

Séries temporais para previsão da comercialização: uma aplicação para as laranjas das Centrais de Abastecimento

Palavras-Chave: Cadeia de suprimentos-1, Comercialização-2, Previsão-3

Autores(as):

Jhonny Ricardo Negri, FEAGRI – UNICAMP

Me Thayane Caroline Oliveira da Conceição (coorientadora), FEAGRI – UNICAMP

Prof^(a). Dr^(a). Andréa Leda Ramos de Oliveira (orientadora), FEAGRI - UNICAMP

INTRODUÇÃO:

O Brasil é responsável por 31% da produção mundial de laranja e por 63,4% do volume global de suco de laranja, sendo que mais de 70% do mercado global de suco de laranja é detido pelo país (USDA, 2022). No que se refere à produção de laranja, as maiores regiões produtoras são o Sudeste e o Nordeste, responsáveis pela produção de aproximadamente 398 milhões e 47,6 milhões de caixas de 40,8 quilos do produto *in natura* na safra de 2011 (FNP, 2012). De acordo com a FNP (2012), o maior produtor nacional de laranja é o Estado de São Paulo, que responde por cerca de 80% da produção total do país.

Entre 2016 a 2020, houve uma redução anual de 2,4% na área cultivada com laranjas no Brasil, mas a produção não diminuiu na mesma proporção, apenas 0,3% ao ano, devido a melhorias contínuas na produtividade da cultura (VIDAL, 2022). Entre os principais fatores que afetam negativamente os indicadores da produção de laranja estão as condições climáticas adversas, como falta de chuvas e geadas; a necessidade de erradicar plantas afetadas pela doença greening e a bionalidade da safra (VIDAL, 2022; Fundecitros, 2022).

No Brasil, a distribuição e comercialização de alimentos é amplamente realizada pelas Centrais de Abastecimento (CEASAs), que são responsáveis pela transação dos maiores volumes de frutas, legumes e verduras (FLV) (CUNHA; CAMPOS, 2008). A rede de CEASAs é descentralizada, composta por cerca de 40 unidades administrativas, 53 unidades comerciais principais e outras unidades de menor porte, assumindo um papel fundamental no suprimento alimentar da população urbana do Brasil (CUNHA; CAMPOS, 2008).

As CEASAs são agentes operacionais importantes nos sistemas de agricultura de governos locais, nucleando, ainda que de forma assistemática, iniciativas pertinentes às políticas agrícolas estaduais, bem como políticas de segurança alimentar. Uma vantagem não explorada neste sentido reside no fato de que as principais centrais coletam sistematicamente dados estatísticos sobre quantidade comercializada e origem dos produtos, o que permite identificar os municípios e microrregiões ofertantes, categorizando sua relevância em quantidade de oferta, a diversificação ou especialização de sua pauta comercial e a sazonalidade dessa oferta (CUNHA, 2015).

Dada a importância da laranja no mercado de alimentos brasileiro, a previsão de sua oferta pode auxiliar governos, produtores e distribuidores a planejar melhor suas estratégias de abastecimento, negociação de preços e ações de segurança alimentar. A disponibilidade de dados históricos de comercialização e origem das laranjas nas CEASAs também pode permitir análises mais detalhadas sobre o comportamento desse mercado em diferentes regiões do país e ao longo do tempo.

Um projeto de modelos de séries temporais para laranjas nas CEASAs pode contribuir significativamente para a compreensão do comportamento desse mercado, fornecendo previsões precisas que auxiliam na tomada de decisões informadas sobre a oferta e demanda de laranjas no Brasil. Esse projeto visa desenvolver um modelo para prever a quantidade de laranjas comercializadas ao longo do tempo, além de descrever o comportamento da série temporal, identificar padrões sazonais, tendências e efeitos de eventos adversos.

METODOLOGIA:

A laranja é uma fruta sazonal, com períodos de safra e entressafra bem definidos, o que a torna um excelente objeto de estudo para a análise de padrões sazonais em séries temporais. Dessa forma, a escolha da fruta laranja como objeto de estudo em um projeto de série temporal pode trazer insights valiosos sobre o comportamento e as tendências dessa *commodity* tão importante para a economia global. A construção do banco de dados se deu pelo site da Companhia Nacional de Abastecimento (Conab) - Programa de Modernização do Mercado Hortigranjeiro (Prohort).

O banco de dados do CEASA São Paulo foi escolhido como base para essa análise por dois fatores: é a central com maior participação em quantidade de quilogramas acumuladas desde 2019, cerca de 35% do total acumulado no território nacional (CONAB,2024) e possui dados históricos completos disponíveis. Os dados foram então classificados e divididos em dois grupos: Amostra de construção com 48 observações (período de 01/2019 a 12/2022), utilizado na construção dos modelos de previsão; e Amostra de validação com 14 observações (01/2023 a 02/2024), como amostras de testes para a capacidade preditiva desses modelos.

Para atingir o objetivo do trabalho, foram utilizados diferentes modelos de série temporal para analisar os dados, tais como o modelo ARIMA (Autoregressive Integrated Moving Average), modelo SARIMA (Seasonal Autoregressive Integrated Moving Average), modelo de suavização exponencial, Holt e Holt-Winters (MORETTIN; TOLOI, 2004), cada modelo possui características específicas que se

adequam melhor a determinados tipos de dados, por isso é importante avaliar diferentes abordagens para escolher o modelo mais adequado para os dados em questão. Além disso, será realizada a análise de diferentes métricas de desempenho dos modelos, como o Erro Médio Absoluto (MASE), Erro Quadrático Médio (EQM) e Erro Médio Absoluto (MAD), para comparar a precisão de cada modelo e selecionar aquele que melhor se ajusta aos dados (KOEHLER; HYNDMAN, 2006).

RESULTADOS E DISCUSSÃO:

A distribuição normal do dados foi verificada pelo histograma dos dados, Q-Q plot dos resíduos e pelos testes *Shapiro-Wilk* ($W = 0.98793$, $p\text{-value} = 0.8034$) e *Kolmogorov-Smirnov* ($D = 0.058102$, $p\text{-value} = 0.97$), não foi rejeitada a hipótese nula de que os dados são normalmente distribuído. Na decomposição da série temporal é observado a tendência de uma queda acentuada desde o início da série até o final de 2020, seguida de um aumento até meados de 2022 e, em seguida, uma nova queda, indicando que a série tem uma tendência de declínio inicial, seguida de um aumento, e depois novamente um declínio. A sazonalidade apresenta oscilações indicando que há variações sazonais consistentes na série, com picos e vales que ocorrem em intervalos semelhantes a cada ano. Os resíduos mostram variações aleatórias, indicando a presença de ruídos.

Para a modelagem do ARIMA foi verificado a estacionariedade, autocorrelação (AC) para identificação dos números de termos $MA(q)$ e autocorrelação parcial (PACF) com o número de termos $AR(p)$ para, então, seleção dos parâmetros do modelo: ARIMA (1,0,0). SARIMA pressupõe as mesmas verificações e sazonalidade multiplicativa sendo representado na forma ARIMA(0, 1, 1, 0, 1, 1)¹². A modelagem dos parâmetros de suavização exponencial, Holt e Holt-Winter (α - Parâmetro de suavização para o nível; β - Parâmetro de suavização para a tendência; γ - Parâmetro de suavização para a componente sazonal) foram feitas com métodos iterativos a partir de um chute inicial. Os valores adequados foram: SES(0.1188), Holt (0.445 e 0,99); Holt-Winters (0,99; 0,01 e 0,554).

A análise de diferentes métricas de desempenho teve como principais resultados o modelo SARIMA com o menor Erro Quadrático Médio (1453320.43) e Erro Médio Absoluto (1207966.60), indicando que se ajustou melhor aos dados de teste do que os outros modelos. O modelo apresentou o menor Erro Percentual Médio Absoluto (5.50%) e Erro Médio Absoluto (0.65) , seguindo por ARIMA (5.96% e 0.69) (Tabela 1).

ARIMA	RMSE	MAE	MAPE	MASE	TheilsU
Training set	1,813E+06	1,445E+06	6,540	0,835	0,040
Test set	1,671E+06	1,281E+06	5,962	0,694	0,037
SARIMA	RMSE	MAE	MAPE	MASE	TheilsU
Training set	1,836E+06	1,462E+06	6,441	0,848	0,042
Test set	1,453E+06	1,208E+06	5,503	0,654	0,032
SES	RMSE	MAE	MAPE	MASE	TheilsU
Training set	1,985E+06	1,604E+06	7,068	0,931	0,045
Test set	2,888E+06	2,597E+06	12,202	1,406	0,062
Holt	RMSE	MAE	MAPE	MASE	TheilsU
Training set	2,143E+06	1,806E+06	8,284	1,048	0,047
Test set	5,447E+06	5,139E+06	23,839	2,782	0,111
Holt-Winters	RMSE	MAE	MAPE	MASE	TheilsU
Training set	1,495E+06	1,170E+06	5,285	0,676	0,033
Test set	1,775E+06	1,521E+06	6,954	0,823	0,039

Tabela 1. Análise de métricas de desempenho com amostragem de construção e validação

Os modelos de previsão SARIMA e Holt -Winters para o período de (01/2023 a 02/2024) apresentaram ajustes mais próximos dos dados reais do período conforme as figuras 2 e 3.

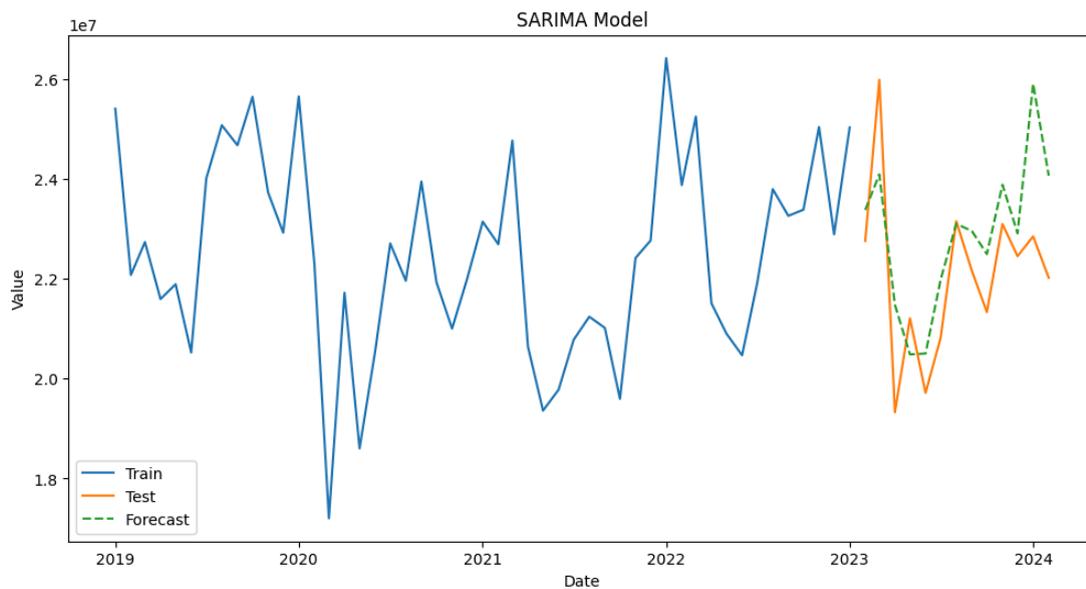


Figura 2. Modelo SARIMA previsão (Forecast) e validação dos dados (test)

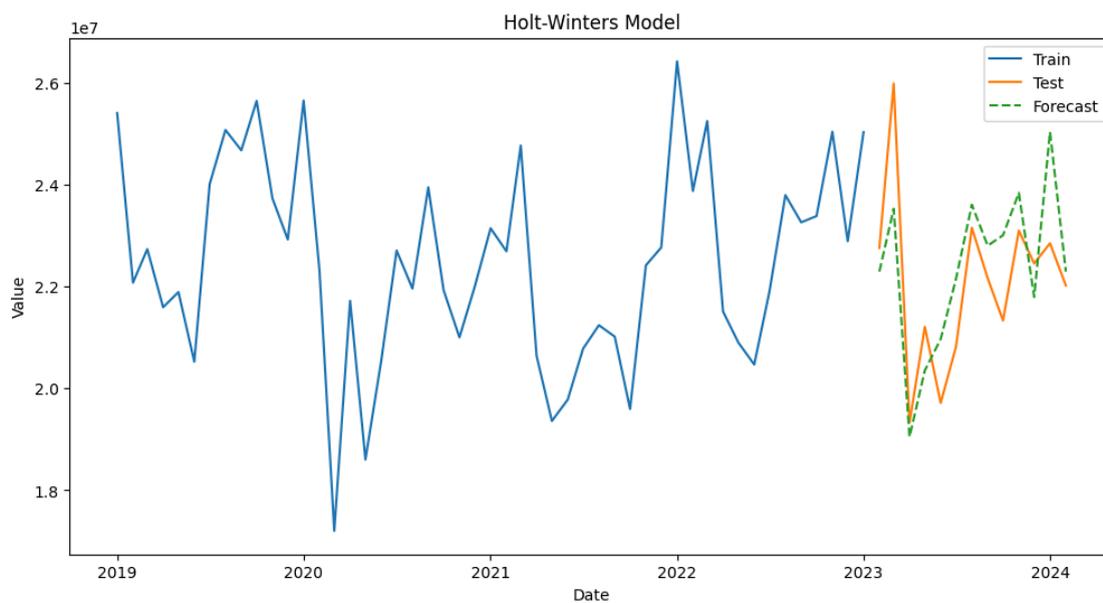


Figura 3. Modelo Holt-Winters previsão (Forecast) e validação dos dados (test)

CONCLUSÕES:

O modelo SARIMA foi o melhor em termos de precisão (menor RMSE e MAE), menores erros percentuais (MAPE) e melhor desempenho preditivo (*Theil's U*). O modelo ARIMA também apresentou um bom desempenho geral, ficando logo atrás do SARIMA. SES e *Holt-Winters* mostraram bons desempenhos em algumas métricas, com destaque para a previsão de *Holt-Winters*, mas não tão consistentes quanto SARIMA e ARIMA. *Holt* apresentou baixa adequação na maioria das métricas, indicando que não se ajustou bem aos dados desta série temporal.

BIBLIOGRAFIA

CONAB. **Programa de Modernização do Mercado Hortigranjeiro (Prohort)**. 2024. Disponível em: <https://www.conab.gov.br/prohort>. Acesso em: 3 mar. 2024.

ADAMI, A.C.O. **Risco e retorno de investimento em citros no Brasil**. 2010. 150 f. Tese (Doutorado) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2010

BOX, G. E. P; JENKINS, G. M.; REINSEL, G. C. **Time Series Analysis: Forecasting and Control**. 3 ed. New Jersey: Prentice Hall, 1994.

CARVALHO Jr., J. G.; COSTA Jr., C. T. **Identification method for fuzzy forecasting models of time series**, Applied Soft Computing. v. 50, p. 166-182, 2017.

CARVALHO, S.A.; GIRARDI, E.A.; MOURÃO FILHO, F.A.A.; FERRAREZI, R.S.; COLETTA FILHO, H.D. **Advances in citrus propagation in Brazil**. Rev. Bras. Frutic.2019, 41, e422.

CITRUSBR (Associação Nacional dos Exportadores de Sucos Cítricos). **A indústria brasileira de suco de laranja**. São Paulo: CitrusBr-APEX, 2014. 34 p. Disponível em:<http://www.citrusbr.com/imgs/biblioteca/CITRUS_APEX_PORTUGUES.pdf>. Acesso em: 20 ago. 2023.

HYNDMAN, R. J.; KHANDAKAR, Y. **Automatic time series forecasting: The forecast package for R**. Journal of Statistical Software, v.27, n.3, p.122, 2008.

KOEHLER, A. B.;HYNDMAN R. J. **Another look at measures of forecast accuracy**, Int. 2006.

MOREIRA, A. S. et al. **Expansão do cinturão citrícola: quais as aptidões e os riscos climáticos?**. Cruz das Almas, BA: Embrapa Mandioca e Fruticultura, 2023. 38 p. il. Color.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**. Edgard Blücher, São Paulo, 2004.

NEGRI, J. A.; KUBOTA, L. C. **Estrutura e Dinâmica do Setor de Serviços no Brasil**. Brasília: IPEA – Instituto de Pesquisa Econômica Aplicada –, 2006.

NEVES, M. F.; JANK, M. S. **Perspectivas da cadeia produtiva da laranja no Brasil: a agenda 2015**. Relatório/Ícone/Markestra/pensa, São Paulo, 2006. Disponível: <http://www.fundace.org.br/arquivos_diversos/agenda_estrategica/Agenda_Citrus_2015_PENSAICON E.pdf>. Acesso em: 20 ago 2023

NEVES, M. F., KALAK, R. B., TROMBIN, V. B. **O retrato da citricultura brasileira**. São Paulo: Citrus, 2010. 137 p.

R CORE TEAM.**R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2023.Disponível em: <<http://www.R-project.org>>. Acesso em: 25 out 2023..