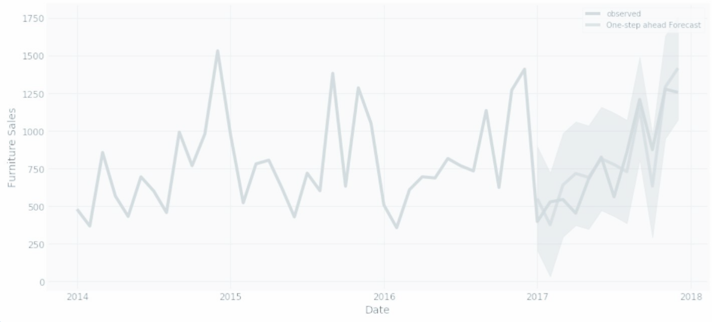


Aula 3

Séries Temporais - Parte II - Modelos Preditivos

Média Móvel

- A **Média Móvel** (vista antes para atenuar a tendência de uma série) também pode ser usada para **gerar predições**

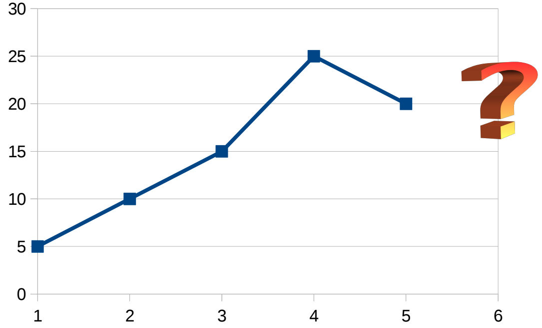
- Neste caso, dado o período corrente t, a predição no período $t + 1$ irá corresponder à média das últimas h observações

$$f_{t+1} = \frac{y_t + y_{t-1} + \cdots + y_{t-h+1}}{h}$$

Exercício - Média Móvel

Considerando a série temporal apresentada abaixo e $h = 3$, calcule a previsão do valor no tempo t_6 considerando o algoritmo da *Média Móvel*

Tempo	Valor
t_1	5
t_2	10
t_3	15
t_4	25
t_5	20
t_6	???



Suavização Exponencial Simples

- A **Suavização Exponencial Simples** (*Simple Exponencial Smoothing - SME*) é baseado na estimativa da variável s_t
- Possui um parâmetro $\alpha \in [0, 1]$ que regula a importância dos mais recentes valores y_t
- A SME é dada pela seguinte expressão recursiva

$$s_t = \alpha y_t + (1 - \alpha) s_{t-1}$$

$$s_1 = y_1$$

- A previsão do período $t + 1$ é obtida por

$$f_{t+1} = s_t$$

Suavização Exponencial Simples

- Se $\alpha \simeq 0$: o modelo tem grande "inércia" $\rightarrow$ peso constante às observações passadas
- Se $\alpha \simeq 1$: atribui mais peso às observações mais recentes
- O parâmetro α é escolhido de forma a minimizar o erro quadrático médio

Suavização Exponencial Simples

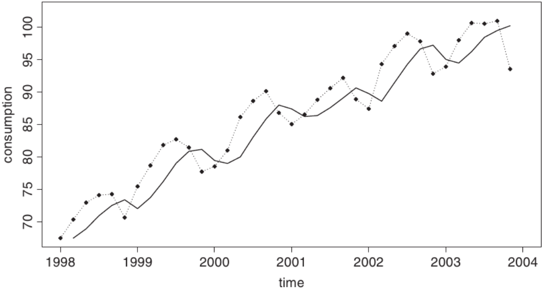


Figure 9.11 Simple exponential smoothing model with parameter $\alpha = 0.5$ for the time series of electricity consumption

Suavização Exponencial Simples

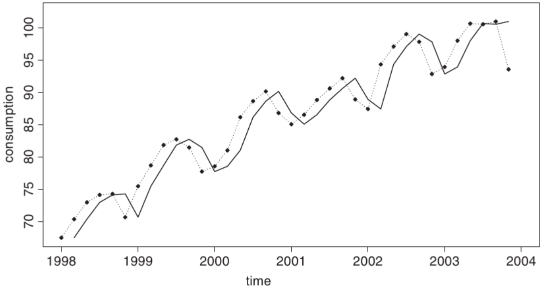
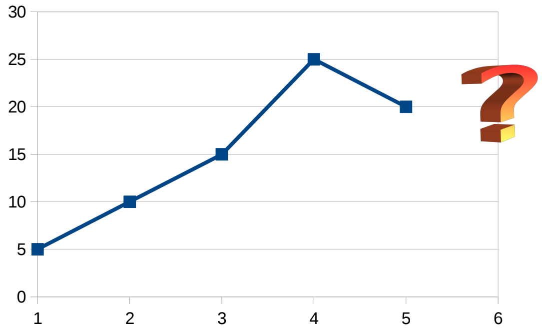


Figure 9.12 Simple exponential smoothing model with parameter $\alpha = 1.0$ for the time series of electricity consumption

Exercício - Suavização Exponencial Simples

Considerando a série temporal apresentada abaixo e $\alpha = 0.5$, calcule a previsão do valor no tempo t_6 considerando o algoritmo da *Suavização Exponencial Simples*

Tempo	Valor
t_1	5
t_2	10
t_3	15
t_4	25
t_5	20
t_6	???



Suavização Exponencial com Ajuste de Tendências

- A SME tende a falhar ao prever valor em séries temporais com tendência $\rightarrow$ efeito de "*lag*"
- É possível estender o SME incluindo um componente de tendência
- Um componente de tendência suavizado linear m_t é incluído na formulação

$$s_t = \alpha y_t + (1 - \alpha)(s_{t-1} + m_{t-1})$$

na qual m_t também é obtido por meio de uma função recursiva

$$m_t = \beta(s_t - s_{t-1}) + (1 - \beta)m_{t-1}$$

na qual $\beta \in [0, 1]$ tem a mesma função e comportamento de α

Suavização Exponencial com Ajuste de Tendências

- A predição para o período $t+1$ é dada por

$$f_{t+1} = s_t + m_t$$

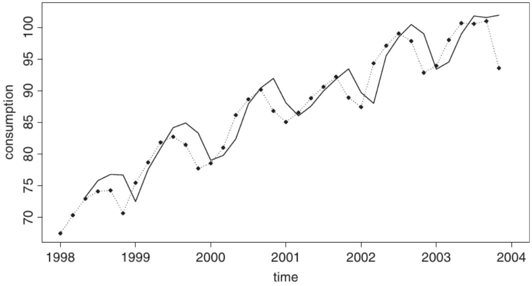
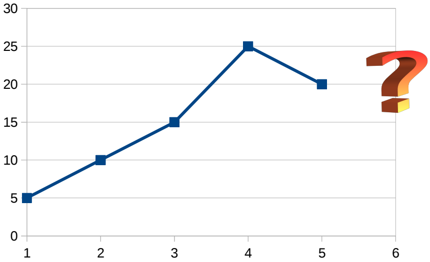


Figure 9.13 *Holt exponential smoothing model with parameters* $\alpha = 1.0$ *and* $\beta = 0.099$ *for the time series of electricity consumption*

Exercício - Suavização Exponencial com Ajuste de Tendência

Considerando a série temporal apresentada abaixo, e $\alpha = 0.5$ e $\beta = 0.5$, calcule a previsão do valor no tempo t_6 considerando o algoritmo da *Suavização Exponencial Simples com Ajuste de Tendência*

Tempo	Valor
t_1	5
t_2	10
t_3	15
t_4	25
t_5	20
t_6	???



Suavização Exponencial com Tendência e Sazonalidade

- Pode-se também incluir um componente de sazonalidade para estender os modelos anteriores
- Assumindo que cada modelo sazonal é composto por L períodos, temos

$$s_t = \alpha \frac{y_t}{q_{t-L}} + (1-\alpha)(s_{t-1} + m_{t-1})$$

$$m_t = \beta(s_t - s_{t-1}) + (1-\beta)m_{t-1}$$

$$q_t = \gamma \frac{y_t}{s_t} + (1-\gamma)q_{t-L}$$

  na qual $\gamma \in [0, 1]$
- A previsão é dada por: $f_{t+1} = (s_t + m_t)q_{t-L+1}$

Suavização Exponencial com Tendência e Sazonalidade

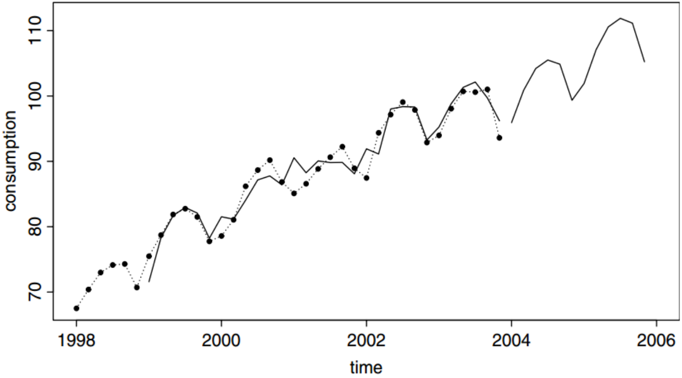


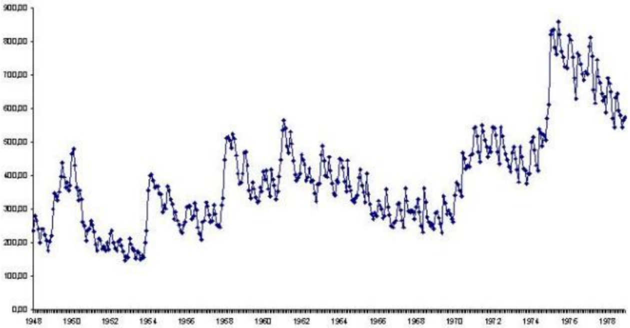
Figure 9.14 Winters exponential smoothing model with parameters $\alpha = 0.8256$, $\beta = 0.0167$ *and* $\gamma = 1.0$ *for the time series of electricity consumption*

Subsequências Mais Próximas

- O objetivo deste método é identificar em todo o histórico subsérie(s) de comportamento(s) próximo(s) ao comportamento atual

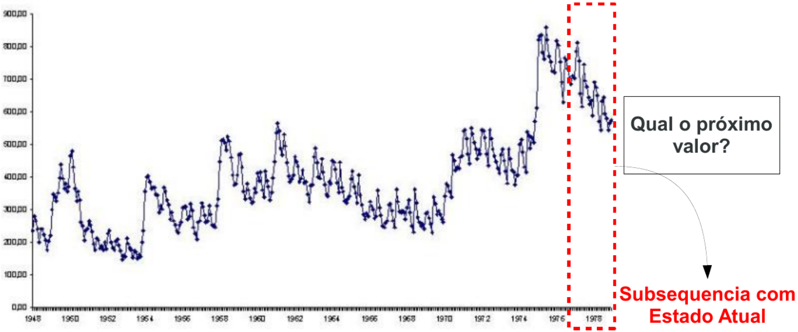
- Com isso, podemos nos basear nos valores posteriores das subséries para prever o valor atual

Subsequências Mais Próximas

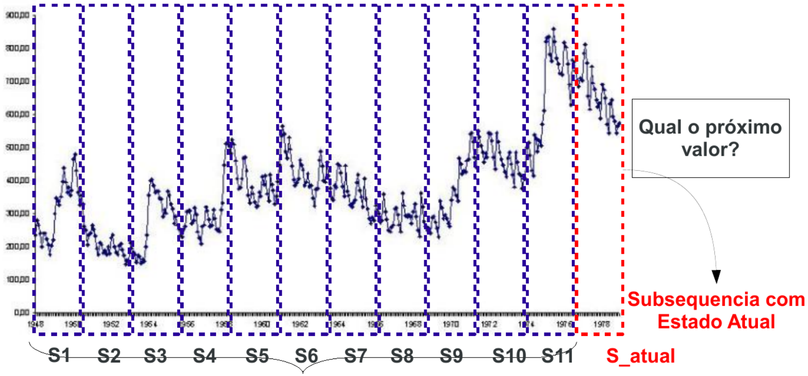


Qual o próximo valor?

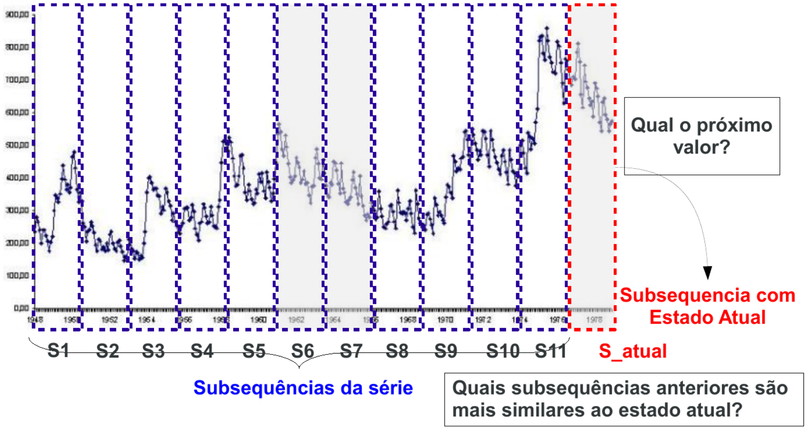
Subsequências Mais Próximas



Subsequências Mais Próximas



Subsequências Mais Próximas



Subsequências Mais Próximas

- Precisamos então **definir o tamanho da subsérie atual** para procurar no passado por subséries próximas

- Consequências do tamanho da subsérie
    - Valores muito pequenos podem fazer com que qualquer subsérie seja parecida
    - Valores muito grandes podem fazer com que nenhuma subsérie seja parecida

Subsequências Mais Próximas

- Precisamo também definir como considerar que uma subsérie é parecida ou próxima a outra subsérie

- Para isso, podemos considerar as medidas tradicionais de proximidade
  - Euclidiana
  - Manhatan
  - Cosseno
  - Correlação de Pearson
  - Correlação de Spearman

Subsequências Mais Próximas

- A tradicional medida de distância Euclidiana é uma boa escolha?

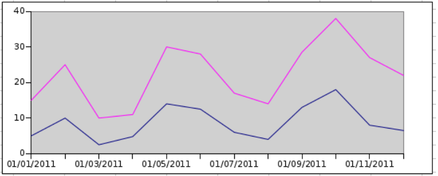
$$d(\mathbf{y}_1, \mathbf{y}_2) = \sqrt{\sum_{i=1}^{h} (y_{1_i} - y_{2_i})^2}$$

- Lembrando que na distância Euclidiana, quanto mais próximo de 0 seu valor, mais parecidos são as duas subséries

Subsequências Mais Próximas

	Série 1	Série 2	d	d^2
Jan-01-2011	5	15	−10	100
Fev-01-2011	10	25	−15	225
Mar-01-2011	2,5	10	−7,5	56,25
Abr-01-2011	4,8	11	−6,2	38,44
Mai-01-2011	14	30	−16	256
Jun-01-2011	12,5	28	−15,5	240,25
Jul-01-2011	6	17	−11	121
Ago-01-2011	4	14	−10	100
Set-01-2011	13	28,5	−15,5	240,25
Out-01-2011	18	38	−20	400
Nov-01-2011	8	27	−19	361
Dez-01-2011	6,5	22	−15,5	240,25

Distância Euclideana 48,76925



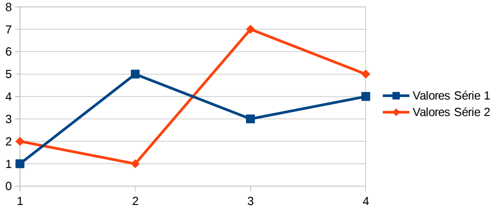
Subsequências Mais Próximas

	Série 1	Série 2		d	d^2
Jan-01-2011	15	20		−5	25
Fev-01-2011	20	16		4	16
Mar-01-2011	21	14		7	49
Abr-01-2011	25	10		15	225
Mai-01-2011	10	25		−15	225
Jun-01-2011	11	26		−15	225
Jul-01-2011	35	9		26	676
Ago-01-2011	36	8		28	784
Set-01-2011	26	16		10	100
Out-01-2011	24	17		7	49
Nov-01-2011	19	26		−7	49
Dez-01-2011	18	27		−9	81
		Distância Euclideana			50,03998

Subsequências Mais Próximas

Calcula a distância Euclidiana para as duas séries apresentadas abaixo

Tempo	Valores Série 1	Valores Série 2
t_1	1	2
t_2	5	1
t_3	3	7
t_4	4	5



Subsequências Mais Próximas

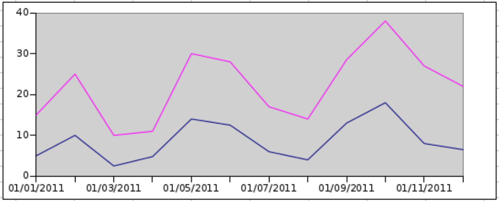
- Correlação de Pearson é uma boa escolha?

$$\rho(\mathbf{y}_1, \mathbf{y}_2) = \frac{\sum_{i=1}^{t}(y_{1_i} - \overline{\mathbf{y}_1})(y_{2_i} - \overline{\mathbf{y}_2})}{\sqrt{\sum_{i=1}^{t}(y_{1_i} - \overline{\mathbf{y}_1})}\sqrt{\sum_{i=1}^{t}(y_{2_i} - \overline{\mathbf{y}_2})}}$$

- Utilizando a Correlação de Pearson
  - Quanto mais próximo de 1, mais correlacionadas (mais similares) são as subséries
  - Quanto mais próxima de -1, mais descorrelacionadas (mais dissimilares) são as subséries

Subsequências Mais Próximas

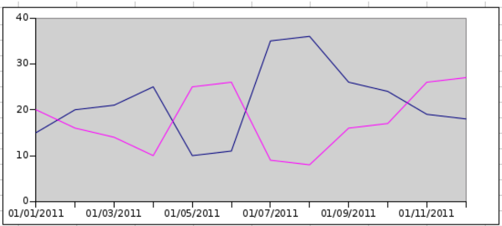
	Série 1	Série 2
Jan-01-2011	5	15
Fev-01-2011	10	25
Mar-01-2011	2,5	10
Abr-01-2011	4,8	11
Mai-01-2011	14	30
Jun-01-2011	12,5	28
Jul-01-2011	6	17
Ago-01-2011	4	14
Set-01-2011	13	28,5
Out-01-2011	18	38
Nov-01-2011	8	27
Dez-01-2011	6,5	22
Correlação Pearson		0,95



Subsequências Mais Próximas

	Série 1	Série 2
Jan-01-2011	15	20
Fev-01-2011	20	16
Mar-01-2011	21	14
Abr-01-2011	25	10
Mai-01-2011	10	25
Jun-01-2011	11	26
Jul-01-2011	35	9
Ago-01-2011	36	8
Set-01-2011	26	16
Out-01-2011	24	17
Nov-01-2011	19	26
Dez-01-2011	18	27

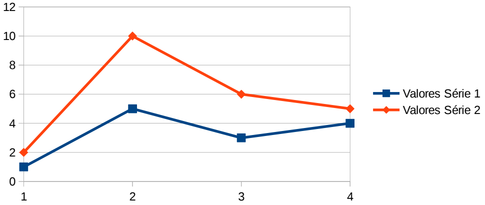
Correlação Pearson	−0,85



Subsequências Mais Próximas

Calcula a Correlação de Pearson para as duas séries apresentadas abaixo

Tempo	Valores Série 1	Valores Série 2
t_1	1	2
t_2	5	10
t_3	3	6
t_4	4	5



Subsequências Mais Próximas

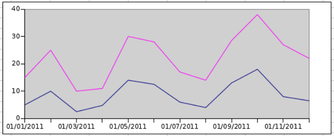
- Como medida de similaridade (proximidade), vamos adotar a **Correlação de Spearman**
- A correlação de Spearman é dada por:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

na qual d_i é a diferença do *ranking* dos valores das duas subséries na i-ésima observação

Subsequências Mais Próximas

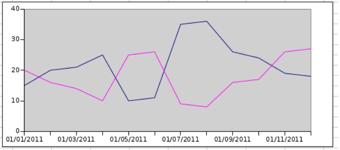
	Série 1	Série 2	Ranking Série 1	Ranking Série 2	d	d^2
Jan-01-2011	5	15	9	9	0	0
Fev-01-2011	10	25	5	6	−1	1
Mar-01-2011	2,5	10	12	12	0	0
Abr-01-2011	4,8	11	10	11	−1	1
Mai-01-2011	14	30	2	2	0	0
Jun-01-2011	12,5	28	4	4	0	0
Jul-01-2011	6	17	8	8	0	0
Ago-01-2011	4	14	11	10	1	1
Set-01-2011	13	28,5	3	3	0	0
Out-01-2011	18	38	1	1	0	0
Nov-01-2011	8	27	6	5	1	1
Dez-01-2011	6,5	22	7	7	0	0
				Correlação de Speraman	**0,99**	



Subsequências Mais Próximas

	Série 1	Série 2	Ranking Série 1	Ranking Série 2	d	d^2
Jan-01-2011	15	20	10	5	5	25
Fev-01-2011	20	16	7	7	0	0
Mar-01-2011	21	14	6	9	−3	9
Abr-01-2011	25	10	4	10	−6	36
Mai-01-2011	10	25	12	4	8	64
Jun-01-2011	11	26	11	2	9	81
Jul-01-2011	35	9	2	11	−9	81
Ago-01-2011	36	8	1	12	−11	121
Set-01-2011	26	16	3	7	−4	16
Out-01-2011	24	17	5	6	−1	1
Nov-01-2011	19	26	8	2	6	36
Dez-01-2011	18	27	9	1	8	64

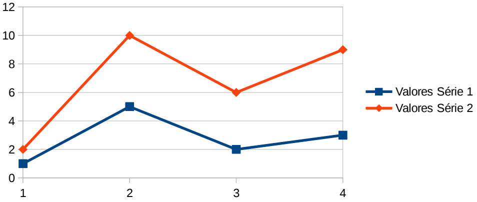
Correlação de Speraman **−0,87**



Subsequências Mais Próximas

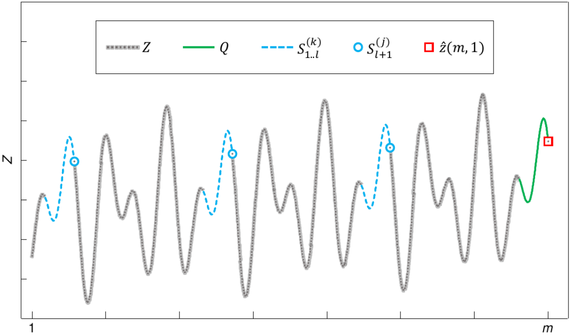
Calcula a Correlação de Spearman para as duas séries apresentadas abaixo

Tempo	Valores Série 1	Valores Série 2
t_1	1	2
t_2	5	10
t_3	2	6
t_4	3	9



Subsequências Mais Próximas

- Agora que já sabemos quais são as k subséries mais similares, devemos definir como iremos utilizar essas subséries para realizar a predição



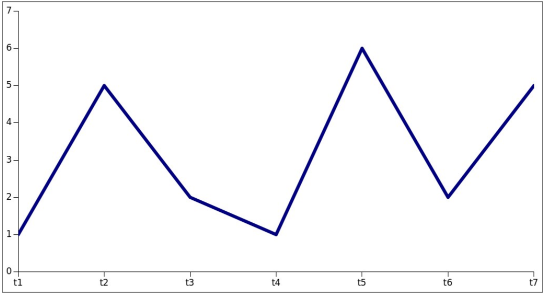
Subsequências Mais Próximas

- Para isso podemos considerar:
  - A média dos próximos pontos de cada uma das k subséries mais próximas
  - Acrescentar a média das diferenças entre os próximos pontos de cada uma das k subséries mais próximas e os últimos pontos das mesmas, e acrescentar essa variação em y_t
  - Acrescentar a média das variações relativas entre os próximos pontos de cada uma das k subséries mais próximas e os últimos pontos da mesma, e acrescentar essa variação em y_t

Exercício - Subsequências Mais Próximas

Calcule o próximo valor da série apresentada abaixo utilizando as duas subsequências mais próximas, uma subsérie de consulta de tamanho 2 e as três abordagens apresentadas no slide anterior para previsão de valores.

Tempo	Valores
t_1	1
t_2	5
t_3	2
t_4	1
t_5	6
t_6	2
t_7	5



Material Complementar

- Predição de séries temporais por similaridade

  http://www.teses.usp.br/teses/disponiveis/55/55134/tde-21112016-150659/pt-br.php

- Distância Euclidiana

  https://pt.wikipedia.org/wiki/Dist%C3%A2ncia_euclidiana

Material Complementar

- Uma comparação dos métodos de correlação de Pearson e Spearman

  https://support.minitab.com/pt-br/minitab/18/help-and-how-to/statistics/basic-statistics/supporting-topics/correlation-and-covariance/a-comparison-of-the-pearson-and-spearman-correlation-methods/

- Descrição de Modelos Estatísticos e de Aprendizado de Máquina para Predição de Séries Temporais

  http://conteudo.icmc.usp.br/CMS/Arquivos/arquivos_enviados/BIBLIOTECA_158_RT_412.pdf

Imagem do Dia