

**DINCON 2017**CONFERÊNCIA BRASILEIRA DE DINÂMICA, CONTROLE E APLICAÇÕES
30 de outubro a 01 de novembro de 2017 – São José do Rio Preto/SP

Previsão via modelos de regressão temporal e de Box-Jenkins para série mensal média de temperatura mínima de Campinas (SP)

Miriam Rodrigues Silvestre¹

Departamento de Estatística, UNESP, Presidente Prudente, SP

Manoel Ivanildo Silvestre Bezerra²

Departamento de Estatística, UNESP, Presidente Prudente, SP

Resumo. O ajuste de séries de dados climatológicos é uma tarefa que pode ser realizada através de diferentes modelos estatísticos e computacionais. Nesse trabalho são ajustados dois modelos estatísticos, o modelo de regressão de decomposição temporal e o modelo SARIMA, à série de médias mensais de temperatura mínima do ar da cidade de Campinas (SP). O período utilizado foi de junho de 1988 a junho de 2016, sendo que do início do período até maio de 2015 foi utilizado para a construção dos modelos e de junho de 2015 a junho de 2016 para avaliar a capacidade preditiva dos mesmos. Os resultados obtidos indicam que o modelo SARIMA obteve melhores resultados para a estatística do Erro Médio Absoluto (EMA) e o modelo de regressão para as estatísticas do Erro Quadrático Médio (EQM) e Raiz do Erro Quadrático Médio (REQM). Ambos os modelos foram adequados, e geraram previsões bem próximas dos verdadeiros valores.

Palavras-chave. Temperatura mínima, Séries Temporais, Modelos de Regressão, Modelo SARIMA.

1 Introdução

O estudo do clima é de bastante interesse para diversas áreas como a agricultura, saúde e economia. Entre as variáveis que compõe os estudos climáticos, a temperatura é uma das mais importantes, e é primordial para o desenvolvimento fenológico das plantas, e manutenção da vida no planeta.

A área de interesse nesse trabalho é o município de Campinas (SP), o qual está localizado na porção leste do Estado de São Paulo, ocupa uma área de 801 km² e conta com uma população de aproximadamente 1 milhão de habitantes. A estação meteorológica da cidade que registra as medidas de temperatura situa-se no campus da Universidade Estadual de Campinas (UNICAMP),

¹ miriam@fct.unesp.br,

² manoel@fct.unesp.br

e o banco de dados é de responsabilidade do Centro de Pesquisa Meteorológicas e Climáticas Aplicadas à Agricultura (CEPAGRI/UNICAMP).

Pretende-se realizar um estudo da série histórica de temperatura mínima mensal média do ar da cidade de Campinas (SP) em busca de um modelo que permita realizar previsões de valores futuros com precisão e avaliar se as indicações sobre a existência de aumento das temperaturas, conforme vem apontando o Painel Intergovernamental de Mudanças Climáticas (IPCC) se verificam para a série em estudo.

2 Métodos

Existem diversos métodos para se ajustar um modelo à um conjunto de dados, nesse trabalho pretende-se abordar os métodos estatísticos e dois modelos de previsão, o modelo de regressão de decomposição temporal e o modelo de Box-Jenkins para séries temporais, os quais são descritos nas seções 2.1 e 2.2.

2.1 Modelo de Regressão de Decomposição Temporal

Para ajustar a sazonalidade mensal observada na série de Temperatura mínima, foi construído um modelo de regressão de decomposição temporal. Maiores detalhes sobre a técnica podem ser encontrados em [1] e [2]. Foi construído um modelo matricial, definido na Equação 1:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad (1)$$

no qual $\mathbf{Y}$ representa o vetor de observações da variável de interesse, constituído de $N=px12$ observações para uma série mensal de p anos; $\mathbf{X} = [\mathbf{C} : \mathbf{D}]$ é uma matriz composta por:

$$\mathbf{C} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & N \end{bmatrix} \text{ e } \mathbf{D} = \begin{bmatrix} \mathbf{D}_1 \\ \vdots \\ \mathbf{D}_k \\ \vdots \\ \mathbf{D}_p \end{bmatrix}, \text{ para } \mathbf{D}_k = \begin{matrix} 12 \times 11 \\ \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ -1 & -1 & -1 & \cdots & -1 & -1 \end{bmatrix} \end{matrix} \begin{matrix} \leftarrow jan \\ \leftarrow fev \\ \vdots \\ \leftarrow nov \\ \leftarrow dez \end{matrix};$$

$$\boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{bmatrix} \text{ é o vetor de parâmetros do modelo e } \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix} \text{ é o vetor de erros. O vetor de parâmetros}$$

$\boldsymbol{\gamma}$ é constituído de duas partes, a primeira parte representando a tendência linear $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$, sendo

que β_0 é a constante e β_1 a coeficiente angular; e a segunda parte a sazonalidade mensal da série dada por $\boldsymbol{\alpha}' = [\alpha_1 \cdots \alpha_{11}]'$. O modelo para cada observação t é definido como:

$$Y_t = \beta_0 + \beta_1 t + \sum_{i=1}^{11} \alpha_i d_{it} + \varepsilon_t, \quad (2)$$

para $d_{it} = \begin{cases} 1, & \text{para } i = \text{janeiro a novembro} \\ -1, & \text{para } i = \text{dezembro} \\ 0, & \text{caso contrário} \end{cases}$. O vetor de observações estimado $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\gamma}}$ é

obtido através do vetor de estimativas de mínimos quadrados dos parâmetros, o qual é dado por: $\hat{\boldsymbol{\gamma}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. É construída uma tabela de análise de variância (ANOVA) a partir da qual é avaliada a significância do modelo ajustado.

2.2 Modelo de Box-Jenkins

A metodologia a ser desenvolvida neste trabalho será baseada em tópicos da área de climatologia utilizando a modelagem de séries temporais, conhecida como metodologia de Box-Jenkins (BJ) [3], ou seja, os modelos auto-regressivos-integrado-médias-móveis sazonal (SARIMA), que são bastante utilizados quando se trabalha com séries temporais sazonais não determinísticas.

3 Resultados

Nesse capítulo serão apresentados os resultados obtidos com o ajuste dos modelos descritos na seção 2.1 e 2.2 aos dados de médias mensais de temperatura mínima de Campinas (SP) utilizando o período de junho de 1988 a maio de 2015.

3.1 Modelo de Regressão de Decomposição Temporal ajustado

Foi realizado o ajuste do modelo de regressão de decomposição temporal e realizada análise dos resíduos do modelo e observou-se que os mesmos eram aleatórios e normalmente distribuídos, como era esperado para que o modelo de regressão fosse aceitável.

As estatísticas do modelo ajustado são apresentadas na Tabela 1. O valor do $R^2(\text{ajust})=89,7\%$ indica que o modelo ajustado consegue explicar 89,7% da variabilidade dos dados, sendo que, modelos razoáveis apresentam estatística $>70\%$. A estatística Cp-Mallows indica o número de parâmetros do modelo ajustado e ajuda a selecionar um bom modelo, e S é o desvio-padrão do modelo ajustado sendo que, quanto menor for seu valor melhor é o ajuste do modelo.

Tabela 1: Estatísticas do modelo de regressão ajustado.

Estatísticas do Modelo	Tmin-med (Regressão)
$R^2(\text{ajust})$	89,7%
Cp-Mallows	11,3
S	0,9293

Fonte: Elaborada pelos autores.

Os parâmetros do modelo podem ser observados na Tabela 2 abaixo. Nota-se que todos os parâmetros são significativos (P-valor $<0,05$) para o modelo exceto a tendência=-0,00029 (P-valor=0,6), que poderia ser retirada e o modelo reajustado. Entretanto, optou-se por deixá-la no modelo.

Tabela 2: Parâmetros estimados, coeficientes, erro padrão dos coeficientes, estatística T e P-valor.

Parâmetros	Coef.	SE Coef.	T	P-valor
b_0 (intercepto)	16,7194	0,1037	161,28	0
b_1 (tendência)	-0,00029	0,000553	-0,52	0,6
a_1	3,1657	0,1714	18,47	0
a_2	3,1941	0,1714	18,63	0
a_3	2,6404	0,1714	15,4	0
a_4	0,8456	0,1714	4,93	0
a_5	-2,3387	0,1715	-13,64	0
a_6	-3,9646	0,1715	-23,12	0
a_7	-4,4771	0,1714	-26,11	0
a_8	-3,1645	0,1714	-18,46	0
a_9	-1,0008	0,1714	-5,84	0
a_{10}	0,8735	0,1714	5,1	0
a_{11}	1,6267	0,1714	9,49	0

Fonte: Elaborada pela autora.

3.2 Modelo de Box-Jenkins (BJ)

A partir da análise das autocorrelações e autocorrelações parciais amostrais foi identificado o modelo SARIMA(2,0,0)(4,0,0)₁₂ para séries sazonais com período $s=12$, ajustado a série mensal média de temperatura mínima. Os resultados com as estimativas dos parâmetros e da variância do ruído branco ($\text{Var}(a_t)$) são apresentados na Tabela 3, e mostra que todos os parâmetros são significativos.

Tabela 3: Estimativas dos Parâmetros.

SARIMA (2,0,0)(4,0,0) ₁₂	Estimativas	Erro Padrão	Estatística-T	p-valor
AR(1)	0,1787	0,0545	3,28	0,001
AR(2)	0,2526	0,0547	4,62	0,000
SAR(1)	0,2378	0,0554	4,29	0,000
SAR(2)	0,1536	0,5400	2,85	0,005
SAR(3)	0,3523	0,0552	6,38	0,000
SAR(4)	0,2558	0,0587	4,36	0,000
$\text{Var}(a_t)$	0,8960	0,9466		

Fonte: Elaborada pelo autor.

A Tabela 4 a seguir, é apresentada para que se possa avaliar a adequacidade do modelo, por meio do Teste de Box-Pierce. Os p-valores apresentados na última linha da

Tabela 4 indica que a hipótese de ruído branco não é rejeitada para nenhum dos Lags apresentados, satisfazendo assim, a condição necessária para o ajuste de um modelo de séries temporais.

Tabela 4: Teste Box-Pierce (Ljung-Box).

Lag-k	12	24	36	48
Qui-Quadrado	8,5	21,6	37,8	54,9
G1	6	18	30	42
p-valor	0,201	0,251	0,155	0,088

Fonte: Elaborada pelo autor.

O gráfico disposto na Figura 1 apresenta as autocorrelações residuais do modelo ajustado, respectivamente. Observa-se que as autocorrelações estão contidas no intervalo de confiança de 95% construído, indicando que as mesmas não são significativas, satisfazendo as pressuposições necessárias para os resíduos do modelo ajustado.

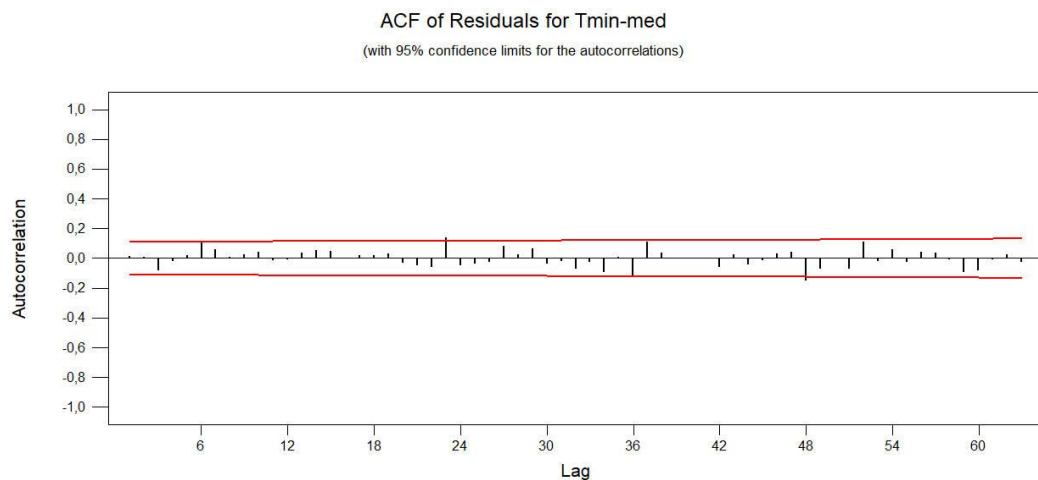


Figura 1: Autocorrelações dos resíduos. Fonte: Elaborada pelo autor.

A Figura 2 apresenta os resíduos versus a ordem dos dados para o modelo ajustado. Observa-se que os resíduos estão distribuídos de forma aleatória, com média em torno de zero como era esperado e variância constante. O Teste de Shapiro-Wilk resultou na estatística $R=0,9988$ com $p\text{-valor}>0,10$, confirmando a veracidade da suposição de normalidade dos resíduos.

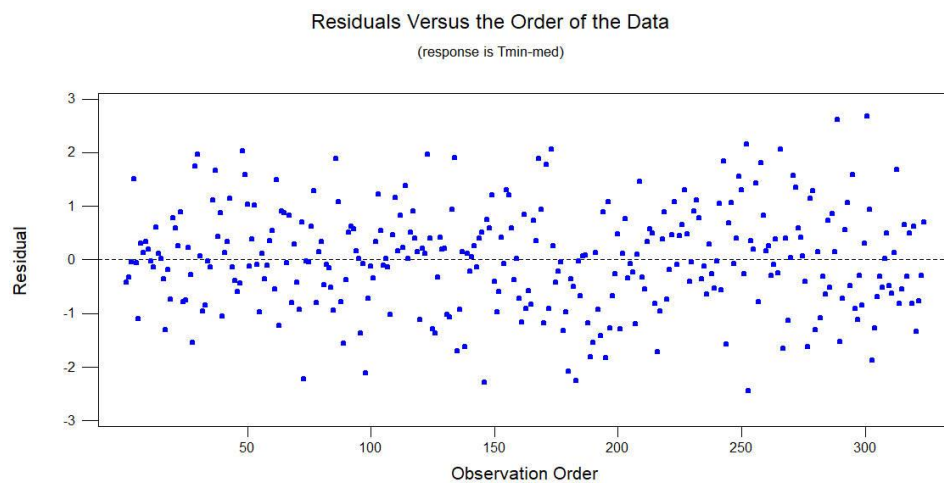


Figura 2: Resíduos versus a ordem. Fonte: Elaborada pelo autor.

3.2 Comparação dos modelos

Para a comparação dos modelos foram calculadas as estatísticas: Erro Médio Absoluto (EMA), Erro Quadrático Médio (EQM) e Raiz do Erro Quadrático Médio (REQM). Os resultados estão dispostos na Tabela 5.

Tabela 5: Estatísticas para a comparação dos modelos de regressão e SARIMA.

Estatísticas dos Modelos	Tmin-med (Regressão)	Tmin-med (SARIMA)
EMA	0,7026	0,66403
EQM	0,6407	0,84473
REQM	0,8004	0,91909

Fonte: Elaborada pelos autores.

As previsões elaboradas para o período junho de 2015 a junho de 2016 foram organizadas num único gráfico (Figura 3), no qual é possível comparar os dados reais com as previsões realizadas pelos dois modelos ajustados.

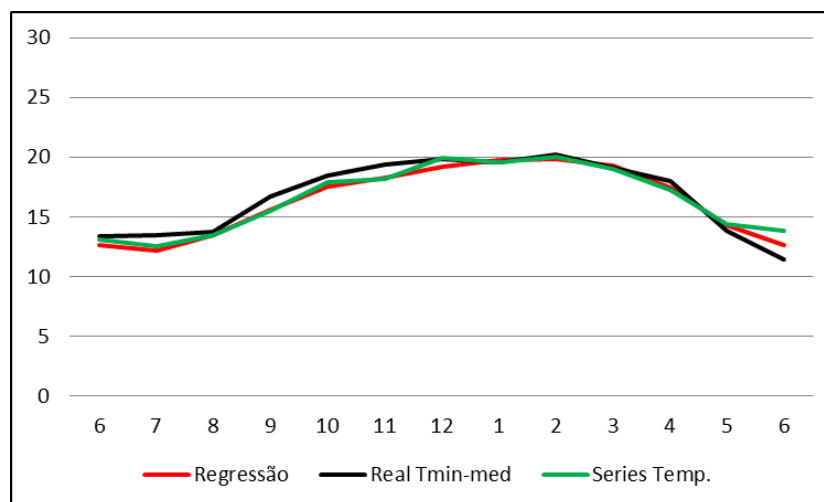


Figura 3: Temperatura mínima-média real (Real Tmin-med), estimada por Regressão (Regressão) e BJ SARIMA (Series Temp.). Fonte: Elaborada pela autora.

4 Discussões

Observou-se que os modelos ajustados obtiveram resultados semelhantes quando comparamos as estatísticas da Tabela 5, destacando o EQM que apresentou um melhor resultado para o modelo de regressão, enquanto que o EMA apresentou melhor resultado para o modelo SARIMA, que se mostra menos sensível a valores discrepantes. Se compararmos as previsões, observa-se que as curvas de previsão (Regressão, SARIMA), estão próximas dos verdadeiros valores.

5 Conclusões

Conclui-se que ambos os modelos ajustados foram adequados para realizar previsões de temperatura mínima média mensal, entretanto, o modelo de regressão é de mais fácil entendimento e utilização que o modelo SARIMA para a área de Climatologia.

Referências

- [1] N. R. Draper, H. Smith, Applied regression analysis, John Wiley & Sons, (1998).
- [2] P. A. Morettin, C. M. C. Toloi, Análise de series temporais, Edgard Blucher, (2004).
- [3] G. E. P. Box, G. M. Jenkins, G. Reinsel, Time series analysis: forecasting and control. 3rd. ed., Prentice Hall, (1994).