

Revisão Bibliográfica

Aplicações do modelo ARIMA para estatística na análise de dados em série temporal

Resumo

A aplicação de métodos estatísticos, pode aumentar a eficiência de processos, cotidianamente. Sendo a previsão um prognóstico de eventos futuros, usado para propósitos de planejamento e decisão, aumentando probabilidades ou mesmo assegurando o êxito a partir de sua aplicabilidade. Uma série temporal é um conjunto de observações devidamente ordenadas no tempo, apresentando a crucial dependência entre as observações. Quanto maior esta série, melhores as chances de um ajuste satisfatório da modelagem matemática. Em estudos considerando a disponibilidade de dados em série temporal, os modelos ARIMA podem ser empregados para descrever, interpretar e compreender comportamentos e tendências, reduzindo a imprecisão de previsões. É evidente a expressiva gama de aplicações do modelo ARIMA e a sua variação sazonal-SARIMA, dentro das ciências agrárias exatas e naturais, engenharias, computação e outras.

Palavras-chave: série temporal, avaliação de tendência, sazonalidade.

Applications of the ARIMA model for time series data analysis

Abstract

The application of statistical methods, can increase the efficiency of processes. Being the forecast a prognosis of future events, used for planning and decision purposes, increasing probabilities or even ensuring success from its applicability. A time series is a set of observations properly ordered in time, presenting the crucial dependence between the observations. How much ist he longer this series, the better the chances of a satisfactory fit of the mathematical modeling. In studies considering the availability of data in time series, ARIMA models can be used to describe, interpret and understand behaviors and trends, reducing the imprecision of predictions. It is evident the expressive range of applications of the ARIMA model and its seasonal variation - SARIMA, within the exact and natural agricultural sciences, engineering, computing and others.

Keywords: time series, trend assessment, seasonality.

Aplicaciones del modelo ARIMA para el análisis de datos en series temporales

Resumen

Con la aplicación de métodos estadísticos, se puede aumentar la eficiencia de los procesos. Siendo una base pronóstica para eventos futuros, utilizada para fines de planificación y decisión, aumentando la probabilidad de sucesión o el éxito de su aplicabilidad. Una serie temporal es un conjunto de observaciones que están ordenadas em el tiempo, con una dependencia crucial entre las observaciones. Cuanto mas larga es la serie de datos, mas grandes son las posibilidades de un ajuste satisfactorio del modelado matemático. Em estudos com serie de datos temporales, los modelos ARIMA se pueden configurar para crear, interpretar y comprender comportamientos y tendencias, reduciendo la imprecisión de los pronósticos. Es evidente la gama expresiva de aplicaciones del modelo ARIMA y su variabilidad estacional - SARIMA, dentro de las ciencias agrícolas exactas y naturales, la ingeniería, la informática y otros.

Palabras clave: series temporales, evaluación de tendencias, estacionalidad.

Received at: 12/08/2019

Accepted for publication at: 28/11/2019

1 - Acadêmica do Curso de Matemática Aplicada e Computacional. Universidade Estadual do Centro Oeste-UNICENTRO. Email: elaine.bandura@gmail.com

2 - Programa de Pós Graduação em Agronomia. Universidade Estadual do Centro Oeste- UNICENTRO. Email: janainametoski@gmail.com

3 - Curso de Engenharia Mecânica. Universidade Tecnológica Federal do Paraná - UTFPR. Campus Pato Branco-PR. Email:Guilhermejadoski@gmail.com

4 - Curso de Tecnologia em Sistemas para Internet. Universidade Tecnológica Federal do Paraná - UTFPR. Campus Guarapuava. Email: gilmarfrjunior@gmail.com

Introdução

Ao se considerar modelos em avaliações estatísticas, é importante considerar as duas classes mais abrangentes, que podem ser modelos univariados e multivariados, Faruk (2010) salienta que sendo ambos são possíveis de serem aplicados para estudos com séries temporais. A inserção da modelagem em sistemas e processos relacionados às ciências agrárias, exatas, engenharias, econômicas, dentre outros, advém do surgimento da necessidade de se compreender o comportamento de determinadas variáveis, assim como da busca de possibilidades de previsão de suas tendências futuras a partir da série histórica de dados. Quanto maior esta série, melhores as chances de um ajuste satisfatório da modelagem matemática.

Historicamente, a ciência agrônoma, como um ramo das ciências naturais, teria a atribuição de estudar e contribuir com o desenvolvimento da agricultura. Esta atribuição pressupunha pensar na agricultura e suas relações, buscando soluções e avanços para a atividade, com base no saber científico (Cavallet, 1999). Mais modernamente, a interdisciplinaridade vem aproximando, as ciências agrárias com outras, como as ciências exatas. Neste contexto, Correa (2003), descreve que a estatística é uma parte da Matemática que fornece métodos para a coleta, organização, descrição, análise e interpretação de dados, viabilizando a utilização dos mesmos na tomada de decisões.

Para Krajewski et al. (2012), uma previsão é um prognóstico de eventos futuros, usado para propósitos de planejamento e decisão, aumentando probabilidades ou mesmo assegurando o êxito a partir de sua aplicabilidade. Assim, nota-se que os métodos estatísticos podem auxiliar de diferentes formas, como uma ferramenta de tomada de decisões. Para Battisti (2008), a aplicação de métodos estatísticos, pode aumentar a eficiência de processos, cotidianamente.

Neste sentido, Sadaei et al. (2014) salientam que na análise de séries temporais fundamentada eminentemente em suas observações passadas, os dados estão dispostos sequencialmente e em intervalos de tempo específicos e predispostos, tornando-se assim, em um método quantitativo de previsão, que realiza a projeção de valores futuros de uma variável. Assim, conforme Box et al (1994), para não recorrer a uma teoria subjacente específica para possibilitar sua interpretação, usa-se um modelo econométrico construído para a previsão da série temporal, possibilitando entender as correlações entre os dados analisados.

A aplicação de técnicas estatísticas justifica sua importância na ciência agrônoma e, assim, também no agronegócio. Para Nogueira (2009) ao se considerar que os dados passados são indicativos do que se esperar no futuro, pode-se então postular um modelo matemático que é representativo do processo, sendo matematicamente considerado como estocástico.

Desenvolvimento

As previsões realizadas antes do início do século XX, eram realizadas com a extrapolação de um valor global, de forma que fosse ajustada em função do tempo, conforme pode ser observado em Tubino (2009). Para Antunes e Cardoso (2015) as séries temporais podem ser definidas como sequências de dados quantitativos inerentes a momentos específicos e avaliados a partir da sua distribuição no tempo. Um conjunto qualquer de observações geradas a partir de dados que dependem do tempo, são denominadas de série temporal. Geralmente, a variável é observada em pontos temporais discretos, normalmente equidistantes, e a análise é realizada na descrição do processo ou fenômeno está associada sequência gerada a partir dos dados coletados. E, segundo Gutierrez (2003), apresenta a crucial dependência entre as observações.

Krajewski et al. (2012), referindo-se a comportamentos de demanda, descrevem que há cinco padrões básicos da maioria das séries temporais de demanda: a) Horizontal: a flutuação de dados em torno de uma média constante; b) Tendencial: o aumento ou a redução sistemática na média das séries ao longo do tempo; c) Sazonal: um padrão de aumentos ou reduções na demanda que pode ser repetido, dependendo da hora, do dia, da semana, do mês ou da estação; d) Cíclico: os aumentos ou reduções graduais menos previsíveis na demanda por períodos mais longos de tempo (anos ou décadas) e, por fim, e) Aleatórios, no quais a variação da demanda é imprevisível. Tubino (2009) acrescenta que o princípio base é de que no futuro, com projeção dos dados passados baseadas em séries temporais, o sistema apresentará dado comportamento predito, sem a influência de outras variáveis.

A associação de dados em série temporal não é necessariamente nova. Conforme descrição de Yule (1927) com o desenvolvimento de ferramentas e métodos de análises de uma série temporal, criou-se um modelo, denominado de auto regressivo (AR), no qual o valor previsto depende dos valores passados. Contudo, esse modelo foi considerado limitado, pois

depende da existência de uma relação linear entre os elementos da sequência, e também, da hipótese de que a série é estacionária.

Dessa forma, os modelos lineares mostraram-se insuficientes para a maioria das análises de séries temporais, já que a maioria das séries reais apresentam forte tendências de não linearidade (CHAVES, 1991). Os autores descrevem ainda que a partir da década de 1950, novos estudos de séries temporais ganharam impulso, com o advento da criação do computador.

Nesta tendência da busca de ajustes satisfatórios para séries de dados, Box e Jenkins (1976) demonstram em seu livro, um sistema matemático que tem como base a Teoria Geral de Sistemas Lineares, culminando na criação do modelo de Box & Jenkins (B&J), que teve como base a suposição de que a passagem de um ruído branco por um filtro linear de memória infinita gera um processo estacionário de segunda ordem, conforme descrevem Souza e Camargo (2004), que acrescentam que a aplicação do modelo B&J segue as seguintes etapas: identificação, estimação e verificação, e, sendo o resultado satisfatório, segue-se com a estimação.

Com os dados coletados, de tal forma que foi estipulado no tempo, espera-se que ela apresente correlação seriada no tempo. De forma genérica, os modelos de Box-Jenkins, são conhecidos por *Auto Regressive Integrated Moving Averages* (ARIMA) ou em português por modelos Auto Regressivos Integrados de Médias Móveis, os seus termos correspondem a série estacionária obtida por diferenciação, os erros aleatórios a partir da defasagens, e ao processo de tornar a série em estacionária pelo método de diferenciação. Segundo Werner e Ribeiro (2003), são modelagens que analisam o comportamento da correlação seriada ou auto correlação entre os valores da série temporal, e com base nesse comportamento realizar previsões futuras.

Na linha temporal de estudos de séries temporais, um modelo que vem sendo destacado é o denominado ARIIMA. Segundo Fava (2000), a combinação das três componentes, a) auto-regressivo (AR), b) filtro de integração (I) e c) componente de médias móveis (MA), formam a sigla usualmente descrita como ARIMA.

Os modelos ARIMA são modelos estatísticos lineares para análise de séries temporais (GUTIERREZ, 2003). Segundo Morettin e Tolo (1981), para que possa aplicar o modelo ARIMA, a série deve ser estacionária, ou seja, conter em torno da distribuição constante uma

média móvel, com isso podendo admitir uma certa estabilidade com o decorrer do tempo.

O modelo ARIMA vem apresentando vasta aplicabilidade na área de Ciências Agrárias. Por exemplo, observa-se em Caballos (2015) que o objetivo foi planejar racionalmente a tomada de decisão com base em previsões confiáveis, considerando os preços variáveis econométricos, neste caso, a metodologia Box-Jenkins foi utilizada para aplicar o modelo econométrico ARIMA, para ajustar o comportamento das séries.

FARUK (2010) demonstra a proposição de um modelo que utiliza ARIMA e a rede neural que possui a capacidade de explorar os pontos fortes das abordagens tradicionais de séries temporais e redes neurais artificiais. No processo a abordagem proposta consistiu em uma metodologia ARIMA associada a uma estrutura de rede de retro propagação de *feed-forward*, inserindo um algoritmo de treinamento conjugado otimizado. A partir da modelagem, os dados previstos de qualidade da água do modelo híbrido são comparados com os da metodologia ARIMA e da arquitetura de rede neural usando as medidas de precisão.

Broz e Viego (2014) aplica o modelo ARIMA para a variação dos preços das florestas, considerando que esta é uma das principais fontes de incerteza no planejamento florestal. Neste caso, o modelo auto regressivo médio móvel ARIMA baseado na metodologia de Box-Jenkins, foi proposto para prever os preços de quatro produtos de *Pinus spp.* para o nordeste da Argentina. Em aplicação similar, Martínez e Chalita (2011) demonstram a utilização da metodologia Box-Jenkins para identificar um modelo de média móvel integrada ARIMA, que se ajusta ao comportamento de séries temporais de preços nominais para atacadista de tomate no México.

Um modelo ARIMA de média móvel autoregressiva fracionada sazonal foi descrito por Montanari, et al. (2000), tendo sido introduzido com componentes periódicos persistentes de curto e longo prazo. A estimação dos parâmetros foi realizada aplicando a aproximação de *Whittle* à função de máxima verossimilhança gaussiana, que produz estimativas assintoticamente consistentes. Nesta forma o modelo foi aplicado aos fluxos mensais de vazão do rio Nilo, a fim de detectar se o que considera como memória longa. Han et al. (2010) desenvolveram trabalho com o monitoramento de seca quase em tempo real, utilizando os modelos ARIMA aplicados na série VTCI e preveem suas mudanças no futuro.

Em processos e fenômenos da natureza, alguns exemplos mais destacado envolvem questões climáticas. Para esta análise, o componente sazonal pode aparecer quando são feitas observações a longo prazo para a série de interesse, isto é, os dados são registrados em frações reduzidas de tempo no dia, diariamente, semanalmente, ou mesmo mensalmente, trimestralmente, por exemplo. As aplicações desta ordem adquirem maior relevância quando se considera a possibilidade de decomposição da série temporal em componentes de tendência, ciclo e sazonalidade, conforme descreve Chatfield (2000).

Para este tipo de série, pode ser exemplificado o trabalho de Silva *et al.* (2008) onde são utilizados os modelos de séries temporais no estudo de variáveis climatológicas, como temperatura e precipitação, com objetivo de analisar a série de temperatura média mensal da cidade de Uberlândia, MG, realizando previsões para períodos subsequentes e descrevendo seus componentes.

No caso de Chen *et al.* (2010) é observada a proposição de um modelo estocástico de energia eólica, baseado em um processo de média móvel integrada auto regressiva ARIMA), em que o modelo leva em consideração os limites não estacionários e físicos da geração estocástica de energia eólica.

Considerando que é comum se encontrar séries, que apresentam a componente sazonal. A partir de Box-Jenkins, foram desenvolvidos os modelos SARIMA, os quais incluem a correlação serial dentro e entre os períodos sazonais. O modelo com a estrutura SARIMA segue os mesmos passos que a modelagem ARIMA, usada para séries não sazonais.

Carvalho e Delgado (2019) apresentam o desenvolvimento de estudo com objetivo de modelar o comportamento da série temporal da evapotranspiração de referência, utilizando redes neurais artificiais (RNA) do tipo GMDH (Método de Grupo de Manipulação de Dados) e comparar com o modelo SARIMA (Sazonal Autor regressivo Integrado de Médias Móveis), demonstrando resultado altamente satisfatórios.

Em pesquisa de Chechy e Bayer (2012) é realizada uma análise de séries temporais dos dados de temperatura mínima e temperatura máxima mensal no Rio Grande do Sul. Neste trabalho é realizada uma comparação de duas classes de modelos tradicionais de previsão, modelos do tipo SARIMA e os modelos Holt-Winters aditivo. Os modelos foram comparados para diferentes horizontes de previsão e resultados muito promissores.

Marchezan e Souza (2010) realizaram previsões referentes ao preço das principais culturas das lavouras temporárias desenvolvidas no Estado do Rio Grande do Sul. A utilização do método de B&J foi para realizar previsões do preço das culturas de soja, arroz e milho, no qual o que melhor se adequou foi o modelo ARIMA para a soja e milho e o modelo SARIMA para o arroz. Já Hernández *et al.* (2019) desenvolveram avaliação com objetivo de gerar informações sobre o comportamento do preço da maçã por meio de uma análise de séries temporais, gerando uma previsão para os 12 meses subsequentes com o modelo SARIMA.

Na área de utilização de máquinas, Kim *et al.* (2013) utilizaram o modelo /ARIMA para o ajuste de máquinas utilizadas no cultivo de arroz, descrevendo a importância do uso destes modelo para análise de dados em série temporal para o planejamento de uso adequado de máquinas agrícolas. Assim como Martin *et al.* (2016) descreveu sobre a importância da análise adequada de série temporal para descrever a evolução do número de veículos em circulação em área geográfica definida concluindo que o modelo ARIMA apresentou adequação para este tipo de análise.

Como pode-se notar, muitas técnicas têm sido testadas com o intuito de se efetuar previsões mais precisas e preferencialmente automatizadas. Concordando com Kaboudan (2002) que assinala vários estudos dirigidos a partir da área da Inteligência Artificial (IA) e assim inspirados na natureza, como Redes Neurais Artificiais (RNA), Algoritmos Genéticos (AG) e Algoritmos Evolutivos (AE) e a Programação Genética.

Considerações Finais

Em estudos considerando a disponibilidade de dados em série temporal, os modelos ARIMA podem ser empregados para descrever, interpretar e compreender comportamentos e tendências, reduzindo a imprecisão de previsões. É evidente a expressiva gama de aplicações do modelo ARIMA e a sua variação sazonal - SARIMA, dentro das ciências agrárias exatas e naturais, engenharias, computação e outras. Sendo esta uma ferramenta importante para a descrição de comportamento de séries temporais, com possibilidade, em muitos casos, de realização de previsões futuras da tendência das variáveis analisadas.

Referências

- AGUADO-RODRÍGUEZ, G. Jr et al. Predicción de variables meteorológicas por medio de modelos ARIMA. **Agrociencia**. vol.50, n.1, pp.1-13,2016.
- BATTISTI, I. D. E.; BATTISTI, G. Métodos estatísticos. Ijuí: Ed. Unijuí, Coleção Educação à Distância, 2008
- BOX, G. E. P e JENKINS G. M., **Time series analysis: forecasting and control**, Ed. Rev.(1976). San Francisco: Holden-Day, 1970.
- BOX, G.E.P; et al. **Time series analysis: forecasting and control**. 3. ed. New Jersey: Prentice Hall, 1994
- BROZ, D. R.; VIEGO, V. N. Predicción de precios de productos de Pinus spp. con modelos ARIMA. **Maderas y Bosques**, v. 20, n. 1, p. 37-46, 2014.
- CARVALHO, R. L. da S., DELGADO, A. R. S.. Estimates of reference evapotranspiration in the municipality of Ariquemes (RO) using neural networks GMDH-type. **Rev. bras. eng. agríc. ambient.** [online]. 2019, vol.23, n.5, pp.324-329. Epub May 06, 2019. ISSN 1415-4366.
- CAVALLET, V. J. A formação do engeheiro agrônomo em questão: A expectativa de um profissional que atenda as demandas do século XXI. 1999.
- CEBALLOS, S., PIRE, R.. Estimación del precio internacional del arroz (*Oryza sativa* L.) bajo el modelo ARIMA. **Revista Mexicana de Ciencias Agrícolas** Pub. Esp. 11: 2083-2089. 2015
- CHATFIELD, C. (2000). **Time-series forecasting**, CRC Press.
- CHAVES, A. N., **Bootstrap em Séries Temporais**. Rio de Janeiro. 208f. Tese (Doutorado em Engenharia Elétrica), Universidade Federal do Rio de Janeiro, 1991.
- CHECHI, L. BAYER, F. M.. Modelos univariados de séries temporais para previsão das temperaturas médias mensais de Erechim, RS. **Rev. bras. eng. agríc. ambient.**, vol.16, n.12, pp.1321-1329. ISSN 1415-4366.2012
- CHEN, P. PEDERSEN, T. BAK-JENSEN B. E CHEN, Z. "Modelo de Série Temporal Baseado em ARIMA para Geração Estocástica de Energia Eólica", em **IEEE Transactions on Power Systems** , vol. 25, n. 2, pp. 667-676, maio de 2010.
- CORREA, S. M. Probabilidade e estatística, 2ª edição. - Belo Horizonte: **PUC Minas Virtual**, 2003.
- FARUK, D.Ö. A hybrid neural network and ARIMA model for water quality time series prediction **Engineering Applications of Artificial Intelligence**, 23 (4), pp. 586-594, 2010
- FAVA, V. Manual de econometria. **In: VASCONCELOS, M. A. S.; D. São Paulo: Editora Atlas, 2000.**
- GUTIERREZ, J. L. C. Monitoring of the Corumbá - i Dam Instrumentation Neural Networks and the Box & Jenkins null Models, 2003.
- HAN P, WANG PX, ZANG SY, DE HAI Z . Drought forecasting based on the remote sensing data using ARIMA models. **Math Comput Model** 51:1398-1403, 2010
- HERNÁNDEZ, J.A., PUENTE, G. B., GÓMEZ, A.A.G, Análisis del precio de la manzana mediante un modelo SARIMA, **Rev. Mex. Cienc. Agríc** vol.10 no.2 Texcoco feb./mar. 2019
- KABOUDAN, M., TSGP - Time Series Genetic Programming. **School of Business**, university of Redlands, 2002.
- Kim, B.; Shin, S. Y.; Kim, W. W.; Yum, S.; Kim, J. Forecasting Demand of Agricultural Tractor, Riding Type Rice Transplanter and Combine Harvester by using an ARIMA Model. **Journal of Biosystems Engineering** , v. 38, n. 1, p. 9-17, 2013. Disponível em: <https://www.e-sciencecentral.org/articles/SC000013522>. Acesso em 20/09/2019.
- KRAJEWSKI, L; et al. Administração de Produção e Operações. 8º ed. Pearson, 2009.
- MARCHEZAN, A. SOUZA, A. M. Previsão do preço dos principais grãos produzidos no Rio Grande do Sul. **Cienc. Rural** . 2010, vol.40, n.11, pp.2368-2374. Epub Dec 03, 2010. ISSN 0103-8478.
- MARTIN, A. C.; HENNING, E.; WALTER, O. M. F. C.; KONRATH, A. C. Análise de séries temporais para previsão da evolução do número de automóveis no Município de Joinville. **Revista Espacios**, v.37, n.6, 2016. Disponível em: https://qualimetria.ufsc.br/files/2016/05/Revista-ESPACIOS_-Vol.pdf. Acesso em: 22/09/2019.

Bandura et al. (2019)

- MARTÍNEZ, G. M. CHALITA, L. E. T. Aplicación de la metodología Box-Jenkins para pronóstico de precios en jitomate. **Rev. Mex. Cienc. Agric.** 2(4):573-577. Nelson, J. P. 1995.
- MONTANARI, A., ROSSO, R., TAQQU, M. S.: A seasonal fractional ARIMA model applied to the Nile River monthly flows at Aswan, **Water Resour. Res.**, 36, 1249-1259, 2000.
- MORETTIN, P. A. TOLOI, Clélia Maria de Castro. Previsão de series temporais, Rio de Janeiro – RJ, 1981.
- RODRIGUES, R. O céu é o limite para o agronegócio brasileiro. **Conjuntura Econômica**, Rio de Janeiro, V.60, n.11, p.14-15, Nov.2006.
- SADAEI, H. J., ENAYATIFAR, R., ABDULLAH, A. H., GANI, A. Short-term load forecasting using a hybrid model with a refined exponentially weighted fuzzy time series and an improved harmony search, **International Journal of Electrical Power & Energy Systems** 62: 118-129. 2014
- SILVA, M.I.S., GUIMARÃES, E.C. & TAVARES, M. Previsão da temperatura média mensal de Uberlândia, MG, com modelos de séries temporais. **Revista Brasileira de Engenharia Agrícola e Ambiental** 12:480-485. 2008
- SOUZA, R. C., CAMARGO, M. E., Análise e Previsão de Séries Temporais : **Os Modelos ARIMA**. 2ª Ed., 2004.
- TUBINO, D.F. Planejamento e controle da produção-TeoriaePrática.SãoPaulo: 2ª. ed. **Atlas**,2009
- WERNER, L., RIBEIRO, J. L. D. Previsão de demanda: uma aplicação dos modelos Box-Jenkins na área de assistência técnica de computadores pessoais. **Gest. Prod.**, vol.10, n.1, pp.47-67. ISSN 0104-530X.. 2003
- YULE, G. U. On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**. 226 (636-646): 267-298. 1927
- ZENG, Q.; et al. A new approach for Baltic Dry Index forecasting based on empirical mode decomposition and neural networks. **Maritime Economics & Logistics**, v.18, n. 2, p. 192-210, 2015.