

Análise e previsão para dados diários de vendas no varejo

Maria Sílvia de Assis Moura ¹,
Antonio Álvaro de Assis Moura ²,
Camilla Emanuelle Marques ³,
Maísa Rodrigues Santos ⁴.

Introdução

Uma empresa de varejo da cidade de Ituiutaba, MG, tem interesse em saber a previsão de seu produto carro chefe, o pão de queijo. Os modelos para previsões de médias mais usados são os modelos *ARIMA*, mas neste caso os dados são diários. E o interesse é também verificar o comportamento das vendas nos dias em que há promoções, ou seja, as quartas-feiras e sábados. Destacamos que aos domingos, a empresa trabalha em horário reduzido.

Este trabalho tem como objetivo o uso dos modelos *ARIMA* em dados diários e com sazonalidade semanal.

1 Material e métodos

A série temporal estudada apresenta o volume de vendas diárias de pães de queijo de um estabelecimento de venda de varejo na cidade de Ituiutaba de 02 de janeiro de 2017 até 31 de dezembro de 2017. Há, na série original oito dados faltantes, a saber nos dias 14 de abril, 01 de maio, 08 de junho, 09 de junho, 15 de agosto, 02 de outubro, 12 de outubro e 25 de dezembro. Como para a construção de modelo SARIMA é necessário que a série seja completa, estes dados foram interpolados, dado a característica sazonal dos dados, optou-se por calcular a média entre o valor posterior e anterior do mesmo dia da semana. Para o dia 25 de dezembro, não foi possível, uma vez não há dado para o primeiro de janeiro de 2018, desta forma, foi calculada a média entre 18 de dezembro e 11 de dezembro. Os dados usados para a construção do modelo foram os dados de 02 de janeiro até 17 de dezembro de 2017, os dados de 18 de dezembro até 31 de dezembro de 2017 foram usados para comparação das previsões.

Para analisar a série foi construído o gráfico de amplitude versus a média, com o intuito de verificar se havia necessidade de aplicar a transformação logarítmica nos dados a fim de tornar aditivo o modelo e constante a variância. Através do gráfico de autocorrelação dos dados investigamos a presença de tendência, observando se há um decaimento lento em direção ao zero, e, se houver ondulações senoidais há indícios de sazonalidade.

Confirmamos a existência de tendência através do teste do sinal (Cox-Stuart) e da sazonalidade através do teste *F* da análise de variância. Após identificada tendência, esta será removida mediante uma diferença simples. E, se houver sazonalidade, a remoção é feita lançando mão de diferença sazonal.

¹DEs-UFSCar. e-mail: msilvia@ufscar.br

²FACES-UFU. e-mail: a3moura@gmail.com

³FACES-UFU. e-mail: camilla.emanuelle@hotmail.com

⁴FACES-UFU. e-mail: maisa12.rodrigues@hotmail.com

Com a série sem tendência e sazonalidade, e com variabilidade constante, isto é, estacionária, foram ajustados modelos *SARIMA*, identificados a partir das funções de autocorrelação *FAC* e autocorrelação parcial *FACP*. O ajuste dos dados ao modelo é aferido através do teste de Box-Pierce, para verificar se os resíduos são independentes e identicamente distribuídos.

Os vários modelos candidatos são comparados por meio do critério de AKAIKE(AIC). Optando pelo modelo que possuir o menor AIC. Para o modelo escolhido, verificamos se há observações atípicas, se houver, introduzimos covariáveis de intervenção nos instantes detectados. Uma observação atípica, em uma série temporal pode ser ajustada como um *AO*, *Additive Outlier*, ou seja, como uma função impulso, que assume valor um apenas em um instante e zero nos demais valores do tempo. Ou como *IO*, *Innovation Outlier*, isto é, como uma função degrau, que é zero e a partir de um valor do tempo assume o valor um e se mantém até o final da série. Os testes para detectar *outliers* estão descritos em Cryer e Sun(2010).

O modelo final, ajustado com as etapas anteriores pode ser usado para prever observações futuras, todos os métodos descritos anteriormente são detalhadamente discutidos em Morettin e Toloi(2006).

Neste trabalho os dados foram analisados com o auxílio do software R Core Team (2018).

2 Resultados

A série temporal observada está mostrada na figura (2). Não destacamos tendência, em uma análise visual, mas, observamos uma tendência, que era esperada, uma vez que os dados são diários e se trata de unidades vendidas.

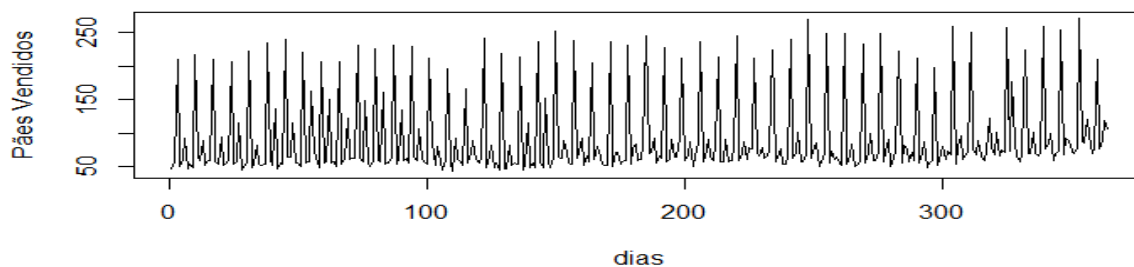


Figura 1: Gráfico da série de vendas diárias de pães de queijo, durante o ano de 2017.

A figura (2) indica que a amplitude depende da média. O teste *t* corrobora esta indicação, pois o valor-*p* é $< 0,0001$, assim, para tornar o modelo aditivo, vamos usar os dados com a transformação logarítmica.

Com a série transformada, verificamos a existência de sazonalidade ao observamos os box-plots para cada um dos dias da semana, como mostrado na figura (2). Notamos que

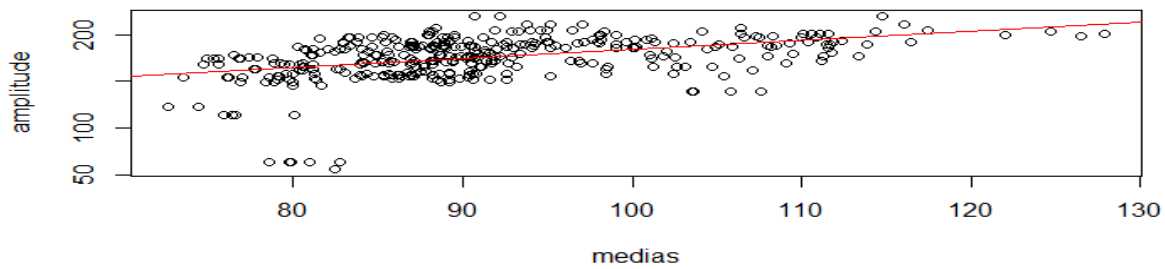


Figura 2: Gráfico da amplitude versus médias da série (Z_t).

às quartas-feira há um aumento nas vendas acentuado, aos sábados há também aumento nas vendas mas em menor escala. Destacamos a existência de *outliers* em particular em uma sexta-feira, que foi indentificado com o dia 24 de novembro.

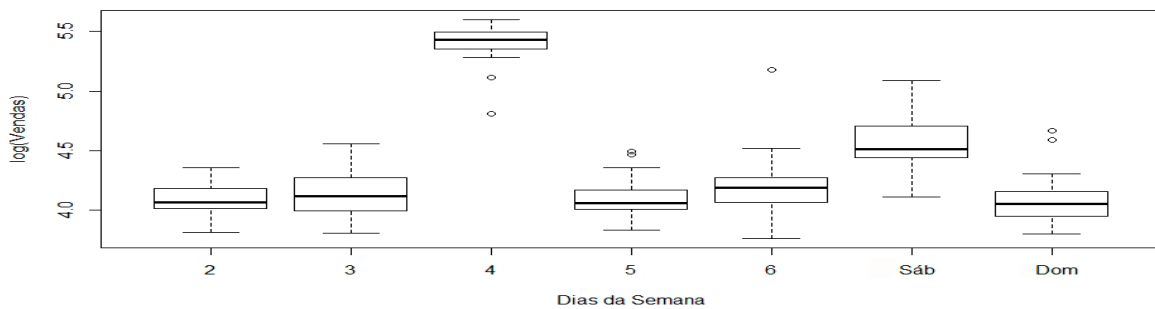


Figura 3: Gráfico Box-Plot para cada dia da semana.

Analisando a figura (2) vemos que a série apresenta tendência, uma vez que a FAC não decresce a zero rapidamente, que foi confirmada através do teste de Cox-Stuart. E, também notamos a existência de sazonalidade, foi ajustada um modelo de análise de variância, tendo os dias da semana como fator, e rejeitou-se a hipótese nula de não-existência de sazonalidade.

Para retirar a tendência, fizemos uma diferença e para retirar a sazonalidade fizemos uma diferença de ordem sete. A FAC e FACP da série livre de tendência e de sazonalidade está mostrada na figura (2).

Com a série estacionária foram ajustados os modelos $SARIMA(0, 1, 1)(1, 1, 0)_7$ e $SARIMA(0, 1, 1)(0, 1, 1)_7$. O teste de Box e Pierce foi realizado, para cada modelo ajustado, com o intuito de verificar se os resíduos satisfazem as condições de independência e de mesma distribuição, o que se verificou em ambos modelos.

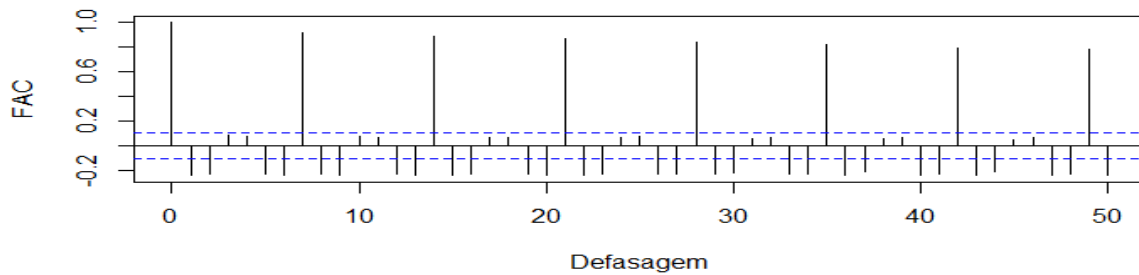


Figura 4: Função de Autocorrelação do logaritmo da série (Z_t) .

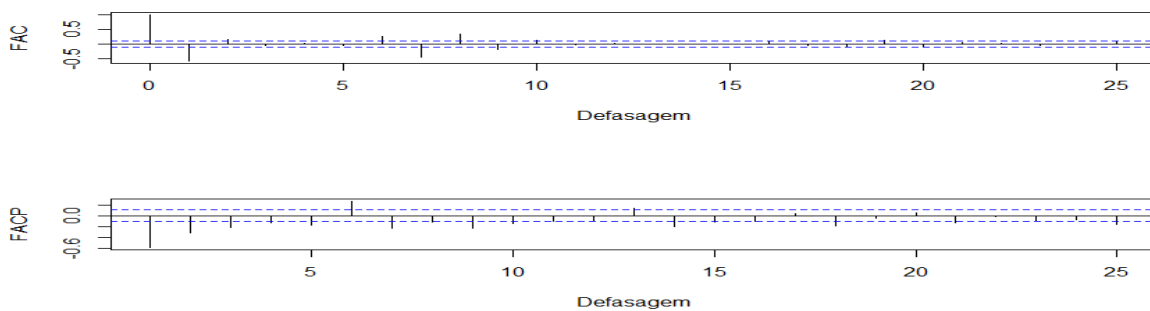


Figura 5: Correlograma da série (Z_t) ajustada para tendência e sazonalidade.

O modelo $SARIMA(0, 1, 1)(0, 1, 1)_7$ apresentou o menor valor de AIC $-2,852$ contra $-2,686$ do modelo $SARIMA(0, 1, 1)(1, 1, 0)_7$, foi feita análise de intervenção no dia 24 de novembro, que resultou significativa. O modelo estimado foi $(1 - B)(1 - B^{12})Z_t = (1 - \Theta B^{12})(1 - \theta B)a_t + \delta I_{\{24 \text{ de novembro}\}}$. As estimativas e seus erros-padrão estão apresentados na tabela 1.

Tabela 1: Estimativas dos parâmetros do modelo ajustado

Coeficiente	Θ_1	θ_1	λ
Estimativa	$-0,9357$	$-0,7400$	$0,8537$
Erro Padrão	$0,0251$	$0,0539$	$0,1258$

De posse do modelo ajustado foram feitas previsões para as duas semanas seguintes, ou seja, os dez primeiros dias de 2018, que estão mostrados na tabela 2.

3 Conclusões

O modelo *ARIMA Sazonal*, *SARIMA*, se mostrou adequado para previsão. A série é diária e o modelo ajustado tem período 7. A característica da série apresentada dois dias da semana com picos de vendas, sabidamente há promoção nesses dias, quarta-feira e sábado. Destacamos a necessidade de verificação de pontos discrepantes na série. Houve em um dia um ponto atípico que foi encontrado na análise descritiva e esse mesmo ponto foi indicado como *AO* durante a modelagem, ao ser incorporado no modelo o *Erro*

Tabela 2: Previsão para os 10 primeiros dias de 2018, dados transformados.

Dia	01/01	02/01	03/01	04/01	05/01	06/01	07/01	08/01	09/01	10/01
Previsão	4,326	4,408	5,561	4,414	4,567	4,748	4,477	4,352	4,434	5,587
Erro Padrão	0,146	0,146	0,146	0,146	0,147	0,147	0,147	0,154	0,154	0,155

Quadrático Médio dos Resíduos foi diminuído de 0,02134 para 0,01894 ou seja, houve uma diminuição de 11,24% na variância dos resíduos após a inclusão do ponto atípico aditivo *AO* no modelo.

Referencias Bibliográficas

CRYER, J. D.; SUN, K.S. *Time Series Analysis: With Applications in R*. 2.ed. Iowa City: Springer. 2010. 491p.

MORETTIN, P. A.; TOLOI, C. M. C. *Análise de séries temporais*. 2.ed. São Paulo: BLUCHER. 2006. 538p.

R CORE TEAM. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2018. URL <http://www.R-project.org/>.