Através da aplicação da diferenciação diversas vezes (geralmente até duas) pode-se obter uma série que pode ser modelada por um processo estacionário

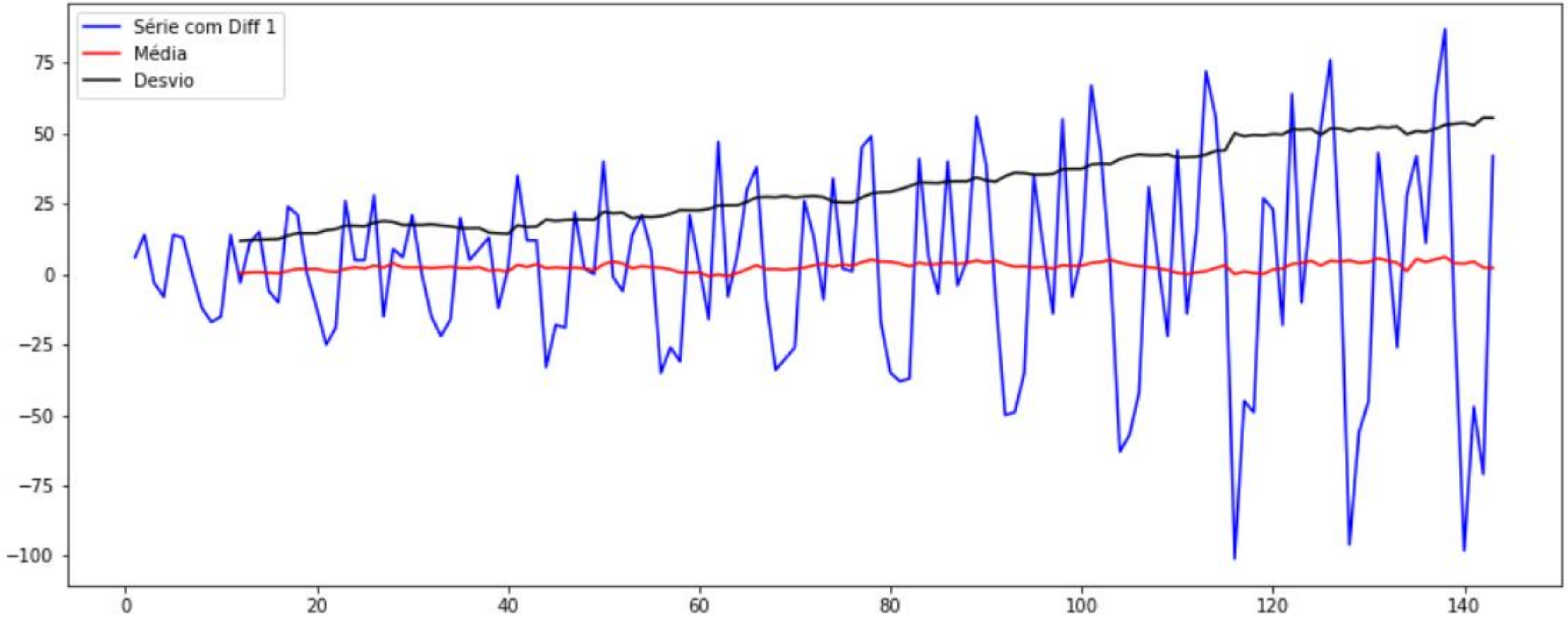
Eliminação da Tendência

- Série Airline com tendência



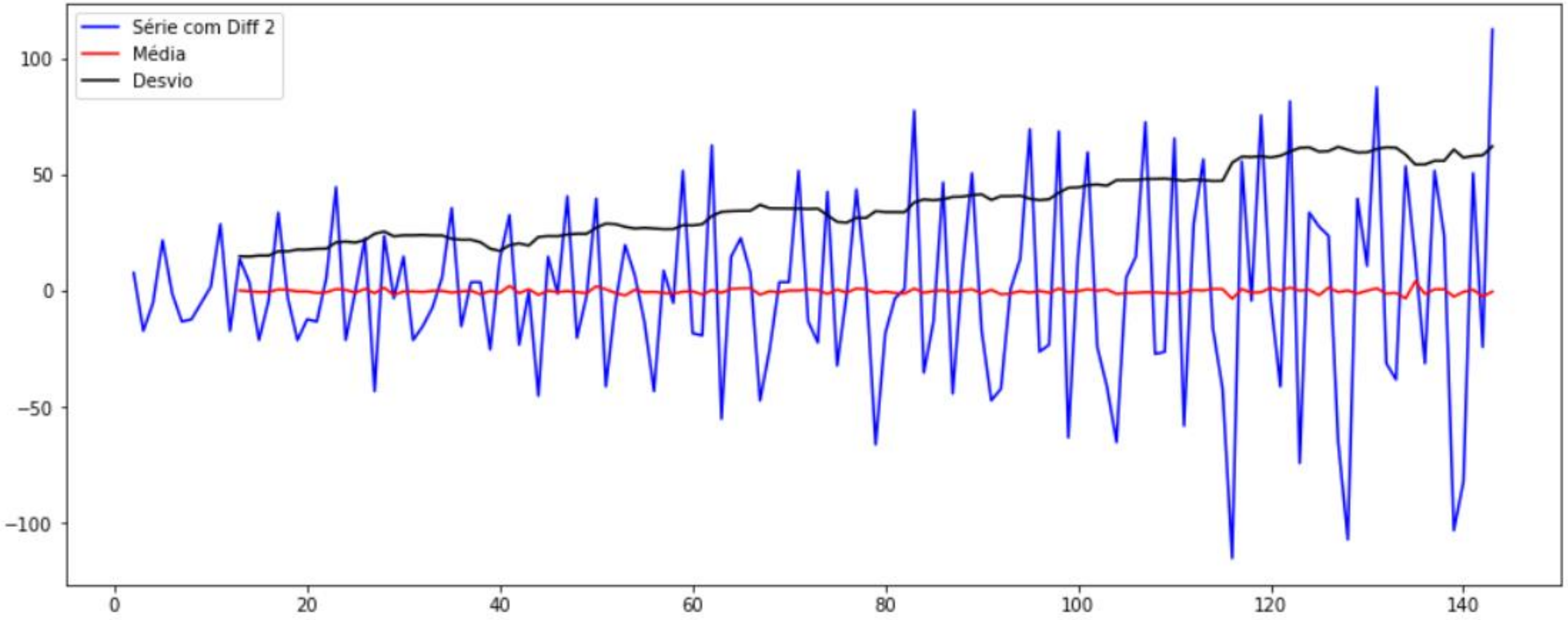
Eliminação da Tendência

- Série Airline com diferenciação de ordem 1



Eliminação da Tendência

- Série Airline com diferenciação de ordem 2



Estimação e Eliminação de ambos: Tendência e Sazonalidade

- Método S1: Estimação dos componentes de tendência e sazonalidade
  - 1º A tendência é estimada inicialmente através de um filtro de suavização eliminando o componente de sazonalidade;
    - Caso o período ***d*** do componente sazonal for um valor par, é dito que $d = 2q$:

$$\hat{m} = (0.5x_{t-q} + x_{t-q+1} + \ldots + x_{t+q-1} + 0.5x_{t-q}) / d, q < t \leq n - q$$

    - Caso seja impar, é dito que $d = 2q + 1$, e é utilizado média móvel simples
  - 2º A sazonalidade é estimada através da média w_k dos desvios para cada período.

$$\{(x_{k+jd} - \hat{m}_{k+jd}), q < k + jd \leq n - q\}, \text{ para cada } k = 1,\ldots,d$$

$$\hat{s}_k = w_k - d^{-1}\sum_{i=1}^{d} w_i, k = 1,\ldots,d, \text{ e } \hat{s}_k = \hat{s}_{k-d}, k > d$$

Estimação e Eliminação de ambos: Tendência e Sazonalidade

- Método S1: Estimação dos componentes de tendência e sazonalidade
  - 3º Remove o componente de sazonalidade:

$$d_t = x_t - \hat{s}_t, t = 1,\ldots,n$$

  - 4º Reajusta o componente de tendência para a série sem a sazonalidade utilizando métodos como o ajuste polinomial;
  - 5º Remove o componente de tendência, restando o componente de ruído:

$$\hat{Y}_t = x_t - \hat{m}_t - \hat{s}_t, t = 1,\ldots,n$$

Estimação e Eliminação de ambos: Tendência e Sazonalidade

- Método S2: Aplicando o operador de Diferenciação sazonal
  - Aplica o operador de diferenciação, sendo o **d** o período sazonal

$$\nabla_d X_t = X_t - X_{t-d} = (1 - B^d) X_t$$

  - Ao aplicar no modelo:

$$X_t = m_t + s_t + Y_t$$

  - Tem como resultado o modelo:

$$\nabla_d X_t = m_t - m_{t-d} + Y_t - Y_{t-d}$$

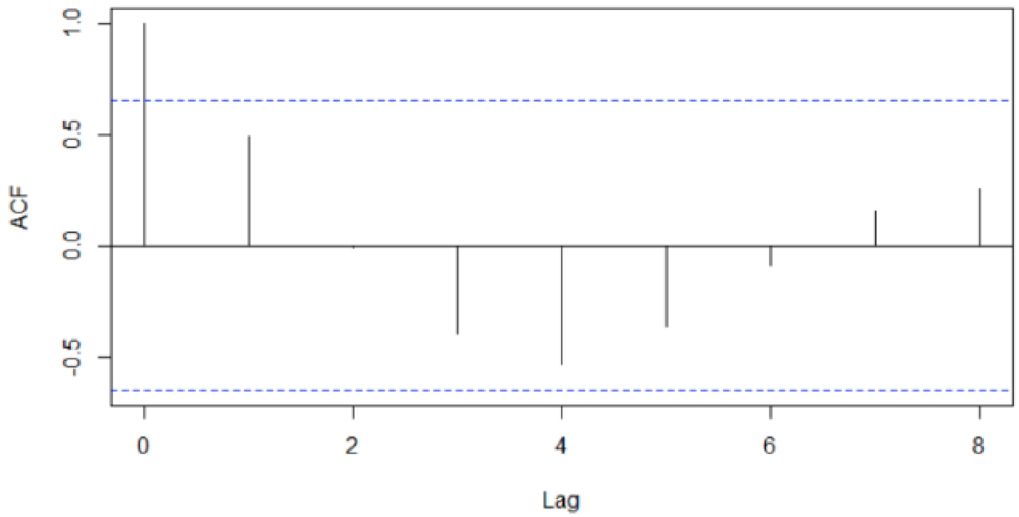
Dessa forma o componente de sazonalidade $\{s_t\}$ é eliminado e o componente $\{m_t\}$ pode ser eliminado utilizando algum método ou diferenciação ∇

Avaliando e Estimando a sequência de ruído

- Após ajustar um modelo para a série temporal que assume estacionariedade sem componentes de sazonalidade e tendência é necessário analisar a sequência de ruído;
  - A fim de constatar que o ruído é um conjunto de observações aleatórias e independentes no tempo sem possuir nenhum padrão temporal;
  - Se a hipótese for verdadeira o modelo está ajustado à série, senão outro modelo deverá ser construído com intuito de representar melhor a série temporal;
- Técnicas:
  - Função de autocorrelação
  - Teste Portmanteau
  - Checagem pela normalidade

Avaliando e Estimando a sequência de ruído

- Função de auto-correlação
  - Se uma sequência é i.i.d., 95% das amostras de autocorrelações devem cair entre os limites $\pm 1.96/\sqrt{n}$
  - Senão a hipótese de sequência i.i.d. é rejeitada



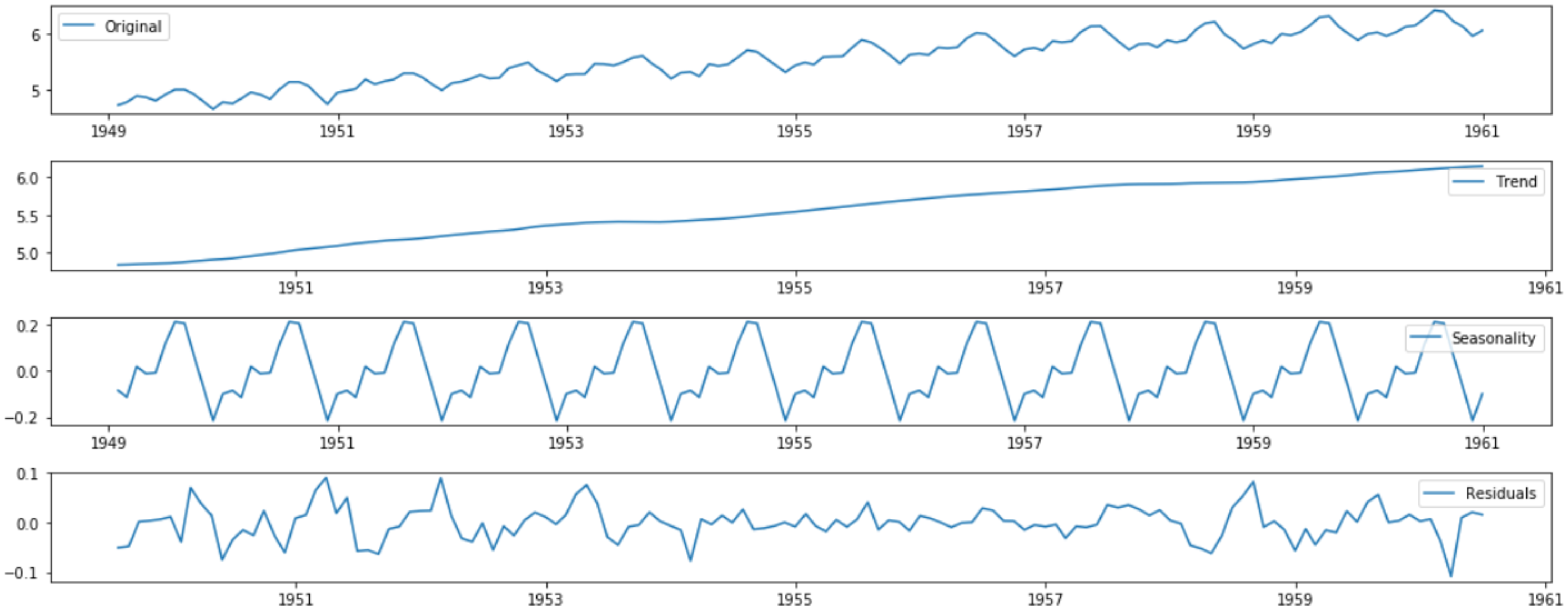
Avaliando e Estimando a sequência de ruído

- Teste portmanteau
  - Verifica a independência dos resíduos
  - É considerada a estatística $Q = n\sum_{j=1}^{h} \hat{\rho}^2(j)$
    - Sendo $\hat{\rho}$ o coeficiente da auto correlação dos resíduos
  - No caso da amostra ser uma sequência i.i.d. a estatística do teste terá aproximadamente uma distribuição X^2 com ***h*** graus de liberdades
  - Se Q é grande ($Q > \chi^2_{1-\alpha}(h)$) sugere que a amostra não seja uma sequência i.i.d.

Avaliando e Estimando a sequência de ruído

- Checagem pela normalidade
  - Caso o ruído seja gaussiano (distribuição normal) é possível concluir que o modelo foi bem ajustado para a série;
  - Testes de normalidade:
    - Shapiro-Wilk test
    - Kolmogorov-Smirnov

Decomposição da Série



Extras

- Código com alguns exemplos em python: https://github.com/EraylsonGaldino/timeseries/blob/master/Aula%2002.ipynb
- Ferramentas para a aula prática:
  - Anaconda: https://www.anaconda.com/download/
  - Jupyter: http://jupyter.org/install

Referências

- BOX, G. E. P. and JENKINS, G. M. (2008). Time series analysis: forecasting and control, 4nd. ed., San Francisco: Holden-Day.
- Brockwell, Peter J. and Davis, Richard A. (2002). Introduction to Time Series and Forecasting, 2nd. ed., Springer-Verlag