

ERHA7016 – Hidrologia Estocástica

# Análise de séries temporais (pt. 2)

Daniel Detzel  
detzel@ufpr.br

Imagem: [bit.ly/2VV3zhV](https://bit.ly/2VV3zhV)





# Agenda

Análise de séries temporais

tendências

transformações numéricas

decomposição



# ANÁLISE DE SÉRIES TEMPORAIS

## tendências

# Análise de séries temporais | tendências

Tendências são as formas mais comuns de manifestação do comportamento não estacionário em séries hidrológicas

Tema em discussão há mais de meio século [*ver Klemes, 1974*]

coincide com o crescimento do uso de modelos do tipo Box & Jenkins (ARMA)  
a partir da década de 1970  
modelos ARMA são estacionários

Causas comumente atribuídas para as tendências:

ações antrópicas nas bacias hidrográficas  
variabilidade e mudanças climáticas

# Análise de séries temporais | tendências

Discussões acerca da **atribuição** das tendências

atribuição: termo utilizado para “explicar mudanças detectadas”

*[Matalas, 1997]*

Tendências são fruto de impactos de mudanças climáticas

*[Salas et al., 2012]*

Efeito não tão evidente em série hidrológicas

*[Kahya e Kalayci, 2004]*

Mudanças na geomorfologia das bacias são lentas, portanto tendências são causadas por mudanças no clima

# Análise de séries temporais | tendências

Não estacionariedade e tendências: falta de consenso na literatura

*[Klemes, 1974]*

Discussões acerca da (não) estacionariedade de séries são “um exercício de futilidade”

*[Koutsoyiannis, 2006, 2011, 2013]*

Tendências são produtos de flutuações de larga escala

*[Milly et al., 2008]*

“Stationarity is dead”

# Análise de séries temporais | tendências

*[Lins & Cohn, 2008]*

“Stationarity: wanted dead or alive?”

*[Poveda e Álvarez, 2012]*

“El colapso de la hipótesis de estacionariedad”

*[Poveda, 2012]*

“Fin al diletantismo (*e.g. amadorismo*) sobre el calentamiento global y su origen antrópico”

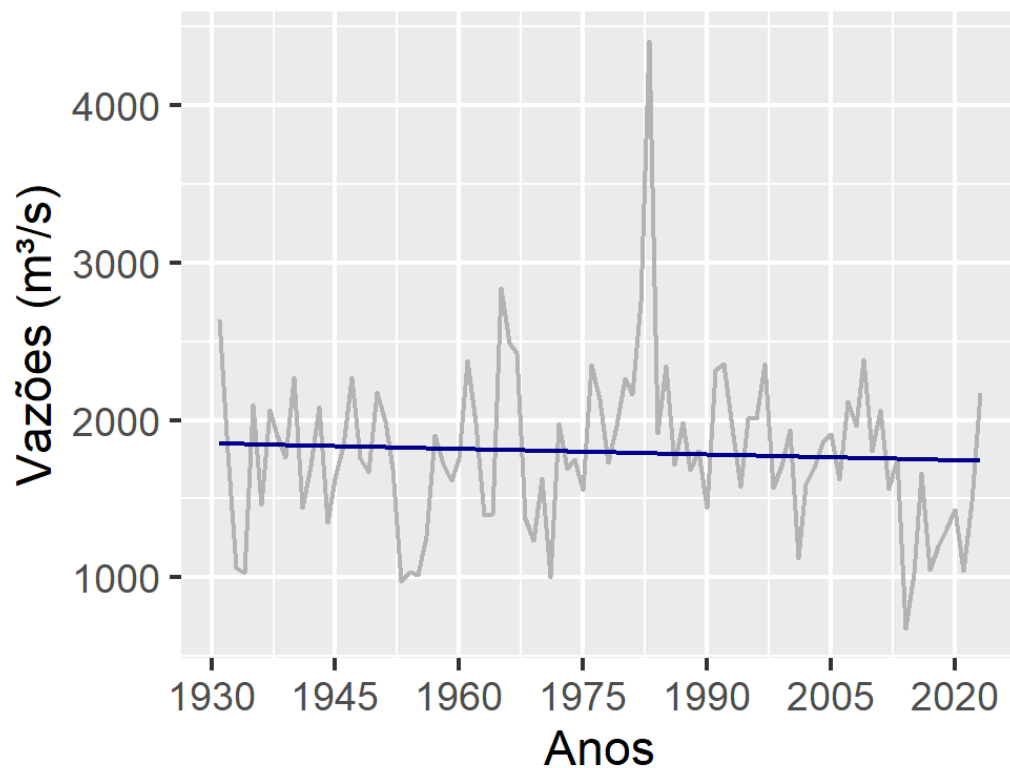
*[Serinaldi e Kilsby, 2015]*

“Stationarity is undead”

# Análise de séries temporais | tendências

## Tendências em séries hidrológicas (anuais)

Marimbondo (rio Grande, MG/SP)



Legenda

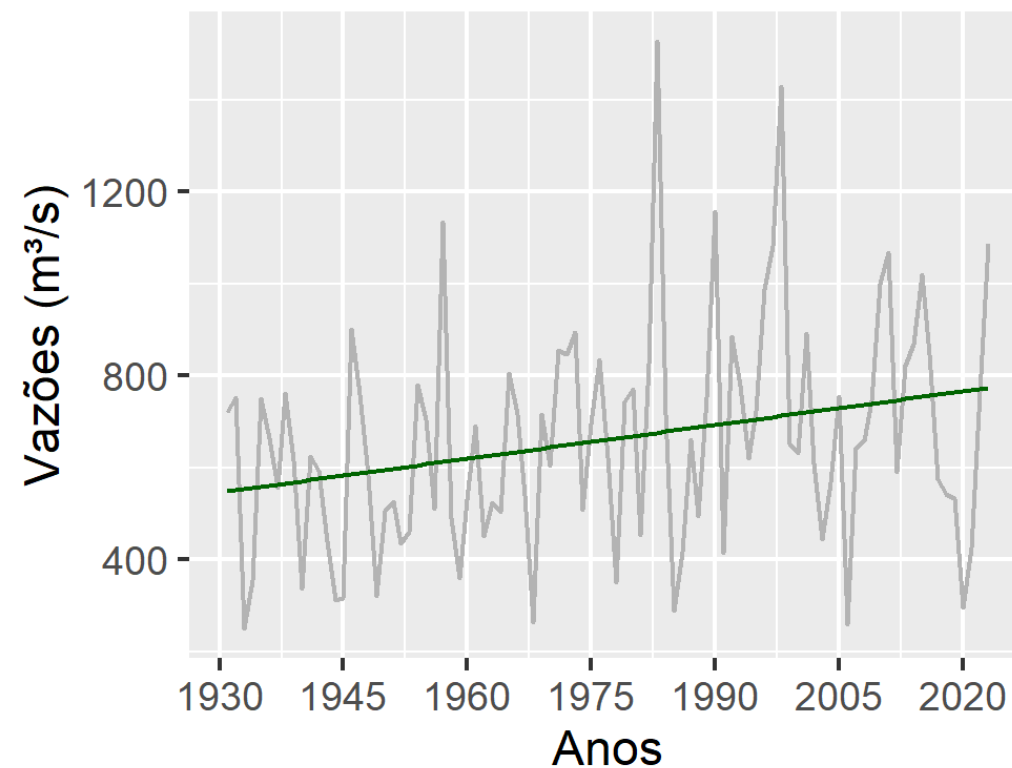


Tendência



Vazões

Foz do Areia (rio Iguaçu, PR)



Legenda



Tendência



Vazões



# Análise de séries temporais | tendências

A detecção das tendências se dá por meio de testes de hipótese

## Teste de Mann-Kendall (MK)

Teste não paramétrico

Tendências não necessariamente lineares, mas sim monotônicas

Hipóteses:

$H_0$ : a série não possui tendência monotônica

$H_1$ : a série possui tendência monotônica

# Análise de séries temporais | tendências

A estatística do teste é calculada por meio de:

$$MK = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}[z_j - z_i]$$

onde  $z_i$  (ou  $z_j$ ) é a observação no instante de tempo  $i$  (ou  $j$ ) e:

$$\text{sgn}[z_j - z_i] = \begin{cases} 1, & \text{se } (z_j - z_i) > 0 \\ 0, & \text{se } (z_j - z_i) = 0 \\ -1, & \text{se } (z_j - z_i) < 0 \end{cases}$$

# Análise de séries temporais | tendências

A variância da variável do teste  $MK$  é dada por:

$$VAR[MK] = \frac{n(n-1)(2n+5)}{18}$$

A variável do teste é calculada por:

$$z = \begin{cases} \frac{MK - 1}{\sqrt{VAR[MK]}}, & \text{se } MK > 0 \\ \frac{MK + 1}{\sqrt{VAR[MK]}}, & \text{se } MK < 0 \end{cases}$$



# Análise de séries temporais | tendências

Rejeita-se  $H_0$  se  $z > z_{\alpha/2}$  para um nível de significância  $\alpha$

Lembra-se que os testes possuem a premissa de que os elementos das amostras devem ser **independentes** entre si

MK em amostras persistentes têm maior probabilidade de indicar tendência quando ela, na realidade, ela não existe [Yue et al., 2002]

É preciso investigar a persistência da série antes de averiguar tendências  
em dados persistentes, o teste de MK deve ser adaptado

# Análise de séries temporais | tendências

## Teste para correlação em série de primeira ordem ( $\hat{r}_1$ )

Hipóteses:

$H_0$ : a série é independente ( $\hat{r}_1 = 0$ )

$H_1$ : a série não é independente ( $\hat{r}_1 \neq 0$ )

Estatística do teste:

$$t_0 = \frac{\hat{r}_1 \sqrt{n-2}}{\sqrt{1 - \hat{r}_1^2}}$$

onde

$n$  tamanho da série

$\hat{r}_1$  coeficiente de correlação de lag 1

# Análise de séries temporais | tendências

Rejeita-se  $H_0$  se:

$$|t_0| > t_{\alpha/2, (n-2)}$$

onde

$t_{\alpha/2, (n-2)}$  variável t-Student com  $n - 2$  graus de liberdade significância de  $\alpha$

O p-valor pode ser calculado por:

$$p = 2 \cdot \Phi^{-1}[1 - |t_0|]$$

onde

$\Phi^{-1}$  inversa da distribuição t-Student com  $n - 2$  graus de liberdade



# Análise de séries temporais | tendências

*[Exemplo]* Significância de  $\hat{r}_1$  nas vazões médias anuais dos rios Grande (usina Marimbondo) e Iguaçu (usina Foz do Areia), considerando  $\alpha = 5\%$ .

	Rio Grande <b>Marimbondo</b>	Rio Iguaçu <b>Foz do Areia</b>
$n$ (anos)	93 (1931 a 2023)	93 (1931 a 2023)
$\hat{r}_1$	0,35	0,25
$t_0$	3,60	2,41
$t_{\alpha/2;(n-2)}$	1,99	1,99
p-valor	< 0,001	0,02

Veredicto: Rejeita-se  $H_0$  em ambos os casos (a série não é independente)

# Análise de séries temporais | tendências

Caso  $H_0$  seja rejeitada, existem algumas opções:

*[Yue et al., 2002]*: adaptar a série temporal

*[Hamed & Rao, 1998]*: adaptar a formulação do teste

*[Yue et al., 2002]* propõem o uso da técnica *pre-whitening*

a partir da série persistente  $z_t$ , obtém-se a série independente  $x_t$  aplicando:

$$x_t = z_t - \hat{r}_1 z_{t-1}$$

onde

$\hat{r}_1$  coeficiente de correlação de primeira ordem (*lag* 1) da série

# Análise de séries temporais | tendências

Entretanto, a mera presença da tendência na série afeta a estimativa de  $\hat{r}_1$   
o estimador passa a ser **tendencioso**

A proposta é, então:

1. Remover a tendência, usando o estimador de Theil-Sen  $b$   
versão não paramétrica do estimador de mínimos quadrados do coeficiente angular de uma reta

$$b = \text{mediana} \left( \frac{z_j - z_i}{j - i} \right), \forall i < j$$



# Análise de séries temporais | tendências

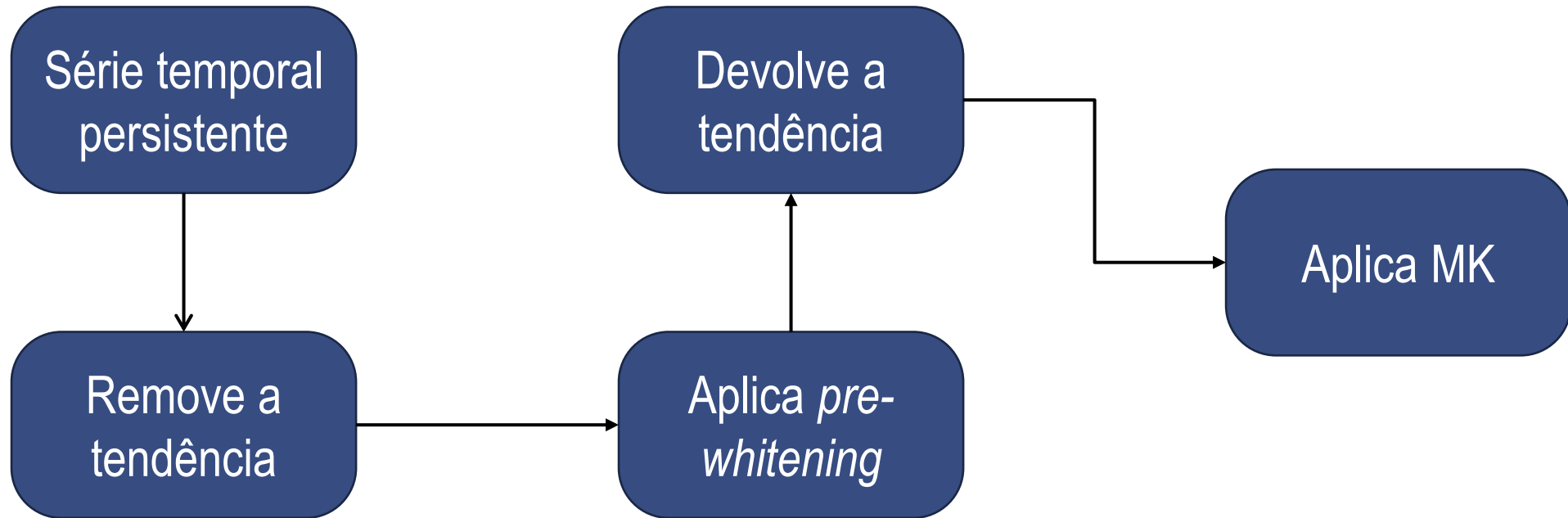
A remoção da tendência é feita a partir de:

$$y_t = z_t - bt$$

2. O *pre-whitening* é aplicado à série  $y_t$ , obtendo-se  $x_t$   
o coeficiente  $\hat{r}_1$  passa a ser estimado a partir da amostra sem tendência  $y_t$
3. Devolve-se a tendência à série (“branca”)  $x_t$
4. Finalmente, aplica-se MK sobre a série  $x_t$

# Análise de séries temporais | tendências

Em resumo:



Por esse motivo, a abordagem de [Yue et al., 2002] é conhecida como *Trend-Free Pre-Whitening (TFPW)*

# Análise de séries temporais | tendências

[Hamed & Rao, 1998] propõe alterar a variância do teste  
séries persistentes tendem a inflar a variância de  $MK$

Lembra-se que a variância é baseada unicamente no tamanho da amostra:

$$VAR[MK] = \frac{n(n-1)(2n+5)}{18}$$

Portanto, o autor sugere usar um estimador denominado **tamanho efetivo da amostra**



# Análise de séries temporais | tendências

O tamanho efetivo da amostra é calculado por:

$$n^* = \frac{n}{1 + 2 \sum_{k=1}^{n-1} (1 - \frac{k}{n}) r_k}$$

onde

$k$  defasagem temporal

$r_k$  autocorrelação de lag  $k$  estimada a partir da amostra

Assim, para séries persistentes,  $n^* < n$  e o valor da variância do teste é reduzido

# Análise de séries temporais | tendências

*[Exemplo]* Tendências nas vazões médias anuais dos rios Grande (usina Marimbondo) e Iguaçu (usina Foz do Areia) – no R, pacote `modifiedmk`

Rio Grande  
Marimbondo

```
mkOriginal <- mkttest(x)  
P-value = 0.463
```

```
mkTFPW <- tfpwmk(x)  
P-value = 0.853
```

```
mkHR <- mmkh(x, ci = 0.95)  
P-value = 0.585
```

Rio Iguaçu  
Foz do Areia

```
mkOriginal <- mkttest(x)  
P-value = 0.013
```

```
mkTFPW <- tfpwmk(x)  
P-value = 0.005
```

```
mkHR <- mmkh(x, ci = 0.95)  
P-value = 0.008
```

# **ANÁLISE DE SÉRIES TEMPORAIS**

transformações numéricas

# Análise de séries temporais | transformações numéricas

Transformações numéricas são técnicas utilizadas em alguns contextos:

- deixar a distribuição dos dados mais simétrica (próxima de uma Normal)
- regularizar a variação da série (controlar sua variância)
- lidar com o comportamento não estacionário

Adicionalmente, são aplicadas para adequar os dados aos requisitos de modelagem, quando necessários

As transformações a serem mostradas são:

- diferenciação
- Box-Cox/Logarítmica



# Análise de séries temporais | transformações numéricas

## Diferenciação:

Utilizada para remover o comportamento não estacionário de uma série  
assume tendências estocásticas

Seja uma série  $z_t$  ( $t = 1, \dots, n$ ), a diferenciação de primeira ordem é dada por:

$$x_t = z_t - z_{t-1}$$

# Análise de séries temporais | transformações numéricas

Séries com comportamentos não estacionários mais complexos (ou seja, com tendências e quebras estruturais), podem requerer diferenciações de segunda ordem:

$$y_t = x_t - x_{t-1} = (z_t - z_{t-1}) - (z_{t-1} - z_{t-2})$$

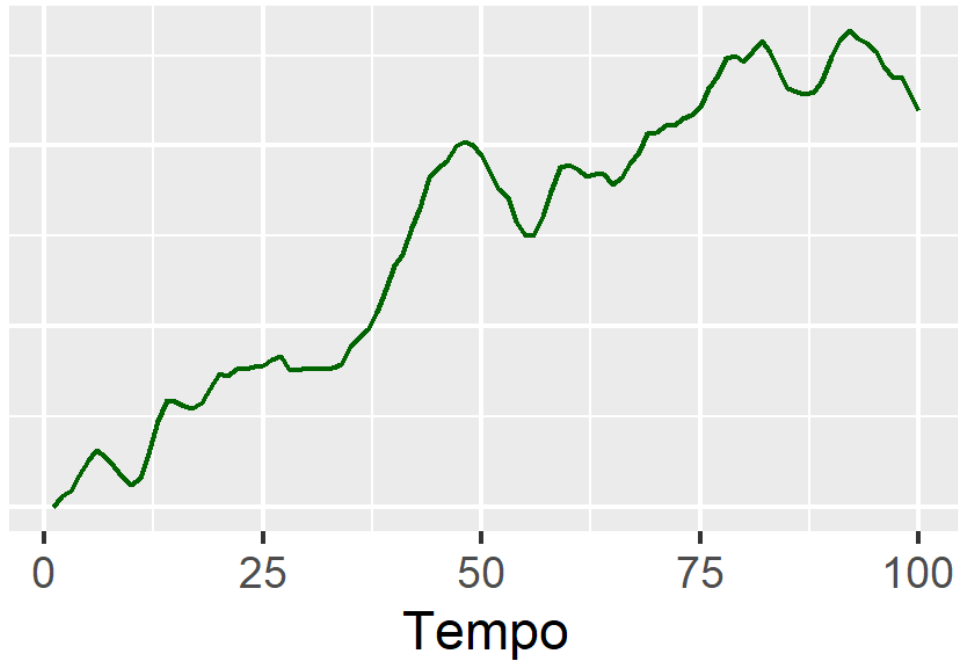
Cuidado deve ser tomado ao usar diferenciações, pois a reintegração da série (ou seja, o retorno à sua escala original) pode não ser possível

o processo requer que o valor inicial  $z_0$  seja conhecido  
para séries sintéticas  $z'_t$  não se conhece o valor  $z'_0$

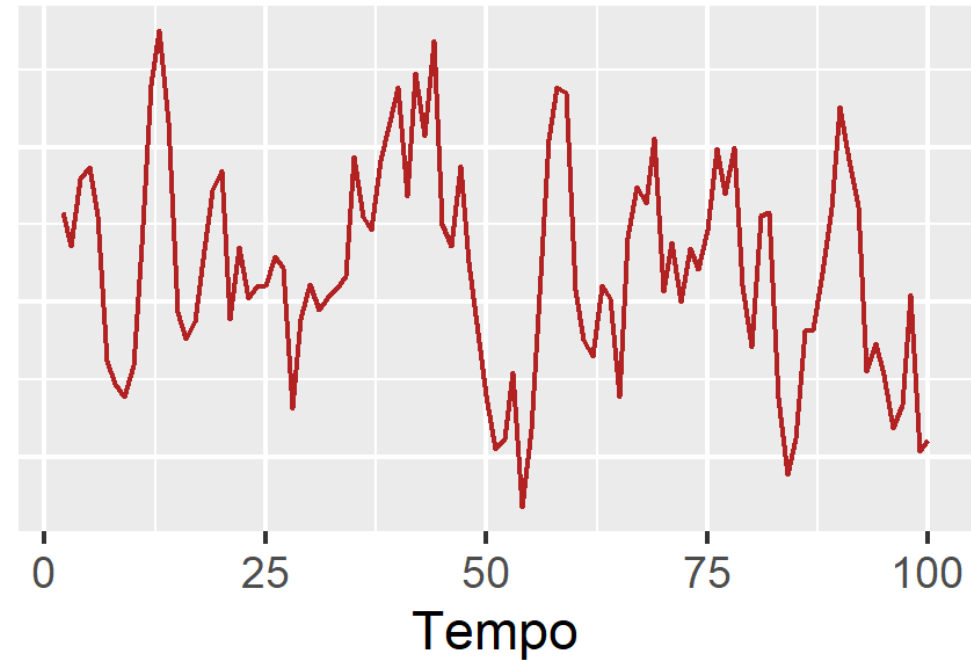
# Análise de séries temporais | transformações numéricas

Se a(s) diferenciação(ões) resultar(em) em um série  $x_t$  ou  $y_t$  estacionária,  $z_t$  é uma série não estacionária **homogênea\***

Original



Diferenciada

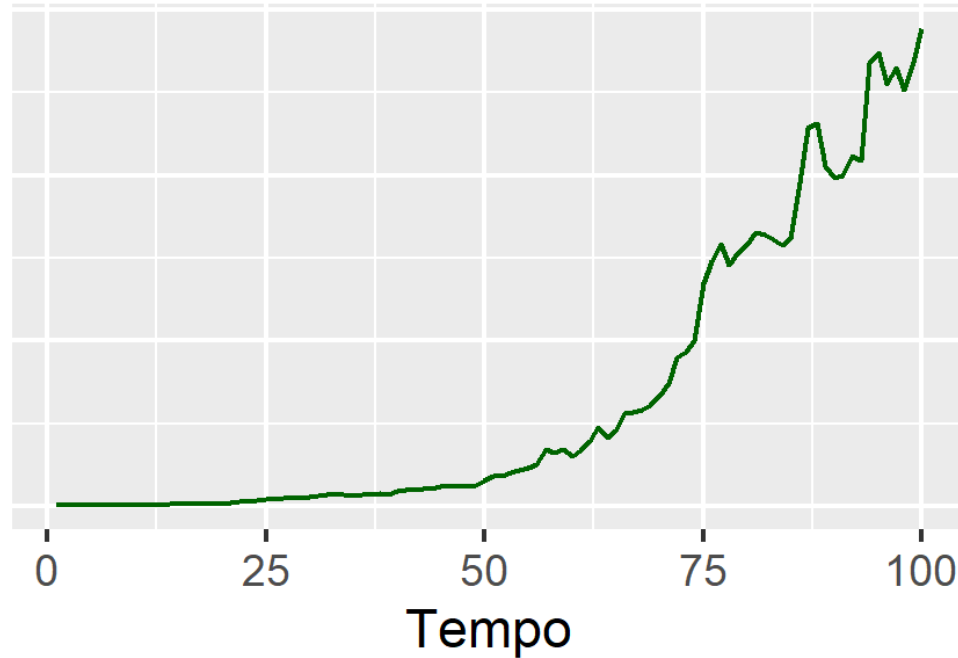


\*o conceito de homogeneidade foi explicado na aula passada

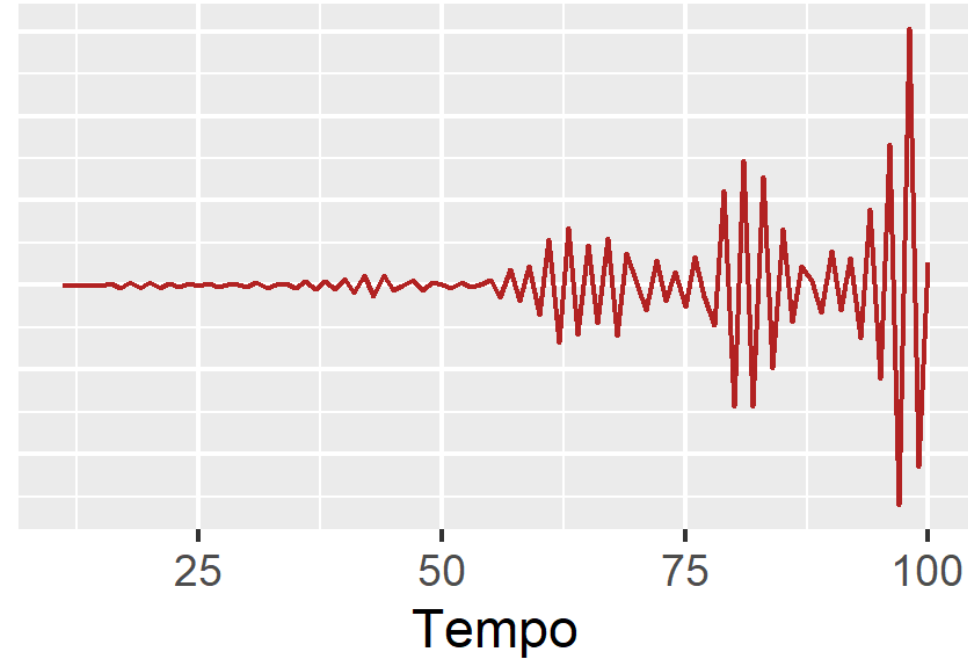
# Análise de séries temporais | transformações numéricas

Caso contrário,  $z_t$  é uma série não estacionária **não homogênea**

Original



Diferenciada (10x)



# Análise de séries temporais | transformações numéricas

## Box-Cox/Log-Normal:

A transformação de Box-Cox tem por objetivo estabilizar a variância de uma série (amenizar a [heteroscedasticidade](#)):

$$x_t = \begin{cases} \frac{(z_t + \epsilon)^\lambda - 1}{\lambda} & ; \lambda \neq 0 \\ \ln(z_t + \epsilon) & ; \lambda = 0 \end{cases}$$

onde

$\lambda$       parâmetro da transformação

$\epsilon$       constante arbitrada quando da presença de números negativos na série



# Análise de séries temporais | transformações numéricas

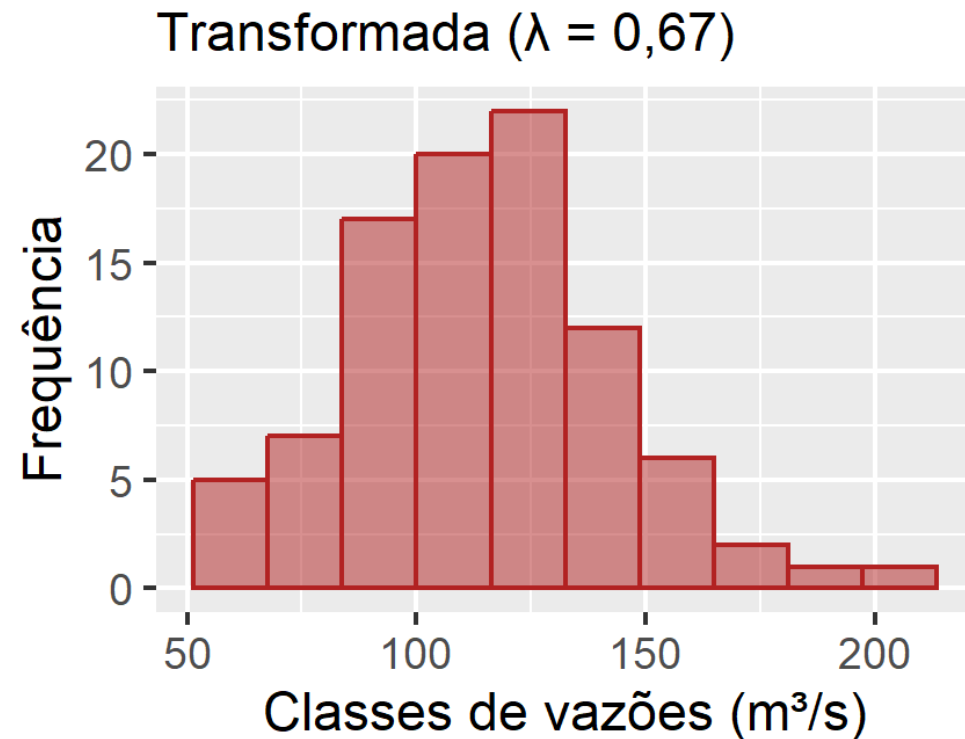
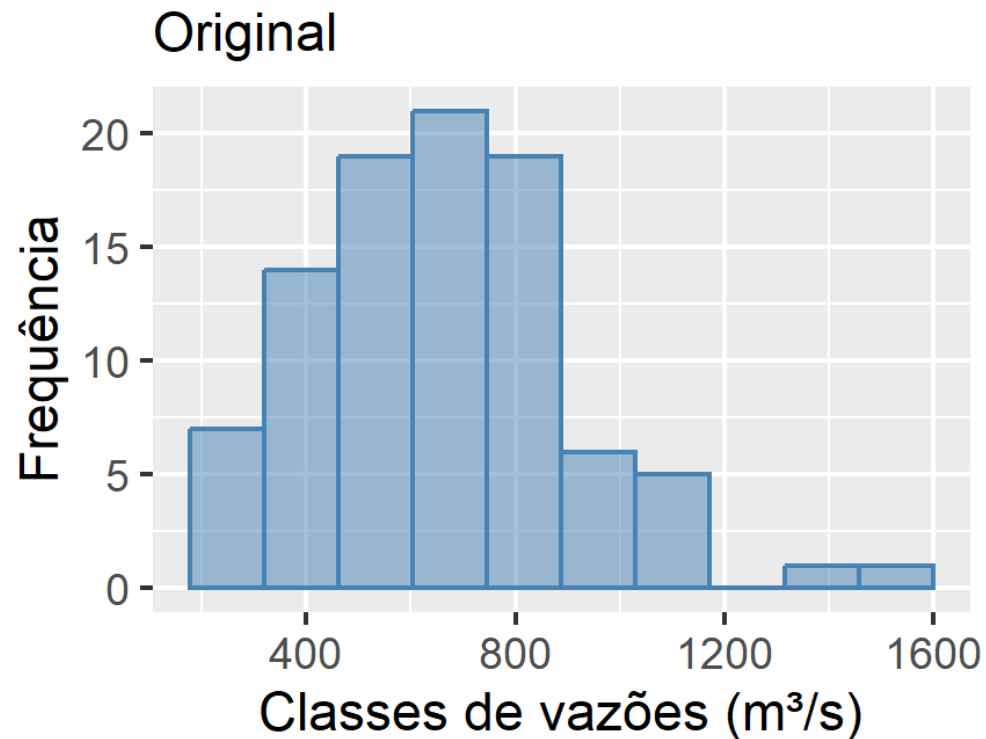
Um efeito secundário importante é que a transformação Box-Cox ajuda a aproximar a distribuição dos dados a uma Normal  
estabilizar a variância reduz a assimetria dos dados

Este efeito é esperado e pode ser utilizado como objetivo para a estimativa do parâmetro  $\lambda$

variam-se os valores até que a série transformada  $x_t$  seja (aprox.) normalmente distribuída

# Análise de séries temporais | transformações numéricas

*[Exemplo]* Transformação Box-Cox aplicada às vazões médias anuais do rio Iguaçu (usina Foz do Areia)



# Análise de séries temporais | transformações numéricas

Contudo, há uma limitação importante quando do uso dessa transformação quando o objetivo é a geração de cenários sintéticos

Na reversão da transformação para o retorno à escala original dos dados, instabilidades numéricas aparecem quando  $\lambda \cong 0$ :

$$z'_t = \begin{cases} (\lambda x'_t + 1)^{1/\lambda} - \epsilon; & \lambda \neq 0 \\ \exp(x'_t + \epsilon) & ; \lambda = 0 \end{cases}$$

# Análise de séries temporais | transformações numéricas

Alternativa: aplicar diretamente a transformação log-normal sobre os dados

$$x_t = \log(z_t)$$

Onde  $x_t \sim N(\mu_x, \sigma_x)$

Assim:

$$x_t^p = \frac{\log(z_t) - \mu_x}{\sigma_x}$$

Resulta em uma série  $x_t^p \sim N(0,1)$

Tendências em séries podem significar alterações nos regimes hidrológicos em bacias hidrográficas

significância estatística pode ser testada, com a devida atenção à premissa de independência entre valores da série

Transformações numéricas são úteis para preparar os dados

remoção de não estacionariedade via diferenciação requer cuidados

transformações Box-Cox e logarítmica aproximam dados de uma distribuição Normal





# ERHA7016 – Hidrologia Estocástica

Daniel Detzel  
detzel@ufpr.br