Imagem: bit.ly/3EggpbU

ERHA7016 – Hidrologia Estocástica

Introdução (pt. 2)

Daniel Detzel
detzel@ufpr.br

Agenda

Revisão de estatística

- testes de hipótese

Processos estocásticos

- definições
- propriedades

Análise de séries temporais

- definição e propriedades
- características de séries hidrológicas
- persistência

Revisão de estatística

testes de hipótese

Revisão de estatística | testes de hipótese

Testes de hipótese:

Regra de decisão para rejeitar, ou não, uma hipótese com base na amostra

Hipótese: suposição que algo pode ser verdade, ou não. Deve ser passível de testes

Hipóteses estatísticas

H_0 – hipótese nula: o que se assume ser verdade

H_1 – hipótese alternativa: o que se assume ser verdade quando da rejeição de H_0

ex. de H_0: a distribuição dos dados é Normal; as amostras não têm a mesma média; a série não tem tendência; etc.

Revisão de estatística | testes de hipótese

Cada teste possui um procedimento matemático específico

Em comum, todos calculam a probabilidade (p-valor) de os dados terem sido observados assumindo que H_0 é verdadeira

$$P(\text{Dados} \mid H_0 \text{ verdadeira}) = \text{p-valor}$$

Por sua vez, o analista assume um risco de se tomar a decisão errada

em particular, rejeitar H_0 quando ela é verdadeira

o risco é quantificado pelo nível de significância α

valores típicos de α: 1%, 5% e 10%

Revisão de estatística | testes de hipótese

Em suma:

$$\textbf{p-valor} \neq \boldsymbol{\alpha}$$

Calculado usando a amostra

Atribuído pelo analista

Se o p-valor:

Baixo (p-valor $\leq \alpha$) → rejeita-se a hipótese nula

Alto (p-valor $> \alpha$) → falha-se em rejeitar a hipótese nula

Revisão de estatística | testes de hipótese

Testes de hipótese são construídos com base em premissas

- os dados são independentes e identicamente distribuídos
- as amostras possuem a mesma variância (homoscedásticas)

Adicionalmente, testes do tipo paramétrico assumem que a distribuição de probabilidade dos dados é conhecida

- o analista precisa adotar uma distribuição para os dados
- perdem poder se a distribuição adotada não for adequada

Testes não paramétricos, não têm essa premissa

- perdem poder se a distribuição de probabilidades dos dados é conhecida

Revisão de estatística | distribuições de probabilidades

Como escolher a distribuição a ser adotada?

1. Avaliação por métodos gráficos
2. Aplicação de testes de hipótese para adequação de ajustes

2. Testes de adequação de ajustes

2.1. Shapiro-Wilk (para normalidade)

2.2. Anderson-Darling

2.3. Filliben/PPCC (*probability plot correlation coeficient*)

Exemplos na sequência usam a série dos níveis mínimos anuais do rio Nilo

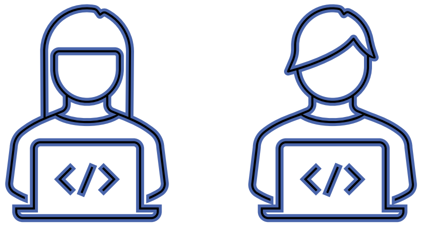
Revisão de estatística | distribuições de probabilidades

2.1. Shapiro-Wilk (para normalidade)

H_0: a amostra vem de uma distribuição normal

H_1: a amostra não vem de uma distribuição normal

```
Shapiro-Wilk normality test
W = 0.99329, p-value = 0.004614
```



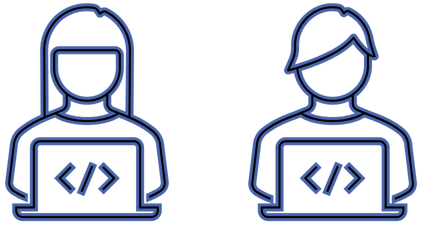
Veredicto: rejeita-se H_0 (a distribuição não é normal)

2.2. Anderson-Darling

H_0: a amostra vem de uma distribuição especificada
H_1: a amostra não vem de uma distribuição especificada

```
Anderson-Darling test of goodness-of-fit
Null hypothesis: Normal distribution
Anmax = 1.5552, p-value = 0.9905

Null hypothesis: log-normal distribution
Anmax = 3.3899, p-value = 0.3714
```



Veredicto: falha-se em rejeitar H_0 (a distribuição é normal)

Revisão de estatística | distribuições de probabilidades

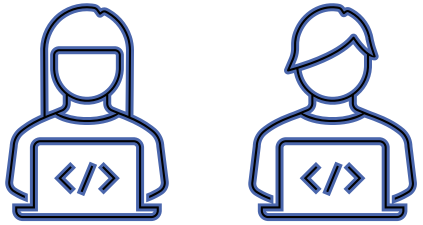
2.3. PPCC

H_0: a amostra vem de uma distribuição especificada

H_1: a amostra não vem de uma distribuição especificada

```
Probability Plot Correlation Coefficient Test
Null hypothesis: Normal distribution
ppcc = 0.99677, n = 663, p-value = 0.0073

Null hypothesis: log-normal distribution
ppcc = 0.82211, n = 663, p-value = 0.0039
```



Veredicto: rejeita-se H_0 (a distribuição não é normal, tampouco log-normal)

PROCESSOS ESTOCÁSTICOS

definições

Processos estocásticos | definições

Processos estocásticos:

Descrevem sistemas cuja evolução ocorre de acordo com leis probabilísticas

Um PE Z é uma família de funções definidas por:

$$Z = \{Z(t, \omega), t \in T; \omega \in \Omega\}$$

onde

t variável pertencente ao espaço de instantes T

ω evento do espaço de probabilidades Ω

Processos estocásticos | definições

Um PE pode ser visto como uma função de 2 variáveis, na qual para cada evento ω é atribuída uma variável t

Se $T \equiv \mathbb{Z} = \{1, \dots, t\}$ → PE discreto $Z_{t,\omega}$ ou simplesmente Z_t

Se $T \in \mathbb{R}$ → PE contínuo $Z(t, \omega)$ ou simplesmente $Z(t)$

Quando aplicado à modelagem de fenômenos naturais, frequentemente assume-se T como sendo finito ($T \equiv \mathbb{Z}$) e tendo o tempo como unidade

Os possíveis valores que a variável Z_t recebe são chamados de estados

Processos estocásticos | definições

Classificação dos modelos estocásticos

[Hipel e McLeod, 1994, p. 23]

		Estados	
		Discretos	Contínuos
Tempo	Discreto	Cadeias de Markov *[Wilks, 1998]*	Modelos de séries temporais
Tempo	Contínuo	Processos pontuais *[Abi-Zeit et al., 2004]*	Equações diferenciais estocásticas *[Breinholt et al., 2021]*

Processos estocásticos | definições

O PE $Z(t, \omega)$ representa 4 casos:

1. Com t e ω variáveis → uma família de funções temporais
2. Com t variável e ω fixo → uma função temporal única
3. Com t fixo e ω variável → uma variável aleatória
4. Com t e ω fixos → um número

Caso 2: definição de série temporal → uma realização infinita do processo estocástico

Processos estocásticos | definições

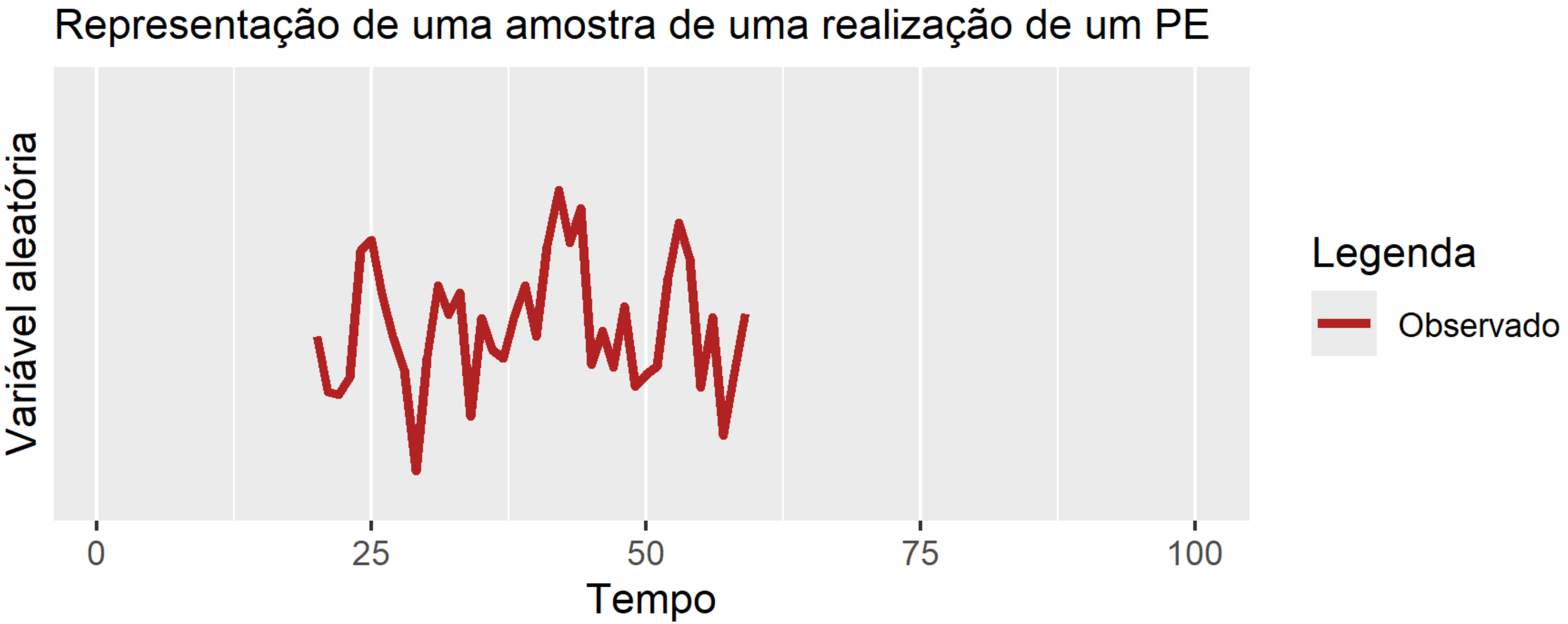
Processos naturais ≠ processos estocásticos
Séries observadas ≠ realizações de um processo estocástico

	Processos (variáveis)	Séries temporais (valores)
Fenômeno natural	Processo natural	Séries observadas (únicas e finitas)
Modelo	Processo estocástico	Realizações (infinitas)

[Koutsoyiannis, 2006]

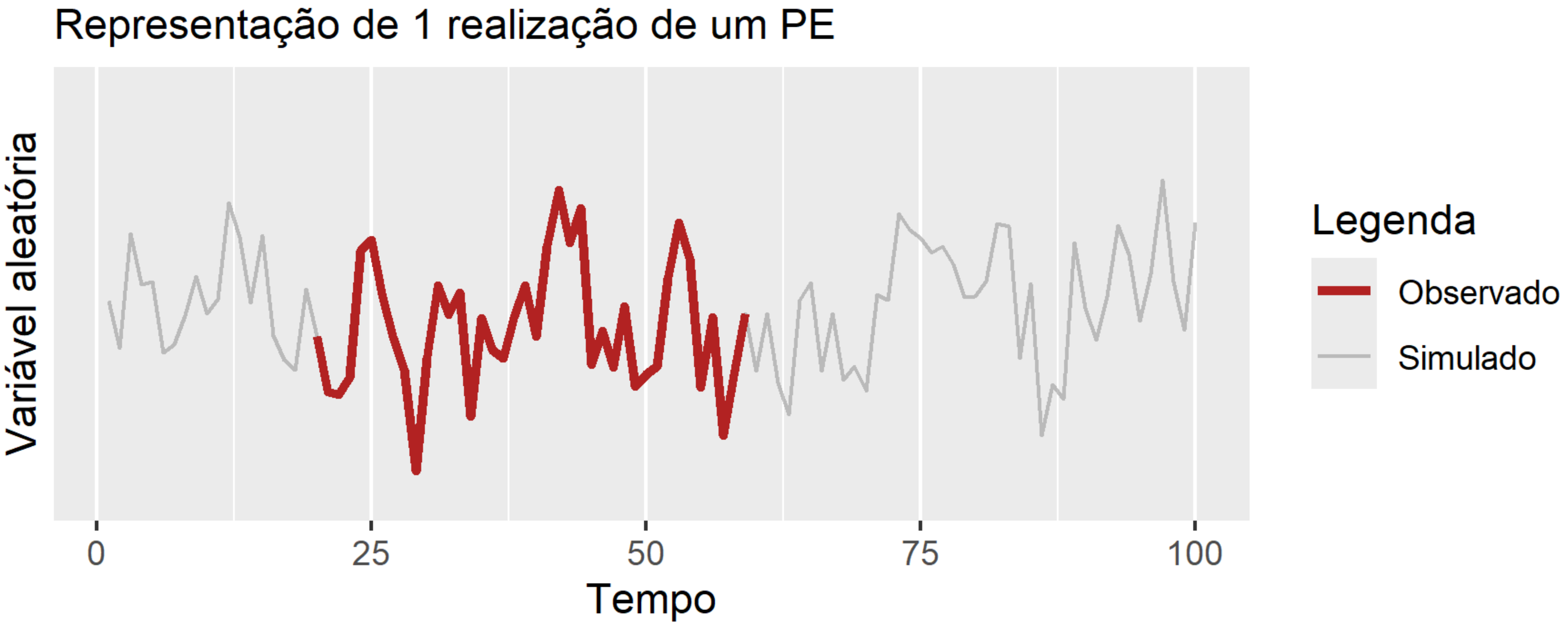
Processos estocásticos | definições

Processos naturais se tratam de amostras de uma realização de um PE



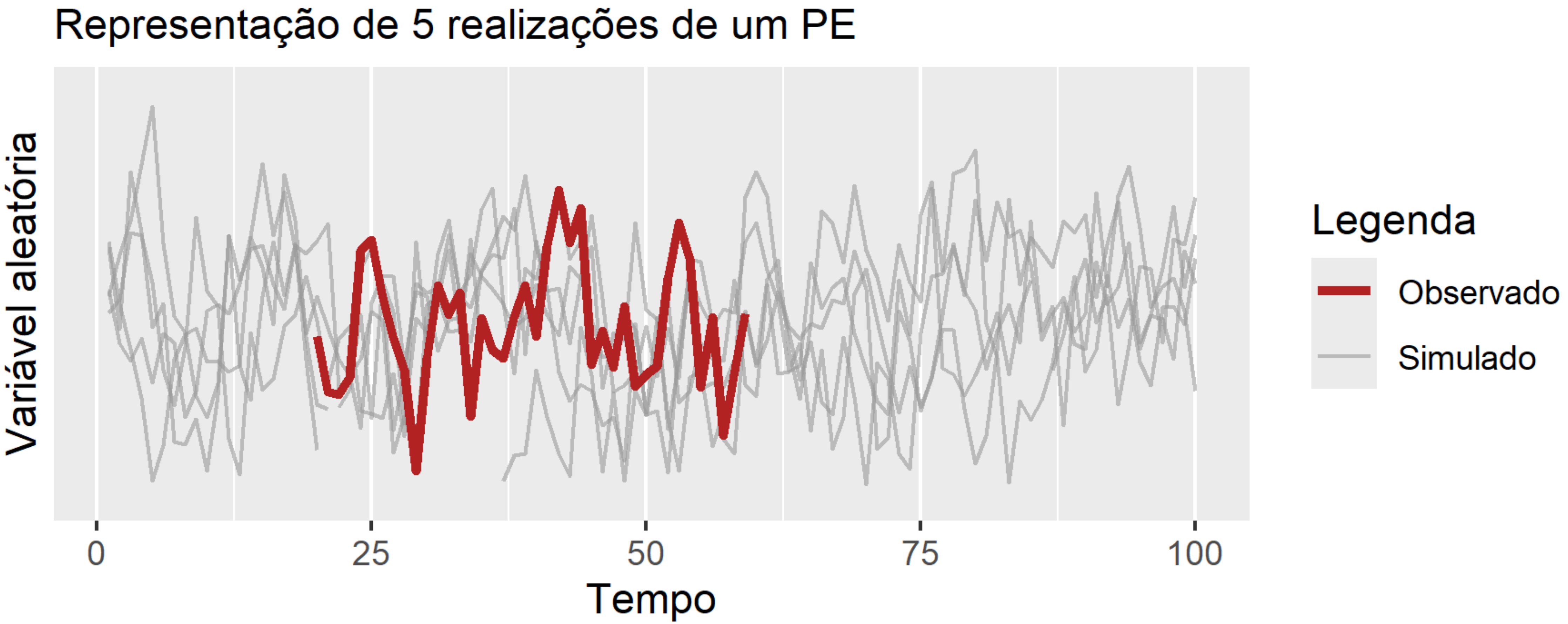
Processos estocásticos | definições

Processos naturais se tratam de amostras de uma realização de um PE



Processos estocásticos | definições

Processos naturais se tratam de amostras de uma realização de um PE



PROCESSOS ESTOCÁSTICOS

propriedades

Processos estocásticos | propriedades

Especificação de um PE:

A determinação de um PE Z envolve o conhecimento da FDA conjunta de todas as suas n (infinitas) realizações z_n

$$F_{Z_1,\dots,Z_n}(z_1, \dots, z_n; t_1, \dots, t_n) = P\{Z(t_1) \leq z_1, \dots, Z(t_n) \leq z_n\}$$

caso unidimensional: a FDA $F_{Z_1}(z_1; t_1)$ é conhecida
caso bidimensional: as FDAs $F_{Z_1,Z_2}(z_1, z_2; t_1, t_2)$ são conhecidas
caso n-dimensional: as FDAs $F_{Z_1,\dots,Z_n}(z_1, \dots, z_n; t_1, \dots, t_n)$ são conhecidas

Processos estocásticos | propriedades

Os parâmetros são também representados por conjuntos

Média conjunta de Z (função média):

$$\mu_t = E[Z_t] = \int_{-\infty}^{\infty} z f_Z(z;t)\, dz$$

Covariância conjunta de Z (função de autocovariância – FACV):

$$\gamma(t_1, t_2) = COV\left[Z_{t_1}, Z_{t_2}\right] = E\left[Z_{t_1} Z_{t_2}\right] - E\left[Z_{t_1}\right] E\left[Z_{t_2}\right]$$

Processos estocásticos | propriedades

Estacionariedade:

Conceito relacionado com um estado de equilíbrio estatístico de um PE

propriedades estatísticas independem do tempo

Estacionariedade estrita (ou forte)

$$F_Z(z_1, \dots, z_n; t_1 + \tau, \dots, t_n + \tau) = F_Z(z_1, \dots, z_n; t_1, \dots, t_n)$$

para quaisquer $t_1, \dots, t_n, \tau \in T$

Processos estocásticos | propriedades

Aplicações práticas estão centradas em distribuições unidimensionais e momentos finitos

tipicamente momentos de primeira e segunda ordens

Estacionariedade ampla (ou fraca, ou de segunda ordem)

atendem às seguintes condições:

i. $E[Z_t] = \mu_t = \mu =$ cte.

ii. $E[Z_t^2] < \infty$

iii. $\gamma(t_1, t_2) = COV\left[Z_{t_1}, Z_{t_2}\right]$ é função apenas de $|t_1 - t_2|$

Processos estocásticos | propriedades

Ergodicidade:

Conceito muito utilizado na teoria de sistemas dinâmicos

estudo da trajetória de partículas em um sistema

sistema ergódico: as trajetórias têm a mesma probabilidade de passar pelos diversos estados que o compõem

as características do sistema inteiro podem ser inferidas com base na trajetória de 1 partícula

Paralelo com as realizações de um PE

um PE é ergódico quando suas estatísticas podem ser inferidas a partir de apenas uma série temporal

Processos estocásticos | propriedades

PE ergódico na média (primeira ordem)

$$\lim_{n\to\infty}\left(\frac{1}{n}\sum_{t=1}^{n} z_t\right) = E[Z_t] = \mu$$

PE ergódico na autocovariância (segunda ordem)

$$\lim_{n\to\infty}\left[\frac{1}{n}\sum_{t=1}^{n-k}(z_t - \mu_z)(z_{t+k} - \mu_z)\right] = COV[Z_t] = \gamma$$

Processos estocásticos | propriedades

PE ergódico na autocovariância (segunda ordem) – condição suficiente

$$\sum_{k=0}^{\infty} |\hat{\gamma}_k| < \infty$$

Fill [2011]: em se tratando de séries hidrológicas, quando o processo é estacionário é usual assumi-lo ergódico

Fill, H.D. (2011). Análise da estacionariedade das vazões do rio Iguaçu em União da Vitória, in: Anais Do XIX Simpósio Brasileiro de Recursos Hídricos. ABRH, Maceió, pp. 1–7.

ANÁLISE DE SÉRIES TEMPORAIS

definição e propriedades

Análise de séries temporais | definição e propriedades

Séries temporais:

Sequências de variáveis observadas e encadeadas temporalmente

Os intervalos de tempo entre registros não precisam ser uniformes

Para intervalos de tempo uniformes, é conveniente escolher uma escala Δt apropriada

a escolha tem relação com o propósito do estudo

ex. 1: cheias em bacias urbanas pequenas → Δt na escala de horas

ex. 2: balanço hídrico em grande bacias → Δt na escala de anos

Análise de séries temporais | definição e propriedades

Estacionariedade:

Importante: a estacionariedade é uma propriedade associada a processos estocásticos, não a séries temporais [ver discussão em *Koutsoyiannis, 2006*]

para séries: falamos em comportamento estacionário

Uma série tem comportamento estacionário quando suas estatísticas não variam com o tempo

tipicamente, a média e o desvio padrão são mantidos constantes

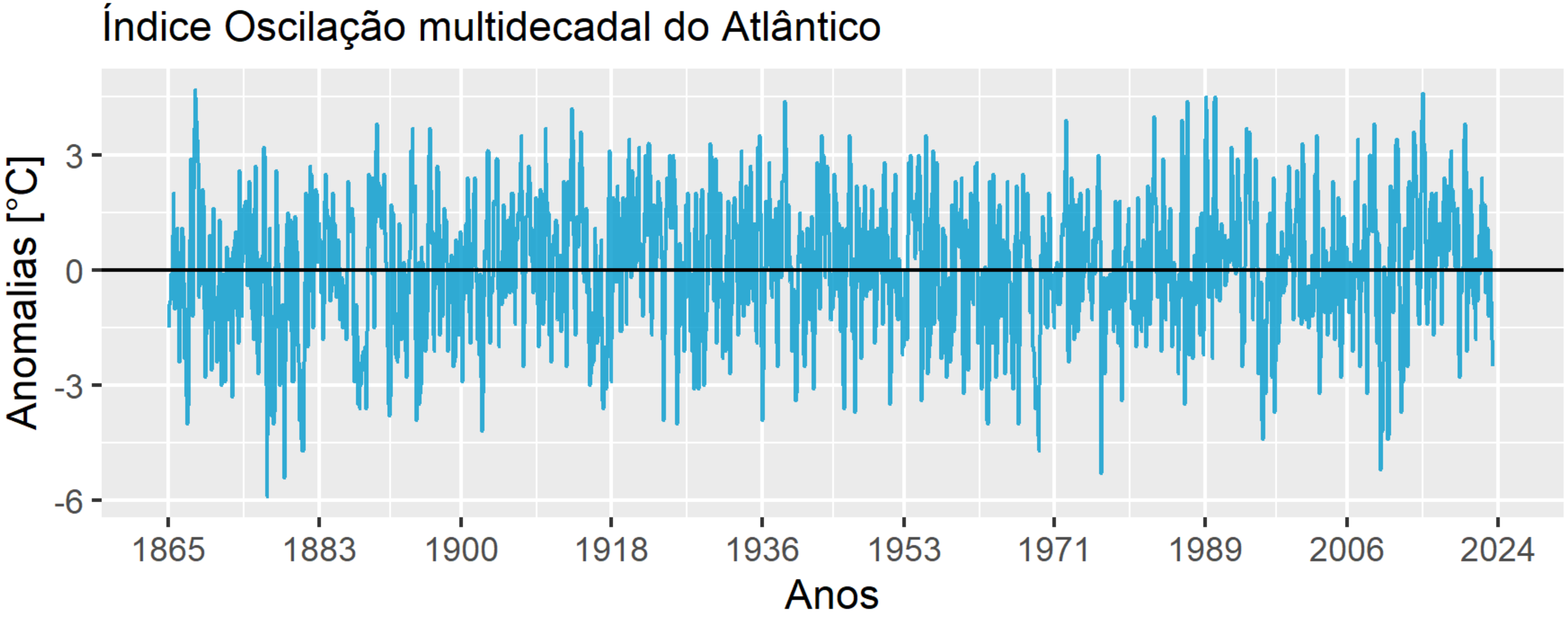
condição frequentemente violada em séries de fenômenos naturais

Koutsoyiannis, D. Nonstationarity versus scaling in hydrology. Journal of Hydrology, v. 324, n. 1-4, p. 239–254, 2006.

Análise de séries temporais | definição e propriedades

Comportamentos não estacionários de séries do cotidiano

sazonalidade, ciclos e tendências



Análise de séries temporais | definição e propriedades

Homogeneidade:

Para *Naghettini & Pinto [2007, p. 265]*

série homogênea: observações extraídas de uma mesma população

série não homogênea: observações advindas de populações diferentes

Séries não homogêneas:

no tempo: tendências, ou quebras de padrão (comportamento não estacionário)

na causa: terremotos, construção de uma barragem, bacias hidrográficas que recebem contribuição de degelo, etc.

Esta é uma definição comum em hidrologia e climatologia

Análise de séries temporais | definição e propriedades

Na literatura clássica de análise de séries temporais *[Box et al., 2008]*, a homogeneidade está atrelada apenas a séries com comportamento não estacionário

Série estacionária	Série não estacionária homogênea	Série não estacionária e não homogênea
	↓	↓
	Não estacionariedade pode ser removida	Não estacionariedade não pode ser removida

(A técnica de diferenciação, usada na remoção da não estacionariedade em séries homogêneas, será apresentada na seção de transformações numéricas)

Análise de séries temporais | definição e propriedades

Homoscedasticidade:

Propriedade associada à constância da variância em relação ao tempo

uma série com variância variável é denominada heteroscedástica

dizer que uma série é heteroscedástica, é equivalente dizer que ela é não estacionária em relação à variância

Propriedade importante na avaliação dos resíduos de um modelo ajustado a uma determinada série

são bons indicadores da qualidade do modelo

(Técnicas de estabilização da variância são apresentada na seção de transformações numéricas)

Análise de séries temporais

características de séries hidrológicas

Análise de séries temporais | séries hidrológicas

Séries hidrológicas possuem particularidades relevantes

norteiam a escolha do modelo estocástico a ser utilizado

1. Têm valores quase sempre limitados a zero

   dados negativos não possuem significado físico para diversas variáveis

   exceções: evaporação líquida de reservatórios, cotas linimétricas (dependendo do referencial), temperatura do ar, etc.

2. Podem apresentar outliers

   valores consideravelmente diferentes do restante da amostra

   recomendável discernimento ao lidar com tais dados

   ex.: é um erro de medição, ou um extremo hidrológico?

Análise de séries temporais | séries hidrológicas

3. Frequentemente não apresentam distribuição Normal

   assimetrias positivas

   requerem cuidados ao usar técnicas que requerem normalidade

   problema minimizado por meio de transformações numéricas

4. São marcados por padrões sazonais

   valores maiores ou menores dependendo do período do ano

   podem ocorrer períodos sazonais distintos

   ex.: estações do ano

Análise de séries temporais | séries hidrológicas

5. Possuem persistência

    dependência de um valor em relação às observações em seu entorno

    no tempo: correlação serial

    no espaço: correlação espacial (proximidade geográfica)

Em geral, uma série hidrológica apresenta todas essas características simultaneamente!

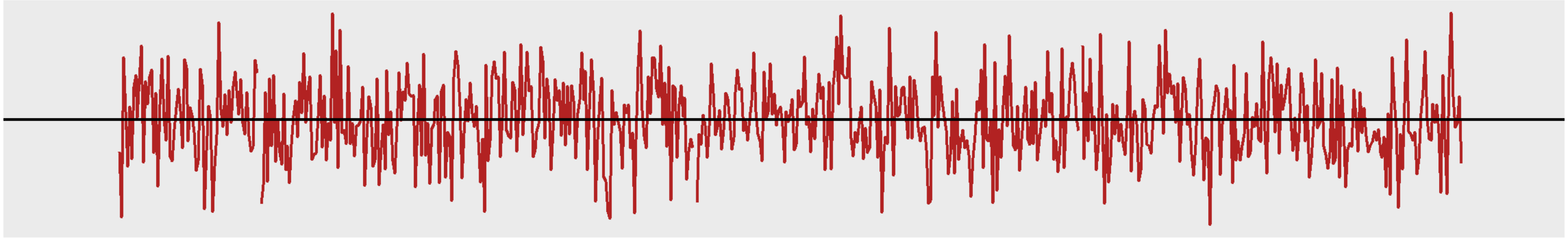
Análise de séries temporais

persistência

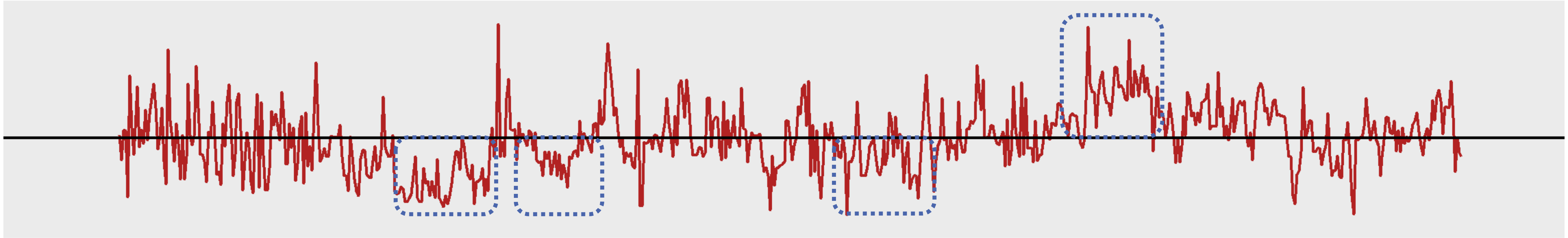
Análise de séries temporais | persistência

As séries mostradas têm estatísticas de média e desvio padrão idênticas

Série independente



Série persistente (níveis mínimos do rio Nilo)



Análise de séries temporais | persistência

Essencialmente: períodos de cheias tendem a ser seguidos por períodos de cheias (o mesmo para estiagens)

"*Joseph Effect*"

[Mandelbrot & Wallis, 1968]

Water Resources Research: 4, 1968, 909-918.

H10

Noah, Joseph, and operational hydrology (M & Wallis 1968)

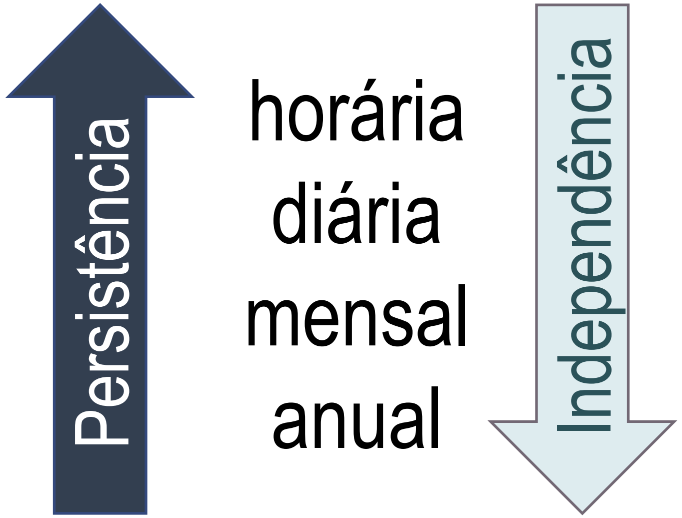
... were all the fountains of the great deep broken up, and the windows of heaven were opened. And the rain was upon the earth forty days and forty nights. *Genesis:* **6**, 11-12.

...there came seven years of great plenty throughout the land of Egypt. And there shall arise after them seven years of famine... *Genesis:* **41**, 29-30.

Dedicated to Harold Edwin Hurst

Análise de séries temporais | persistência

Em geral, quanto menor a escala temporal, mais persistente é a série



A persistência é quantificada por meio de correlações

Coeficiente de correlação de Pearson: a mais bem conhecida medida de correlação

Mede a associação linear entre as variáveis

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

onde

$\bar{x}$ e $\bar{y}$	médias de x e y, respectivamente
s_x e s_y	desvios padrão de x e y, respectivamente
n	tamanho da amostra

Análise de séries temporais | persistência

Para avaliar a persistência: correlação serial (ou correlação em série, ou autocorrelação)

medida de dependência linear entre os elementos de uma série:

$$\hat{r}_k = \frac{1}{n-1} \frac{\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}$$

onde

k *lag* (defasagem temporal)

$$-1 \leq \hat{r}_k \leq 1$$

Análise de séries temporais | persistência

O gráfico $\hat{r}_k$ vs. k é chamado de correlograma ou função de autocorrelação amostral (FAC)

- em geral calculam-se os valores $\hat{r}_k$ sequencialmente para $k = 1,2,\dots,\max k$
- o correlograma de uma série forma uma série
- os valores de $\hat{r}_k$ decaem progressivamente com o aumento de k

Sobre o *lag* máximo a ser considerado na FAC

- atribuído pelo analista, de acordo com (i) o propósito da análise e com (ii) a escala temporal da série
- alternativamente, pode-se adotar de 10% a 30% de n

Análise de séries temporais | persistência

Procura-se o *lag* para o qual $\hat{r}_k = 0$ → marca o fim da dependência

porém, devido a variações amostrais, os valores apenas flutuam em torno de 0

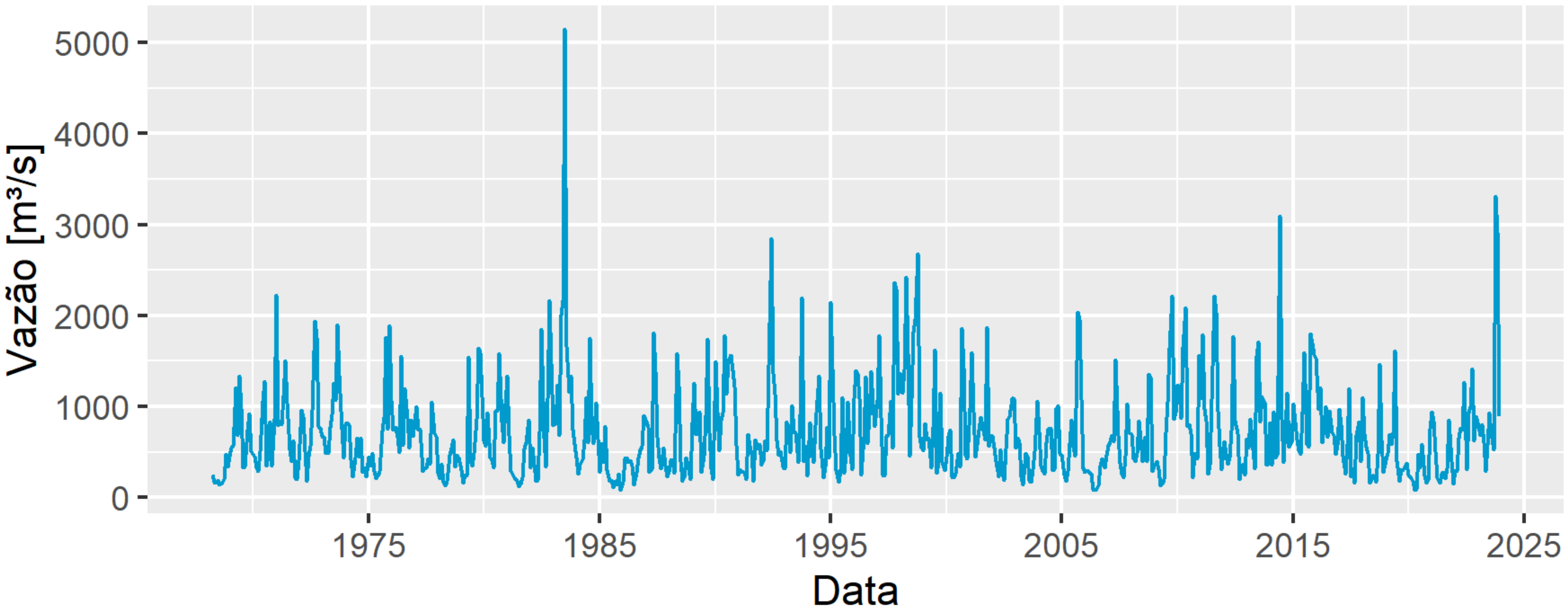
por isso definem-se intervalos de confiança IC_k:

$$IC_k = \pm z \sqrt{\frac{1}{n}\left(1 + 2\sum_{t=1}^{k} \hat{r}_t^2\right)}$$

onde $z = 1{,}96$ (para 95% conf.) ou $z = 2{,}58$ (para 99% conf.)

Análise de séries temporais | persistência

[Exemplo 1] Montagem do correlograma para as vazões médias mensais em Foz do Areia, no rio Iguaçu, para *lag* máximo de 36 meses.



Análise de séries temporais | persistência

Série histórica de abr./1968 a dez./2023

$$\hat{r}_0 = 1$$
$$\hat{r}_1 = 0{,}47$$
$$\hat{r}_2 = 0{,}19$$
$$\vdots$$
$$\hat{r}_{36} = 0{,}03$$

$\bar{Q}$ (m³/s)	$k = 0$
253	253
167	167
159	159
186	186
...	...
524	524
3315	3315
2920	2920
891	891

$\bar{Q}$ (m³/s)	$k = 1$
253	-
167	253
159	167
186	159
...	...
524	709
3315	524
2920	3315
891	2920
-	891

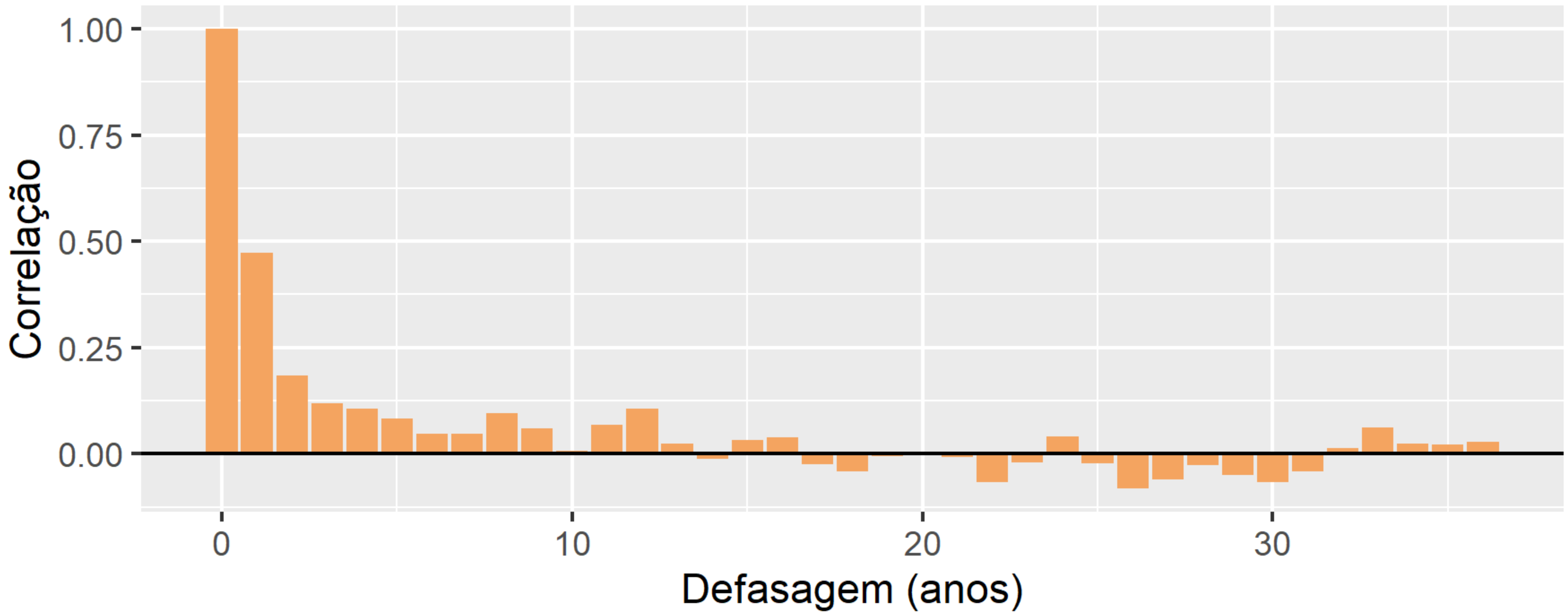
$\bar{Q}$ (m³/s)	$k = 2$
253	-
167	-
159	253
186	167
...	...
524	935
3315	709
2920	524
891	3315
-	2920
-	58

...

Análise de séries temporais | persistência

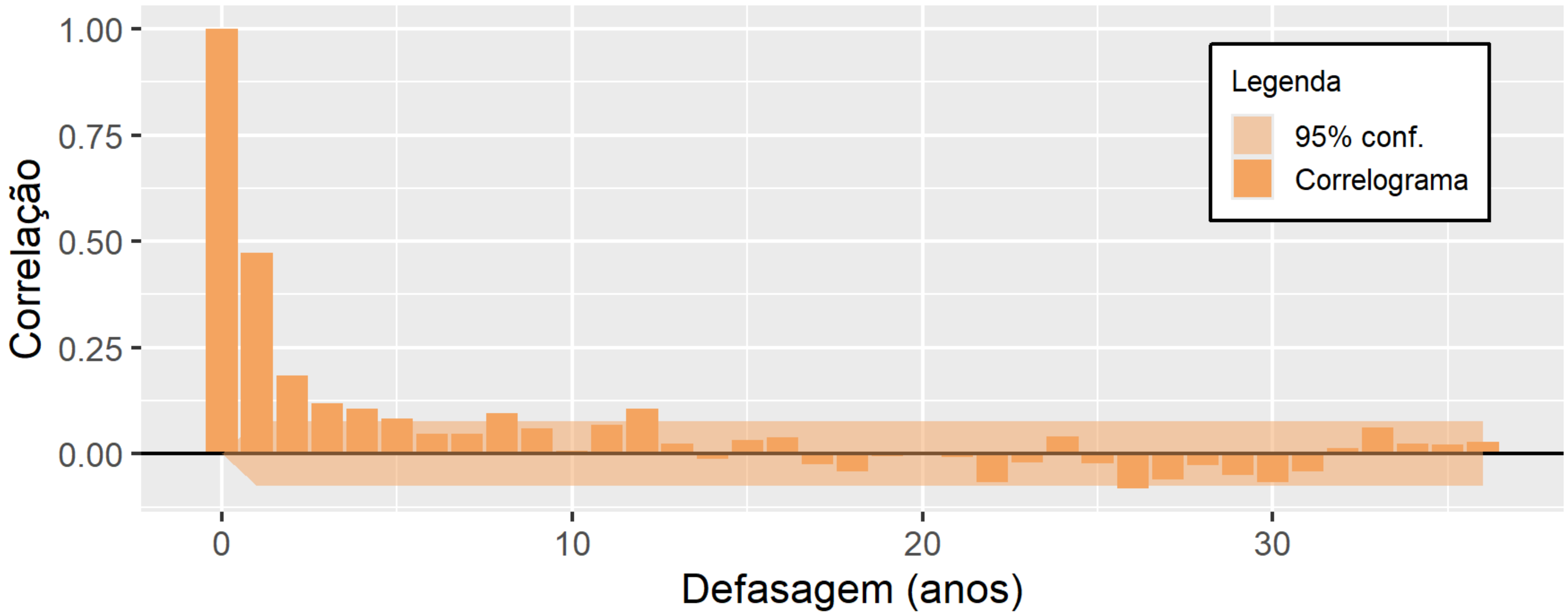
Função de autocorrelação

não possui informação suficiente para determinar quando $\hat{r}_k = 0$



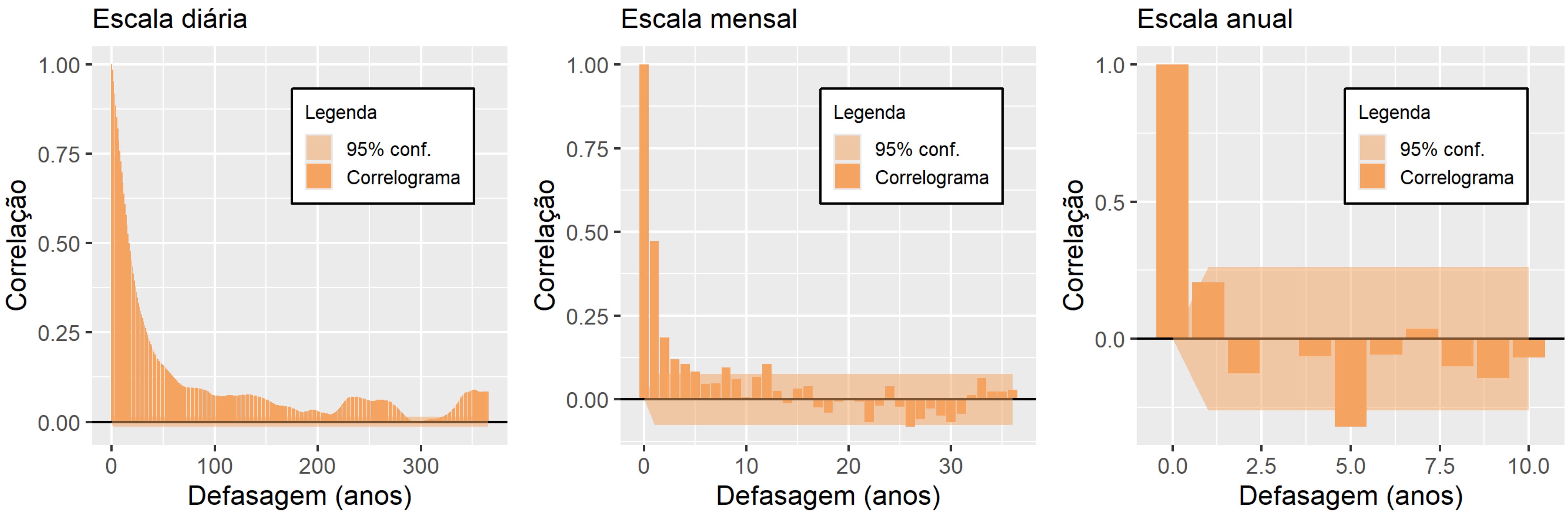
Função de autocorrelação após inclusão dos intervalos de confiança

correlação estatisticamente nula para $k = 5$



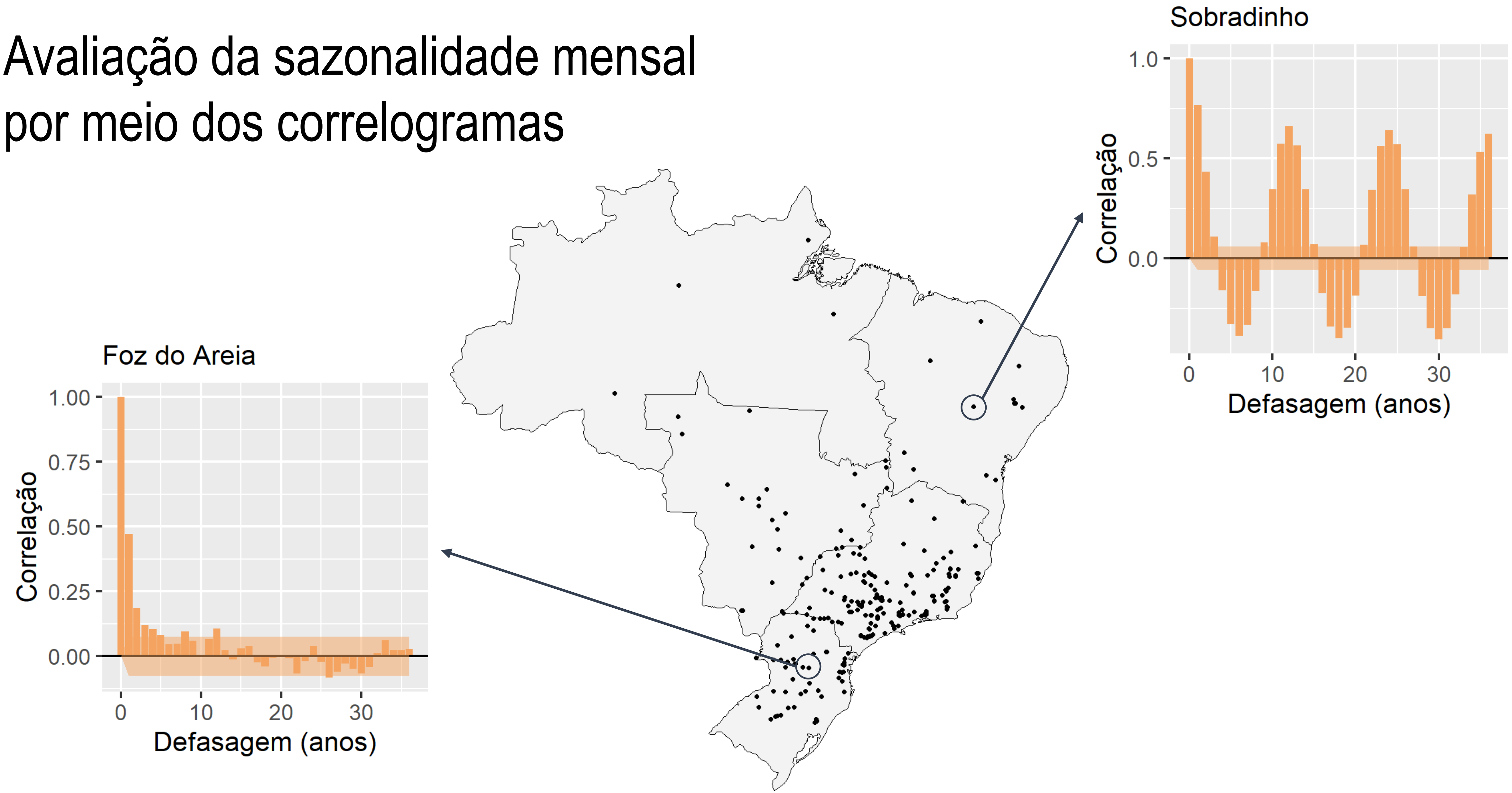
Análise de séries temporais | persistência

Correlogramas para vazões médias em diferentes escalas em Foz do Areia



Análise de séries temporais | persistência

Avaliação da sazonalidade mensal por meio dos correlogramas



Resumo

Testes de hipótese podem ser entendidos como regras de decisão para averiguações estatísticas

atenção para a diferença entre α e p-valor

Processos estocásticos são usados para a modelagem de sistemas que evoluem no tempo e que seguem premissas probabilísticas

séries temporais são amostras de uma das infinitas possíveis realizações de um processo estocástico

Séries hidrológicas apresentam características peculiares

propriedades relevantes: estacionariedade, homoscedasticidade, persistência

ERHA7016 – Hidrologia Estocástica

Daniel Detzel
detzel@ufpr.br