

Modelos empíricos:

Aplicação de Modelo Autorregressivo

Prof.^a Andrea de Oliveira Cardoso

Modelos empíricos

São modelos construídos através do ajuste de funções aos dados observados. As relações entre as variáveis são estabelecidas sem considerar os mecanismos físicos envolvidos.

Características:

- ✓ Mais simples do que os modelos conceituais;
- ✓ Dependem da abrangência e qualidade dos dados considerados em representar um sistema;
- ✓ Precisam de variáveis com fortes relações, capturadas pela função de ajuste;
- ✓ São ajustáveis para condições específicas e facilmente atualizados, a partir de novos dados;
- ✓ Úteis para representar processos caracterizados por fortes relações entre variáveis observáveis, cujo modelo conceitual é complexo, de difícil aplicação ou desconhecido.

Exemplos de modelos empíricos

Modelo Autorregressivo

Aplicado para variáveis cujas séries históricas apresentam uma forte autocorrelação temporal que se associa ao fenômeno (ou processo) modelado.

- 1 variável (histórico temporal);
- série estacionária (sem tendência);
- valores futuros obtidos pela combinação dos valores no passado.

Modelo de Regressão Linear Múltipla

Aplicado a casos em que uma variável (dependente) possui fortes relações com outras variáveis (independentes), que podem ser utilizadas para estimar a variável dependente.

- Várias variáveis: 1 variável dependente e várias variáveis independentes;
- a variável dependente possui correlação significativa com as variáveis independentes;
- a partir dessas relações estima-se a variável dependente.

Exemplos de aplicação

Modelo autorregressivo:

- ❑ Modelagem e simulação de séries temporais de vazões (Ex: Cosser et al. 2010);
- ❑ Estimativa de cheias (Ex: Vinieri Filho, 2004);
- ❑ Comportamento de nível de lençol freático (Ex: Manzione, 2017);
- ❑ Estudo de indicadores e índices de qualidade de água e previsão de consumo urbano de água (Ex: Leite et al. 2013).

Modelo de regressão linear múltipla:

- ❑ Estimativa séries de vazão: a partir de variáveis chuva e vazão (Ex: Silveira e Desbordes, 1999) e a partir de índices climáticos (Ex: Cardoso e Silva Dias, 2005);
- ❑ Estudo de efeito da poluição na saúde humana (Ex: Esteves et al., 2004);
- ❑ Estudos de irrigação localizada com esgoto sanitário tratado (Ex: Batista et al., 2008).

Modelo autorregressivo AR(p)

Chama-se de Autorregressivo de Ordem P , $AR(p)$, o modelo que:

Os valores da série Z_t são uma combinação linear dos p valores passados mais um termo a_t , no qual incorpora uma componente da série até o tempo t que não é explicada pelos valores passados.

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t$$

coeficientes autoregressivos

Outros exemplos são os modelos ARMA (modelo autorregressivo de média móvel) e ARIMA (modelo autorregressivo integrado de média móvel)

Ajuste do modelo (calibração)

O coeficiente de determinação R^2 é uma medida do poder explicativo de um modelo regressivo. Fornece a proporção da variação da variável estimada que é explicada pelo modelo.

$$R^2 = \frac{\frac{n \sum_{j=1}^N y_{ic}^2 - (\sum_{j=1}^N y_i)^2}{n(n-1)}}{\frac{n \sum_{j=1}^N y_i^2 - (\sum_{j=1}^N y_i)^2}{n(n-1)}}$$

onde y_{ic} é o valor calculado através da função de regressão e y_i é o valor observado; n é o número de observações.

Os valores de R^2 podem variar entre 0 e 1, quando este valor for 1 a regressão foi perfeita, sendo que, o inverso ocorre quando se aproxima de 0.

Ajuste do modelo (calibração)

O valor de R^2 tende a aumentar, com o aumento das variáveis preditoras, mesmo que não melhore o desempenho do modelo.

Assim, é importante calcular o coeficiente de determinação ajustado (R^2 **ajustado**), que é uma medida utilizada para avaliar a porcentagem da explicação da regressão, ajustada para o número de preditores do modelo em relação ao número de observações, sendo útil para a escolha do melhor modelo.

*Ao adicionar muitas variáveis preditoras pode ocorrer um superajuste. O R^2 **ajustado** pode ajudar a compensar o aumento natural de explicação provocado pelo elevado número de variáveis preditoras, relativo ao tamanho da amostra, sendo calculado com a expressão:*

$$R^2_{ajustado} = 1 - \left(\frac{n-1}{n-k} \right) (1 - R^2)$$

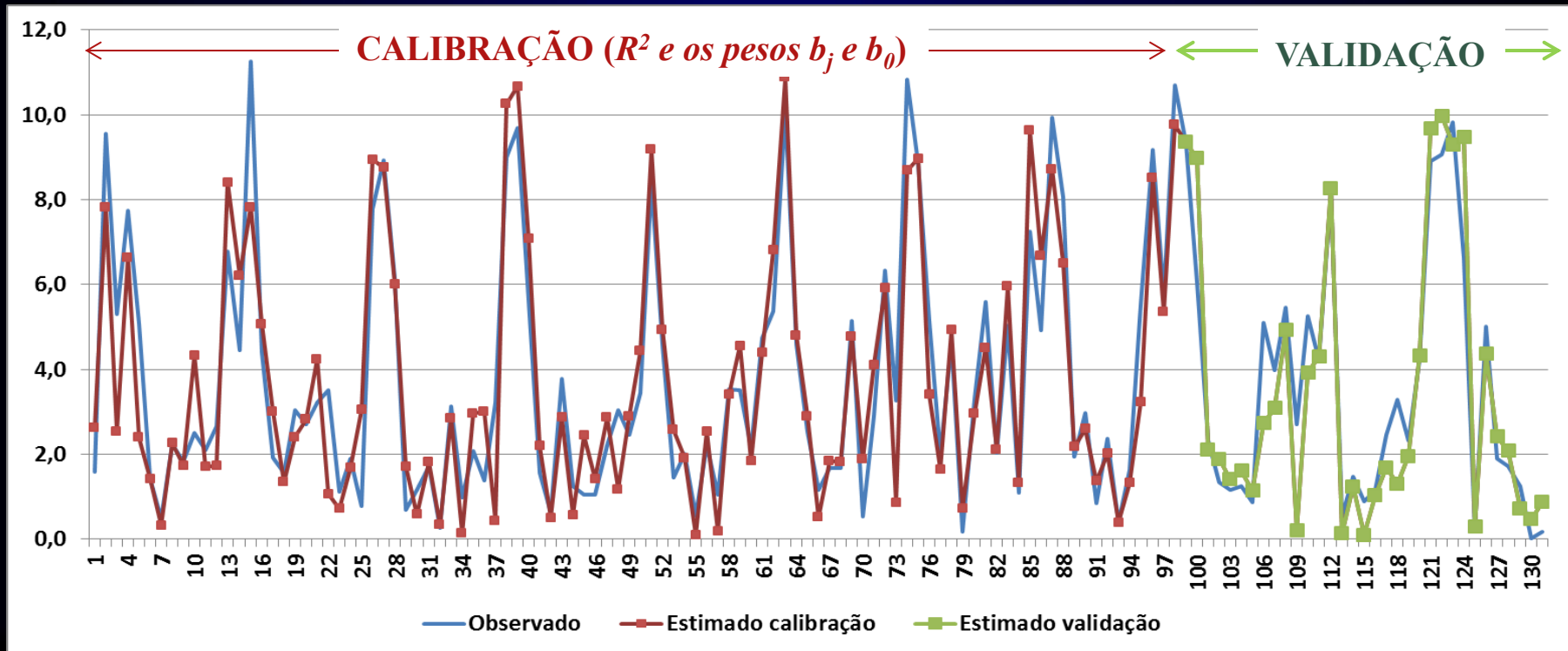
sendo k é o número de variáveis independentes e n o número de observações utilizadas no ajuste do modelo.

Calibração e validação:

Aplica-se a qualquer modelo empírico

Como o objetivo da modelagem é estimar a variável dependente com o menor erro, são necessários dois processos:

- 1) Calibração:** Obtem-se o melhor ajuste considerando um longo período da série e avaliando o R^2 ;
- 2) Validação:** Aplica-se o mesmo modelo para outro período da série, considerando os coeficientes ajustados na calibração (1).



Desenvolvimento do modelo autorregressivo AR(p)

Para desenvolver o modelo é importante saber a ordem “p” dos valores passados a serem considerados:

Exemplos:

Ordem 1: Para um modelo de ordem 1 com variável mensal, considera-se o mês anterior para estimar o mês de interesse.

Ordem 2: Para um modelo de ordem 2 com variável mensal, considera-se os dois meses anteriores para estimar o mês de interesse.

e assim por diante...

**Portanto, é necessário analisar a série para
identificar a ordem (p)**

Autocorrelação

Autocorrelação temporal, que significa a correlação da variável com ela mesma num tempo passado ou futuro. Algumas vezes, estas correlações são chamadas “*lag correlations*”, sendo a palavra *lag* utilizada para indicar esta variação temporal.

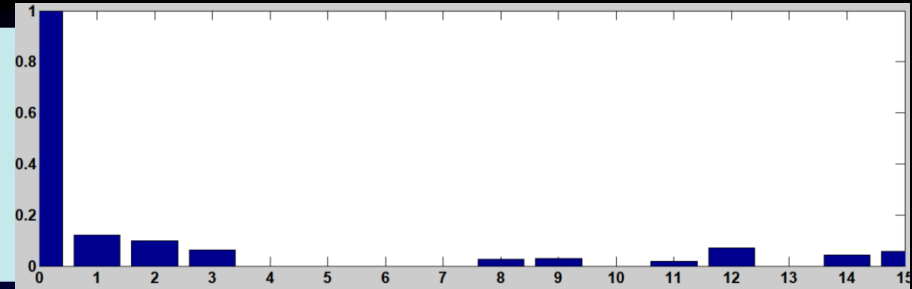
Exemplo de defasagem diária			
s/ defasagem Lag 0 (t)	defasagem de 1 dia Lag 1 (t-1)	defasagem de 2 dias Lag 2 (t-2)	defasagem 3 dias Lag 3 (t-3)
12/07/2018	11/07/2018	10/07/2018	09/07/2018
13/07/2018	12/07/2018	11/07/2018	10/07/2018
14/07/2018	13/07/2018	12/07/2018	11/07/2018
15/07/2018	14/07/2018	13/07/2018	12/07/2018
16/07/2018	15/07/2018	14/07/2018	13/07/2018
17/07/2018	16/07/2018	15/07/2018	14/07/2018
18/07/2018	17/07/2018	16/07/2018	15/07/2018
19/07/2018	18/07/2018	17/07/2018	16/07/2018
20/07/2018	19/07/2018	18/07/2018	17/07/2018
21/07/2018	20/07/2018	19/07/2018	18/07/2018
22/07/2018	21/07/2018	20/07/2018	19/07/2018
23/07/2018	22/07/2018	21/07/2018	20/07/2018
24/07/2018	23/07/2018	22/07/2018	21/07/2018
25/07/2018	24/07/2018	23/07/2018	22/07/2018
26/07/2018	25/07/2018	24/07/2018	23/07/2018
27/07/2018	26/07/2018	25/07/2018	24/07/2018
28/07/2018	27/07/2018	26/07/2018	25/07/2018
29/07/2018	28/07/2018	27/07/2018	26/07/2018
30/07/2018	29/07/2018	28/07/2018	27/07/2018
31/07/2018	30/07/2018	29/07/2018	28/07/2018

O conjunto de autocorrelações calculadas para várias defasagens (lags) é chamado “função de autocorrelação”, sendo apresentadas graficamente através de um gráfico conhecido como **correlograma**.

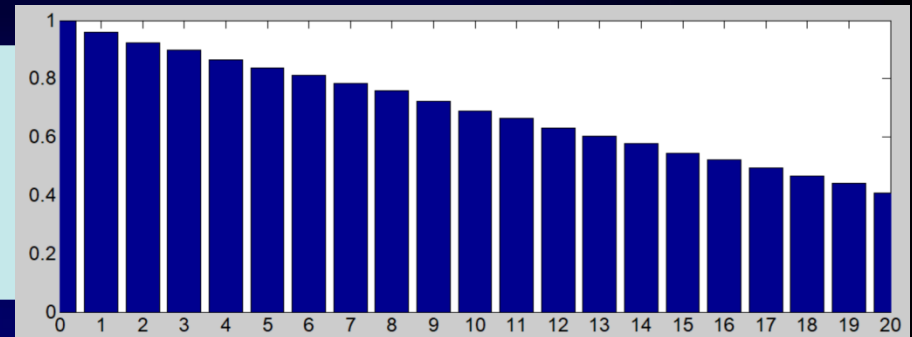
Padrão do correlograma

Considerando uma série de tamanho N:

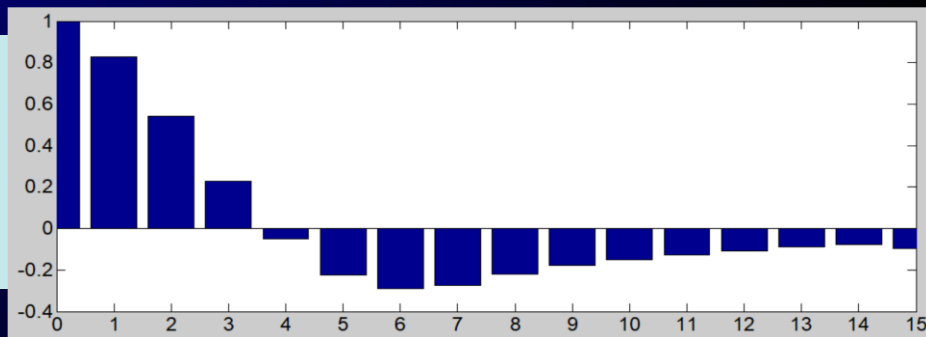
Série completamente aleatória,
para N grande,
 $r(k) \cong 0, k > 0$



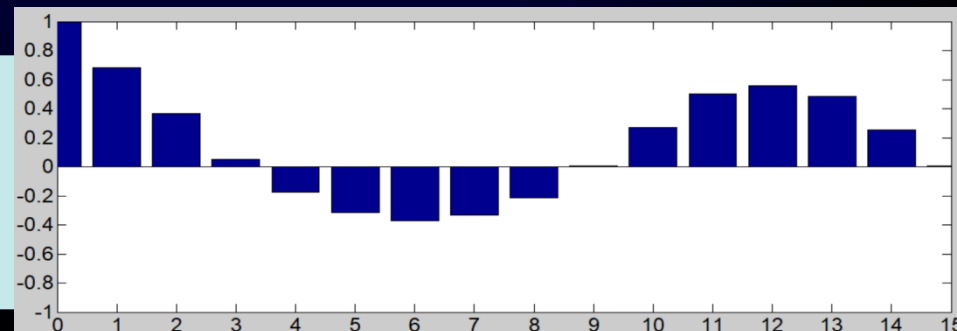
Série com tendência
valores de r não caem para 0,
exceto para lag muito alto



Série estacionária
em curto prazo r(k) é alto
lags menores



Série com flutuação sazonal
picos r(6) e r(12)
acompanhando o ciclo



Etapas para desenvolver o modelo AR(p)

1. Selecionar o período dos dados: separando uma parte para calibração (ajuste) e outra parte para validação (avaliação)
2. Verificar se há tendência linear nos dados e se houver remover a tendência (mesmo período de calibração)
3. Montar a matriz defasada (diversos lags)
4. Fazer a autocorrelação e o correlograma
5. Escolher a defasagem (p) para montar o modelo AR(p)
6. Realizar a regressão estatística (período de calibração), fornecendo o conjunto de variáveis (Z_{tp}) para obter os pesos do modelo
7. Validar o modelo (equação calibrada) aplicado ao período de validação.

Atividade

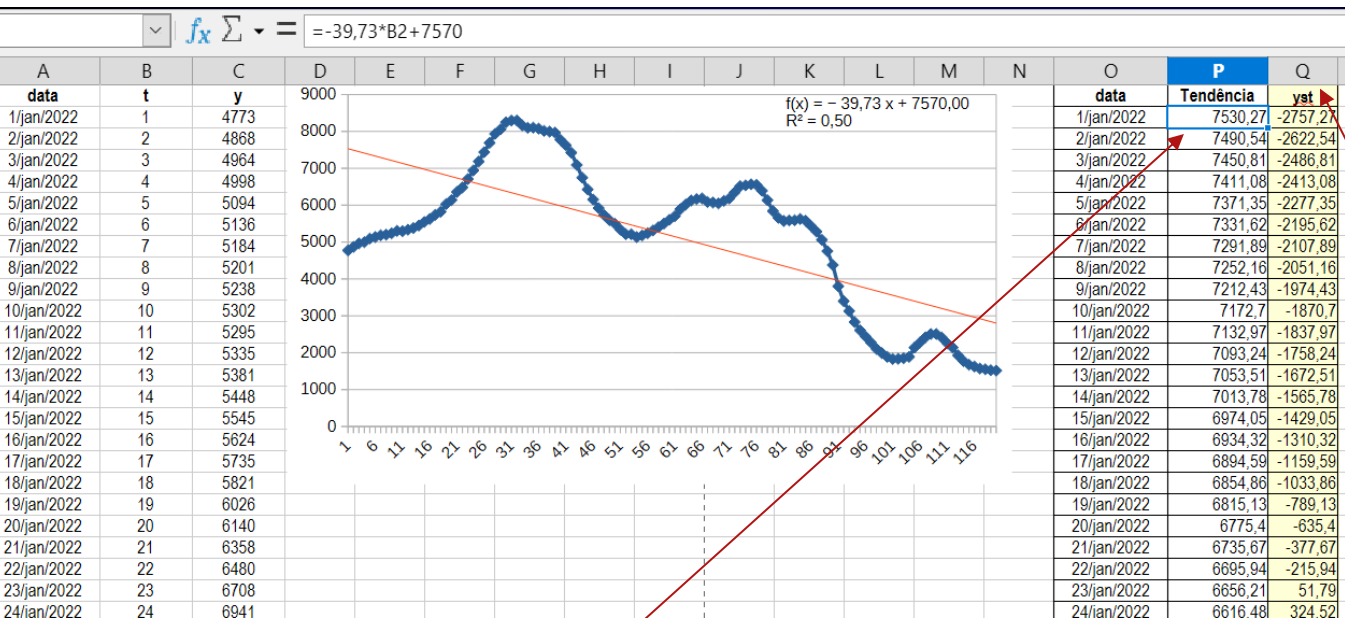
O arquivo dados “vaz_sobradinho_01a04_2022e2023.xlsx” contém dados diários de vazão no posto Sobradinho, de janeiro a abril de 2022 e 2023. Utilize essa série histórica para elaborar um Modelo Autoregressivo AR(p), considerando o ano 2022 para calibração e 2023 para validação. Realize as seguintes etapas utilizando o Calc:

- a) Para o período de calibração, identifique a tendência linear (mostrar gráfico);
- b) Remova a tendência linear de toda a série, considerando a equação identificada na calibração, item a (mostre o print da planilha);
- c) Calcule a autocorrelação defasada em até 10 dias (lag 10) e o faça o gráfico correlograma para o período de calibração (mostre o correlograma);
- d) Defina a ordem p do modelo através da menor autocorrelação superior a 0,97;
- e) Desenvolva o modelo autoregressivo AR(p), no período de calibração. Apresente a equação deste modelo, incluindo os valores dos coeficientes estimados e o R^2 .
- f) Apresente um gráfico comparando os valores estimados pelo modelo com os valores observados, no período de validação.
- g) Calcule a correlação e a REQM através da comparação das séries estimada pelo modelo e observada, no período de validação.

Submeter o PDF pelo moodle até 23h do dia 29/10/2026

Remoção de tendência no Calc

Utilizando a série de vazão de Sobradinho de janeiro a abril de 2022 e 2023 (vaz_sobradinho_01a04_2022e2023.xlsx), faça um gráfico de vazão para o período de 2022 pelo tempo corrido (1 a ...) e inclua a linha de tendência linear, a equação e o R^2 . Exemplo:



$y_{st} = y - \text{tendência}$

- Gere a série a série de tendência e a série de vazão sem tendência (y_{st}), ou seja vazão - tendência.

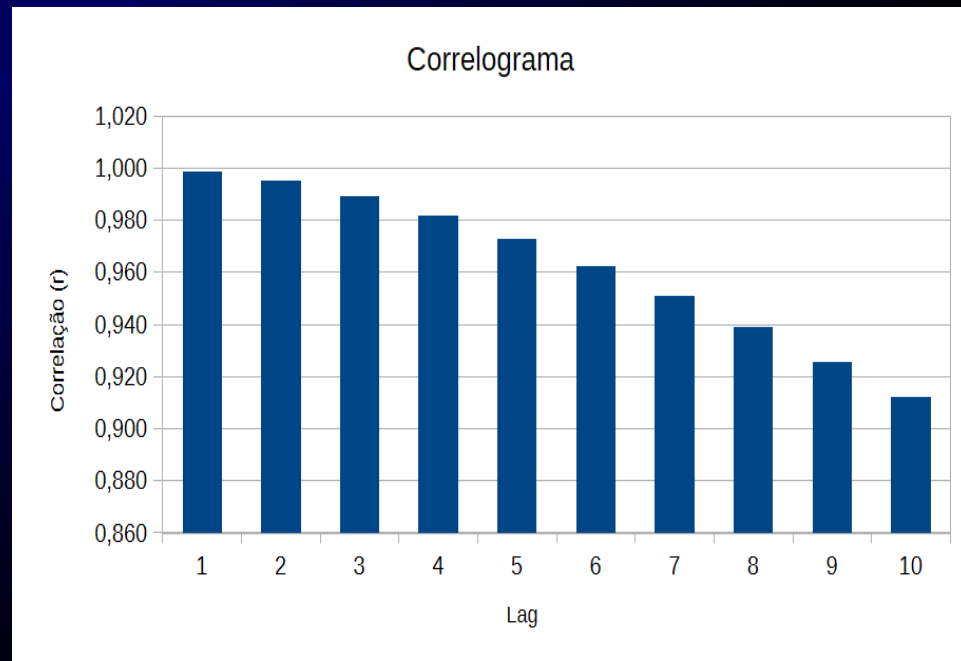
Obtenção da equação de tendência no Calc

- 1.Organizar e Selecionar os Dados:** Organize os dados em duas colunas adjacentes: a primeira com a variável temporal e a segunda com os valores numéricos correspondentes. Em seguida, selecione o intervalo de dados que deseja plotar (incluindo os cabeçalhos).
- 2.Inserir o Gráfico:** No menu superior, vá em Inserir e selecione Gráfico.... Na janela do Assistente de Gráficos, escolha o tipo Dispersão (XY). Clique em Concluir.
- 3.Inserir a Linha de Tendência:** Dê um duplo clique sobre o gráfico para ativá-lo (entrar no modo de edição). Em seguida, clique com o botão direito sobre a linha ou pontos de dados da série e selecione **Inserir Linha de Tendência**: Na janela de opções, certifique-se de que o tipo Linear está selecionado.
- 4.Exibir a Equação e o R^2 :** Na mesma janela de configuração da linha de tendência (ou clicando com o botão direito sobre a linha de tendência recém-criada e escolhendo Formatar Linha de Tendência), marque as caixas de seleção "Mostrar equação" e "Mostrar coeficiente de determinação (R^2)".

No Calc (Libre Office) a autocorrelação com defasagem (lag) é obtida através do deslocamento das linhas, seleção do período coincidente e interesse e aplicação da função **correl**, entre a série Z_t e as séries Z_{t-lag} . Exemplo:

data	Zt	Zt-1	Zt-2	Zt-3	Zt-4	Zt-5	Zt-6	Zt-7	Zt-8	Zt-9	Zt-10
1/jan/2022	-2757,27										
2/jan/2022	-2622,54	-2757,27									
3/jan/2022	-2486,81	-2622,54	-2757,27								
4/jan/2022	-2413,08	-2486,81	-2622,54	-2757,27							
5/jan/2022	-2277,35	-2413,08	-2486,81	-2622,54	-2757,27						
6/jan/2022	-2195,82	-2277,35	-2413,08	-2486,81	-2622,54	-2757,27					
7/jan/2022	-2107,89	-2195,82	-2277,35	-2413,08	-2486,81	-2622,54	-2757,27				
8/jan/2022	-2051,16	-2107,89	-2195,82	-2277,35	-2413,08	-2486,81	-2622,54	-2757,27			
9/jan/2022	-1974,43	-2051,16	-2107,89	-2195,82	-2277,35	-2413,08	-2486,81	-2622,54	-2757,27		
10/jan/2022	-1870,7	-1974,43	-2051,16	-2107,89	-2195,82	-2277,35	-2413,08	-2486,81	-2622,54	-2757,27	
11/jan/2022	-1837,97	-1870,7	-1974,43	-2051,16	-2107,89	-2195,82	-2277,35	-2413,08	-2486,81	-2622,54	-2757,27
12/jan/2022	-1758,24	-1837,97	-1870,7	-1974,43	-2051,16	-2107,89	-2195,82	-2277,35	-2413,08	-2486,81	-2622,54
13/jan/2022	-1672,51	-1758,24	-1837,97	-1870,7	-1974,43	-2051,16	-2107,89	-2195,82	-2277,35	-2413,08	-2486,81
14/jan/2022	-1565,78	-1672,51	-1758,24	-1837,97	-1870,7	-1974,43	-2051,16	-2107,89	-2195,82	-2277,35	-2413,08
15/jan/2022	-1429,05	-1565,78	-1672,51	-1758,24	-1837,97	-1870,7	-1974,43	-2051,16	-2107,89	-2195,82	-2277,35
16/jan/2022	-1310,32	-1429,05	-1565,78	-1672,51	-1758,24	-1837,97	-1870,7	-1974,43	-2051,16	-2107,89	-2195,82
17/jan/2022	-1159,59	-1310,32	-1429,05	-1565,78	-1672,51	-1758,24	-1837,97	-1870,7	-1974,43	-2051,16	-2107,89
18/jan/2022	-1033,86	-1159,59	-1310,32	-1429,05	-1565,78	-1672,51	-1758,24	-1837,97	-1870,7	-1974,43	-2051,16
19/jan/2022	-789,13	-1033,86	-1159,59	-1310,32	-1429,05	-1565,78	-1672,51	-1758,24	-1837,97	-1870,7	-1974,43
20/jan/2022	-635,4	-789,13	-1033,86	-1159,59	-1310,32	-1429,05	-1565,78	-1672,51	-1758,24	-1837,97	-1870,7
21/jan/2022	-377,67	-635,4	-789,13	-1033,86	-1159,59	-1310,32	-1429,05	-1565,78	-1672,51	-1758,24	-1837,97
22/jan/2022	-215,94	-377,67	-635,4	-789,13	-1033,86	-1159,59	-1310,32	-1429,05	-1565,78	-1672,51	-1758,24
23/jan/2022	51,79	-215,94	-377,67	-635,4	-789,13	-1033,86	-1159,59	-1310,32	-1429,05	-1565,78	-1672,51
24/jan/2022	324,52	51,79	-215,94	-377,67	-635,4	-789,13	-1033,86	-1159,59	-1310,32	-1429,05	-1565,78

9/abr/2022	2475,43	2440,33	2451,66	2461,02	2469,75	2478,9	2487,17	2497,7	2508,49	2538,54	2568,6
10/abr/2022	2464,72	2475,43	2451,66	2461,02	2469,75	2478,9	2487,17	2497,7	2508,49	2538,54	2568,6
11/abr/2022	2489,11	2464,72	2475,43	2440,33	2451,66	2461,02	2470,75	2479,8	2487,17	2497,7	2508,49
12/abr/2022	2488,99	2489,11	2464,72	2475,43	2440,33	2451,66	2462,02	2470,75	2479,8	2487,17	2497,7
13/abr/2022	2516,89	2488,99	2489,11	2464,72	2475,43	2440,33	2451,66	2462,02	2470,75	2479,8	2487,17
14/abr/2022	2548,98	2516,89	2488,99	2489,11	2464,72	2475,43	2440,33	2451,66	2462,02	2470,75	2479,8
15/abr/2022	2586,26	2548,98	2516,89	2488,99	2489,11	2464,72	2475,43	2440,33	2451,66	2462,02	2470,75
16/abr/2022	2684,39	2586,26	2548,98	2516,89	2488,99	2489,11	2465,72	2476,43	2440,33	2451,66	2462,02
17/abr/2022	2793,38	2684,39	2586,26	2548,98	2516,89	2488,99	2489,11	2465,72	2476,43	2440,33	2451,66
18/abr/2022	2852,04	2793,38	2684,39	2586,26	2548,98	2488,99	2489,11	2465,72	2476,43	2440,33	2451,66
19/abr/2022	2903,52	2852,04	2793,38	2684,39	2586,26	2548,98	2517,89	2488,99	2489,11	2465,72	2476,43
20/abr/2022	2954,2	2903,52	2852,04	2793,38	2684,39	2586,26	2549,98	2488,99	2489,11	2465,72	2476,43
21/abr/2022	2999,68	2954,2	2903,52	2852,04	2793,38	2684,39	2586,26	2517,89	2488,99	2489,11	2465,72
22/abr/2022	3041,53	2999,68	2954,2	2903,52	2852,04	2793,38	2684,39	2586,26	2549,98	2488,99	2489,11
23/abr/2022	3106,49	3041,53	2999,68	2954,2	2903,52	2852,04	2794,36	2684,39	2586,26	2549,98	2488,99
24/abr/2022	3150,46	3106,49	3041,53	2999,68	2954,2	2903,52	2853,04	2794,36	2684,39	2586,26	2549,98
25/abr/2022	3189,46	3150,46	3041,53	2999,68	2954,2	2903,52	2853,04	2794,36	2684,39	2586,26	2549,98
26/abr/2022	3310,35	3189,46	3150,46	3106,49	3041,53	2999,68	2954,2	2903,52	2853,04	2794,36	2684,39
27/abr/2022	3346,2	3310,35	3245,46	3150,46	3106,49	3041,53	2999,68	2954,2	2903,52	2853,04	2794,36
28/abr/2022	3380,77	3346,2	3310,35	3245,46	3150,46	3106,49	3041,53	2999,68	2954,2	2903,52	2853,04
29/abr/2022	3469,13	3380,77	3346,2	3310,35	3245,46	3150,46	3106,49	3041,53	2999,68	2954,2	2903,52
30/abr/2022	3505,51	3469,13	3380,77	3346,2	3310,35	3245,46	3150,46	3106,49	3041,53	2999,68	2954,2



Instruções

Exemplo de obtenção da equação

Para ajustar o modelo AR(p), utiliza-se a função **proj.lin** do Calc. Esta função fornece os coeficientes autorregressivos (coeficientes angulares) e o coeficiente linear (intercepto).

Exemplo: sendo a coluna B a série Z_t e as colunas C a G as séries Z_{t-1} a Z_{t-5} , para obter os coeficientes da equação do **AR(3)** aplica-se a função **PROJ.LIN(B2:B111; C2:E111; 1; 1)**, pressionando o Ctrl + Shift + Enter para calcular a matriz de coeficientes. O resultado é conforme:

Coeficientes (angulares e intercepto em ordem inversa)	-0,0079328331	-0,8846632841	1,8910017953	-15,943877519
Erro padrão dos coeficientes	0,0985615283	0,1896010121	0,0970142916	13,478278582
Coeficiente de determinação (R2) e Erro padrão de y	0,9994526281	75,063554211	#N/DISP	#N/DISP
Estatística F de Fisher	64515,543196	106	#N/DISP	#N/DISP
Soma dos quadrados da regressão	1090545678,7	597260,9401	#N/DISP	#N/DISP

Aplicando a equação do modelo e substituindo os valores observados dos Z_{t-lags} estima-se o valor de Z_t

Equação do modelo:

$$Z_t = 1,89 \cdot Z_{t-1} - 0,884 \cdot Z_{t-2} - 0,007 \cdot Z_{t-3} - 15,94$$

$$R^2 = 0,99$$

Referências

- BATISTA, R. O.; SOARES, A.; MANTOVANI, E. C., MATOS, A. T., SOUZA, J. A. R.; BATISTA, R. O. (2008) Modelos Empíricos da irrigação localizada com esgoto sanitário tratado. Engenharia na Agricultura, Viçosa, MG, v.16, n.3, 369-377.
- Chatfield C., 1996: The Analysis of Time Series: An introduction. Chapman & Hall, fifth edition, NY. 283 pp.
- COSER M. C; MENDONÇA, A. S. F.; ZORZAL, R.; FASSARELLA, J. A. C. *Aplicabilidade de modelos de geração de vazões sazonais apresentando dependência de longo termo. X Simpósio de Recursos Hídricos do Nordeste, 2010. PAP004334.*
- CARDOSO, A. O., SILVA DIAS, P. L. O uso de índices de padrões de teleconexões e modos de TSM na previsão empírica de vazão no rio São Francisco. In: Simpósio Internacional de Climatologia, 2005, Fortaleza, Brasil. Anais.
- CARDOSO, A. O; SILVA DIAS, P. L. Assessing Forecast of the empirical model to forecast precipitation in the South and Southeast Regions Brazil. In: 8th International Conference on Southern Hemisphere Meteorology and Oceanography, 2006, Foz do Iguaçu, Brazil. Proceedings ... São José dos Campos: INPE, 2006. p. 517-520. CD-ROM. ISBN 85-17-00023-4.
- MANZIONE, R. L. Impactos do clima no comportamento de níveis freáticos em área de conservação utilizando um modelo autoregressivo de séries temporais. Brazilian Journal of Biosystems Engineering v. 11(2): 185-198, 2017.
- STEVES, G. R. T.; BARBOSA, S. R.; SILVA, E. P.; ARAÚJO, P. D. Estimativa dos efeitos da Poluição Atmosférica sobre a Saúde Humana: alguns possibilidades metodológicas e teóricas para a cidade de São Paulo Associação Nacional de Pós-Graduação e Pesquisa em Ambiente e Sociedade. Encontro da Associação Nacional de Pós Graduação e Pesquisa em Ambiente e Sociedade. São Paulo, ANPPAS, 2004. p.1-20.
- VINIERI FILHO, C. IMPLEMENTAÇÃO DE UM PROGRAMA PARA INTEGRAR MODELOS DE PREVISÃO DE CHEIAS FLUVIAIS. 2004. 154 f. Tese de Doutorado. Instituto Tecnológico de Aeronáutica (ITA). São José dos Campos, SP.
- LEITE, Dilza Aparecida Nalin de Oliveira et al. Avaliação dos parâmetros do índice de qualidade de água segundo o modelo ARIMA. Holos Environment, p. 24-39, 2013.
- SILVEIRA, A. L. L., DESBORDES, M., (1999). Modelo Hidrológico Distribuído Urbano com Poucos Parâmetros. Revista Brasileira de Recursos Hídricos V. 4, n. 1, p 35-48.
- OLIVO, A. A; MILIONI, A. Z. *Modelos matemáticos para a previsão de cheias fluviais.* http://www.abrh.org.br/novo/ii_simp_rec_hidric_centro_oeste_campo_grande15.pdf
- MANZIONE, R. L. Impactos do clima no comportamento de níveis freáticos em área de conservação utilizando um modelo autoregressivo de séries temporais. Brazilian Journal of Biosystems Engineering v. 11(2): 185-198, 2017.
- XAVIER, L.; GENTON, M. G. Predictive spatio-temporal models for spatially sparse environmental data. Statistica Sinica, Vol. 15, s. 547-568, 2005. <http://www3.stat.sinica.edu.tw/statistica/oldpdf/A15n213.pdf>