

Curso de extensão em jurimetria

Alexandre Chibebe Nicolella
José de Jesus Filho
Ricardo Feliz Okamoto

Objetivos da aula

1. Compreender o modelo de detecção de ponto de mudança
2. Discutir formas de monitoramento

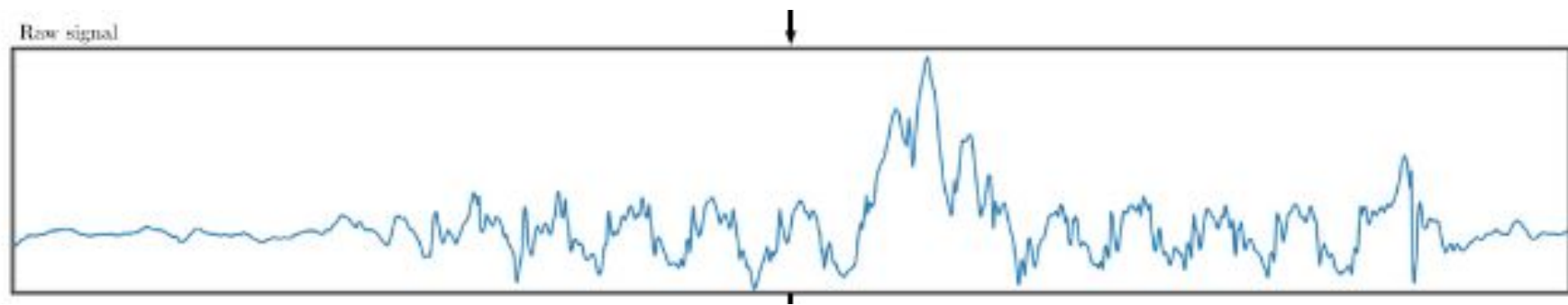
Análise de ponto de mudança (*change point analysis*)

Percurso

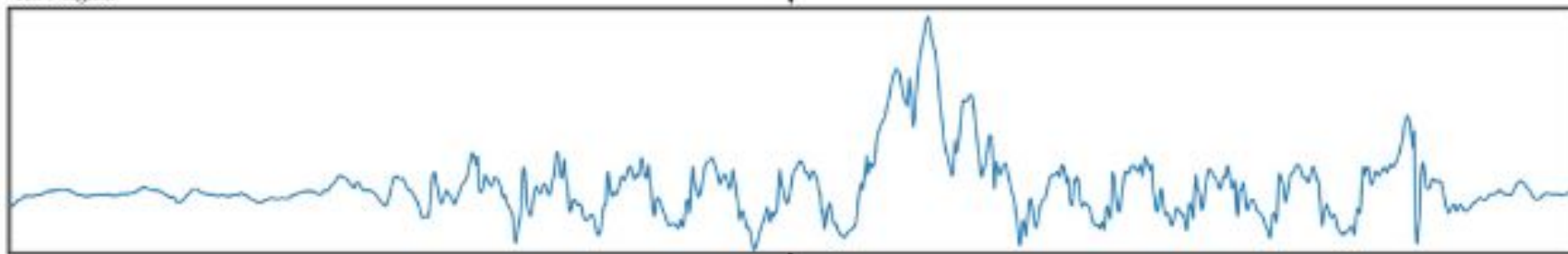
1. Preliminares
2. O que é o modelo de CPT?
3. Cuidados metodológicos importantes

Preliminares

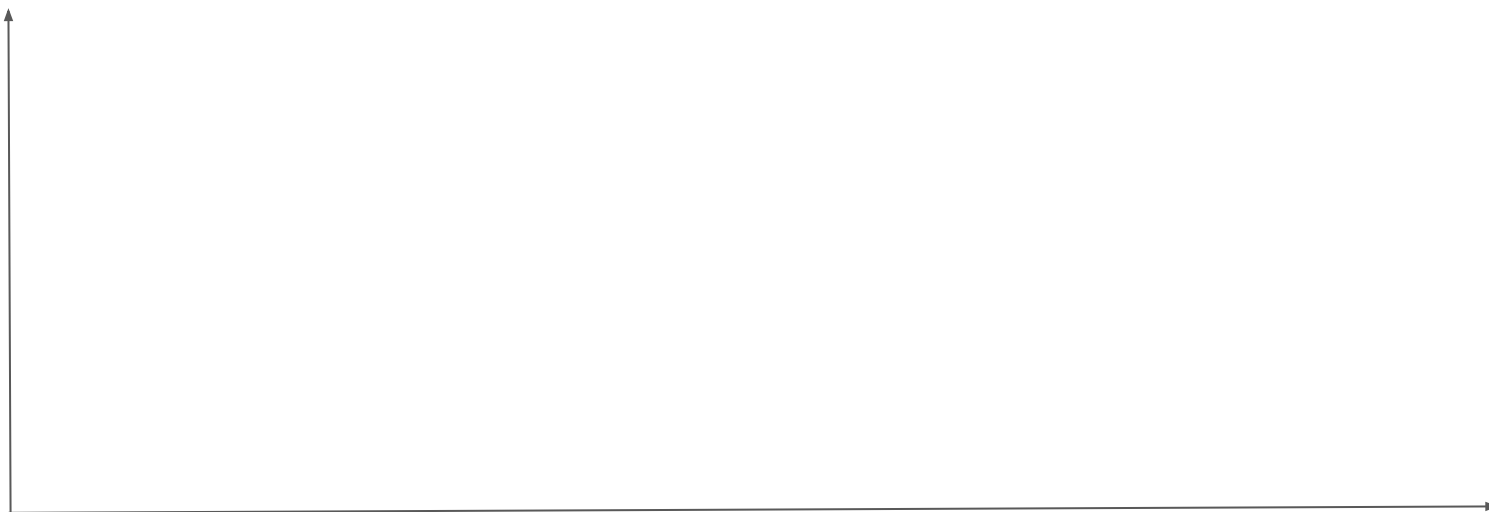
Raw signal



Raw signal



$y = \text{variável}$



$x = \text{tempo}$

Raw signal

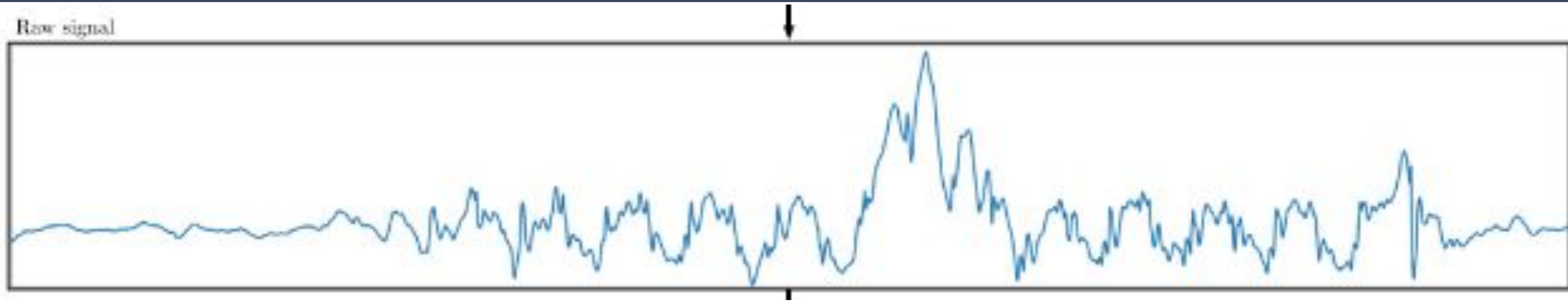


y = variável

Vamos tentar imaginar como
é a base de dados subjacente
a esse gráfico

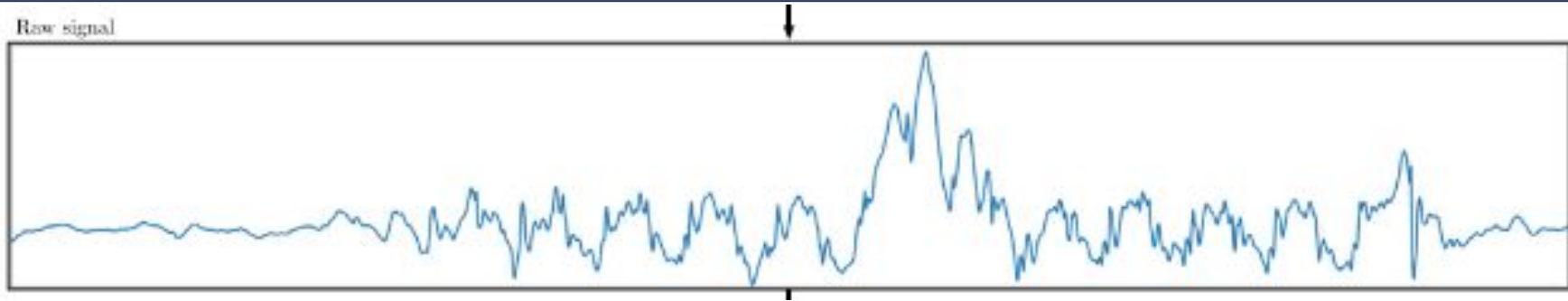
x = tempo

Vamos tentar imaginar como é a base de dados subjacente a esse gráfico



- O que estamos observando? Isto é, qual é a unidade de observação deste gráfico?
- Quais são as informações sobre essa unidade de observação?
- Portanto, o que significa cada linha e cada coluna?

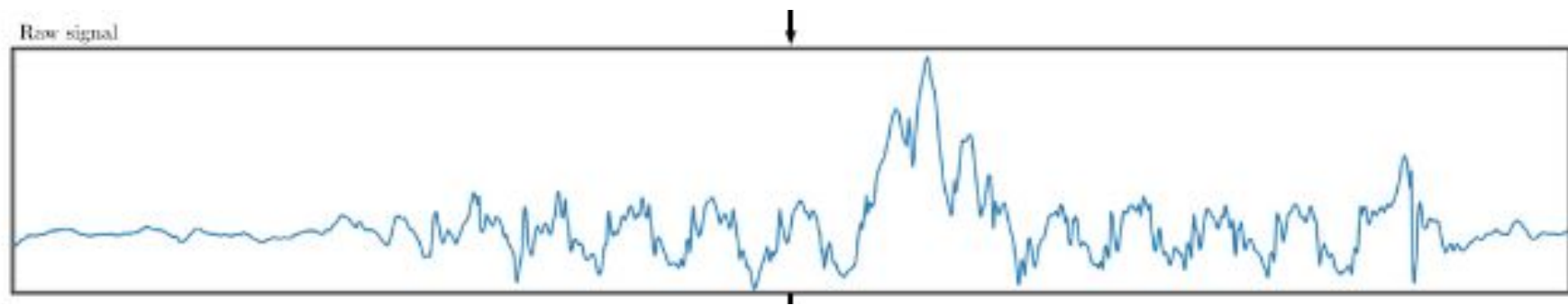
Vamos tentar imaginar como é a base de dados subjacente a esse gráfico



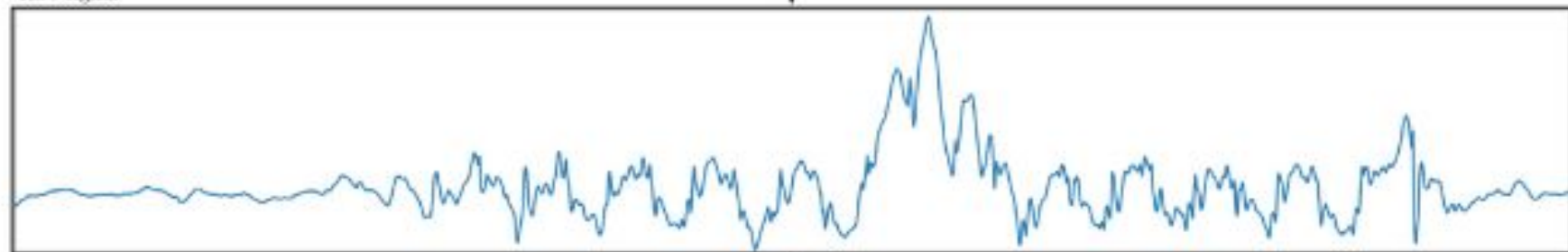
- O que estamos observando? Isto é, qual é a unidade de observação deste gráfico? **Uma data/hora**
- Quais são as informações sobre essa unidade de observação? **A variável que estamos observando**
- Portanto, o que significa cada linha e cada coluna? **Cada linha é uma data e cada coluna é uma informação que coletamos nesta data**

Data	Variável
01/01/2021	20
02/01/2021	28
03/01/2021	26
04/01/2021	19
...	...
24/11/2024	23

Raw signal

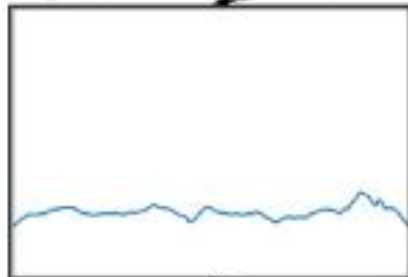


Raw signal

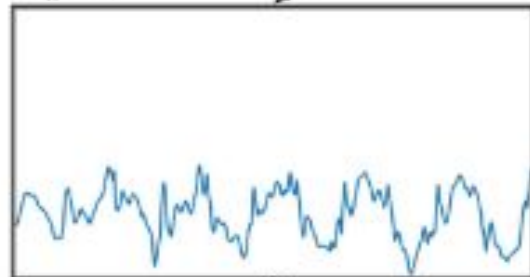


Change point detection

Regime 1



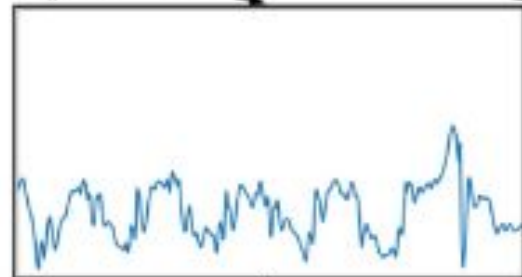
Regime 2



Regime 3



Regime 4



Regime 5

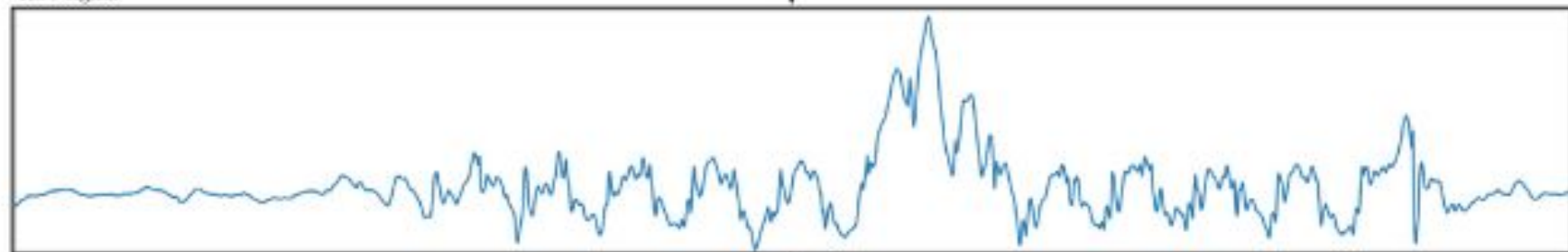


O modelo de CPT

O que o modelo de ponto de mudança quer responder?

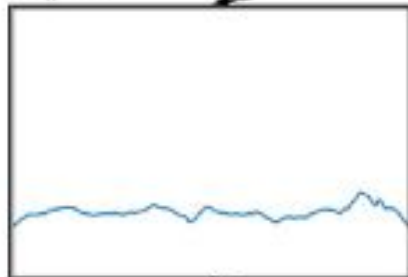
1. **Houve alguma mudança nos dados ao longo do tempo?** Queremos identificar se houve mudança ou não.
2. **Se sim, quando foi essa mudança?** Queremos determinar o ponto exato desta mudança.
3. **Qual é a diferença entre os dados antes e depois da mudança?** Uma vez que determinamos *quando* foi a mudança, queremos comparar o antes e o depois.
4. **Qual a nossa certeza sobre a localização do ponto de mudança?** Queremos acessar também a confiança que temos no ponto de mudança detectado.
5. **Quantas mudanças ocorreram?** Queremos saber também quantas mudanças ocorreram

Raw signal

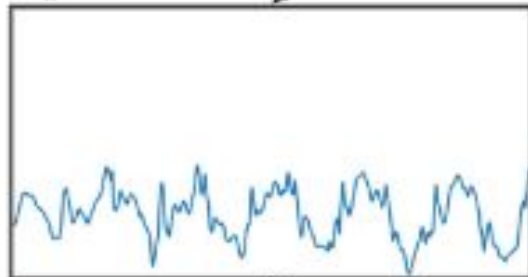


Change point detection

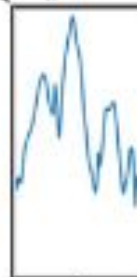
Regime 1



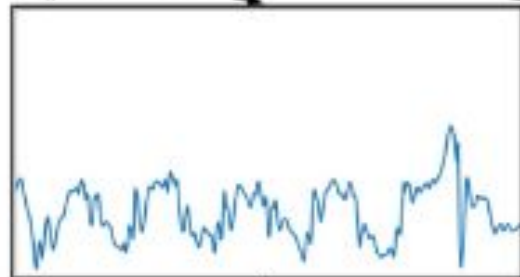
Regime 2



Regime 3



Regime 4

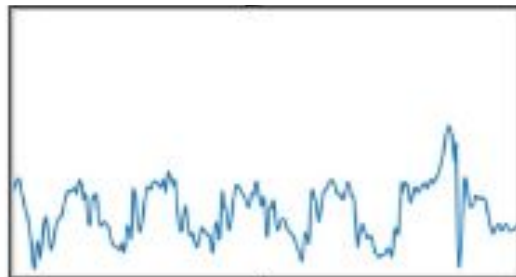
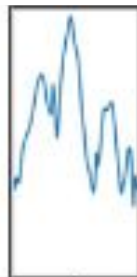
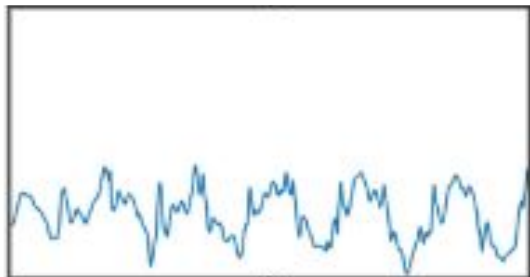
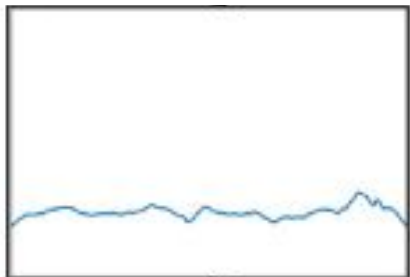


Regime 5



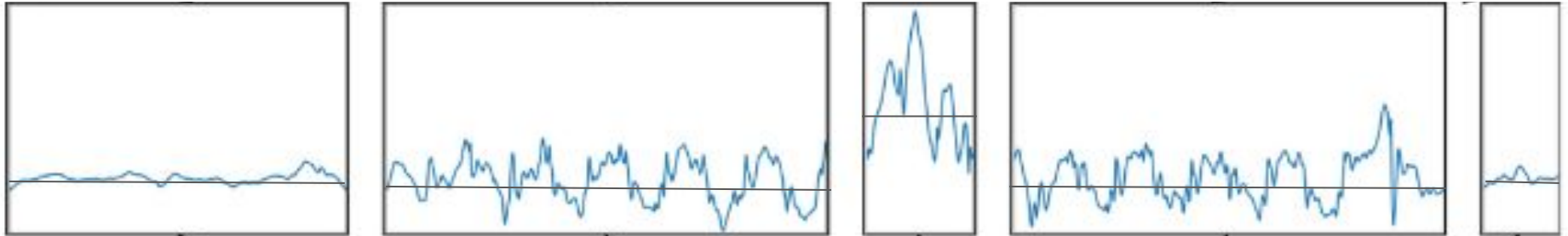
A análise de ponto de mudança (changepoint analysis) está buscando um ponto em que ocorre uma “mudança”. **Mas é uma mudança em relação ao quê?**

- Média
- Variância
- Coeficiente (*slope*)

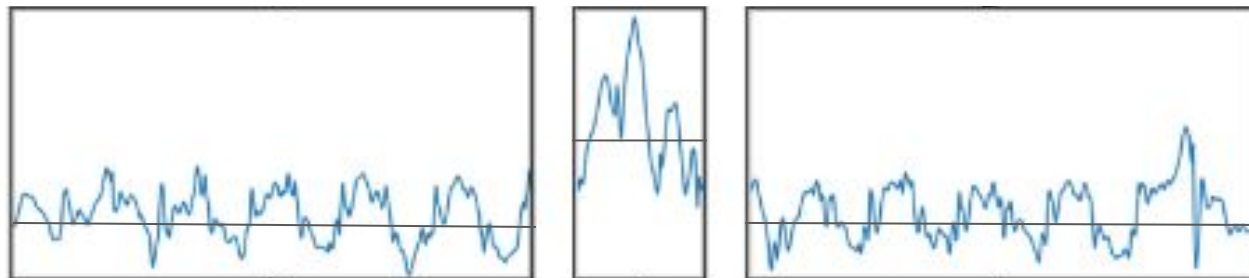


A análise de ponto de mudança (changepoint analysis) está buscando um ponto em que ocorre uma “mudança”. Mas é uma mudança em relação ao quê?

- Média
- Variância
- Coeficiente (*slope*)

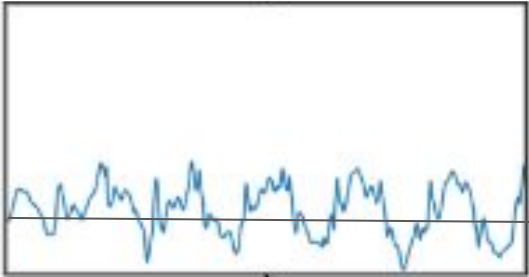


Mudança de média



Data	Variável
08/03/2022	20
09/03/2022	28
10/03/2022	26
11/03/2022	19
...	...
18/09/2022	23

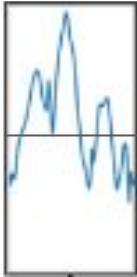
22



08/03/2022 18/09/2022

Data	Variável
19/09/2022	35
20/09/2022	36
21/09/2022	39
22/09/2022	40
...	...
07/10/2022	33

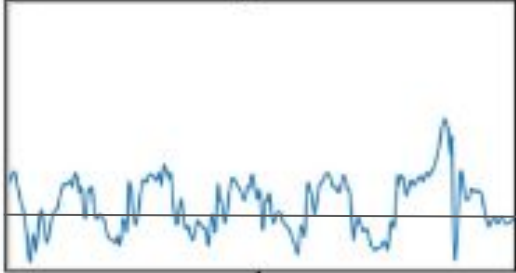
35



19/09/2022 07/10/2022

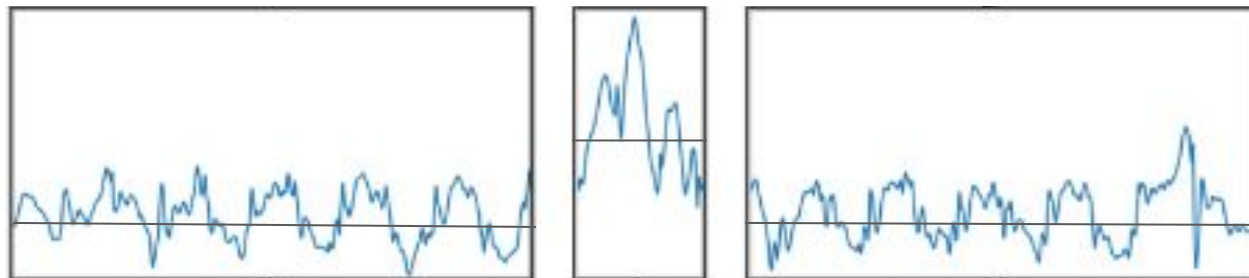
Data	Variável
08/10/2022	21
09/10/2022	22
10/10/2022	25
11/10/2022	17
...	...
09/02/2023	23

22

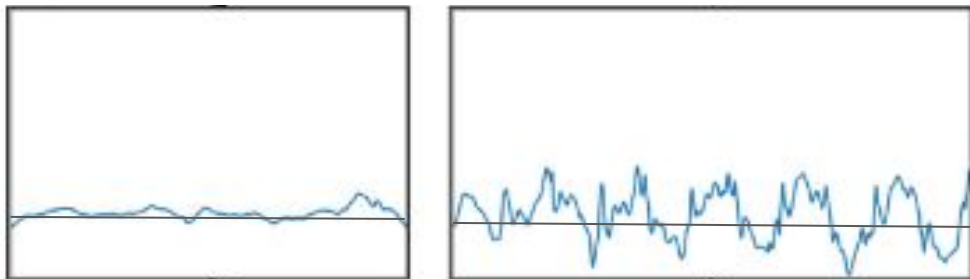


08/10/2022 09/02/2023

Mudança de média



Mudança de variância

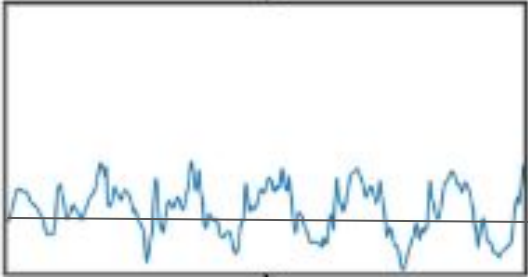


Data	Variável	22	Desvio	1
01/01/2022	23		1	
02/01/2022	20		-2	
03/01/2022	22		0	
04/01/2022	23		1	
...	
07/03/2022	22		0	



01/01/2022 17/03/2022

Data	Variável	22	Desvio	4
08/03/2022	20		-2	
09/03/2022	28		6	
10/03/2022	26		4	
11/03/2022	19		-3	
...	
18/09/2022	23		1	



08/03/2022 18/09/2022

Gráfico 1

média = 3,79
desvio padrão = 1,38

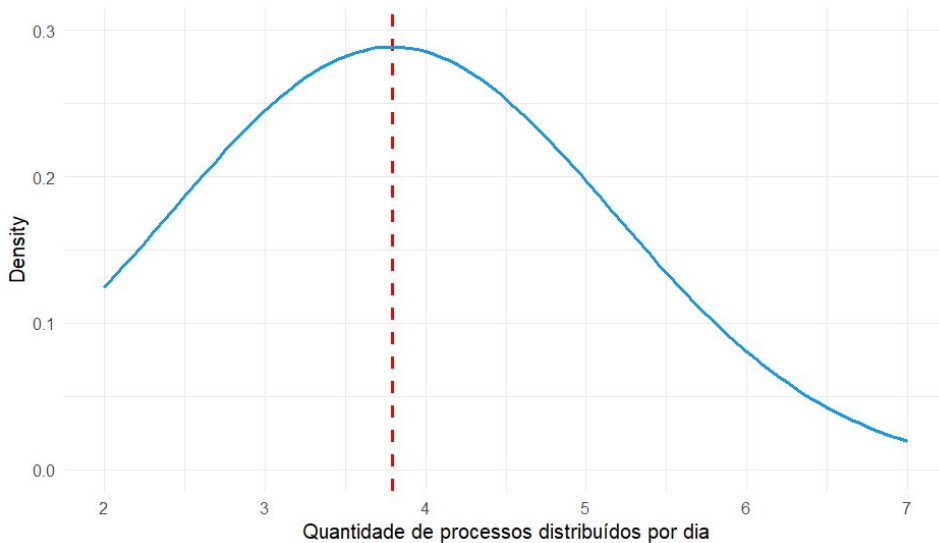
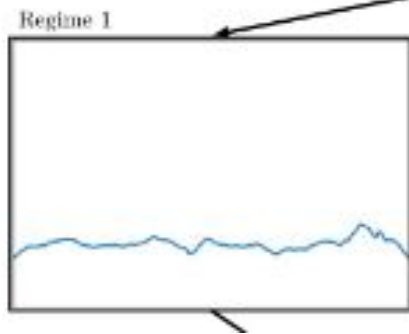
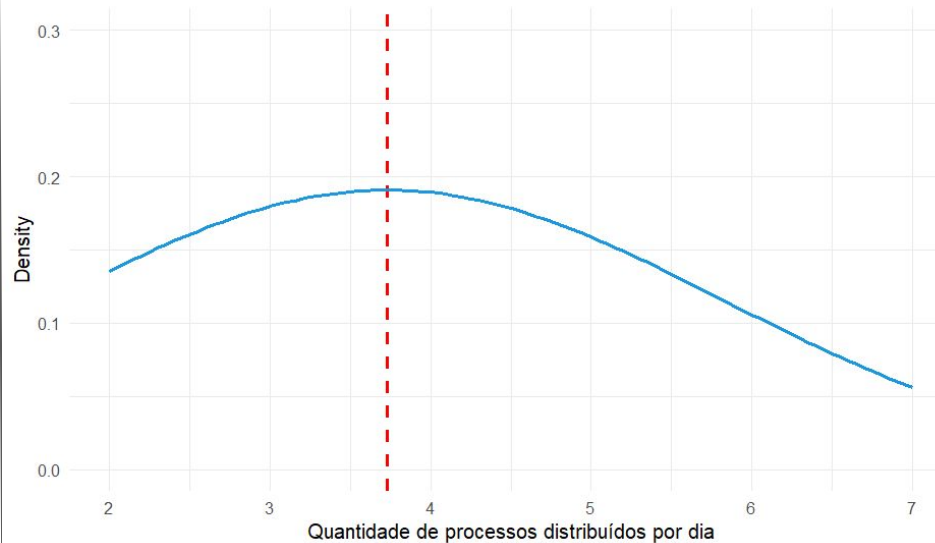
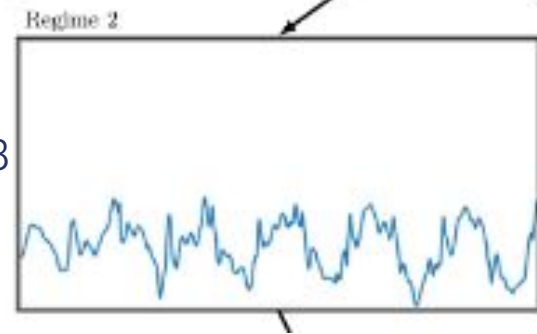


Gráfico 2

média = 3,73
desvio padrão = 3,18

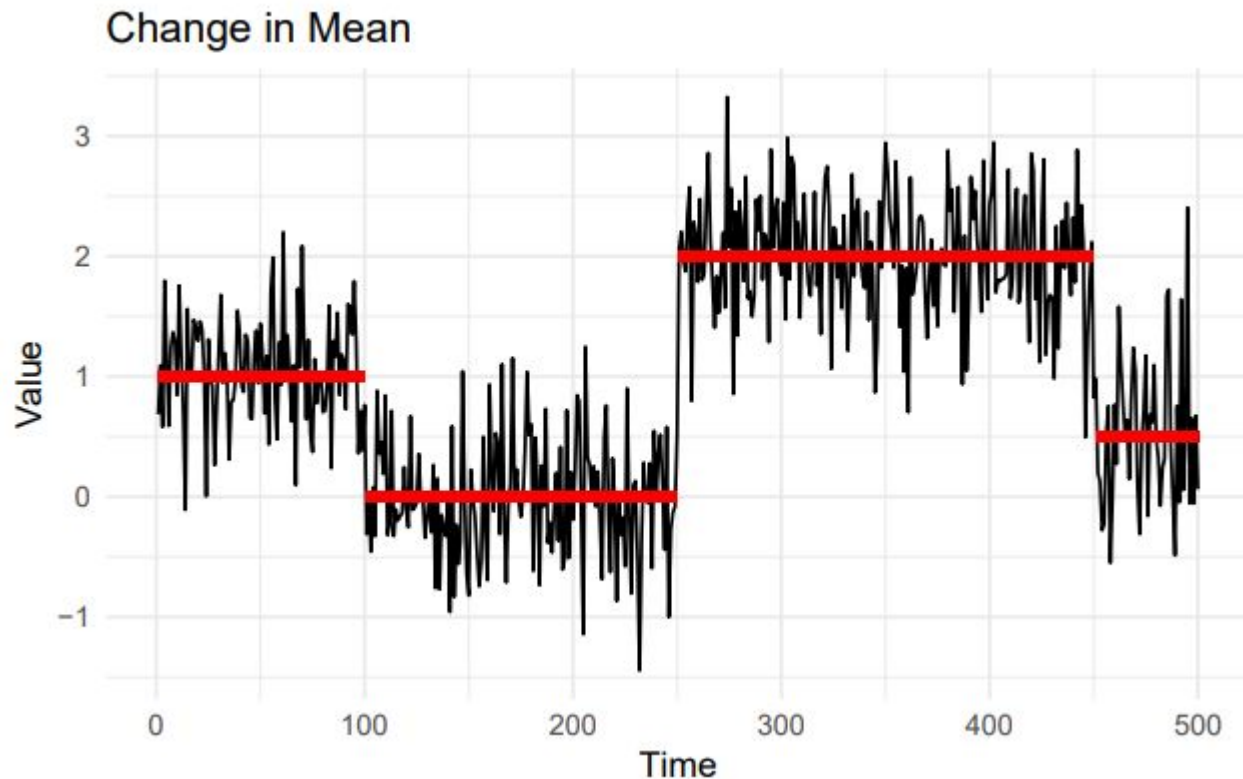


Tipos de mudança

- Média
- Variância
- Regressão

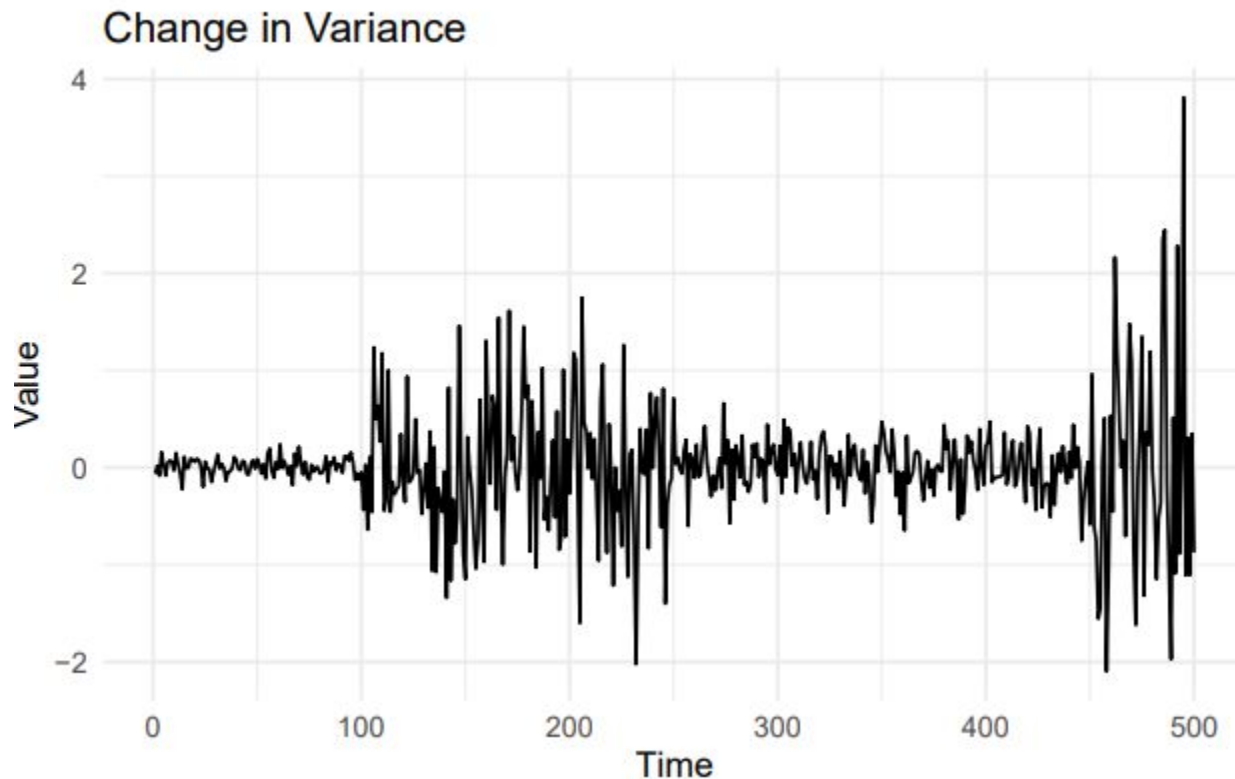
Tipos de mudança

- Média
- Variância
- Regressão



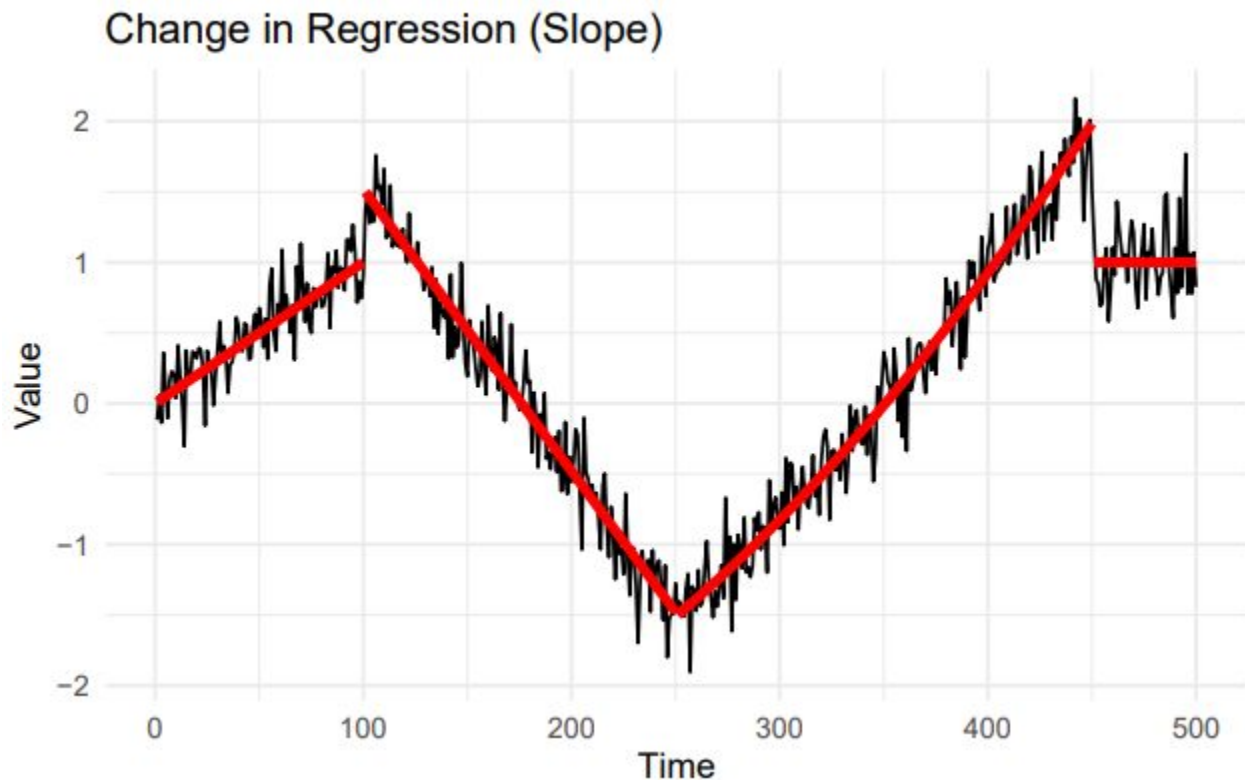
Tipos de mudança

- Média
- Variância
- Regressão



Tipos de mudança

- Média
- Variância
- Regressão



Tipos de análises

O objetivo do *changepoint* é analisar quando uma série temporal tem um ponto de virada. Essa análise pode ser online ou offline.

- Na análise online, a análise é feita em tempo real, a cada momento em que chegam dados novos. Neste caso, o objetivo do changepoint é conseguir identificar um ponto de virada o mais rápido possível.
- Na análise offline, a análise é feita com todos os dados já prontos. Neste caso, o objetivo do changepoint é conseguir identificar o ponto de virada da forma mais acurada possível.

Análise offline

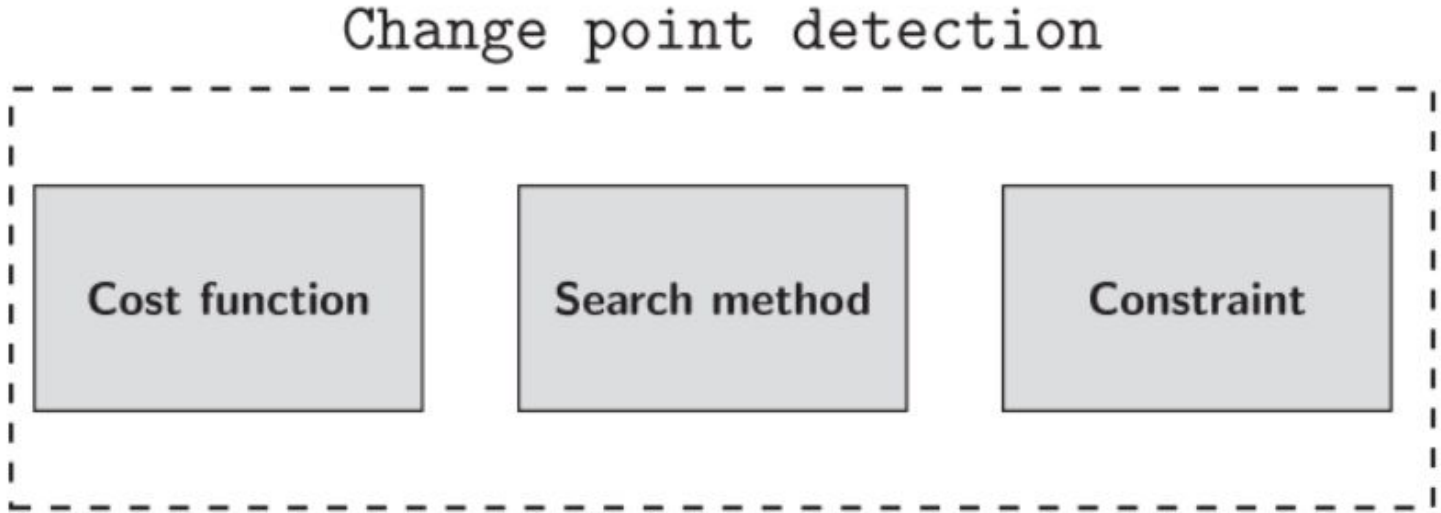


Fig. 2. Typology of change point detection methods described in this article. Reviewed algorithms are defined by three elements: a cost function, a search method and a constraint (on the number of change points).

Função de custo

- A função de custo é uma medida de homogeneidade. Aqui, escolhemos o tipo de mudança que queremos detectar.
 - Se o segmento for homogêneo, então a função de custo terá um sinal baixo.
 - Se o segmento for heterogêneo, então a função de custo terá um sinal alto
- Que tipo de custo é esse? Vamos ver o custo do “resíduo”

Análise offline

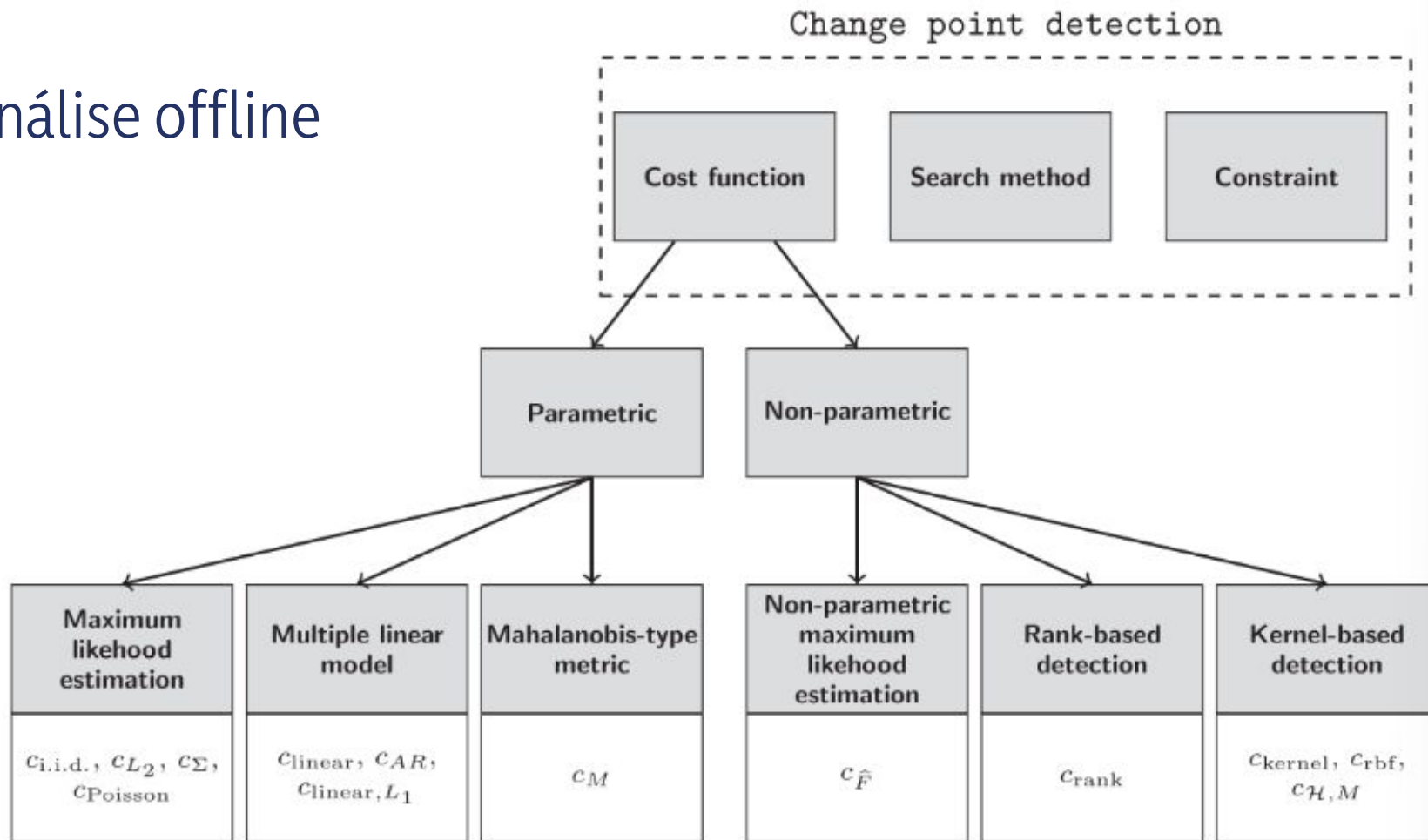


Fig. 6. Typology of the cost functions described in [Section 4](#).

Análise offline

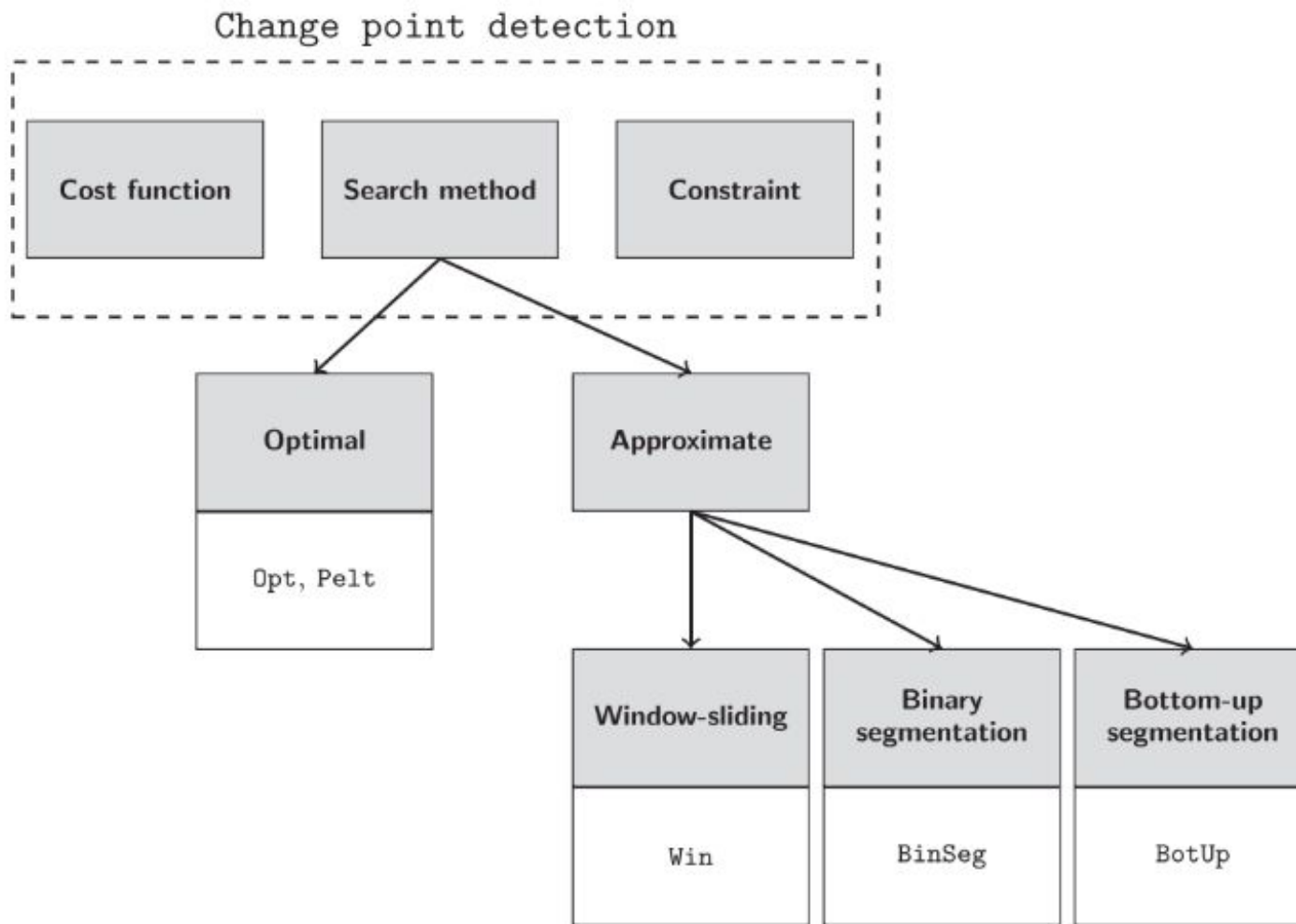


Fig. 7. Typology of the search methods described in [Section 5](#).

Análise offline

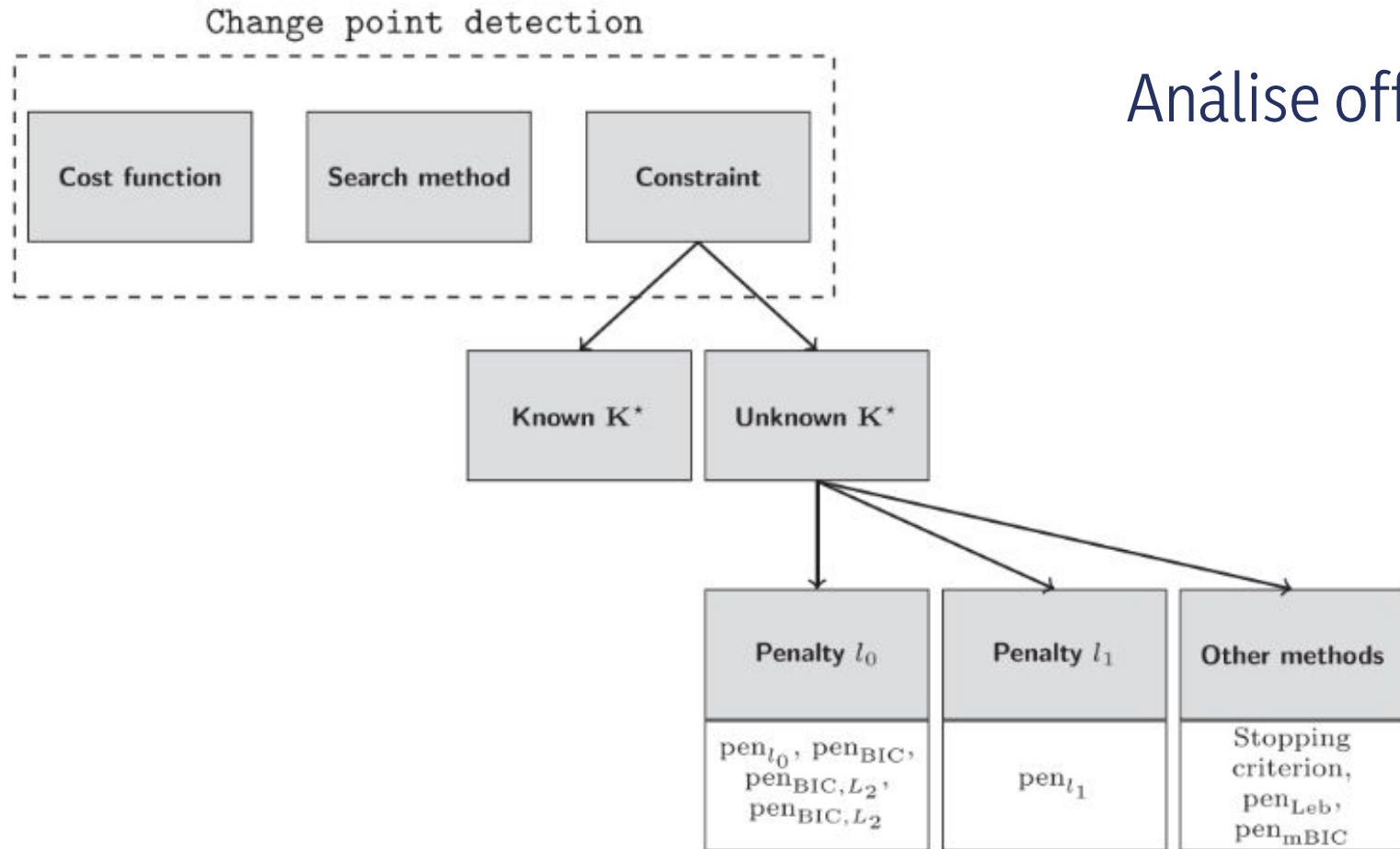
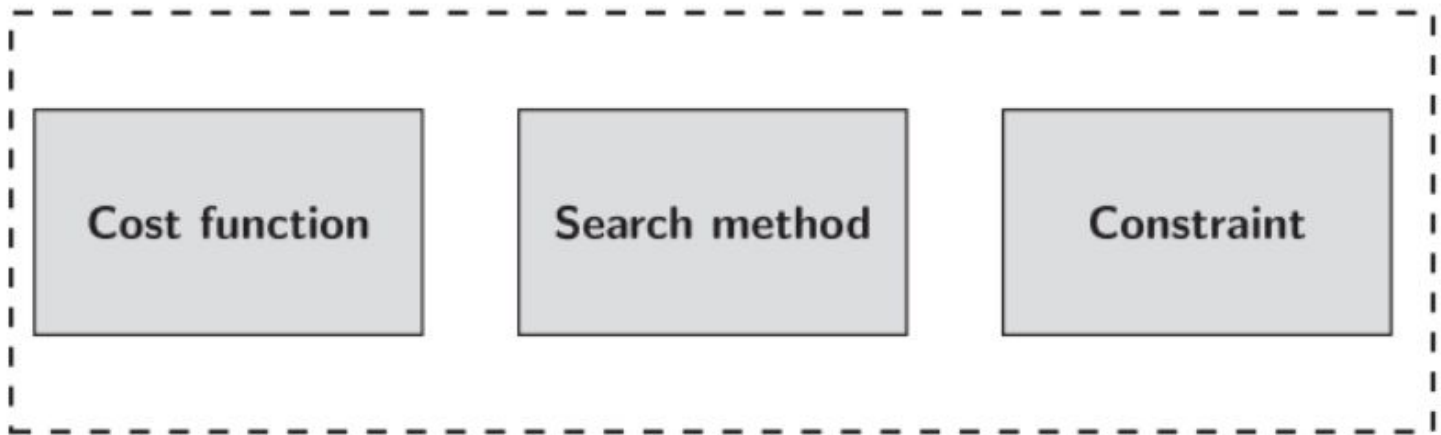


Fig. 11. Typology of the constraints (on the number of change points) described in [Section 6](#).

Análise offline

Change point detection



Abordagem intuitiva da análise de mudança de ponto (*cpt*)

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$n = ?$$

$$\tau = ?$$

Abordagem intuitiva da análise de mudança de ponto (*cpt*)

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$n = 4$$

$$\tau = 1, 2, 3$$

Abordagem intuitiva da análise de mudança de ponto (*cpt*)

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$n = 4$$

$$\tau = 1, 2, 3$$

Para cada τ , calculamos:

$$\bar{y}_{1:\tau}$$

$$\bar{y}_{(\tau+1)}$$

Abordagem intuitiva da análise de mudança de ponto (*cpt*)

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$n = 4$$

$$\tau = \textcircled{1}, 2, 3$$

$$\tau=1$$

$$\bar{y}_{1:1}$$

$$\bar{y}_{2:4}$$

Abordagem intuitiva da análise de mudança de ponto (*cpt*)

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$n = 4$$

$$\tau = (1, 2, 3)$$

$$\tau=1$$

$$\bar{y}_{1:1}$$

$$\bar{y}_{2:4}$$

Abordagem intuitiva da análise de mudança de ponto (*cpt*)

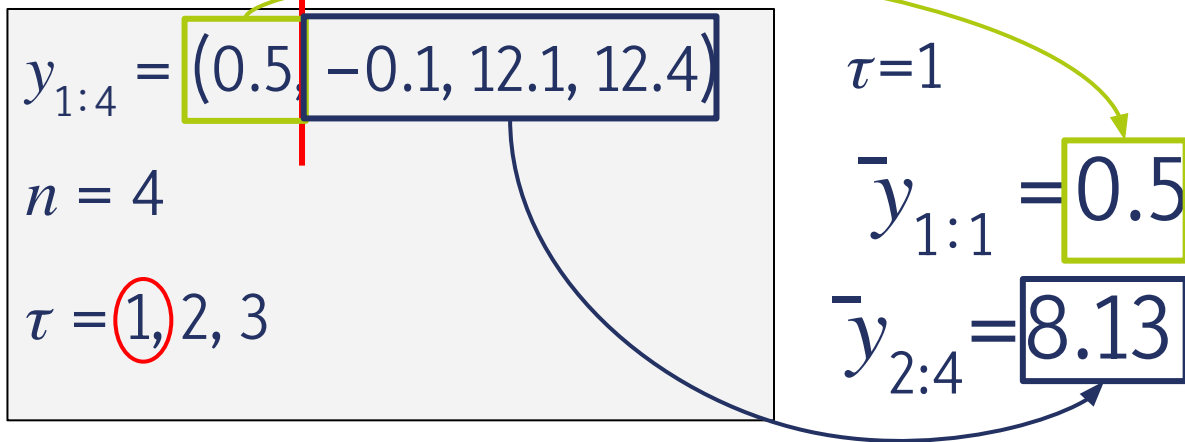
$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$
$$n = 4$$
$$\tau = \textcircled{1}, 2, 3$$

$$\tau = 1$$

$$\bar{y}_{1:1} = 0.5 / 1$$

$$\bar{y}_{2:4} = (-0.1 + 12.1 + 12.4) / 3$$

Abordagem intuitiva da análise de mudança de ponto (*cpt*)



Abordagem intuitiva da análise de mudança de ponto (*cpt*)

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$n = 4$$

$$\tau = 1, \textcircled{2}, 3$$

$$\tau=2$$

$$\bar{y}_{1:2}$$

$$\bar{y}_{3:4}$$

Abordagem intuitiva da análise de mudança de ponto (*cpt*)

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$n = 4$$

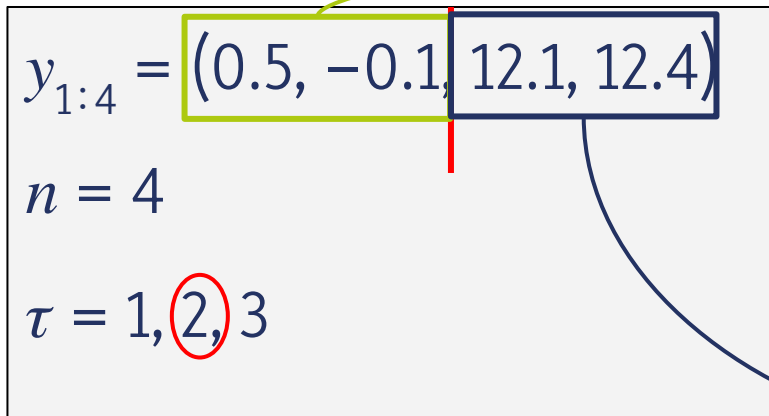
$$\tau = 1, \textcircled{2}, 3$$

$$\tau=2$$

$$\bar{y}_{1:2}$$

$$\bar{y}_{3:4}$$

Abordagem intuitiva da análise de mudança de ponto (*cpt*)

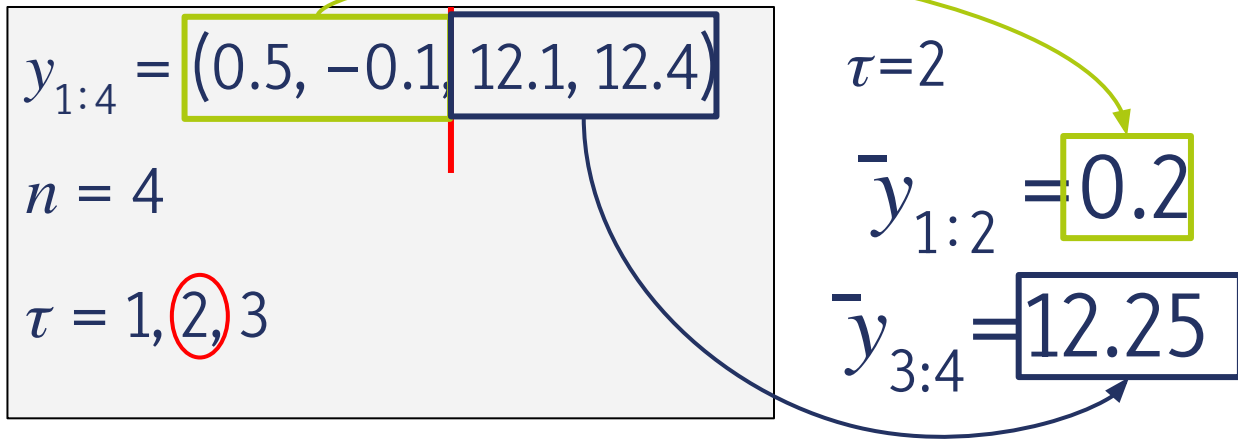
$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$
$$n = 4$$
$$\tau = 1, \textcircled{2}, 3$$


$$\tau = 2$$

$$\bar{y}_{1:2} = (0.5 - 0.1) / 2$$

$$\bar{y}_{3:4} = (12.1 + 12.4) / 2$$

Abordagem intuitiva da análise de mudança de ponto (*cpt*)



Abordagem intuitiva da análise de mudança de ponto (*cpt*)

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$n = 4$$

$$\tau = 1, 2, \textcircled{3}$$

$$\tau=3$$

$$\bar{y}_{1:3}$$

$$\bar{y}_{4:4}$$

Abordagem intuitiva da análise de mudança de ponto (*cpt*)

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$n = 4$$

$$\tau = 1, 2, \textcircled{3}$$

$$\tau=3$$

$$\bar{y}_{1:3}$$

$$\bar{y}_{4:4}$$

Abordagem intuitiva da análise de mudança de ponto (*cpt*)

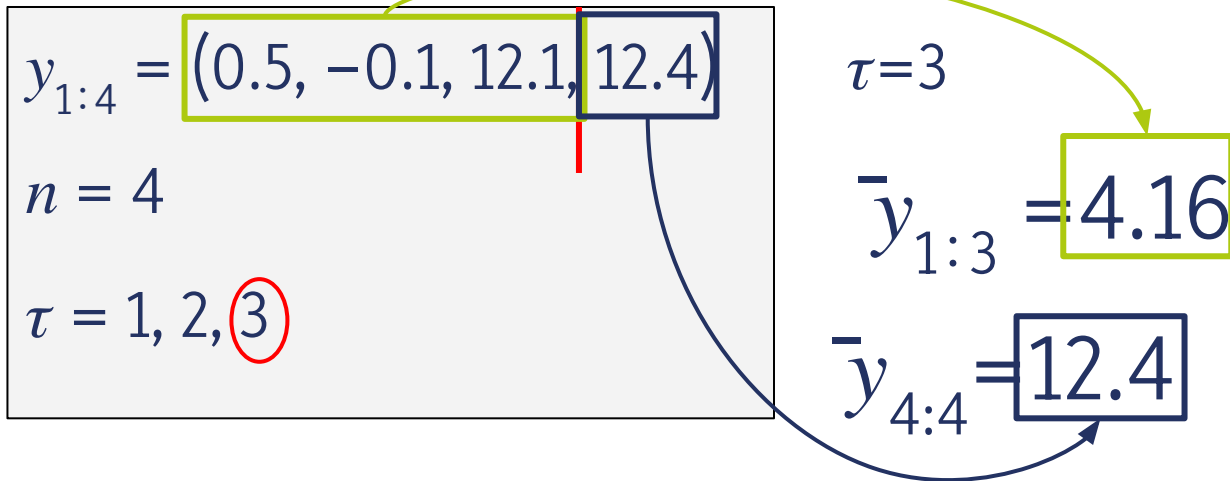
$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$
$$n = 4$$
$$\tau = 1, 2, \textcircled{3}$$

$$\tau = 3$$

$$\bar{y}_{1:3} = (0.5 - 0.1 + 12.1) / 3$$

$$\bar{y}_{4:4} = 12.4 / 1$$

Abordagem intuitiva da análise de mudança de ponto (*cpt*)



Abordagem intuitiva da análise de mudança de ponto (*cpt*)

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

$$\tau=2$$

$$\bar{y}_{1:2} = 0.2$$

$$\bar{y}_{3:4} = 12.25$$

$$\tau=3$$

$$\bar{y}_{1:3} = 4.16$$

$$\bar{y}_{4:4} = 12.4$$

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$n = 4$$

$$\tau = 1, 2, 3$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

$$\tau=2$$

$$\bar{y}_{1:2} = 0.2$$

$$\bar{y}_{3:4} = 12.25$$

$$\tau=3$$

$$\bar{y}_{1:3} = 4.16$$

$$\bar{y}_{4:4} = 12.4$$

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$n = 4$$

$$\tau = 1, 2, 3$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

$$\tau=2$$

$$\bar{y}_{1:2} = 0.2$$

$$\bar{y}_{3:4} = 12.25$$

$$\tau=3$$

$$\bar{y}_{1:3} = 4.16$$

$$\bar{y}_{4:4} = 12.4$$

Alguma dúvida em
algum desses valores?

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$n = 4$$

$$\tau = 1, 2, 3$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

$$\tau=2$$

$$\bar{y}_{1:2} = 0.2$$

$$\bar{y}_{3:4} = 12.25$$

$$\tau=3$$

$$\bar{y}_{1:3} = 4.16$$

$$\bar{y}_{4:4} = 12.4$$

Próximos passos:

1. Queremos encontrar o ponto τ que gera dois segmentos com a maior **diferença de médias**.
2. Mas temos que tomar cuidado, porque quando um segmento é muito pequeno a diferença pode ficar gritante. Então precisamos de um “**fator de escala**”.

Próximos passos:

$$|\bar{y}_{1:\tau} - \bar{y}_{(\tau+1):n}|$$

1. Queremos encontrar o ponto τ que gera dois segmentos com a maior **diferença de médias**.
2. Mas temos que tomar cuidado, porque quando um segmento é muito pequeno a diferença pode ficar gritante. Então precisamos de um “**fator de escala**”.

Próximos passos:

$$|\bar{y}_{1:\tau} - \bar{y}_{(\tau+1):n}|$$

1. Queremos encontrar o ponto τ que gera dois segmentos com a maior **diferença de médias**.
2. Mas temos que tomar cuidado, porque quando um segmento é muito pequeno a diferença pode ficar gritante. Então precisamos de um “**fator de escala**”.

- A diferença de média é simplesmente a média do segmento antes com a média do segmento depois.
- Essa diferença está em módulo, porque não nos importa o sinal, apenas o valor absoluto da diferença!

Próximos passos:

$$\sqrt{\frac{\tau(n - \tau)}{n}}$$

1. Queremos encontrar o ponto τ que gera dois segmentos com a maior **diferença de médias**.
2. Mas temos que tomar cuidado, porque quando um segmento é muito pequeno a diferença pode ficar gritante. Então precisamos de um “**fator de escala**”.

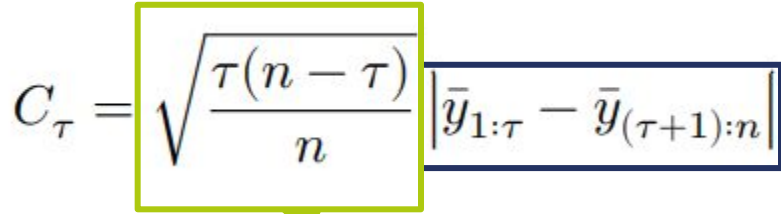
$$\sqrt{\frac{\tau(n - \tau)}{n}}$$

- Para cada sequência, o valor de n é fixo e o valor de t muda. A sequência se inicia no $t = 1$ e se encerra no $t = n$.
- Small t = Factor near 1: Quando t está muito próximo de 1, então o valor tende a ser $n/n = 1$
- Big t = Factor near 1: Quando t está muito próximo de n , então também o valor tende a ser $n/n = 1$
- Highest factor? Quando o valor de t está na metade do n , então esse é o valor máximo do fator de escala

Próximos passos:

1. Queremos encontrar o ponto τ que gera dois segmentos com a maior **diferença de médias**.
2. Mas temos que tomar cuidado, porque quando um segmento é muito pequeno a diferença pode ficar gritante. Então precisamos de um “**fator de escala**”.

Próximos passos:

$$C_{\tau} = \sqrt{\frac{\tau(n-\tau)}{n}} \left| \bar{y}_{1:\tau} - \bar{y}_{(\tau+1):n} \right|$$


1. Queremos encontrar o ponto τ que gera dois segmentos com a maior **diferença de médias**.
2. Mas temos que tomar cuidado, porque quando um segmento é muito pequeno a diferença pode ficar gritante. Então precisamos de um “**fator de escala**”.

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$n = 4$$

$$\tau = 1, 2, 3$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

$$\tau=2$$

$$\bar{y}_{1:2} = 0.2$$

$$\bar{y}_{3:4} = 12.25$$

$$\tau=3$$

$$\bar{y}_{1:3} = 4.16$$

$$\bar{y}_{4:4} = 12.4$$

$$C_\tau = \sqrt{\frac{\tau(n-\tau)}{n}} \left| \bar{y}_{1:\tau} - \bar{y}_{(\tau+1):n} \right|$$

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$n = 4$$

$$\tau = 1, 2, 3$$

$$C_\tau = \sqrt{\frac{\tau(n-\tau)}{n}} |\bar{y}_{1:\tau} - \bar{y}_{(\tau+1):n}|$$

For $\tau = 1$:

$$C_1 = \sqrt{\frac{1(4-1)}{4}} |0.5 - 8.13\bar{3}| = 0.866 \times 7.63\bar{3} = 6.61.$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$n = 4$$

$$\tau = 1, 2, 3$$

$$C_\tau = \sqrt{\frac{\tau(n-\tau)}{n}} |\bar{y}_{1:\tau} - \bar{y}_{(\tau+1):n}|$$

For $\tau = 2$:

$$C_2 = \sqrt{\frac{2(4-2)}{4}} |0.2 - 12.25| = 1 \times 12.05 = 12.05.$$

$$\tau=2$$

$$\bar{y}_{1:2} = 0.2$$

$$\bar{y}_{3:4} = 12.25$$

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$n = 4$$

$$\tau = 1, 2, 3$$

$$C_\tau = \sqrt{\frac{\tau(n-\tau)}{n}} |\bar{y}_{1:\tau} - \bar{y}_{(\tau+1):n}|$$

For $\tau = 3$:

$$C_3 = \sqrt{\frac{3(4-3)}{4}} |4.1\bar{6} - 12.4| = 0.866 \times 8.23\bar{3} = 7.13.$$

✓ 2:4

✓ 3:4

$$\tau=3$$

$$\bar{y}_{1:3}=4.16$$

$$\bar{y}_{4:4}=12.4$$

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$n = 4$$

$$\tau = 1, 2, 3$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

$$\tau=2$$

$$\bar{y}_{1:2} = 0.2$$

$$\bar{y}_{3:4} = 12.25$$

$$\tau=3$$

$$\bar{y}_{1:3} = 4.16$$

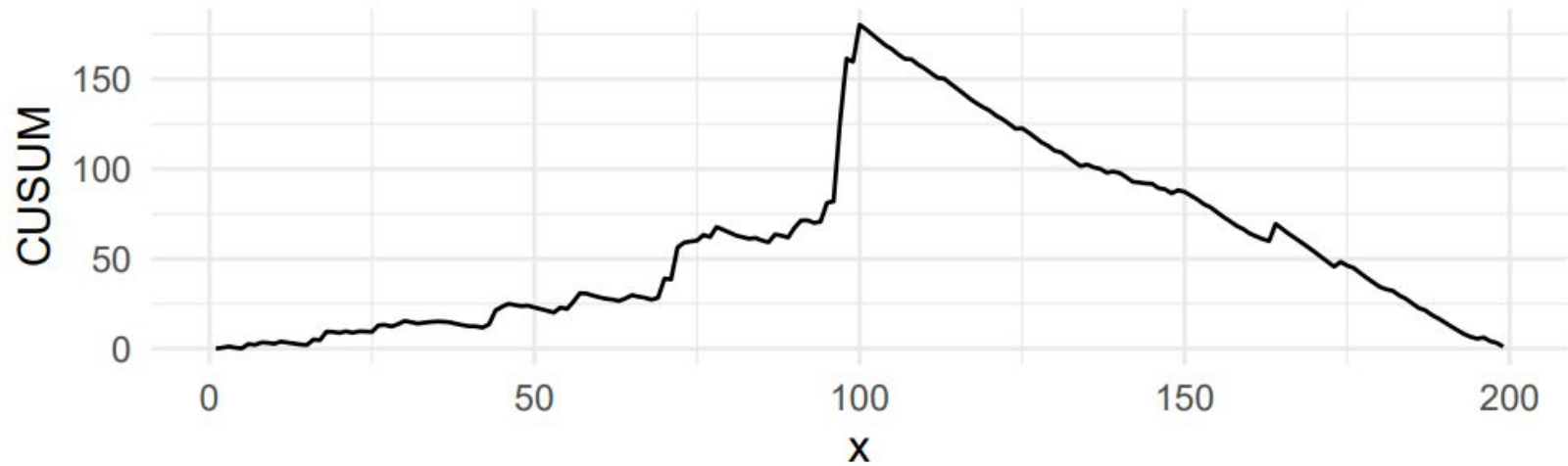
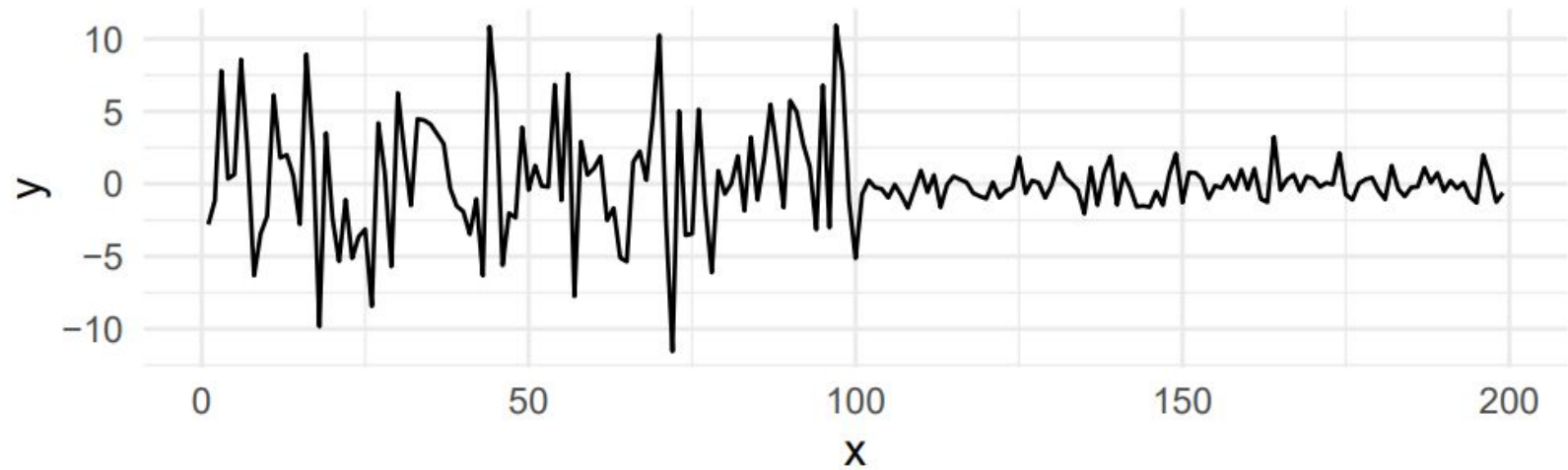
$$\bar{y}_{4:4} = 12.4$$

$$C_{\tau} = \sqrt{\frac{\tau(n-\tau)}{n}} |\bar{y}_{1:\tau} - \bar{y}_{(\tau+1):n}|$$

$$C_1 = 6.61$$

$$C_2 = 12.05$$

$$C_3 = 7.13$$



Outra função de custo

Vamos voltar um pouco e ver outra função de custo...

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$n = 4$$

$$\tau = 1, 2, 3$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

$$\tau=2$$

$$\bar{y}_{1:2} = 0.2$$

$$\bar{y}_{3:4} = 12.25$$

$$\tau=3$$

$$\bar{y}_{1:3} = 4.16$$

$$\bar{y}_{4:4} = 12.4$$

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$n = 4$$

$$\tau = 1, 2, 3$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

$$\tau=2$$

$$\bar{y}_{1:2} = 0.2$$

$$\bar{y}_{3:4} = 12.25$$

$$\tau=3$$

$$\bar{y}_{1:3} = 4.16$$

$$\bar{y}_{4:4} = 12.4$$

Próximos passos:

1. Vamos calcular o resíduo em relação a cada segmento
2. E depois vamos somar todos os resíduos
3. Como o resíduo precisa estar ao quadrado, vamos controlar isso colocando, ao fim, tudo em raiz quadrada

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

Valor (y)	Média

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

Valor (y)	Média

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

Valor (y)	Média
0.5	
-0.1	
12.1	
12.4	

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

Valor (y)	Média
0.5	
-0.1	
12.1	
12.4	

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

Valor (y)	Média
0.5	0.5
-0.1	8.13
12.1	8.13
12.4	8.13

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

Valor (y)	Média	Resíduo^2
0.5	0.5	
-0.1	8.13	
12.1	8.13	
12.4	8.13	

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

Valor (y)	Média	Resíduo^2
0.5	0.5	
-0.1	8.13	
12.1	8.13	
12.4	8.13	

$$\text{Resíduo}^2 = (\text{valor} - \text{média})^2$$

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

Valor (y)	Média	Resíduo ²
0.5	0.5	$(0.5-0.5)^2$
-0.1	8.13	$(-0.1-8.13)^2$
12.1	8.13	$(12.1-8.13)^2$
12.4	8.13	$(12.4-8.13)^2$

$$\text{Resíduo}^2 = (\text{valor} - \text{média})^2$$

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

Valor (y)	Média	Resíduo ²
0.5	0.5	0 ²
-0.1	8.13	(-8.14) ²
12.1	8.13	(3.97) ²
12.4	8.13	(4.27) ²

$$\text{Resíduo}^2 = (\text{valor} - \text{média})^2$$

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

Valor (y)	Média	Resíduo^2
0.5	0.5	0
-0.1	8.13	66.26
12.1	8.13	15.76
12.4	8.13	18.23

1. Vamos calcular o resíduo em relação a cada segmento
2. E depois vamos somar todos os resíduos
3. Como o resíduo precisa estar ao quadrado, vamos controlar isso colocando, ao fim, tudo em raiz quadrada

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

Valor (y)	Média	Resíduo^2
0.5	0.5	0
-0.1	8.13	66.26
12.1	8.13	15.76
12.4	8.13	18.23

- ~~Vamos calcular o resíduo em relação a cada segmento OK~~
- E depois vamos somar todos os resíduos**
- Como o resíduo precisa estar ao quadrado, vamos controlar isso colocando, ao fim, tudo em raiz quadrada

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

Valor (y)	Média	Resíduo^2
0.5	0.5	0
-0.1	8.13	66.26
12.1	8.13	15.76
12.4	8.13	18.23

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

Valor (y)	Média	Resíduo^2
0.5	0.5	0
-0.1	8.13	66.26
12.1	8.13	15.76
12.4	8.13	18.23
Total		100.25

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

Valor (y)	Média	Resíduo^2
0.5	0.5	0
-0.1	8.13	66.26
12.1	8.13	15.76
12.4	8.13	18.23
Total		100.25

- ~~Vamos calcular o resíduo em relação a cada segmento OK~~
- ~~E depois vamos somar todos os resíduos OK~~
- Como o resíduo precisa estar ao quadrado, vamos controlar isso tirando a raiz**

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

Valor (y)	Média	Resíduo^2
0.5	0.5	0
-0.1	8.13	66.26
12.1	8.13	15.76
12.4	8.13	18.23
Total		100.25

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

Valor (y)	Média	Resíduo^2
0.5	0.5	0
-0.1	8.13	66.26
12.1	8.13	15.76
12.4	8.13	18.23
Total		100.25

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

Valor (y)	Média	Resíduo^2
0.5	0.5	0
-0.1	8.13	66.26
12.1	8.13	15.76
12.4	8.13	18.23
Total		100.25

$$R1 = \sqrt{100.25}$$

$$R1 = 10.01$$

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

$$R1 = 10.01$$

Valor (y)	Média	Resíduo ²
0.5	0.5	0
-0.1	8.13	66.26
12.1	8.13	15.76
12.4	8.13	18.23
Total		100.25

$$R1 = \sqrt{100.25}$$

$$R1 = 10.01$$

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

$$R1 = 10.01$$

Valor (y)	Média	Resíduo^2
0.5	0.5	0
-0.1	8.13	66.26
12.1	8.13	15.76
12.4	8.13	18.23
Total		100.25

- ~~Vamos calcular o resíduo em relação a cada segmento OK~~
- ~~E depois vamos somar todos os resíduos OK~~
- ~~Como o resíduo precisa estar ao quadrado, vamos controlar isso tirando a raiz OK~~

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$n = 4$$

$$\tau = 1, 2, 3$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

$$R1 = 10.01$$

$$C1 = 6.61$$

$$\tau=2$$

$$\bar{y}_{1:2} = 0.2$$

$$\bar{y}_{3:4} = 12.25$$

$$R2 = ?$$

$$C2 = 12.05$$

$$\tau=3$$

$$\bar{y}_{1:3} = 4.16$$

$$\bar{y}_{4:4} = 12.4$$

$$R3 = ?$$

$$C3 = 7.13$$

Valor (y)

Média

Resíduo²

0.5

-0.1

12.1

12.4

Total

Agora vocês fazem o resto!

1. Calculem os resíduos totais para $t=2$ e $t=3$
2. Discutam o que representa os “resíduos totais”. O que esse valor está nos indicando?
3. Discutam qual dos resíduos totais ($R1$, $R2$ ou $R3$) melhor indica o ponto de mudança
4. Qual é a diferença dos R s para os C s?

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$n = 4$$

$$\tau = 1, 2, 3$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

$$R1 = 10.01$$

$$C1 = 6.61$$

$$\tau=2$$

$$\bar{y}_{1:2} = 0.2$$

$$\bar{y}_{3:4} = 12.25$$

$$R2 = 0.47$$

$$C2 = 12.05$$

$$\tau=3$$

$$\bar{y}_{1:3} = 4.16$$

$$\bar{y}_{4:4} = 12.4$$

$$R3 = ?$$

$$C3 = 7.13$$

Valor (y)	Média	Resíduo^2
0.5	0.2	0.09
-0.1	0.2	0.09
12.1	12.25	0.0225
12.4	12.25	0.0225
Total		0.225

Agora vocês fazem o resto!

1. Calculem os resíduos totais para $t=2$ e $t=3$
2. Discutam o que representa os “resíduos totais”. O que esse valor está nos indicando?
3. Discutam qual dos resíduos totais ($R1$, $R2$ ou $R3$) melhor indica o ponto de mudança
4. Qual é a diferença dos R s para os C s?

$$y_{1:4} = (0.5, -0.1, 12.1, 12.4)$$

$$n = 4$$

$$\tau = 1, 2, 3$$

$$\tau=1$$

$$\bar{y}_{1:1} = 0.5$$

$$\bar{y}_{2:4} = 8.13$$

$$R1 = 10.01$$

$$C1 = 6.61$$

$$\tau=2$$

$$\bar{y}_{1:2} = 0.2$$

$$\bar{y}_{3:4} = 12.25$$

$$R2 = 0.47$$

$$C2 = 12.05$$

$$\tau=3$$

$$\bar{y}_{1:3} = 4.16$$

$$\bar{y}_{4:4} = 12.4$$

$$R3 = 9.72$$

$$C3 = 7.13$$

Valor (y)	Média	Resíduo^2
0.5	4.16	13.4
-0.1	4.16	18.15
12.1	4.16	63.04
12.4	12.4	0
Total		94.6

Agora vocês fazem o resto!

1. Calculem os resíduos totais para $t=2$ e $t=3$
2. Discutam o que representa os “resíduos totais”. O que esse valor está nos indicando?
3. Discutam qual dos resíduos totais ($R1$, $R2$ ou $R3$) melhor indica o ponto de mudança
4. Qual é a diferença dos R s para os C s?

Abordagem intuitiva da análise de mudança de ponto (*cpt*)

[Demonstração ao vivo no R]

Análise online

ϵ -real-time algorithm: “Na prática, **nenhum algoritmo de detecção de pontos de mudança opera em tempo real perfeito, pois é necessário inspecionar novos dados antes de determinar se um ponto de mudança ocorreu entre os dados antigos e os novos.** No entanto, diferentes algoritmos online requerem **quantidades diferentes de novos dados antes que a detecção de pontos de mudança possa ocorrer.** Com base nessa observação, definiremos um novo termo para usar ao longo deste artigo. Denotaremos como um algoritmo de tempo quase real (**ϵ -real-time**) um algoritmo online que precisa de pelo menos ϵ amostras de dados no novo lote de dados para ser capaz de encontrar pontos de mudança” (Aminikhanghahi; Cook, 2016).

Problema metodológico não resolvido

Para monitorar taxa de vitória, devemos monitorar quando ela se desvia dos 50%. Entretanto, vimos que o Teorema de Priest & Klein não explica a taxa da vitória em situações muito micro. Então é possível (e na verdade altamente provável) que em um único dia a taxa de vitória para um tipo de litigante seja muito superior a 50%. Então surge a pergunta (ainda não resolvida):

Qual é o tamanho do ϵ que precisamos para fazer uma detecção de ponto de mudança online? Isto é, qual é o mínimo de processos para o qual o Teorema de Priest & Klein ainda se aplica? Sabemos que ele não explica muito em fragmentações por vara, mas e fragmentações temporais?

Cuidados metodológicos

Pressupostos do modelo

Não podemos rodar o modelo de CPT em qualquer tipo de dados. Eles precisam cumprir três requisitos:

1. Independência dos dados;
2. Normalidade dos pontos antes e depois da mudança;
3. Variância constante ao redor dos dados

Discussão

Reflexão

Se queremos fazer um modelo de *cpt* para acompanhar processos, podemos pensar em dois acompanhamentos:

1. Taxa de vitória ao longo do tempo
2. Quantidade de processos

O que significam as taxas para cada um desses valores? Como vocês “fatiariam” essa análise? Por exemplo, pensem em “fatiar” essas medidas por coisas como: “taxa de vitória **por litigante**” ou “quantidade de processos **por assunto**”...

Muito obrigado!

ric.feliz@gmail.com