

# Trabalho No.2

CE017 - Séries temporais e modelos dinâmicos

Fernando Lucambio Pérez

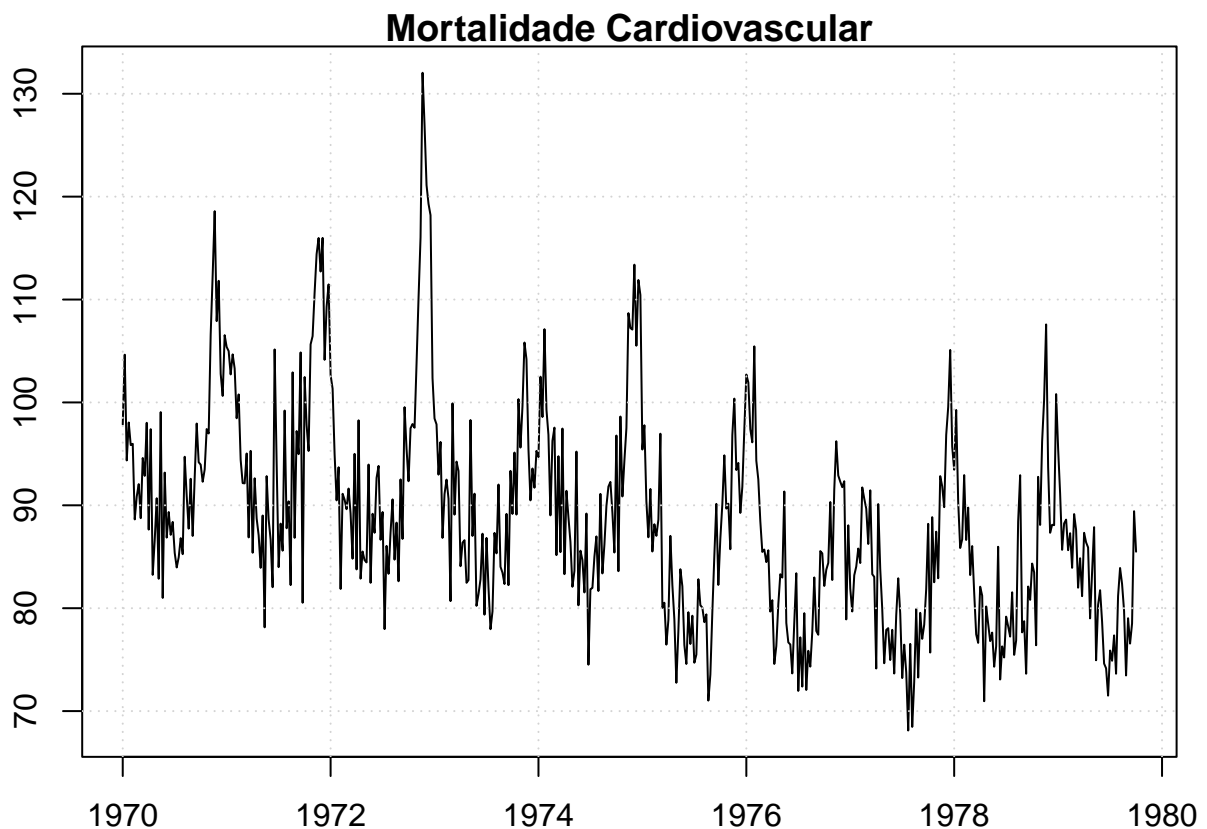
2026-03-26

Redigir de maneira individual e entregar na área correspondente no sistema **Microsoft Teams** um relatório eletrônico com as respostas até o dia **05 de maio de 2026**.

## Modelos ARIMA

1- Seja  $X_t$  um processo que represente a série de mortalidade cardiovascular `cmort`:

```
par(mar=c(3,3,1,1))  
plot(astsa::cmort, main="Mortalidade Cardiovascular", xlab="", ylab="")  
grid()
```



Ajuste um modelo AR(2) à serie  $X_t$ . Assumindo que esse modelo seja o modelo verdadeiro, encontre as previsões ao longo de um horizonte de quatro semanas  $X_{n+m}^n$  para  $m = 1, 2, 3, 4$  e os intervalos de previsão de 95% correspondentes.

2- Ajustar um modelo AR(2) para a série de mortalidade cardiovascular `cmort` discutida no exercício 1, desta vez, usando regressão linear e usando Yule-Walker.

- a. Compare as estimativas dos parâmetros obtidos pelos dois métodos.
- b. Compare os erros padrão estimados das estimativas dos coeficientes obtidos por regressão linear com suas aproximações assintóticas correspondentes, como dado no Teorema 10.

3- Suponha

$$Y_t = \beta_0 + \beta_1 t + \cdots + \beta_q t^q + X_t, \quad \beta_q \neq 0,$$

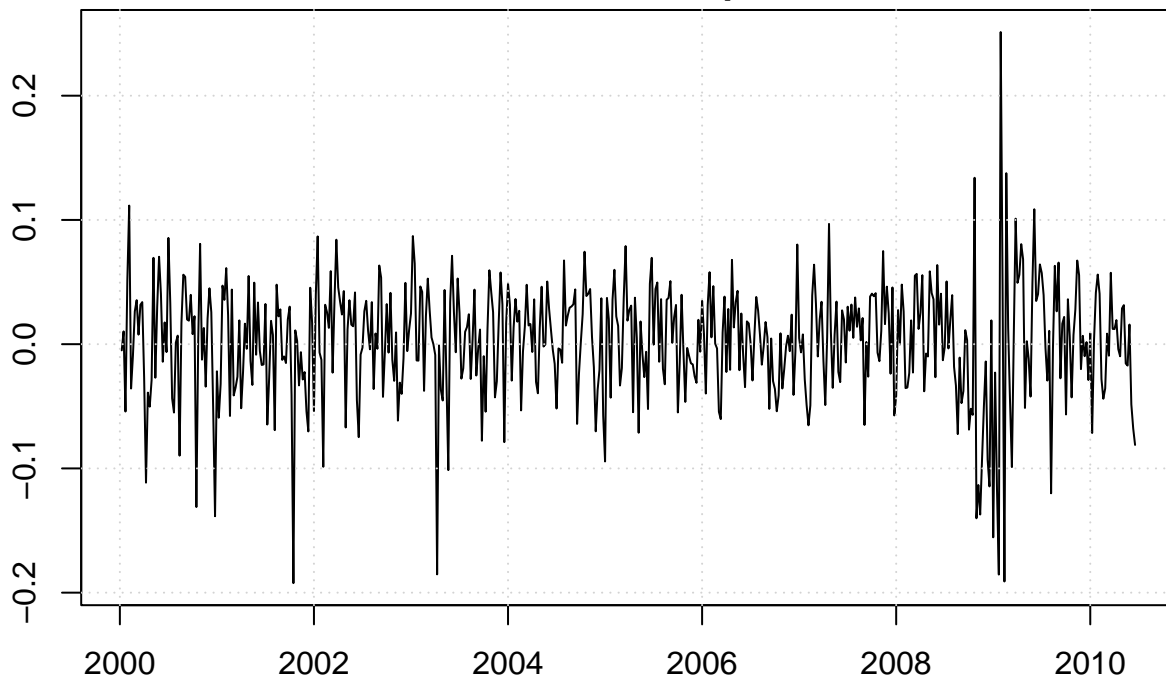
onde  $X_t$  é estacionário. Mostre que  $\nabla^k Y_t$  não é estacionário para  $k < q$ , mas é estacionário para  $k \geq q$ .

4- Os preços do petróleo bruto em dólares por barril são guardados no arquivo de dados `oil`.

```
par(mar=c(3,3,4,1))
oil = axtsa::oil
dl.oil = diff(log(oil))
plot(dl.oil,main="Taxa de crescimento do preço do petróleo \n
bruto em dólares por barril")
grid()
```

## Taxa de crescimento do preço do petróleo

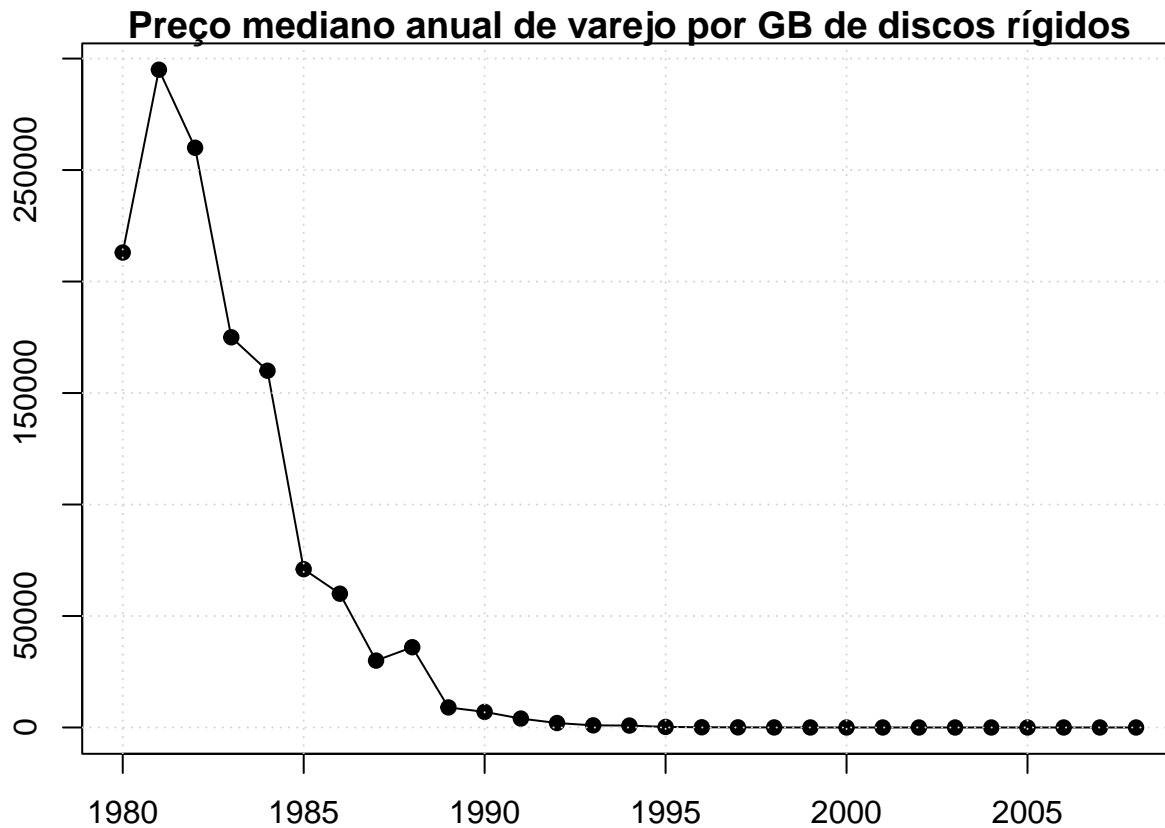
### bruto em dólares por barril



Ajuste um modelo ARIMA(p,d,q) para a taxa de crescimento realizando todos os diagnósticos necessários. Comente.

5- Um dos notáveis desenvolvimentos tecnológicos na indústria de computadores tem sido a capacidade de armazenar informações densamente em um disco rígido. Além disso, o custo de armazenamento diminuiu constantemente, causando problemas de excesso de dados, em vez de big datas. O conjunto de dados para esta tarefa é o `cpg`, que consiste no preço mediano anual de varejo por GB de discos rígidos, digamos  $C_t$ , de uma amostra de fabricantes de 1980 a 2008.

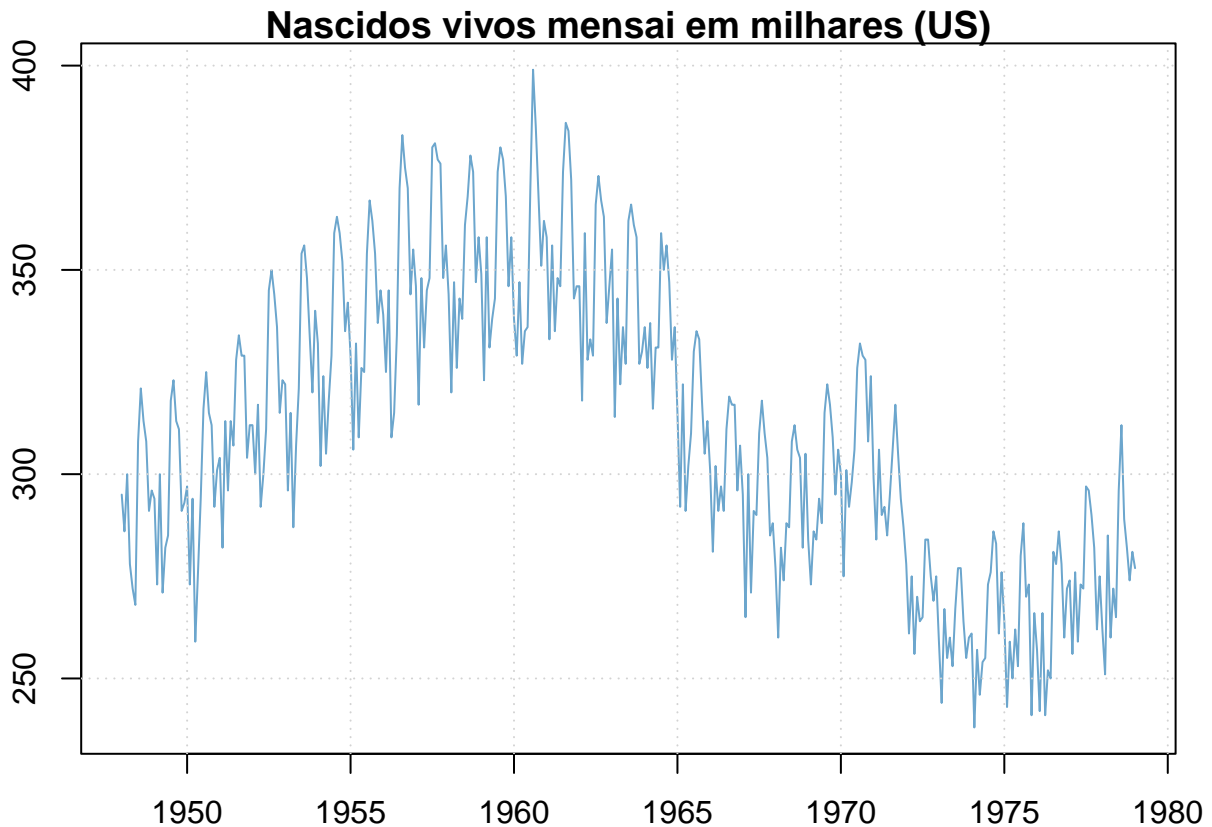
```
par(mar=c(3,3,1,1))
plot(astsa::cpg, type="o", xlab="", ylab="", pch=19,
     main="Preço mediano anual de varejo por GB de discos rígidos")
grid()
```



- Descreva o que você vê no gráfico acima.
- Argunte que a curva da resposta  $C_t$  versus o tempo  $t$  se comporta como  $C_t \approx \alpha e^{\beta t}$  ajustando uma regressão linear de  $\log(C_t)$  em  $t$  e então mostrando graficamente a linha ajustada para compará-la aos dados registrados. Comente.
- Inspeccione os resíduos do ajuste de regressão linear e comente.
- Ajuste a regressão novamente, mas agora usando o fato de que os erros são autocorrelacionados. Comente.

6- Ajuste um modelo sazonal ARIMA (SARIMA) de sua escolha para a série de nascidos vivos dos EUA birth. Use o modelo estimado para prever os próximos 12 meses.

```
par(mar=c(3,3,1,1))
plot(astsa::birth, xlab="Tempo", main="Nascidos vivos mensai em milhares (US)",
     ylab = "", pch=19, col="skyblue3")
grid()
```



## SSA - Análise de espectro Singular

7- O conjunto de dados `TS.csv` contém informações do preço de um produto vendido em uma loja de varejo entre 12-jul-2016 a 20-mar-2018, variável `price`. Esses são dados de séries temporais.

- Apresente graficamente a série. Vemos que os dados são caracterizados por alguma sazonalidade. Devido a essa razão, usaremos diversos valores do comprimento de janela  $L$ . Escolher como valores de  $L$ : 12, 24, 36 e 48.
- Selecionar os 70% primeiros valores da série como conjunto de dados de treino e estime a série utilizando a metodologia SSA.
- Utilize os restantes 30% valores da série como dados de teste. Nestes encontre os valores preditos por cada uma dos diferentes modelos, cada um obtido pelos diferentes comprimentos de janela.
- Calcule o Erro Absoluto Médio (MAE), Erro Percentual Médio Absoluto (MAPE) e o Erro Quadrático Médio (MSE) em cada caso e selecione o comprimento de janela mais adequado

Apresentamos a continuação as definições das métricas para avaliar a qualidade de ajuste nos modelos apresentados no Exercício 7, item d.

### Erro Absoluto Médio

O erro absoluto médio (MAE) mede a média da diferença entre o valor real com o valor predito. Mas por haver valores positivos e negativos, adiciona-se o módulo entre a diferença dos valores. Esta métrica não é afetada por valores discrepantes (outliers).

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|.$$

O valor obtido pelo MAE tem a mesma escala dos dados utilizados para previsão.

### Erro Percentual Absoluto Médio

O erro percentual absoluto médio (MAPE) é uma métrica que mostra a porcentagem de erro em relação aos valores reais.

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{\max\{\epsilon, |y_i|\}}.$$

O número  $\epsilon$  é utilizado para evitar a divisão por zero, escolhe-se um valor pequeno mas arbitrário, por exemplo,  $\epsilon = 0.001$ . Esta métrica é uma das mais utilizadas para reportar a performance do modelo, trazendo uma compreensão mais abrangente do resultado do MAE.

### Erro Quadrático Médio

O erro quadrático médio (MSE) calcula a média da diferença ao quadrado entre o valor predito com o valor real. Desta forma penalizam-se valores que sejam muito diferentes entre o previsto e observado. Significa que, quanto maior o valor do MSE pior o modelo performou.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

Apesar de sua ideia ser poderosa, a métrica MSE apresenta um problema de interpretabilidade. Por elevar ao quadrado as diferenças entre observados e preditos, a unidade fica distorcida.