

# Especialização em *Data Science* e Estatística Aplicada

## Módulo IV - Análise de Séries Temporais

Prof. Dr. Eder Angelo Milani

Goiânia, 2025



# Conteúdo Programático

- Principais conceitos de séries temporais. (Aula 1)
- Funções de autocovariância e de autocorrelação. (Aula 1)
- Definição e estimação da tendência e da sazonalidade.
- Métodos de suavização.
- Modelagem Box-Jenkins: modelos AR, MA, ARMA, ARIMA e SARIMA.
- Análise de séries temporais interrompidas.



# Conteúdo - Aula 2

## 1. Definição e estimação de tendência e sazonalidade

- Introdução
- Tendência
  - Tendência polinomial
- Sazonalidade
  - Método de regressão



# Introdução

## O que é tendência?

# Introdução

## O que é tendência?

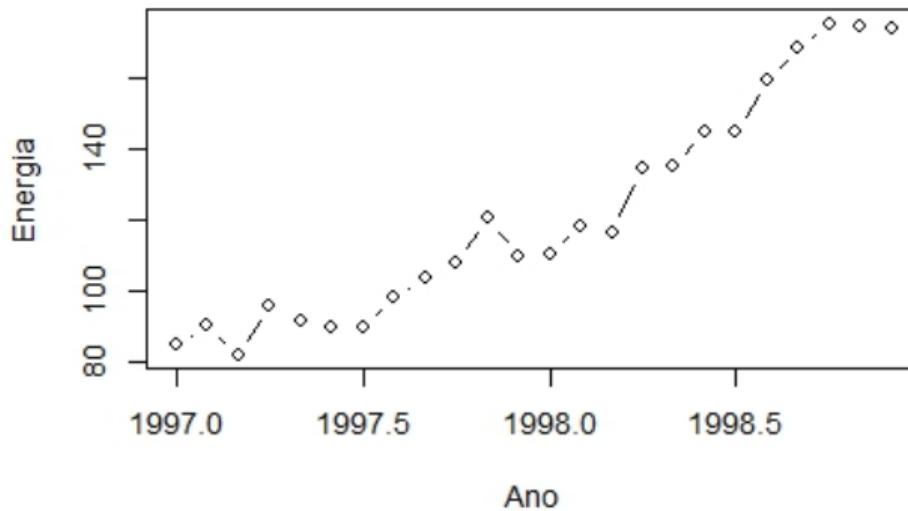
Tendência é o movimento de longo prazo de uma série temporal (crescente, decrescente ou estável). Identificá-la ajuda a responder perguntas como:

- A demanda por determinado serviço de saúde está aumentando?
- Há uma redução no número de hospitalizações ao longo dos anos?

Suponha que você tenha uma série temporal com o número de casos de gripe registrados mensalmente ao longo de 5 anos. A tendência mostra que o número de casos está crescendo ao longo dos anos, isso pode indicar:

- falha em campanhas de vacinação;
- mudança climática afetando a disseminação de vírus;
- aumento da população ou urbanização.

# Introdução



**Figura 1:** Valores mensais do consumo de energia elétrica no Estado do Espírito Santo, referentes aos anos 1997 e 1998. Série temporal retirada de Morettin e Toloi (2006)

# Introdução

O que é sazonalidade?

# Introdução

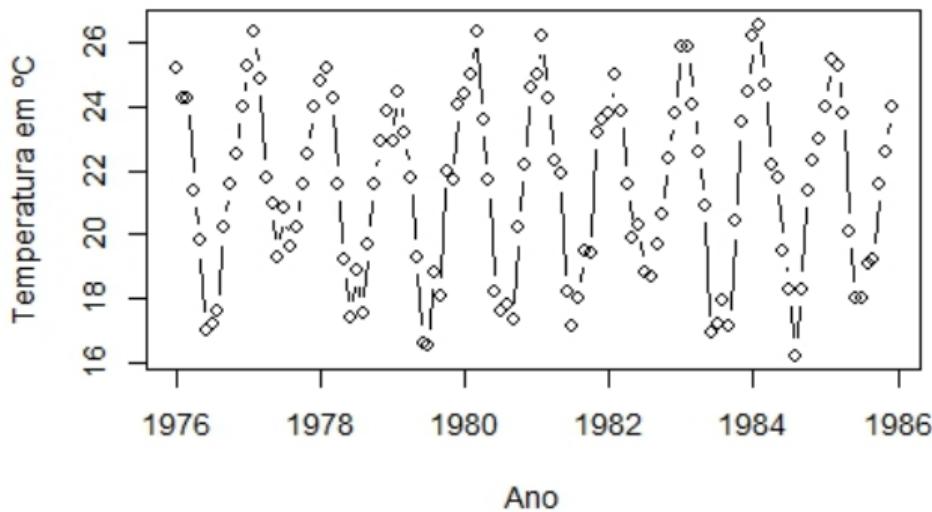
## O que é sazonalidade?

Sazonalidade é um padrão que se repete em intervalos regulares, como semanal, mensal ou anual. Um exemplo é o aumento de casos de gripe no inverno ou internações por desidratação no verão.

Suponha que você tenha uma série temporal com o número de casos de gripe registrados mensalmente ao longo de 5 anos. A sazonalidade mostra picos regulares no inverno, por exemplo. Isso permite:

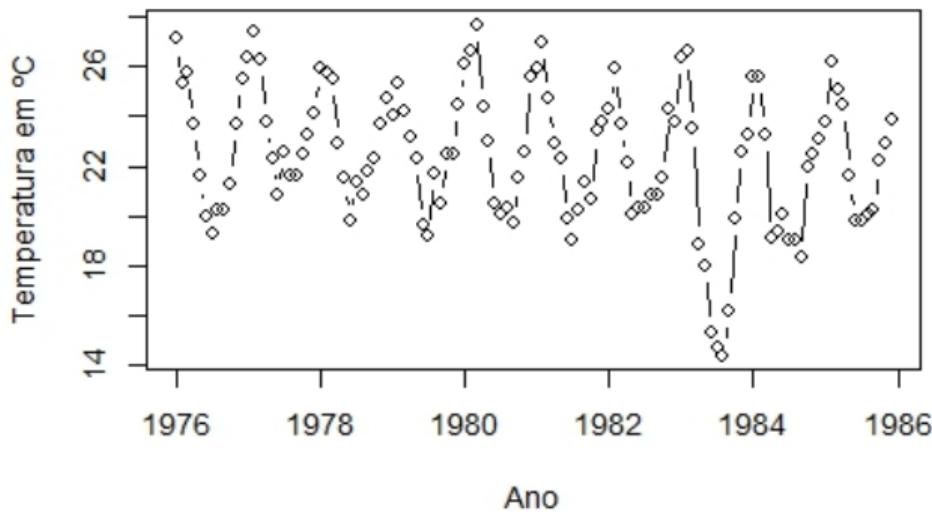
- planejar estoques de medicamentos;
- aumentar a capacidade hospitalar em meses críticos;
- intensificar campanhas de prevenção antes do pico.

# Introdução



**Figura 2:** Temperaturas médias mensais, em graus centígrados, da cidade de Cananéia, de janeiro de 1976 a dezembro de 1985. Série temporal retirada de Morettin e Toloi (2006)

# Introdução



**Figura 3:** Temperaturas médias mensais, em graus centígrados, da cidade de Ubatuba, de janeiro de 1976 a dezembro de 1985. Série temporal retirada de Morettin e Toloi (2006)

# Introdução

Em resumo, identificar tendência e sazonalidade permite:

- entender o comportamento do fenômeno;
- identificação de eventos fora do padrão;
- melhorar previsões;
- apoiar a tomada de decisões estratégicas em saúde pública, logística e comunicação.

# Introdução

Consideremos as observações  $\{Z_t, t = 1, \dots, N\}$  de uma série temporal. Um modelo de decomposição consiste em escrever  $Z_t$  como uma soma de três componentes não-observáveis,

$$Z_t = T_t + S_t + a_t, \quad (1)$$

sendo que  $T_t$  e  $S_t$  representam a tendência e a sazonalidade, respectivamente, enquanto  $a_t$  é uma componente aleatória, de média zero e variância constante  $\sigma_a^2$ .

Se  $\{a_t\}$  for um ruído branco, então  $E(a_t a_s) = 0, s \neq t$ ; mas poderemos, eventualmente, relaxar esta suposição, tomando  $\{a_t\}$  como um processo estacionário. Em geral,  $\{Z_t\}$  é uma série não-estacionária.

**Definição:** Dizemos que  $\{\epsilon_t, t \in \mathbb{Z}\}$  é ruído branco discreto se as v.a.  $\epsilon_t$  são não correlacionadas, isto é,  $Cov(\epsilon_t, \epsilon_s) = 0, t \neq s$ .

# Introdução

O interesse principal em considerar um modelo do tipo (1) será o de estimar  $S_t$  e construir a série livre de sazonalidade ou sazonalmente ajustada. Isto é, se  $\hat{S}_t$  for uma estimativa de  $S_t$ ,

$$Z_t^{SA} = Z_t - \hat{S}_t,$$

será a série sazonalmente ajustada.

Em geral, as componentes  $T_t$  e  $S_t$  são bastante relacionadas e a influência da tendência sobre a componente sazonal pode existir.

Estimando-se  $T_t$  e  $S_t$  e subtraindo de  $Z_t$  obteremos uma estimativa da componente aleatória  $a_t$ .

# Introdução

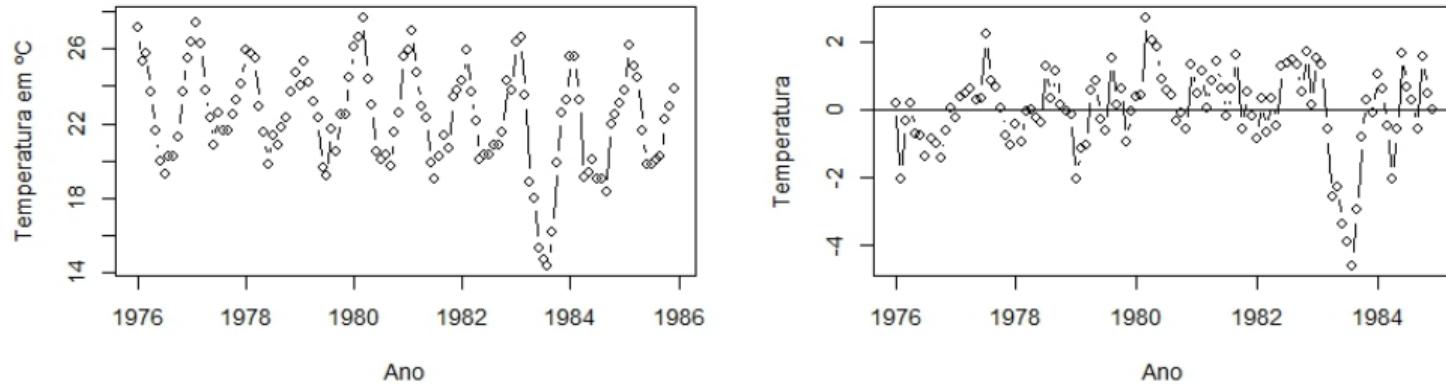


Figura 4: Série de temperatura de Ubatuba e a série do componente aleatório.

# Tendência

Inicialmente vamos supor que a componente sazonal  $S_t$  não esteja presente. O modelo que consideraremos será

$$Z_t = T_t + a_t,$$

sendo que  $a_t$  é um ruído branco, com variância  $\sigma_a^2$ .

Há vários métodos para estimar  $T_t$ . Os mais utilizados consistem em:

- ajustar uma função do tempo, como um polinômio, uma exponencial ou outra função suave de  $t$ ;
- suavizar (ou filtrar) os valores da série ao redor de um ponto, para estimar a tendência naquele ponto.

# Tendência

Estimando-se a tendência através de  $\hat{T}_t$ , podemos obter a série ajustada para tendência ou livre de tendência,

$$Y_t = Z_t - \hat{T}_t.$$

Obs.: um procedimento que é também utilizado para eliminar a tendência de um série é aquele de tomar diferenças, como foi definido anteriormente.

# Tendência polinomial

Um procedimento muitas vezes utilizado é ajustar uma curva aos valores observados da série para estimar  $T_t$  e fazer previsões.

O problema mais sério que se encontra ao estimar  $T_t$  através de um polinômio é que, embora ele possa ajustar-se bem ao conjunto de valores observados, extrações futuras podem não ser tão boas.

# Tendência polinomial

Suponha então que

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_m t^m,$$

onde o grau  $m$  do polinômio é bem menor que o número de observações  $N$ . Para estimar os parâmetros  $\beta_j$  utilizamos o método dos mínimos quadrados, ou seja, minimizamos

$$f(\beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_m t^m) = \sum_{t=1}^N (Z_t - \beta_0 - \beta_1 t - \beta_2 t^2 - \dots - \beta_m t^m)^2,$$

obtendo-se os estimadores de mínimos quadrados usuais  $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_m$ .

# Tendência polinomial

**Aplicação no software R.**

# Tendência polinomial

Utilizando o modelo estimado para  $T_t$  podemos prever valores futuros da série.

De modo geral, o valor previsto  $h$  passos à frente, dadas as observações até o instante  $t = N$ , é  $\hat{Z}_N(h)$ , e o erro de previsão correspondente é

$$e_N(h) = Z_{N+h} - \hat{Z}_N(h),$$

que por sua vez,

$$\hat{Z}_N(h) = \hat{T}_{N+h},$$

para  $h = 1, 2, 3, \dots$

# Tendência polinomial

Por exemplo, na Tabela 1, temos os valores reais para janeiro, fevereiro, março e abril de 1979.

$h$	$Z_t$
1	179,8
2	185,8
3	270,3
4	196,9

**Tabela 1:** Valores mensais do consumo de energia elétrica no Estado do Espírito Santo

# Tendência polinomial

**Aplicação no software R.**

# Sazonalidade

Nosso objetivo será ajustar uma série para a componente sazonal, ou seja, estimar  $S_t$  e subtrair a série estimada de  $Z_t$  no modelo

$$Z_t = T_t + S_t + a_t, t = 1, 2, \dots, N.$$

Desta maneira, um procedimento de ajustamento sazonal consiste em:

- obter estimativas  $\hat{S}_t$  de  $S_t$ ;
- calcular  $Z_t^{SA} = Z_t - \hat{S}_t$ .

# Sazonalidade

Se o modelo for multiplicativo, da forma

$$Z_t = T_t S_t a_t,$$

que é muitas vezes adequado para séries econômicas, que apresentam um crescimento exponencial, tomando-se logaritmos, obtemos o modelo aditivo para os logaritmos.

Ou seja,

$$\ln(Z_t) = \ln(T_t S_t a_t) = \ln(T_t) + \ln(S_t) + \ln(a_t),$$

portanto, em vez de utilizar o modelo multiplicativo pode-se utilizar o modelo aditivo para o logaritmo da série observada.

# Sazonalidade

Séries sazonais são caracterizadas por apresentarem correlação alta em “lags sazonais”, isto é, lags que são múltiplos do período  $s$ . Um procedimento de ajustamento sazonal será tal que esta correlação seja eliminada, ou pelo menos, removida em grande parte.

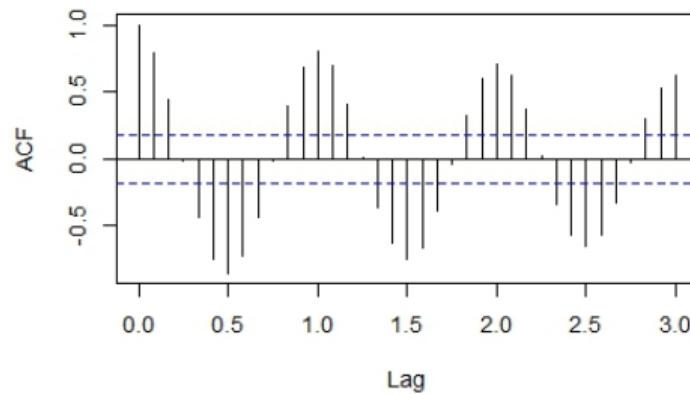


Figura 5: ACF da série de temperatura da cidade de Cananéia

# Sazonalidade

Sem perda de generalidade, consideremos o caso que temos dados mensais e o número total de observações,  $N$ , é múltiplo de 12, isto é,  $N = 12p$ ,  $p$  = número de anos, de modo que os dados podem ser representados como na Tabela 2.

# Sazonalidade

Anos	Meses						Médias
	jan 1	fev 2	mar 3	...	dez 12		
1	$Z_{11}$	$Z_{12}$	$Z_{13}$	...	$Z_{1,12}$	$\bar{Z}_{1.}$	
2	$Z_{21}$	$Z_{22}$	$Z_{23}$	...	$Z_{2,12}$	$\bar{Z}_{2.}$	
:	:	:	:	...	:	:	
$p$	$Z_{p1}$	$Z_{p2}$	$Z_{p3}$	...	$Z_{p,12}$	$\bar{Z}_{p.}$	
Médias	$\bar{Z}_{.1}$	$\bar{Z}_{.2}$	$\bar{Z}_{.3}$	...	$\bar{Z}_{.12}$	$\bar{Z}$	

Tabela 2: Observações mensais de uma série temporal com  $p$  anos

Temos que

$$\bar{Z}_{i.} = \frac{1}{12} \sum_{j=1}^{12} Z_{ij}, \quad i = 1, \dots, p, \quad \bar{Z}_{.j} = \frac{1}{p} \sum_{i=1}^p Z_{ij}, \quad i = 1, \dots, 12, \quad \bar{Z} = \frac{1}{12p} \sum_{i=1}^p \sum_{j=1}^{12} Z_{ij}$$

# Sazonalidade

Anos	jan	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez	Médias
1976	25	24	24	21	20	17	17	18	20	22	22	24	21
1977	25	26	25	22	21	19	21	20	20	22	22	24	22
1978	25	25	24	22	19	17	19	18	20	22	23	24	21
1979	23	24	23	22	19	17	16	19	18	22	22	24	21
1980	24	25	26	24	22	18	18	18	17	20	22	25	22
1981	25	26	24	22	22	18	17	18	20	19	23	24	22
1982	24	25	24	22	20	20	19	19	20	21	22	24	22
1983	26	26	24	23	21	17	17	18	17	20	24	24	21
1984	26	27	25	22	22	20	18	16	18	21	22	23	22
1985	24	26	25	24	20	18	18	19	19	22	23	24	22
Médias	25	25	25	22	21	18	18	18	19	21	23	24	22

Tabela 3: Temperatura média mensal da cidade de Cananéia

# Sazonalidade

Existem vários procedimentos para estimar  $S_t$ , sendo que os mais usuais são:

- (a) método de regressão;
- (b) método de médias móveis.

Um outro enfoque é incorporar a variação sazonal e a tendência em um modelo ARIMA, a ser estudado mais à frente.

# Sazonalidade - Método de regressão

Os métodos de regressão são ótimos para séries que apresentam sazonalidade determinística, ou seja, que pode ser prevista perfeitamente a partir de meses anteriores.

No modelo

$$Z_t = T_t + S_t + a_t, \quad t = 1, 2, \dots, N$$

temos que

$$T_t = \sum_{j=0}^m \beta_j t^j \text{ e } S_t = \sum_{j=1}^{12} \alpha_j d_{jt},$$

sendo que  $d_{jt}$  são variáveis periódicas (senos, co-senos ou variáveis sazonais “dummies”) e  $a_t$  é ruído branco, com média zero e variância  $\sigma_a^2$ .

# Sazonalidade - Método de regressão

Como estamos supondo sazonalidade constante,  $\alpha_j$  não depende de  $t$ . Podemos ter, por exemplo,

$$d_{jt} = \begin{cases} 1, & \text{se o período } t \text{ corresponde ao mês } j, \ j = 1, \dots, 12; \\ 0, & \text{caso contrário.} \end{cases}$$

Neste caso,  $d_{1t} + d_{2t} + \dots + d_{12,t} = 1, t = 1, 2, \dots, N$ , de modo que a matriz de regressão não é de posto completo, mas de posto  $m+12$  (observe que temos  $m+13$  parâmetros:  $\alpha_1, \alpha_2, \dots, \alpha_{12}, \beta_0, \beta_1, \dots, \beta_m$ ).

# Sazonalidade - Método de regressão

Impondo-se a restrição adicional  $\sum_{j=1}^{12} \alpha_j = 0$ , obtemos um modelo de posto completo

$$Z_t = \sum_{j=0}^m \beta_j t^j + \sum_{j=1}^{11} \alpha_j D_{jt} + a_t,$$

sendo que agora

$$D_{jt} = \begin{cases} 1, & \text{se o período } t \text{ corresponde ao mês } j; \\ -1, & \text{se o período } t \text{ corresponde ao mês 12;} \\ 0, & \text{caso contrário: } j = 1, \dots, 11. \end{cases}$$

# Sazonalidade - Método de regressão

Um exemplo de matriz  $D$  pode ser vista a seguir.

$$D_{11,t} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \\ -1 & -1 & -1 & -1 & \dots & -1 & -1 \\ 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \end{bmatrix} \rightarrow \begin{array}{l} \text{jan} \\ \text{fev} \\ \text{mar} \\ \vdots \\ \text{out} \\ \text{nov} \\ \text{dez} \\ \text{jan} \\ \text{fev} \\ \text{mar} \\ \vdots \end{array}$$

Mas ao invés de usar a matriz  $D$  como definido em Morettin e Toloi (2006), vamos adotar a matriz  $D$  da seguinte forma.

# Sazonalidade - Método de regressão

$$D_{11,t} = \begin{bmatrix} 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \end{bmatrix} \rightarrow \begin{array}{l} \text{jan} \\ \text{fev} \\ \text{mar} \\ \vdots \\ \text{nov} \\ \text{dez} \\ \text{jan} \\ \text{fev} \\ \text{mar} \\ \vdots \end{array}$$

Note que há mudanças nos meses de janeiro e dezembro, e um deslocamento nos demais meses. A mudança da formatação da matriz  $D$  não altera o resultado final, mas facilita, pois é possível usar rotinas já programadas no software R.

# Sazonalidade

Deste modo, podemos utilizar a teoria usual de mínimos quadrados e obter os estimadores de  $\alpha_j$ ,  $\beta_j$ , ou seja, para uma amostra  $Z_1, \dots, Z_N$ , obtemos o modelo

$$\mathbf{Z} = \mathbf{C}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\alpha} + \mathbf{a},$$

onde

$$\mathbf{Z}_{N \times 1} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_N \end{bmatrix},$$

$$\mathbf{C}_{N \times (m+1)} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 2^m \\ \vdots & \vdots & & \vdots \\ 1 & N & \cdots & N^m \end{bmatrix}, \quad \boldsymbol{\beta}_{(m+1) \times 1} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix},$$

# Sazonalidade

$$\mathbf{D}_{N \times 11} = \begin{bmatrix} D_{11} & D_{21} & \cdots & D_{11,1} \\ D_{12} & D_{22} & \cdots & D_{11,2} \\ \vdots & \vdots & & \vdots \\ D_{1N} & D_{2N} & \cdots & D_{11,N} \end{bmatrix}, \quad \boldsymbol{\alpha}_{11 \times 1} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{11} \end{bmatrix},$$
$$\mathbf{a}_{N \times 1} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}.$$

# Sazonalidade

A equação pode ser escrita na forma

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{a},$$

onde

$$\mathbf{X} = [\mathbf{C} : \mathbf{D}], \quad \boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{bmatrix},$$

de modo que

$$\hat{\boldsymbol{\gamma}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Z}.$$

são os estimadores usuais de mínimos quadrados.

# Aplicação

**Aplicação no software R.**

# Sazonalidade

O modelo que acabamos de estudar para séries sazonais consiste em 12 parâmetros independentes (intercepto + 11) e não leva em consideração a forma da tendência sazonal. Por exemplo, o fato de as médias de março e abril serem bastante semelhantes (e diferentes das médias de junho e julho) não se reflete na modelagem. Em alguns casos, a sazonalidade pode ser modelada com curvas de cosseno, requerendo uma quantidade menor de parâmetros, mas que incorpora a variação suave esperada de um período para o outro, preservando a sazonalidade.

# Sazonalidade

Um modelo simples para a sazonalidade envolvendo funções seno e cosseno é dado por

$$S_t = \beta_0 + \beta_1 \cos(2\pi ft) + \beta_2 \sin(2\pi ft),$$

sendo que para dados mensais com tempo indexado como  $1, 2, \dots$ , a frequência mais importante é  $f = 1/12$ , porque a onda cosseno se repetirá a cada 12 meses, dizemos neste caso que o período é 12.

À medida que  $t$  varia, a curva oscila entre um máximo e um mínimo. Como a curva se repete exatamente a cada  $1/f$  unidades de tempo,  $1/f$  é chamado de período da onda cosseno.

# Sazonalidade

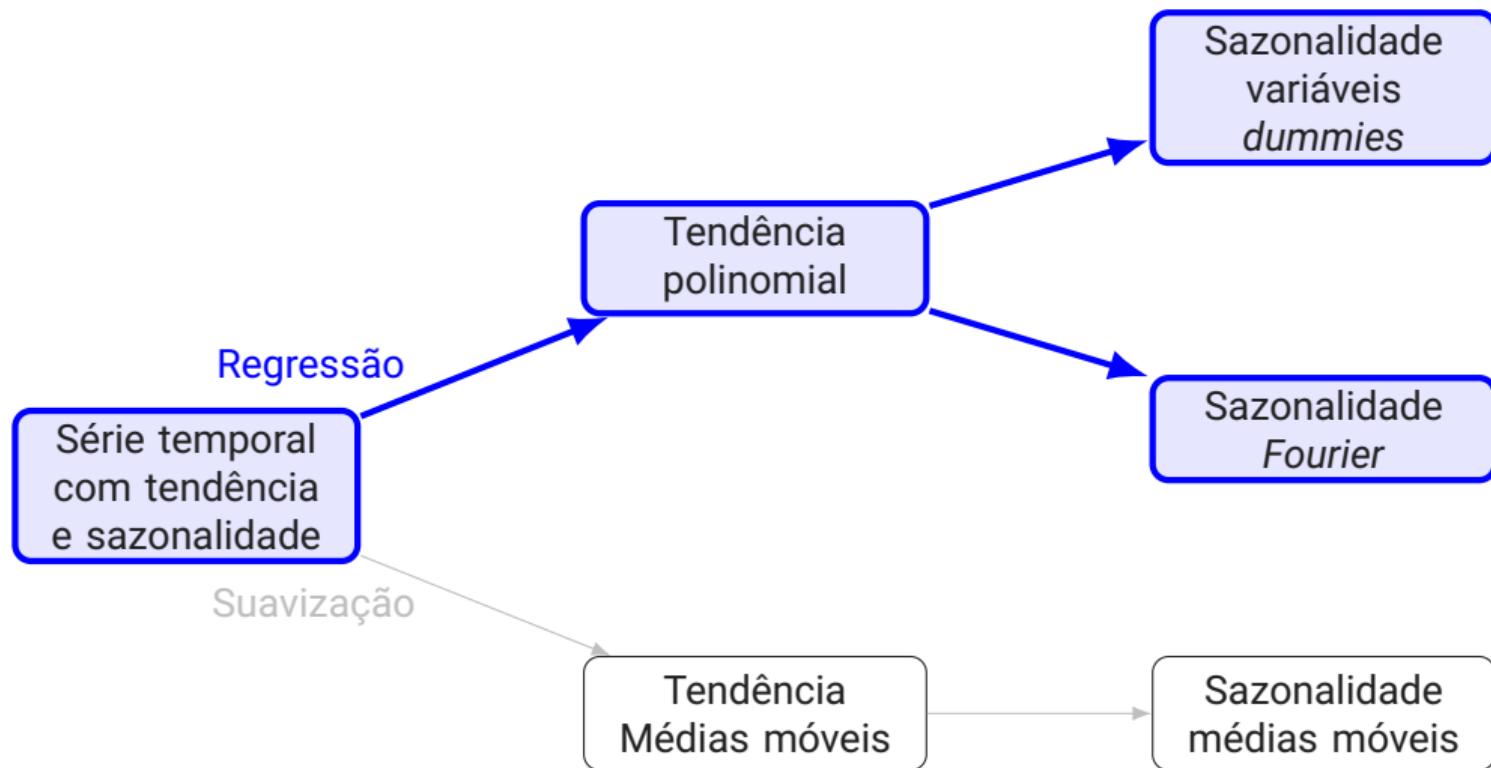
Funções adicionais de seno e cosseno em outras frequências serão frequentemente utilizadas na modelagem. Para séries mensais, as frequências harmônicas mais altas, como 2/12 e 3/12, são especialmente pertinentes e, às vezes, melhoraram o ajuste em detrimento da adição de mais parâmetros ao modelo.

De fato, pode-se demonstrar que qualquer sazonalidade com período 12 pode ser expressa exatamente pela soma de seis pares de funções seno e cosseno, o que provoca um *overfitting* (ou sobreajuste).

# Aplicação

**Aplicação no software R.**

## Em resumo



# Especialização em *Data Science* e Estatística Aplicada

## Módulo IV - Análise de Séries Temporais

Prof. Dr. Eder Angelo Milani

[edermilani@ufg.br](mailto:edermilani@ufg.br)

