



UFC

**UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS AGRÁRIAS
DEPARTAMENTO DE ECONOMIA AGRÍCOLA
CURSO DE AGRONOMIA**

Francisco Libério César Feitosa de Araújo

**PREVISÃO DE PREÇO DO MILHO COM A APLICAÇÃO DE UM MODELO
AUTORREGRESSIVO**

FORTALEZA

2023

Francisco Libério César Feitosa de Araújo

**PREVISÃO DE PREÇO DO MILHO COM A APLICAÇÃO DE UM MODELO
AUTORREGRESSIVO**

Monografia submetida ao curso de Graduação em Agronomia, do Centro de Ciências Agrárias da Universidade Federal do Ceará, como requisito parcial para obtenção do título de Engenheiro(a) Agrônomo(a).

Orientador: Prof. Dr. Vitor Hugo Miro Couto Silva.

FORTALEZA

2023

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

A689p Araújo, Francisco Libério César Feitosa de.
Previsão de preço do milho com a aplicação de um modelo autorregressivo / Francisco Libério César Feitosa de Araújo. – 2023.
30 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Ciências Agrárias, Curso de Agronomia, Fortaleza, 2023.
Orientação: Prof. Dr. Vitor Hugo Miro Couto Silva.

1. Preço. 2. Série temporal. 3. Milho. I. Título.

CDD 630

FRANCISCO LIBÉRIO CÉSAR FEITOSA DE ARAÚJO

PREVISÃO DE PREÇO DO MILHO COM A APLICAÇÃO DE UM MODELO
AUTORREGRESSIVO

Monografia submetida ao curso de Graduação em Agronomia, do Centro de Ciências Agrárias da Universidade Federal do Ceará, como requisito parcial para obtenção do título de Engenheiro(a) Agrônomo(a).

Aprovada em: 28/06/2023.

BANCA EXAMINADORA

Prof. Dr. Vitor Hugo Miro Couto Silva (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. José De Jesus Sousa Lemos
Universidade Federal do Ceará (UFC)

Prof. Dr. Francisco José Silva Tabosa
Universidade Federal do Ceará (UFC)

Dedico este trabalho a Deus e aos meus pais,
Conceição(in memoriam) e Libério

AGRADECIMENTOS

A Deus por toda força e fé que me ajuda a seguir minha caminhada com determinação inabalável, apesar das adversidades às quais são características da vida.

À minha amada mãe, Maria da Conceição Alves Feitosa De Araújo (*in memoriam*), que me deu todo amor se dedicando em minha educação e formação, transmitindo seu interesse pela ciência e na vida acadêmica como veterinária, docente e pesquisadora, sendo minha eterna professora e inspiração.

Ao meu pai, Francisco Libério Bezerra De Araújo por todo apoio e confiança nos momentos mais difíceis sempre junto de mim sendo eternamente não só um professor pessoal como docente ao qual ele foi em sua vida profissional mas como um mentor e inspiração para mim.

À minha querida e afável irmã, Maria Jesita Juliana Feitosa de Araújo, por todo carinho, amor, dedicação e sacrifícios que ela já fez por mim, sempre junto a mim, ouvindo sempre meus temores e tristezas e me apoiando fazendo sempre eu acreditar que posso mais.

A meus avós, Lídio Bezerra (*in memoriam*), por todo amor, gentileza e carinho, juntamente com sua biografia a qual dedicou parte significativa de sua vida à agricultura e pecuária no estado do Ceará, sendo parte importante no meu interesse na área de Agronomia e Maria Araújo (*in memoriam*), por toda amor gentileza e sentimentos aos quais me transmitiu.

A meus avós, Francisco Das Chagas Feitosa, pelos bons momentos e apoio e Maria Marilene Alves (*in memoriam*), por todo apoio, carinho e amor e sentimentos transmitidos em vida.

A minhas tias, Maria Marilan e Maria Merylan por terem sido boas amigas e me dado apoio e motivação nos momentos complicados.

A meus tios Alberto Jorge (*in memoriam*) e Antônia Araújo (*in memoriam*), pelos bons momentos vividos, carinho e felicidade que transmitia.

A meu tio José Araújo pelo apoio na minha formação na época escolar e pela motivação transmitida.

Ao Prof. Dr. Vitor Hugo Miro Couto Silva, pelo período como professor e orientação ajudando a desenvolver este trabalho.

Ao Prof. PHD José de Jesus Sousa Lemos, por ter me aceito e me supervisionado no Laboratório do Semiárido (LabSar), neste período final de curso, numa experiência bastante

enriquecedora e também ter aceito participar da banca dando sua contribuição para finalização deste trabalho

Ao Prof. Dr. Francisco José Silva Tabosa por ter aceito participar da banca e dado sua contribuição para finalização deste trabalho.

Aos Professores da Universidade Federal do Ceará que me ensinaram e inspiraram demonstrando grande dedicação na docência, produção científica ajudando no desenvolvimento não só da Instituição mas da sociedade também.

Ao Ruan que me ajudou muito durante o curso sempre solícito e dando apoio em desafios além da própria Graduação.

Ao valoroso amigo Lucas Levi, de tanto tempo sempre transmitindo boas vibrações e sempre me apoiando.

Aos meus amigos da UFC que me ajudaram durante toda essa caminhada na graduação sendo bons companheiros: Matheus, Lucas, João Victor, Victor, Maquimahom, Rodrigo, Cássio, Josias, Iana, Bianca, Américo, Raul, Jonas, Davi, Marcelle, Emanuel, Lucas Rondon, Marcela, Cláudio, Italo, Lenildo, Jorge, Gabriel, Igor e outros não citados.

Aos meus amigos de vida que estão junto a mim a tanto tempo em momentos difíceis e alegres: Davi, Arthur, João Victor, João Felipe, Beatriz, Matheus, Valdecírio, Bruno Henrique, Rebeca, Elizabeth, Sônia, Regina e outros não citados.

A todos os trabalhadores que contribuem para a construção e fortificação de um ensino público libertador e edificante.

A todos os funcionários da UFC pelos serviços prestados.

RESUMO

O presente trabalho analisa e projeta o preço do milho no Brasil, que é um dos maiores produtores mundiais de matérias primas agrícolas e cujo uma parte importante do PIB advém da agricultura. Ademais, a produção brasileira de milho é ampla e possui uma diversificada cadeia produtiva, incluindo, alimentação, produção de ração e etanol. Outrossim, o estudo foi elaborado com séries temporais para analisar e projetar os preços do milho para ajudar os produtores a maximizar sua eficiência tendo mais dados para melhorar seus planejamentos de comercialização, produção e aumentar a eficiência na rentabilidade da escolha de como usar seu capital. A análise de séries temporais é importante para compreender as tendências. Padrões sazonais ou cíclicos e flutuações aleatórias nas variáveis exemplo dos preços do milho. Existem vários métodos análise das séries temporais no trabalho se utilizou o teste de Dickey Fuller para verificar a viabilidade da série ser utilizada para efetuar uma projeção de preços sendo necessário que ela fosse estacionária, uma série estacionária é uma série temporal que tem uma média, variância e autocorrelação constantes ao longo do tempo. Dessa forma, os valores são estatisticamente consistentes e mudam em torno de uma média constante, caso contrário para projeção ser eficiente e necessário se faça transformações nas variáveis tornando a série estacionária, modelo ARIMA (Autoregressive Integrated Moving Average) foi utilizado para se efetuar as previsões se separando uma % dos dados para treino e outra para teste. Portanto, tendo como objetivo do trabalho a aplicação técnicas de estimação e previsões de séries temporais para analisar séries de preço da saca de milho no Brasil. Se demonstrou factível e viável e execução de projeções com modelos autorregressivos no preço do milho seguindo os devidos protocolos, outrossim, se conclui que analisar as séries temporais de preços permite uma compreensão melhor do comportamento dos preços e a mitigação de riscos.

Palavras-chave: preço 1; milho 2; série temporal 3.

ABSTRACT

The present work analyzes and projects the price of corn in Brazil, which is one of the world's largest producers of agricultural commodities and whose agriculture contributes significantly to the GDP. Furthermore, Brazilian corn production is extensive and has a diversified production chain, including feed, feed production, and ethanol. Moreover, the study was developed using time series to analyze and project corn prices to help producers maximize their efficiency by having more data to improve their marketing and production planning and increase efficiency in the profitability of capital allocation. The analysis of time series is important to understand trends, seasonal and cyclical patterns, and random fluctuations in variables, such as corn prices. There are several methods for analyzing time series, and in this work, the Dickey Fuller test was used to verify the viability of using the series to make price projections. It was necessary for the series to be stationary, which means having constant mean, variance, and autocorrelation over time. In this way, the values are statistically consistent and change around a constant mean. Otherwise, transformations in the variables are required to make the series stationary. The ARIMA (Autoregressive Integrated Moving Average) model was used to make predictions, with a portion of the data separated for training and another for testing. Therefore, the objective of this work is to apply techniques for estimation and forecasting of time series to analyze the price series of corn in Brazil. It has been demonstrated that executing projections with autoregressive models in corn prices is feasible and viable, following the appropriate protocols. Furthermore, it is concluded that analyzing time series of prices allows for a better understanding of price behavior and risk mitigation.

Keywords: price 1; corn 2; time series 3.

LISTA DE FIGURAS

Figura 1- Séries de preço do milho (R\$/saca), média mensal (jan/2013 – dez/2022). Fonte: CEPEA.....	22
Figura 2- Séries de preço do milho (R\$/saca): dados de treino e teste.....	23
Figura 3- Teste Dickey Fuller sem transformação na série.....	23
Figura 4- 5 primeiros dados em primeira diferença.....	24
Figura 5- Séries de preço do milho (R\$/saca), média mensal (jan/2013 – dez/2022). Fonte: CEPEA.....	25
Figura 6- Séries de preço do milho (R\$/saca) com destaque para divisão dos dados em conjunto de dados de treino e dados de teste.Fonte: Resultados da pesquisa.....	25
Figura 7- Funções de autocorrelação e autocorrelação parcial (dados de treino). Fonte: Resultados da pesquisa.....	26
Figura 8- Série de preços observados e preços previstos (dados de treino). Fonte: Resultados da pesquisa.....	28
Figura 9- Série de preços observados e preços previstos (dados de teste). Fonte: Resultados da pesquisa.....	28

LISTA DE GRÁFICOS

Gráfico 1- Séries de preço do milho (R\$/saca), média mensal (jan/2013– dez/2022). Fonte: CEPEA	24
Gráfico 2- Séries de preço do milho (R\$/saca): dados de treino e teste	25
Gráfico 3- Séries de preço do milho (R\$/saca) com destaque para divisão dos dados em conjunto de dados de treino e dados de teste	27
Gráfico 4- Funções de autocorrelação e autocorrelação parcial (dados de treino).....	29
Gráfico 5- Série de preços observados e preços previstos (dados de teste). Fonte: Resultados da pesquisa	30
Gráfico 6- Série de preços observados e preços previstos (dados de teste). Fonte: Resultados da pesquisa	30

LISTA DE TABELAS

Tabela 1- Estatísticas descritivas da série de preços (dados de treino).....	27
Tabela 2- Resultados do teste de Dickey Fuller (dados de treino)	28

LISTA DE ABREVIATURAS E SIGLAS

IBGE	Instituto Brasileiro de Geografia e Estatística
CONAB	Companhia Nacional de Abastecimento
CEPEA	Centro de Estudos Avançados em Economia Aplicada
ANEC	Associação Nacional dos Exportadores de Cereais
AR	Autorregressivo
USP	Universidade de São Paulo
ARIMA	Auto Regressive Integrated Moving Average
ADF	Augmented Dickey-Fuller
DEA	Data Envelopment Analysis
IA	Inteligência artificial

LISTA DE SÍMBOLOS

\$	Dólar
%	Porcentagem
R\$	Reais

SUMÁRIO

1	INTRODUÇÃO	16
2	REFERENCIAL TEÓRICO	17
2.1	Análise de dados e agricultura	17
2.2	<i>O agronegócio e o milho</i>	19
2.3	As séries temporais	19
2.4	Projeções e modelos autorregressivos	21
2.5	Séries estacionárias e teste Dickey Fuller	22
3	METODOLOGIA	23
3.1	Série Temporal	24
3.2	Treino e Teste	24
3.3	Estacionariedade e teste de raiz unitária e ajuste do modelo	25
4	RESULTADOS E DISCUSSÕES	26
4.1	A Série do milho	27
4.2	Treino e teste do modelo	27
4.3	Ajuste do modelo	28
5	CONSIDERAÇÕES FINAIS	31
	REFERÊNCIAS	32

1 INTRODUÇÃO

O PIB brasileiro possui dentre os seus componentes uma parte significativa oriunda do agronegócio, segundo projeções do CEPEA em 2022 foi cerca de 25% do produto interno bruto teve participação do agronegócio sendo o Brasil global trader no setor com notório destaque na produção de commodities agrícolas, como no caso do milho ao qual é o foco do trabalho presente, compreende-se que é estratégico e necessário que se efetue estudos na área acerca dos dados aos quais são apurados no decorrer dos anos para que possa se traçar estratégias cada vez mais precisas auxiliando que produtores consigam maximizar sua eficiência de mercado tendo mais competitividade com concorrentes, assim sendo o que irá pautar esta pesquisa é análise juntamente com projeções referentes ao preço do milho, por intermédio de séries temporais.

A produção de milho, no Brasil, é efetuada em alta escala e com uma grande competitividade internacional, trazendo a posição de terceiro maior produtor do mundo segundo levantamento da Statista. Trata-se de uma cadeia produtiva bastante ampla e com diferentes segmentos, desde produção de ração para alimentação de animais como a produção de etanol, com maior valor agregado e rentabilidade.

Assim, tratando-se de um mercado de suma importância que envolve interesses do setor público e da iniciativa privada, o estudo acerca do comportamento de preços é de grande importância, podendo-se aplicar técnicas de séries temporais na análise da variável com o objetivo de gerar previsões e auxiliando na elaboração de projetos agropecuários. A análise de séries temporais de preço constitui, assim, uma ferramenta importante para o planejamento da produção, além de poder ser muito útil em análises financeiras.

O conceito de séries temporais, bem como o conjunto de ferramentas deste campo de estudo, é muito útil, com aplicações em diferentes áreas como agricultura, finanças e meteorologia. O estudo de séries temporais é importante para a compreensão de tendências, padrões sazonais, ciclos e flutuações aleatórias. Neste campo existem vários métodos, entre as funções e utilidades mais recorrentes estão a modelagem da série temporal juntamente com técnicas para realizar previsões.

Ademais, pode-se efetuar análises com a intenção de identificar pontos de mudança ou anomalias nos dados. Por exemplo, no caso do preço do milho, pode-se analisar como a pandemia de COVID-19 teve influência na demanda e na oferta por esta commodities agrícola, acarretando mudanças nos preços ao longo do tempo. Esse tipo de análise é importante para entender o impacto de eventos externos na economia e na produção agropecuária, possibilitando

a elaboração de estratégias mais adequadas para lidar com a possibilidade de crises. Outrossim, para a análise de séries temporais de preços do milho, pode ser utilizada uma ampla gama de ferramentas e métodos, dos quais se destacam os modelos derivados da modelo ARIMA (Autoregressive Integrated Moving Average), o modelo de suavização exponencial, entre outros. Cada uma das técnicas existentes possui vantagens e limitações, e a escolha da mais adequada para cada situação depende dos dados e do objetivo da análise.

Tratando-se em específico da análise dos preços do milho, objeto deste trabalho, é importante salientar as diversas variáveis que podem interferir no mercado de commodities agrícolas, relacionados com a oferta e a demanda, como a política econômica e agrícola do governo, as condições climáticas, o custo dos insumos, que muitas vezes importados e relacionados intrinsecamente ao valor do câmbio, além de outros fatores. Desta forma, a análise de séries temporais deve ser aplicada de forma atrelada a outras ferramentas e técnicas de análise, considerando todas as variáveis relevantes para a tomada de decisões estratégicas.

Considerando a justificativa apresentada acima, esta monografia busca analisar o comportamento do preço médio da saca de milho no Brasil durante o período de uma década, entre 2012 e 2013, bem como realizar projeções com a aplicação e ajuste de um modelo autorregressivo.

2. REFERENCIAL TEÓRICO

2.1. Análise de dados e agricultura

Segundo Gomes e Mangabeira (2004), o uso de Análise de Envoltória de Dados (DEA) para medir a eficiência relativa de agricultores pode servir como apoio à decisão na produção agrícola, ao indicar as fontes de ineficiência e as unidades que podem servir de referência às práticas adotadas. O emprego de modelos DEA, do inglês (Data Envelopment Analysis) em agricultura pode apoiar as decisões dos agricultores, ao indicar as fontes de ineficiência e as unidades que podem servir de referência às práticas adotadas (benchmarks).

A análise de dados é um processo que pode extrair informações valiosas de informações coletadas de várias fontes. Esse conhecimento pode ajudar a impulsionar a tomada de decisões, melhorar o processo, aprimorar a experiência do cliente e desenvolver novos produtos. Por isso, a análise de dados é crucial para o desenho de políticas públicas efetivas, especialmente em regiões exemplo de semiáridas que possuem uma oscilação característica do clima.

As vulnerabilidades econômicas que decorrem da geração e da apropriação desigual da renda estão associadas a um conjunto de sinergias. Em geral, se atribui à instabilidade climática as causas das dificuldades observadas na maioria dos municípios do Nordeste, sobretudo aqueles posicionados no Semiárido (reconhecido ou não pelo governo federal). (LEMOS, 2019, p. 14).

O clima tem participação substancial ao desenvolvimento das regiões do semiárido brasileiro, em decorrência, da atividade econômica das cidades muitas vezes serem oriundas de agricultura em muitos casos de subsistência. Os dados são importantes fontes de informação que permitem análises, estudos e até mesmo o desenvolvimento de políticas públicas baseadas em evidências que possam garantir melhor acesso da população aos seus direitos. A análise de políticas públicas pode contribuir para os estudos da ciência da informação sobre políticas de informação pública. Isso requer uma compreensão do papel do Estado como produtor e acumulador de informações e suas implicações para a sociedade ganha cada vez mais importância e é aplicada cada vez mais em áreas estratégicas das mais diversas áreas possíveis, devido à alta competitividade e eficiência almejada. Um exemplo é a agricultura que usa análise de dados para otimizar o cultivo e comercialização de milho.

A agricultura é um importante atividade econômica no Brasil, respondendo por 24,8 % do produto interno bruto (PIB) do país. A agricultura brasileira se desenvolve sob uma estrutura agrária concentrada e desigual, praticada de diferentes formas: tradicional, moderna, familiar, empresarial e orgânica. A modernização da agricultura brasileira resultou na mecanização do campo, aumento da produtividade e aumento da produção de produtos agrícolas, como soja, milho, café, cana-de-açúcar e algodão. A agricultura familiar é a principal responsável pelo abastecimento de alimentos e matérias-primas para o mercado interno brasileiro.

A análise de dados agrícolas pode ajudar os agricultores a entender melhor o presente e poder prever o futuro. A combinação de ferramentas e registro detalhado de informações agrega valor à tomada de decisão. Dados e tecnologias como GPS, drones, inteligência artificial (IA) podem monitorar a saúde do solo, clima, pragas, colheitas e equipamentos, além de projetar preços para saber analisar as possibilidades de alocação de capital. A análise de dados pode revelar gargalos e instabilidades que precisam ser resolvidos por meio de projetos específicos. O uso e análise de grandes volumes de dados (big data) na agricultura também permite a avaliação do ciclo da cultura, o que garante melhores decisões do uso do solo.

2.2. O agronegócio e a produção de milho no Brasil

O agronegócio brasileiro é um dos setores mais importantes da economia do país, gerando empregos e contribuindo para o abastecimento interno e externo de alimentos e biocombustíveis. Dentre as culturas cultivadas no país, o milho é uma das mais relevantes, sendo utilizado tanto na alimentação animal quanto na produção de etanol.

De acordo com a Conab tem como perspectivas para a safra de milho de 2022/2023 Para o ano, estima produção total de 123,7 milhões de toneladas de milho, volume 9% acima da safra anterior. O Brasil é o terceiro maior produtor mundial da cultura, atrás apenas dos Estados Unidos e da China. A importância do milho é substancial na alimentação animal tendo em vista que cerca de 60% da produção brasileira é utilizada na fabricação de ração. Os setores de avicultura, suinocultura e bovinocultura são os principais consumidores de ração à base de milho no país. Além disso, a cultura é utilizada na alimentação de animais de produção como peixes e camarões tendo uma contribuição no papel do Brasil de maior produtor de proteína animal do mundo.

No setor de biocombustíveis, o milho é uma importante matéria-prima para a produção de etanol. De acordo com a União Brasileira do Biodiesel e Bioquerosene (Ubrabio), o milho responde por cerca de 10% da produção de etanol no país. A utilização do milho como matéria-prima para a produção de biocombustíveis tem crescido nos últimos anos, incentivada pelo aumento da demanda pelo combustível renovável.

A exportação de milho é mais um aspecto importante do setor. O Brasil é um dos principais exportadores da cultura, com destaque para países da Ásia como Japão, Coreia do Sul e Taiwan. De acordo com As exportações brasileiras de milho alcançaram o recorde de 43,17 milhões de toneladas em 2022, e mais que dobraram em relação aos 20,6 milhões embarcados em 2021, mostraram dados da Anec (Associação Nacional dos Exportadores de Cereais), Além da exportação, o mercado interno de milho é robusto e diversificado. A cultura é utilizada na alimentação humana em diversos produtos, exemplo de pães e bolos juntamente com uma gama de alimentos processados. Ademais, o milho também é utilizado na produção de amido, utilizado em diversos segmentos industriais, como na fabricação de papel, plásticos, têxteis e alimentos.

O Brasil é um dos principais produtores de sementes de milho do mundo, com destaque para a região Centro-Oeste do país. Essa atividade também possui grande relevância para o setor. De acordo com dados da Associação Brasileira de Sementes e Mudas (Abrasem), o mercado de sementes movimenta R\$ 10 bilhões ao ano no Brasil, sendo o terceiro maior produtor do mundo, atrás apenas de Estados Unidos e China.

Estes números apresentados demonstram a importância do agronegócio e do cultivo de milho para a economia brasileira. Além disso, destacam a diversidade e a relevância do setor, abrangendo desde a produção de alimentos até a fabricação de biocombustíveis e produtos industriais.

2.3. Análise de séries temporais

Segundo Sharpe et al. (2011), sempre que tivermos dados registrados sequencialmente ao longo do tempo e considerarmos o tempo um importante aspecto dos dados, temos uma série temporal. A maioria das séries temporais são igualmente espaçadas em intervalos aproximadamente regulares, como mensal, trimestral e anual.

Ademais, de acordo com Vasconcelos (2015, p. 25), dentro da chamada teoria neoclássica ou marginalista, o objetivo da firma é sempre maximizar o lucro total. Portanto, o planejamento se torna um ponto estratégico para seguir as melhores práticas e verificar a viabilidade de investimento do produtor agrícola, analisando a viabilidade e a forma de investir, com projeções que são úteis para se ter uma ideia do custo real de oportunidade na cultura do milho, por exemplo.

Nessa perspectiva, as empresas utilizam dados de séries temporais para monitorar se um processo em particular está estável ou instável e auxiliar na previsão do futuro, um processo chamado de previsão (DOANE e SEWARD, 2014, p. 597).

Além disso, dados de séries temporais econômicos também estão presentes em publicações como The Wall Street Journal, BusinessWeek, USA Today, Time e na navegação na Web.

É costumeiro visualizar dados de séries temporais como um gráfico de linhas ou de barras, com o tempo no eixo horizontal para mostrar como a variável de interesse muda. Em um gráfico de linha, os pontos X-Y estão conectados a segmentos de linha para facilitar a visualização de flutuações. (DOANE e SEWARD, 2014 p.598).

Tratando-se da análise de variáveis temporais, uma primeira análise é realizada pela avaliação gráfica, obtendo-se, por conseguinte, uma visualização mais ampla do comportamento da série. Outra análise importante para a compreensão do comportamento da série é a decomposição deste, observando suas componentes como tendência, sazonalidade e ciclos.

Conforme Sharpe et al. (2011), uma série temporal é composta por três componentes principais: sazonal, cíclica e aleatória. A componente sazonal refere-se à variação regular que ocorre ao longo do tempo, seguindo padrões estáveis e previsíveis. A componente

cíclica envolve ciclos de variação com períodos mais longos que um ano, geralmente relacionados a fenômenos econômicos ou de negócios. Já a componente aleatória representa a variação não sistemática e imprevisível da série temporal, decorrente de fatores não controláveis ou eventos aleatórios. Esses componentes desempenham papéis distintos na análise e modelagem das séries temporais, permitindo compreender e capturar os diferentes padrões e variações presentes nos dados.

2.4. O modelo autorregressivo

Na análise de dados as projeções podem ser efetuadas de diversas formas e seguir um amplo espectro de metodologias que devem ser definidas de acordo com as especificidades e necessidades de estudo. A definição do modelo de previsão é de suma importância. Além disso, deve-se verificar se a série de dados é condizente com os objetivos da análise e se são necessárias transformações sobre os dados.

Tratando-se de séries temporais e projeções, dentre os mais diversos modelos, um dos mais adotados é modelo autorregressivo. Conforme Sharpe et al. (2011, p.629), o modelo autorregressivo é um modelo de regressão baseado em uma média de valores anteriores da série, ponderados de acordo com a regressão em termos de valores defasados da própria variável. Ainda de acordo com estes autores, a correlação entre a série e a versão (defasada) da mesma série que é contrabalançada por um número de períodos de tempo é chamada de autocorrelação.

De acordo com Enders (2014), um modelo autoregressivo (AR) de ordem p , com p defasagens, possui a seguinte estrutura geral:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

Nesse modelo, y_t representa o valor da série temporal no momento t , μ é uma constante, $\phi_1, \phi_2, \dots, \phi_p$ são os coeficientes que indicam a relação entre o valor atual e os valores passados da série, e ε_t é o termo de erro aleatório no momento t . Para o modelo AR(1), a equação se simplifica para:

$$y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t$$

No modelo autorregressivo se observa que o valor atual da série é uma função linear dos valores passados, juntamente somado a um resíduo aleatório. Os coeficientes do modelo demonstram a influência relativa dos valores passados na previsão do valor atual. A ordem p

do modelo AR ou seja, o número de valores passados usados na previsão, é geralmente definido com a análise da função de autocorrelação da série.

2.5. Séries estacionárias e teste Dickey-Fuller

Antes de se efetuar uma projeção de série temporal em modelos autorregressivos é importante que se verifique a adequação da série em questão para a modelagem. Para uma projeção eficiente existe uma premissa de estacionariedade da série. Caso a condição de estacionariedade não esteja presente, serão necessárias algumas transformações na série para que a modelagem e a projeção sejam eficientes.

A estacionariedade é um atributo fundamental em séries temporais. Conforme Sharpe et al. (2011), uma série temporal que apresenta uma variação relativamente constante é chamada de estacionária na variância. Quando uma série temporal é considerada simplesmente estacionária, isso geralmente significa que ela possui uma variância estacionária, pois denota que as propriedades da série permanecem inalteradas ao longo do tempo. Desse modo, para que uma série seja considerada estacionária, é necessário que ela mantenha um comportamento consistente em relação às suas características estatísticas, como média e variância constantes ao longo do tempo.

A série não deve manifestar uma inclinação, tratando-se de uma alteração contínua na média ao decorrer do tempo, nem regularidade sazonal ou seja, uma variação também contínua na dispersão no decorrer do tempo. A premissa da estabilidade desempenha um papel crucial no modelo AR, pois possibilita que a relação entre os valores passados e presentes da série tenham uma constância temporal. Isso é importante pois permite previsões que almejam maior precisão. No que tange uma série ao qual não é estacionária, o modelo AR pode não detectar comportamentos importantes nas flutuações dos dados, e assim previsões menos eficientes e de menor precisão.

Um modo para avaliar se os dados têm estabilidade e compõe uma série estacionária é verificar se a média e a variância permanecem constantes ao longo do tempo. Outra abordagem ainda mais sofisticada e de maior precisão, consiste em analisar a função de autocorrelação da série, que mensura a correlação entre os valores da série em diferentes intervalos. O critério formal para analisar a estacionariedade de uma série consiste na aplicação de testes estatísticos, dos quais se destacam os testes de Dickey-Fuller (DF) e de Dickey Fuller Aumentado (ADF – Augmented Dickey–Fuller), como mencionado por Greene (2003).

Nestes testes, a primeira etapa é a definição de uma hipótese nula, que assume a presença de uma raiz unitária, indicando que a série de dados é não estacionária. Segundo, a hipótese alternativa sugere a ausência de raiz unitária, indicando que a série é estacionária. Essas hipóteses são essenciais para a aplicação correta e interpretação adequada do teste de Dickey Fuller Aumentado na análise de séries temporais.

O teste de Dickey Fuller Aumentado (ADF) é um teste estatístico que verifica se a série possui raiz unitária. Quando a série é estacionária, rejeita-se a hipótese nula. Assim sendo, as duas hipóteses do teste são definidas da seguinte forma:

- Hipótese nula (H_0): A série possui uma raiz unitária e, portanto, é não estacionária.
- Hipótese alternativa (H_A): A série é estacionária e não possui raiz unitária.

A interpretação do teste de ADF é baseada no valor da estatística do teste. Se o valor da estatística de teste for menor que o valor crítico, a hipótese nula (H_0) é rejeitada, o que significa que a série é estacionária. Se o valor da estatística de teste for maior que o valor crítico, a hipótese nula não é rejeitada, e a série não é estacionária.

Ademais, há a possibilidade de calcular o p-valor (valor-p) do teste, que representa a probabilidade de obter um valor do estatístico de teste tão ou mais significativo que o valor observado, assumindo que a hipótese nula seja verdadeira. Se o valor-p for menor que um nível de significância determinado na análise, geralmente de 5%, a hipótese nula é rejeitada e a série é considerada estacionária.

3. METODOLOGIA

Nesta seção será explanado acerca do aspecto metodológico do trabalho para levantamento dos dados, juntamente com os métodos aplicados na análise, modelagem e construção de previsões de preço.

Foram utilizados dados secundários, obtidos em sites aos quais efetuam os registros históricos, formando assim a série temporal de preço do milho. De forma específica, se utilizou o banco de dados do Cepea/USP, obtendo-se um recorte temporal de 10 anos, entre janeiro de 2013 e dezembro de 2022. Os dados são referentes ao preço médio mensal à vista, em Reais, da saca de milho de 60 quilos. Após a coleta, os dados foram organizados e armazenados por intermédio do programa Microsoft Excel.

Para efetuar a análise, executar os testes estatísticos, projeções, formatação de tabelas e a produção de gráficos, foi empregada a linguagem de programação Python, bem como bibliotecas direcionadas à análise de dados como Pandas e Statsmodels. Utilizou-se também a

plataforma Google Colab, que oferece um ambiente gratuito, baseado em nuvem, e que permite a elaboração e execução de códigos em um navegador da web.

3.1. Série temporal 2013-2022

Dando-se início à análise, dentre as primeiras etapas está a plotagem dos dados para se obter uma perspectiva mais ampla quanto ao comportamento da série; no caso em questão, o preço médio mensal do milho entre 2013 e 2022, correspondendo a 120 valores.

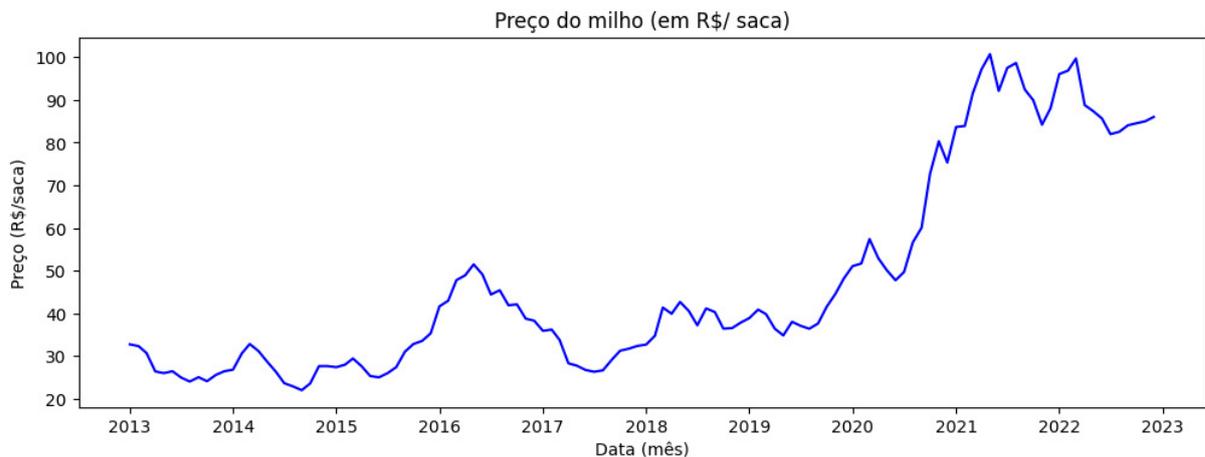


Figura 1. Séries de preço do milho (R\$/saca), média mensal (jan/2013 – dez/2022).
Fonte: CEPEA.

3.2. Treino e teste

O processo de modelagem empregou a técnica de divisão dos dados em conjunto de treino e conjunto de teste, muito comum em modelos preditivos e de aprendizado de máquina. A partir do conjunto de dados original, foi utilizado 75% para o conjunto de treino, obtendo-se 96 observações e 25 % para o conjunto de teste, perfazendo 24 observações.

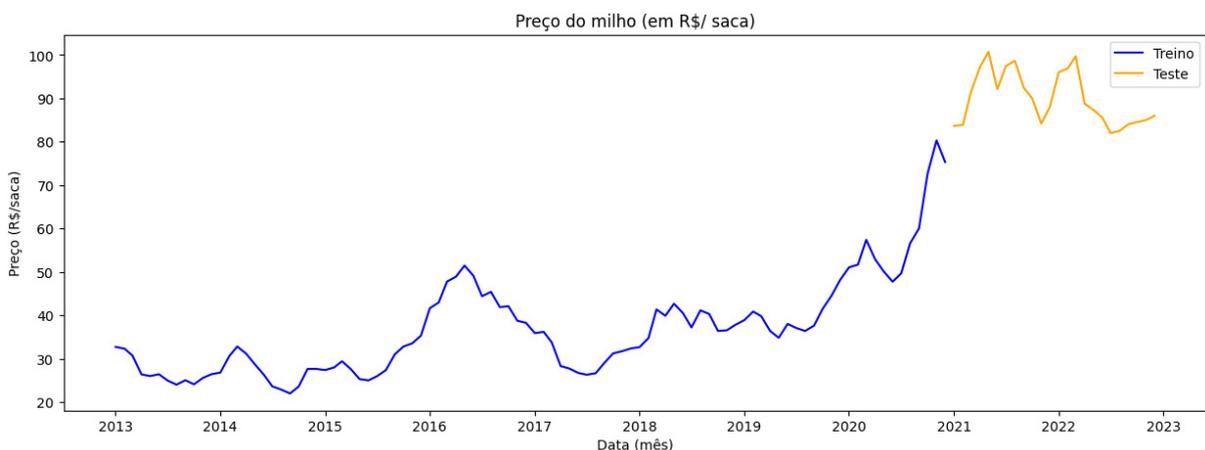


Figura 2. Séries de preço do milho (R\$/saca): dados de treino e teste.

Fonte: Elaboração do autor.

3.3. Estacionariedade e teste de raiz unitária e ajuste do modelo

Na sequência da pesquisa se efetuou o teste de Dickey Fuller sobre os dados de treino, para verificar se a série temporal é estacionária.

```

Teste Estatístico Dickey Fuller    -0.4141
Valor-P                          0.9077
Lags Usados                       1.0000
Número de observações usadas     94.0000
Valores Críticos (1%)            -3.5019
Valores Críticos (5%)            -2.8928
Valores Críticos (10%)           -2.5835
dtype: float64

```

Figura 3. Resultados do teste Dickey Fuller sem transformação na série.

Fonte: Elaboração do autor.

O que se verificou foi que a série de preços, em escala original, não é estacionária. Observando um valor-p maior que o nível de significância, decide-se por aceitar a hipótese nula do teste. Assim para se ter continuidade ao processo de modelagem, é necessário que se efetue alguma transformação na série para que se obtenha uma série estacionária. A transformação aplicada foi a primeira diferença.

A Figura a seguir mostra a aplicação do código em linguagem Python e a visualização de uma amostra dos dados em primeira diferença.

```

# Gerar dados em 1ª diferença
treino_dif = treino[['dif_preco']].dropna()
treino_dif.head()

```

dif_preco	
data	
2013-02-01	-0.41
2013-03-01	-1.63
2013-04-01	-4.30
2013-05-01	-0.39
2013-06-01	0.43

Figura 4. Processo de primeira diferença nos dados.

Fonte: Elaboração do autor.

Após essa transformação, aplicou-se mais uma vez o teste de Dickey-Fuller. Também, se analisou o ajuste do modelo pelos aspectos de mensuração de autocorrelação e autocorrelação parcial. Demais, o modelo matemático estimado para se obter a equação que descreve e prevê os dados de série temporal foi feito através do modelo AR(1).

4 RESULTADOS E DISCUSSÕES

4.1. A Série do milho

O presente trabalho teve como foco efetuar a previsão de preços futuros da saca de milho, assim, séries completas apresentaram 120 observações no período de janeiro de 2013 a dezembro de 2022. Figura 1 apresenta as séries de preços da saca de milho (60 kg) para o período de análise. Desta forma verificou-se que com os ajustes de primeira diferença foi possível se obter um conjunto de dados estacionário e desta forma ser eficiente aplicação de um modelo autoregressivo.

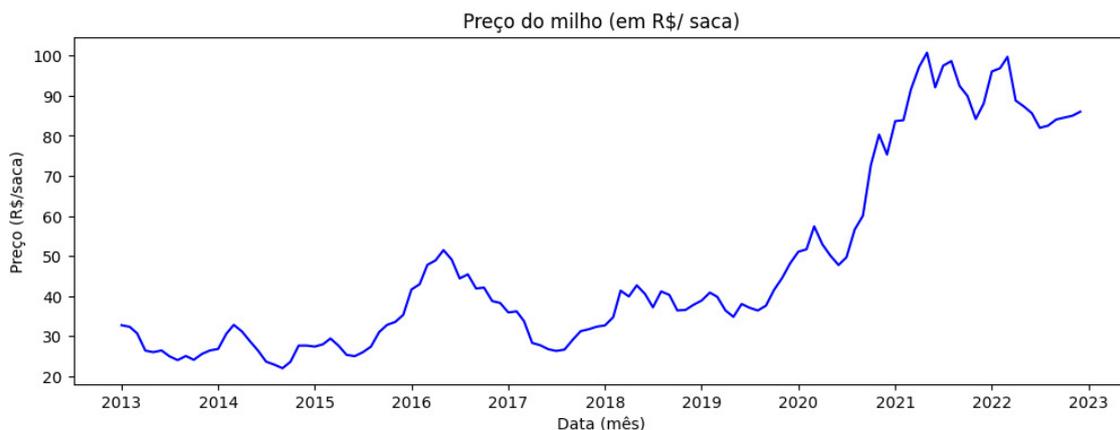


Figura 5. Séries de preço do milho (R\$/saca), média mensal (jan/2013 – dez/2022).

Fonte: CEPEA.

4.2 Treino e teste do modelo

Para estimar o modelo e previsão vamos adotar a estratégia de dividir os dados da amostra em um conjunto de dados de treino e um conjunto de dados de teste adotando-se este procedimento, estima-se o modelo empregando os dados de treino e, com o modelo estimado e ajustado, realiza previsões que terão sua acurácia verificada nos dados de teste. Na presente aplicação foi definido que os últimos 24 meses (equivalente a 25% da amostra) iriam compor o conjunto de dados de teste, ou seja, o período correspondente aos períodos dos últimos 2 anos da série ao qual se coletou.

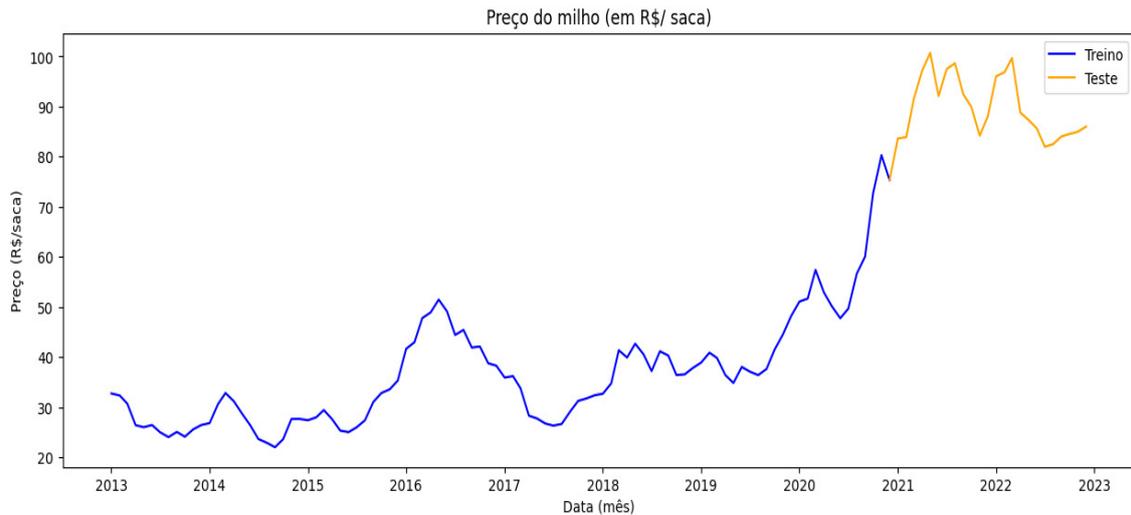


Figura 6. Séries de preço do milho (R\$/saca) com destaque para divisão dos dados em conjunto de dados de treino e dados de teste.

Fonte: Elaboração do autor.

Na Tabelas 1, a seguir, é possível observar as estatísticas descritivas da série de preços, seja da série em escala original, seja em primeira diferença.

Tabela 1. Estatísticas descritivas da série de preços (dados de treino).

	Dados em nível	Dados em 1ª diferença
Nº de observações	96.000000	95.000000
Média	36.861146	0.448211
Desvio-padrão	11.433706	2.995984
Valor mínimo	22.020000	-5.450000
Percentil 25	27.647500	-1.515000
Percentil 50 (Mediana)	35.085000	0.340000
Percentil 75	41.715000	1.960000
Valor máximo	80.310000	12.650000

Fonte: Elaboração do autor.

Na Tabela 2 estão presentes os resultados do teste de Dickey-Fuller para as séries em nível e em primeira diferença. observa-se que os dados em nível não são estacionários, entretanto ao tomar a primeira diferença da série obtém uma série estacionária.

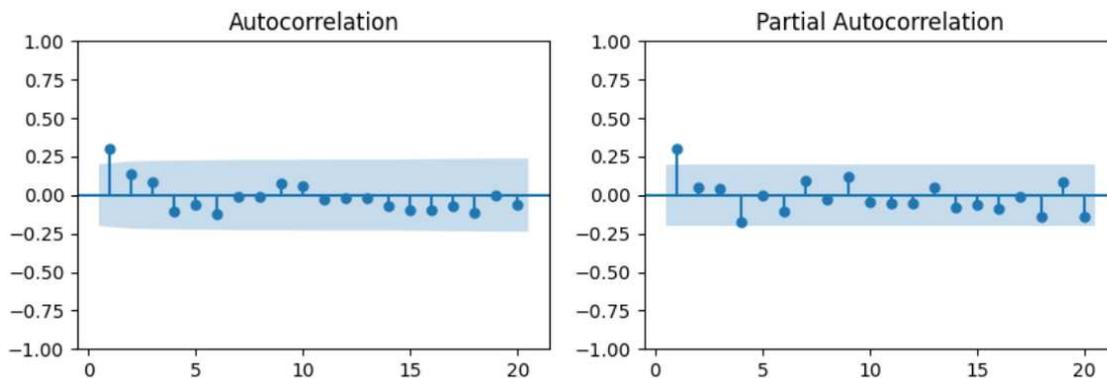
Tabela 2. Resultados do teste de Dickey-Fuller (dados de treino).

	Dados em nível	Dados em 1ª diferença
Valor do teste	0.4141	-6.8275
Valor-p	0.9077	0.0000
Lags usados	1.0000	0.0000
Nº de observações usado	94.0000	94.0000
Valor Crítico de referência (1%)	-3.5019	
Valor Crítico de referência (5%)	-2.8928	
Valor Crítico de referência (10%)	-2.5835	

Fonte: Elaboração do autor.

4.3. Ajuste do modelo

Na Figura 7 temos a série de preços em primeira diferença, os gráficos das funções de autocorrelação e autocorrelação parcial, e a distribuição da série de preços em primeira diferença. Com a obtenção de uma série estacionária, a análise dos gráficos das funções de autocorrelação (ACF - Autocorrelation Function) e autocorrelação parcial (PAC - Partial Autocorrelation Function) possibilita inferir a respeito de diferentes possibilidades de modelagem.

**Figura 7.** Funções de autocorrelação e autocorrelação parcial (dados de treino).

Fonte: Elaboração do autor.

Na presente aplicação vamos optar por uma opção parcimoniosa e aplicar uma estrutura de um modelo autorregressivo de ordem 1, ou seja, um modelo AR(1). A ACF apresenta apenas a 1ª defasagem significativa. O mesmo se observa na PAC, em decorrência de se ter oscilações com as lags, ora subindo, ora descendo. Caso ocorresse um decaimento maior e contínuo poderia se perceber um comportamento de médias móveis. Se o comportamento tivesse uma um decaimento pequeno mais contido poderia se caracterizar como um modelo ARMA sendo uma junção de autorregressivo e de médias móveis.

Dessa forma, foi estimado um modelo AR(1) usando a biblioteca Statsmodels. A Tabela 3, a seguir, mostra os resultados do modelo AR(1) cuja especificação de ajustou melhor aos dados. Nesta versão, apresentada na Tabela 3, adotando um nível de significância de 5% percebe-se que o intercepto e o coeficiente autorregressivo são estatisticamente significantes. Também foram estimadas versões com intercepto, mas estes não se mostraram estatisticamente significantes ao nível de significância adotado.

Tabela 3. Sumário do modelo AR(1) - dados de treino.

	coeficiente	erro-padrão	z	P> z 	[0,025	0,975]
Intercepto	-	-	-	-	-	-
ar.L1	0,3236	0,079	4,103	0,000	0,169	0,478
sigma	8,1385	0,839	9,705	0,000	6,495	9,782
N. de obs	95					
Log Likelihood	-234,443					
AIC	472,887					
BIC	477,995					

Fonte: Elaboração do autor.

Com este resultado tem-se uma equação estimada dada por:

$$\Delta p_t = 0,3236 \Delta p_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0; 8,1385)$$

Na presente análise realizou-se previsões “mês a mês”. Dessa forma, com base nesse modelo, em que o valor previsto no tempo t é igual 0,3236 vezes o preço no tempo t menos 1 mais o erro, pode-se fazer estimações de preços para cada período (mês) t da seguinte forma:

$$p_t = p_{t-1} + \Delta \hat{p}_t$$

O gráfico da Figura 5 mostra as séries de preços observados e preços previstos aplicando o modelo estimado e os dados de treino.

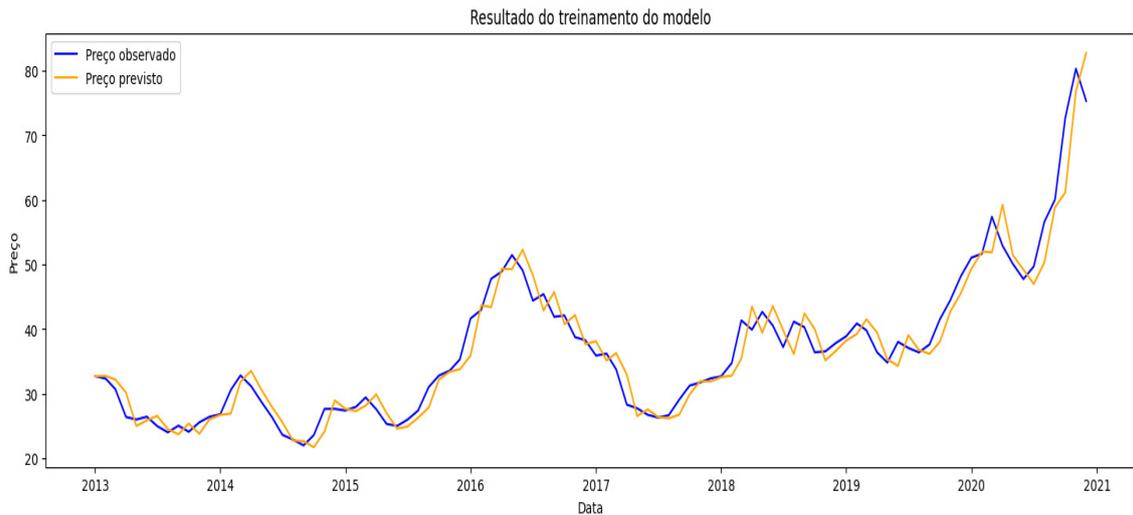


Figura 9. Série de preços observados e preços previstos (dados de treino).
Fonte: Elaboração do autor.

No conjunto de treino temos uma medida de erro quadrático médio de 8,1387 e de erro absoluto médio de 2,1517.

A verdadeira performance preditiva do modelo estimado deve ser avaliada no conjunto de dados de teste. Conforme indicado outrora, os valores observados nos últimos dois anos foram destacados da base como dados de teste.

O gráfico da Figura 9 explana a série de preços médios observados no conjunto de dados de teste e a série prevista, aplicando-se o modelo estimado anteriormente.

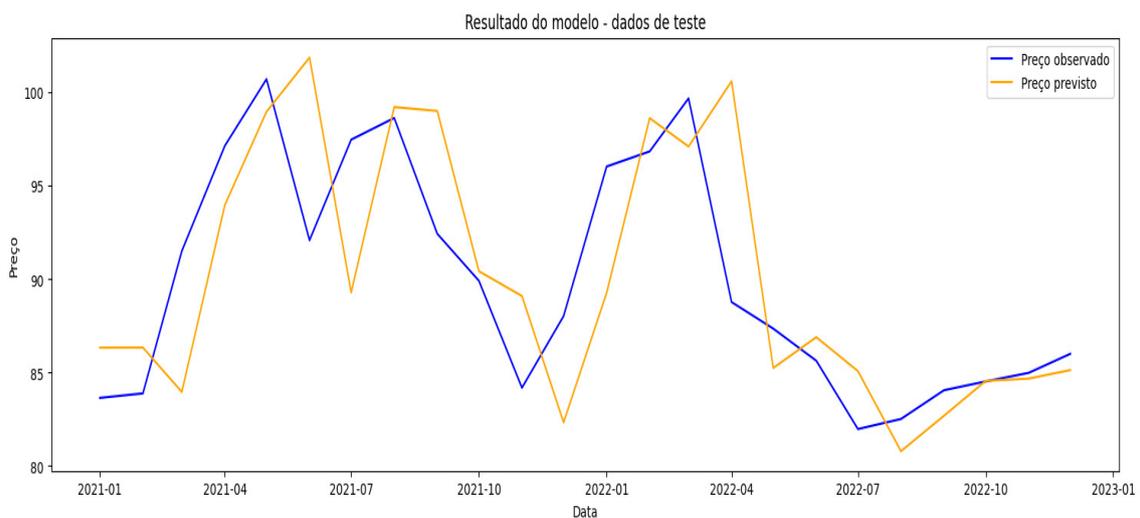


Figura 8. Série de preços observados e preços previstos (dados de teste).
Fonte: Elaboração do autor.

Dessa forma, como nas previsões geradas no conjunto de dados de treino, deve-se avaliar o ajuste das previsões ao conjunto de dados de teste; que são as previsões que realmente importam para a análise. Quando se trata em ocasiões às quais se comparam diferentes modelos, estas estatísticas permitem avaliar qual deles apresenta o melhor resultado em termos de acurácia das previsões.

Nesta situação em específico as previsões geradas sobre a série definida como teste, obteve uma medida de erro quadrático médio de 23.4710 e de erro absoluto médio de 3.6492. Ademais, é possível verificar com a aplicação do modelo aos dados de teste, obteve-se valores de erro mais elevados quando se aplicou o modelo aos dados de treino. Essa diferença é um resultado comum, uma vez que o modelo foi ajustado aos dados de treino, mas não aos dados de teste. Dessa forma, ao aplicar o modelo aos dados de teste, tem-se uma simulação do comportamento do modelo ao ser aplicado em dados desconhecidos. Situação que verdadeiramente ocorre nos processos de análise e de previsões com séries temporais.

5 CONSIDERAÇÕES FINAIS

O objetivo do presente trabalho foi aplicar técnicas de estimação e previsões de séries temporais para analisar e projetar séries de preço da saca de milho no Brasil. Para esse objetivo se aplicou um modelo autorregressivo de ordem 1, bastante tradicional na análise estatística de séries temporais.

O milho é um produto de grande importância, consistindo em um dos principais insumos para a produção de alimentos, seja para o consumo humano ou alimentação animal. As técnicas de previsão são de grande valia para as atividades de planejamento dos agentes que atuam no mercado de milho, como produtores, consumidores e beneficiadores. Portanto, o modelo autorregressivo se demonstrou adequado para atividades de caráter preditivo dos preços desse produto agrícola. Esse tipo de análise permite uma compreensão melhor do comportamento dos preços e, com um modelo bem ajustado, pode auxiliar na tomada de decisões e assim mitigação de riscos.

REFERÊNCIAS

CARVALHO, Carla. O que esperar da safra de milho no Brasil em 2023? **Terra Investimentos**. Disponível em: <<https://blog.terrainvestimentos.com.br/o-que-esperar-da-safra-de-milho-no-brasil-em-2023/>>. Acesso em: 11 de junho de 2023.

DOANE, David P.; SEWARD, Lori E. **Estatística Aplicada à Administração e Economia**. AMGH; 4ª edição (2014).

ENDERS, W. *Applied Econometric Time Series*. 4th ed. Editora: Wiley, 2014.

ETANOL de milho é alternativa sustentável de biocombustível e pode conquistar 20% do mercado. **UNEM**. Disponível em: <<https://etanoldemilho.com.br/2021/07/06/etanol-de-milho-e-alternativa-sustentavel-de-biocombustivel-e-pode-conquistar-20-do-mercado/>>. Acesso em: 17 de abril de 2023.

FIGUEIREDO, Nayara. Exportação de milho do Brasil dispara até 2ª semana do mês e já supera fevereiro de 2022. **Novacana**, 2023. Disponível em: <https://www.novacana.com/noticias/exportacao-milho-brasil-dispara-2-semana-mes-supera-fevereiro-2022-140223>

GREENE, William H. *Econometric Analysis*. 5. ed. New Jersey: Prentice Hall, 2003.

GOMES, E. G.; MANGABEIRA, J. A. C. Uso de análise de envoltória de dados em agricultura: o caso de Holambra. Embrapa Meio Ambiente, Jaguariúna, 2001. 22 p. (Embrapa Meio Ambiente. Documentos, 34). Disponível em: <https://periodicos.uff.br/engevista/article/view/8755/6223>. Acesso em: 20 maio. 2023.

INDICADOR do milho ESALQ/BM & FBOVESPA <<https://www.cepea.esalq.usp.br/br/indicador/milho.aspx>>. **Cepea USP**. Acesso em: 03 de março de 2023.

LEMOS, José de Jesus Sousa. **Vulnerabilidades induzidas no semiárido**, Fortaleza: Imprensa Universitária, 2020.

MERCADO de sementes movimenta R\$ 10 bi ao ano no Brasil. **ABRATES**, 2018. Disponível em: <<https://www.abrates.org.br/noticia/mercado-de-sementes-movimenta-r-10-bi-ao-ano-no-brasil/>>. Acesso em: 17 de de abril de 2023.

SHARPE, Norean R.; VEAUX, Richard D De; VELLEMAN, Paul F. **Estatística Aplicada**. Porto Alegre: Grupo A Editora, 2011.

TOSI, Marcos. Etanol de milho: “patinho feio” americano virou indústria das que mais crescem no Brasil. **Gazeta**, 2022. Disponível em: <<https://www.gazetadopovo.com.br/agronegocio/etanol-de-milho-industria-cresce-no-brasil/>>. Acesso em: 19 de maio de 2023.