

MARCELO NEVES COSSI JORGE

**Adequação de modelos de previsão à variação dos dados de
séries históricas da Conab**

Marcelo Neves Cossi Jorge

**Adequação de modelos de previsão à variação dos dados de
séries históricas da Conab**

Trabalho de Graduação apresentado ao Conselho de Curso de Graduação em Engenharia de Produção Mecânica da Faculdade de Engenharia do Campus de Guaratinguetá, Universidade Estadual Paulista, como parte dos requisitos para obtenção do diploma de Graduação em Engenharia de Produção Mecânica.

Orientador: Prof. Dr. Fernando Augusto Silva Marins.

2021

J82a	Jorge, Marcelo Neves Cossi Adequação de modelos de previsão à variação dos dados de séries históricas da Conab / Marcelo Neves Cossi Jorge – Guaratinguetá, 2021. 89 f : il. Bibliografia: f. 87-89 Trabalho de Graduação em Engenharia de Produção Mecânica – Universidade Estadual Paulista, Faculdade de Engenharia de Guaratinguetá, 2021. Orientador: Prof. Dr. Fernando Augusto Silva Marins 1. Análise de séries temporais. 2. Processo decisório. 3. Mercadorias. I. Título. CDU 65.012.4
------	--

MARCELO NEVES COSSI JORGE

ESTE TRABALHO DE GRADUAÇÃO FOI JULGADO ADEQUADO COMO
PARTE DO REQUISITO PARA A OBTENÇÃO DO DIPLOMA DE
“GRADUADO EM ENGENHARIA DE PRODUÇÃO MECÂNICA”

APROVADO EM SUA FORMA FINAL PELO CONSELHO DE CURSO DE
GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO MECÂNICA

Prof. Dr. Fernando Augusto Silva Marins.

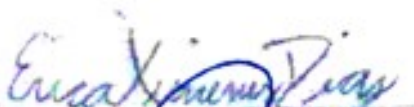
BANCA EXAMINADORA:



Prof. Dr. Fernando Augusto Silva Marins
Orientador/UNESP-FEG



Prof. Dr. Aneirson Francisco da Silva
UNESP-FEG



Prof. Dra. Erica Ximenes Dias
UNESP-FEG

Janeiro de 2021

DADOS CURRICULARES

MARCELO NEVES COSSI JORGE

NASCIMENTO 24.01.1997 – São José dos Campos / SP

FILIAÇÃO Samuel Tadeu Jorge
Nilma da Silva Neves

2012/2014 Ensino Médio
Escola EMBRAER Juarez Wanderley

AGRADECIMENTOS

Em primeiro lugar agradeço pela minha saúde, minha inteligência, e pelo apoio que tive durante os anos de graduação;

ao meu orientador, *Prof. Dr. Fernando Augusto Silva Marins* que jamais deixou de me incentivar. Sem a sua orientação, dedicação e auxílio, o estudo aqui apresentado não seria possível;

aos meus pais *Samuel, Ângela e Nilma*, que apesar das dificuldades enfrentadas, sempre incentivaram meus estudos e me garantiram saúde;

às funcionárias e funcionários da Biblioteca do Campus de Guaratinguetá pela dedicação, presteza e principalmente pela vontade de ajudar, sem eles o trabalho realizado não existiria;

aos funcionários da Faculdade de Engenharia do Campos de Guaratinguetá que se dedicaram a disseminar conhecimento e pensamento crítico da maneira mais empática possível.

“... Pois, quando estou em meus sonhos,
sou todos os personagens, mesmo que veja
através dos olhos de apenas um.”

Leiteira

RESUMO

O trabalho identifica e discute os principais fatores que influenciam na precisão de um modelo de previsão e apresenta uma análise sobre séries históricas da Conab. Utilizando modelos de previsão conhecidos (média móvel, alisamento exponencial simples, ARIMA e SARIMA) e com o auxílio de softwares é verificada a precisão de cada modelo de acordo com o intervalo de dados escolhido, estabelecendo uma metodologia que permite encontrar de maneira rápida um modelo coerente. A análise de séries temporais, ou históricas, é um instrumento estatístico de pesquisa que proporciona autoconhecimento e inteligência para criação de planos de ação nas empresas. Fazer uma previsão consistente ao se analisar uma série é uma qualidade decisiva para empresa que busca absorver de maneira rápida as mudanças do seu ecossistema, entretanto fazer uma previsão rápida e de qualidade necessita de um modelo adequado. Escolher um modelo adequado, por sua vez, necessita de uma metodologia que garanta coerência na escolha. Foi observado que os modelos de alisamento exponencial e média móvel se mostram suscetíveis a falhas quando a série não é estacionária, entretanto a seleção dos parâmetros e o intervalo analisado podem aumentar a eficiência desses modelos. O modelo ARIMA se mostrou de aplicação ágil e precisa, garantindo a melhor descrição das séries e melhores resultados. O estudo dos erros se mostrou importante para se verificar se a descrição da série foi adequada. O modelo SARIMA não se mostrou interessante para as séries estudadas devido ao fato de o intervalo entre as observações ser de um ano, o que prejudica a análise da sazonalidade, além disso, os dados obtidos em intervalos menores que um ano não apresentaram período constante.

PALAVRAS-CHAVE: Modelos de previsão. Séries históricas. Precisão. *Commodities*. Conab.

ABSTRACT

The work identifies and discusses the main factors that influence the accuracy of a forecasting model and presents an analysis of Conab's historical series. Using known forecasting models (moving average, simple exponential smoothing, ARIMA and SARIMA) and with the aid of software, the accuracy of each model is verified according to the chosen data range, establishing a methodology that allows to quickly find a coherent model. The analysis of time series, or historical, is a statistical research tool that provides self-knowledge and intelligence for creating action plans in companies. Making a consistent forecast when analyzing a series is a decisive quality for a company that seeks to quickly absorb changes in its ecosystem, however making a quick and quality forecast requires an appropriate model. Choosing an appropriate model, in turn, requires a methodology that ensures consistency in the choice. It was observed that the exponential smoothing and moving average models are susceptible to failures when the series is not stationary, however the selection of parameters and the analyzed range can increase the efficiency of these models. The ARIMA model proved to be agile and precise, ensuring the best description of the series and the best results. The study of errors proved to be important to verify whether the series description was adequate. The SARIMA model was not interesting for the series studied due to the fact that the interval between observations is one year, which hinders the analysis of seasonality, in addition, the data obtained at intervals shorter than one year did not present a constant period.

KEYWORDS: Prediction models. Historical series. Precision. Commodities. Conab.

LISTA DE ILUSTRAÇÕES

Figura 1 – Principais redes de citação de acordo com os filtros indicados	18
Figura 2 – Fluxograma de atividades para organização do trabalho	19
Figura 3 – Cronograma das etapas	20
Figura 4 – Valor da média (laranja) de acordo com a variação dos dados (azul)	24
Figura 5 – Valor da média (laranja) de acordo com a variação dos dados (azul)	24
Figura 6 – Número de passageiros ao longo dos anos	26
Figura 7 – Interface para construção de gráficos no Excel	36
Figura 8 – Variação da produção anual e da média da produção anual de feijão no Brasil ..	37
Figura 9 – Trecho de tabela feita no Excel	37
Figura 10 – Seleção de abas para utilização dos testes de raiz unitária	38
Figura 11 – Exemplo de correlograma	39
Figura 12 – Valores para FAC e FACP	40
Figura 13 – Interface para aplicação do modelo ARIMA	42
Figura 14 - Coeficientes obtidos na modelagem ARIMA	42
Figura 15 – Variação da área plantada de trigo 1977 a 2019	44
Figura 16 - Variação da área plantada de trigo de 2001 a 2019	45
Figura 17 – p-valor para o teste Dickey-Fuller aumentado	45
Figura 18 – p-valor para o teste ADF-GLS	46
Figura 19 – Estatística de teste e valores críticos para o teste KPSS	46
Figura 20 – correlograma da série referente a área plantada de trigo	47
Figura 21 – Previsão ARIMA da área plantada de trigo para 2020	48
Figura 22 – Distribuição dos resíduos ARIMA para a área plantada de trigo	48
Figura 23 - Variação da produtividade em relação ao trigo de 1977 a 2019	51
Figura 24 – Teste ADF-GLS para a série de produtividade em relação ao trigo	52
Figura 25 – Correlograma da série de produtividade em relação ao trigo	53
Figura 26 – Aplicação do modelo ARIMA à série de produtividade em relação ao trigo ..	53
Figura 27 - Distribuição dos resíduos ARIMA para a produtividade em relação ao trigo ..	54
Figura 28 - Variação da produção de relação ao trigo de 1977 a 2019	57
Figura 29 – Teste Dickey-Fuller para a produção de trigo	58
Figura 30 – Aplicação do teste ADF-GLS para a produção de trigo	59
Figura 31 – Correlograma para a série de produção de trigo	59
Figura 32 – Gráfico das primeiras diferenças em relação à série de produção de trigo	60

Figura 33- Teste Dickey-Fuller aumentado para as primeiras diferenças	60
Figura 34 – Teste ADF-GLS para as primeiras diferenças	61
Figura 35 – Teste KPSS para as primeiras diferenças	61
Figura 36 – Aplicação do Modelo ARIMA à série de produção de trigo	62
Figura 37 – Distribuição dos resíduos ARIMA para a produção de trigo	63
Figura 38 – Previsão do modelo ARIMA em relação a produção de trigo	63
Figura 39 – Variação da produção de soja de 1976 a 2019	65
Figura 40 – Teste ADF-GLS para a série de produção de trigo	66
Figura 41 – Correlograma para a série de produção de trigo	66
Figura 42 – Aplicação do modelo ARIMA para a série de produção de soja	67
Figura 43 – Dispersão dos erros para o modelo ARIMA da produção de soja	68
Figura 44 - Previsão do modelo ARIMA em relação a produção de soja	68
Figura 45 – Média móvel para $r=10$ aplicada a produção de soja	69
Figura 46 - Média móvel para $r=3$ aplicada a produção de soja	70
Figura 47 – Alisamento exponencial em relação à série de produção de soja	70
Figura 48 – Dispersão dos erros para o alisamento exponencial	71
Figura 49 – Variação da produção de milho de 1976 a 2019	72
Figura 50 – Teste ADF-GLS para a série de produção de milho	73
Figura 51 – Correlograma da série de produção de milho	74
Figura 52- Aplicação do modelo ARIMA para a série de produção de milho	75
Figura 53 - Dispersão dos erros para o modelo ARIMA da produção de milho	75
Figura 54 - Previsão do modelo ARIMA em relação a produção de milho	76
Figura 55 – Distribuição dos erros para o modelo de médias móveis	77
Figura 56 – Distribuição dos erros para o modelo de alisamento exponencial	78
Figura 57 – Variação da produção de feijão de 1976 a 2019	79
Figura 58 – Variação da produção de feijão de 1993 a 2019	79
Figura 59 – Teste Dickey-Fuller aumentado para a produção de feijão de 1993 a 2019	80
Figura 60 – Teste ADF-GLS para a produção de feijão de 1993 a 2019	80
Figura 61 - Teste Dickey-Fuller aumentado para a produção de feijão de 1976 a 2019	81
Figura 62 – Correlograma para a série de produção de feijão de 1976 a 2019	82
Figura 63 – Aplicação do modelo ARIMA à série de produção de feijão	82
Figura 64 – Teste da normalidade dos resíduos para a previsão da produção de feijão	83
Figura 65 – Períodograma da produção de feijão	85

LISTA DE TABELAS

Tabela 1 – Médias móveis para a área plantada de trigo	49
Tabela 2 – Previsão da área plantada de trigo utilizando alisamento exponencial	50
Tabela 3 – Erro absoluto e desvio médio absoluto para área plantada de trigo	50
Tabela 4 – Média móvel aplicada à produtividade em relação ao trigo	55
Tabela 5 – Alisamento exponencial aplicado à produtividade em relação ao trigo	55
Tabela 6 – Erros dos modelos para a produtividade em relação ao trigo	56
Tabela 7 – Erros para o modelo ARIMA aplicado à série de produção de trigo	64
Tabela 8 – Erros para as previsões do modelo ARIMA para a produção de soja	69
Tabela 9 - Erros para as previsões do modelo ARIMA para a produção de milho	76
Tabela 10 – Erros para aplicação do modelo ARIMA à série de produção de feijão	83

LISTA DE ABREVIATURAS E SIGLAS

ARIMA	Modelo Autorregressivo Integrado de Médias Móveis
CFP	Companhia de Financiamento da Produção
Cibrazem	Companhia Brasileira de Armazenamento
Cobral	Companhia Brasileira de Alimentos
Conab	Companhia Nacional de Abastecimento
FAC	Função autocorrelação
FACP	Função autocorrelação parcial
MAPA	Ministério da Agricultura e Pecuária
SARIMA	ARIMA sazonais multiplicativos
ADF-GLS	Teste de raiz unitária desenvolvido por Elliot, Rothengberg e Stock.
KPSS	Teste de raiz unitária criado por Kwiatkowski, Phillips, Schmidt e Shin.

SUMÁRIO

1	INTRODUÇÃO	15
1.1	CONTEXTUALIZAÇÃO E QUESTÃO DE PESQUISA	15
1.2	OBJETIVOS E DELIMITAÇÃO	16
1.3	JUSTIFICATIVA	17
1.4	ETAPAS E CLASSIFICAÇÃO DE PESQUISA	18
1.5	ORGANIZAÇÃO DO TRABALHO	21
2	FUNDAMENTAÇÃO TEÓRICA	22
2.1	SÉRIES TEMPORAIS	22
2.2	DECOMPOSIÇÃO CLÁSSICA E SÉRIES ESTACIONARIAS	23
2.3	TENDÊNCIA	24
2.4	SAZONALIDADE	25
2.5	MÉTODO GRÁFICO E TESTES DE RAIZ UNITÁRIA	26
2.6	MÉTODOS ESTATÍSTICOS DE PREVISÃO	27
2.6.1	Média móvel	27
2.6.2	Alisamento exponencial simples	28
2.6.3	ARIMA	29
2.6.4	SARIMA	30
2.7	ERROS DE PREVISÃO	30
2.7.1	Erro simples e absoluto de previsão	31
2.7.2	Desvio padrão da previsão	31
2.7.3	Desvio médio absoluto	31
2.7.4	Intervalo de confiança	32
2.8	SOFTWARES PARA PREVISÃO E ANÁLISE DE SÉRIES TEMPORAIS	32
2.8.1	Excel	32
2.8.2	Gretl	33
3	DESENVOLVIMENTO DA PESQUISA	34
3.1	DESCRIÇÃO DO PROBLEMA A SER TRATADO	34
3.2	UNIDADE DE PESQUISA	34
3.3	INSTRUMENTOS DE PESQUISA	36
3.3.1	Ferramentas do Excel	36
3.3.1.1	Plotagem de gráficos	36

3.3.1.2 Tabelas, funções e contas.....	37
3.3.2 Ferramentas do Gretl.....	38
3.3.2.1 Testes de raiz unitária.....	38
3.3.2.2 Identificação dos parâmetros ARIMA.....	39
3.3.2.3 Identificação dos parâmetros SARIMA.....	40
3.3.2.4 Modelagem.....	41
3.4 TRATAMENTO DOS DADOS.....	43
4 RESULTADOS.....	44
4.1 TRIGO.....	44
4.1.1 Série histórica da área de plantação de trigo.....	44
4.1.1.1 Teste da estacionariedade	45
4.1.1.2 Modelo ARIMA	47
4.1.1.3 Média móvel e Alisamento exponencial	49
4.1.1.4 Cálculo dos erros	50
4.1.1.5 Discussão dos resultados	50
4.1.2 Série histórica da produtividade em relação ao trigo	51
4.1.2.1 Modelo ARIMA	52
4.1.2.2 Média móvel e alisamento exponencial	54
4.1.2.3 Cálculo dos erros	56
4.1.2.4 Discussão dos resultados	56
4.1.3 Série histórica da produção de trigo	57
4.1.3.1 Modelo ARIMA	62
4.1.3.2 Discussão dos resultados	64
4.2 SOJA	65
4.2.1 Série histórica de produção de soja	65
4.2.1.1 Modelo ARIMA	67
4.2.1.2 Média móvel	69
4.2.1.3 Alisamento exponencial	70
4.2.1.4 Discussão dos resultados	71
4.3 MILHO	72
4.3.1 Série histórica da produção de milho	72
4.3.1.1 Modelo ARIMA	74
4.3.1.2 Média Móvel	76
4.3.1.3 Alisamento exponencial	77

4.3.1.4	Discussão dos resultados	78
4.4	FEIJÃO	78
4.4.1	Série histórica da produção de feijão	79
4.4.1.1	Testes de estacionariedade	79
4.4.1.2	Modelo ARIMA	81
4.4.1.3	Discussão dos resultados	84
4.5	MODELO SARIMA	84
5	CONSIDERAÇÕES FINAIS	86
5.1	VERIFICAÇÃO DAS RESPOSTAS AS QUESTÕES DE PESQUISA	86
5.2	VERIFICAÇÃO DO ATENDIMENTO AOS OBJETIVOS	87
5.3	SUGESTÕES PARA A CONTINUIDADE DO TRABALHO	88
	REFERÊNCIAS	90

1 INTRODUÇÃO

Neste capítulo serão apresentados o contexto, a questão e os objetivos da pesquisa, contendo as justificativas de pesquisa e a organização do trabalho. Os métodos, a caracterização e os conceitos utilizados ao decorrer do trabalho serão apresentados no desenvolvimento dos capítulos.

1.1 CONTEXTUALIZAÇÃO E QUESTÃO DE PESQUISA

Segundo Araujo e Souza (2015) séries temporais ou históricas são conjuntos de observações equidistantes, tendo como aspecto importante a dependência serial entre os dados observados. A maioria dos métodos estatísticos estão voltados para a análise de observações independentes, o estudo temporal, por sua vez, requer de técnicas específicas para a modelagem da dependência entre os dados, as quais são o objeto de estudo do trabalho.

O estudo de técnicas para a modelagem da dependência entre os dados se justifica pela importância das previsões de demanda e de produção, que tem como base fundamental a análise de séries temporais. Martin et al. (2015) afirmam que tais previsões estimam valores que podem influenciar na estratégia de uma organização devido ao fato de sugerirem condições futuras de trabalho. Utilizar um modelo de previsão adequado e validado para a situação em estudo pode acarretar numa economia de recursos.

Em um outro aspecto, Martin et al. (2015) diz que a análise e previsão de demanda contribui para que a empresa produza a quantidade adequada no momento oportuno, reduzindo a necessidade de estoque, sendo um diferencial na competitividade da empresa. Além disso, Yang et al (2017) evidenciam que previsões de demanda tem aplicabilidade abrangente, citando alguns exemplos, como na área econômica (como na previsão do PIB), na área de controle e qualidade (acompanhando as variações nas medidas) e na área da energia (prevendo o consumo energético de cidades e edifícios), além de serem essenciais no desenvolvimento de planos de produção.

Dentre suas diversas aplicações pode-se destacar a aplicabilidade na agropecuária, como, por exemplo, na previsão de safras de *commodities* (produtos que funcionam como matéria prima), que permite que os produtores tenham embasamento para tomar decisões no que se diz respeito a precificação, distribuição e planejamento de gastos.

No cenário agropecuário brasileiro tem-se a Conab (Companhia Nacional de

Abastecimento), uma empresa pública vinculada ao Mapa (Ministério da Agricultura, Pecuária e Abastecimento), que tem destaque devido a sua importância nas tomadas de decisão no setor primário do país. Sua missão é “promover a garantia de renda ao produtor rural, a segurança alimentar e nutricional e a regularidade do abastecimento, gerando inteligência para a agropecuária e participando da formulação e execução das políticas públicas” (CONAB, 2017). Para isso a empresa levanta dados e faz previsões sobre a produção de *commodities* no Brasil.

Para que uma empresa utilize previsões de maneira competitiva é necessário rapidez. Nessa perspectiva, Oliveira et al. (2016) destaca que no cotidiano das empresas existem planilhas eletrônicas e *softwares* especializados que auxiliam no cálculo da previsão de demanda ou produção para se garantir agilidade e, assim, a viabilidade do estudo.

Pensando nesse contexto o trabalho tem como questões de pesquisa:

- a) Como determinar, em uma série histórica, um modelo mais aderente à realidade?
- b) Quais são os principais fatores que influenciam na precisão de um modelo de previsão?
- c) Quais modelos de previsão descrevem adequadamente séries históricas da Conab?

1.2 OBJETIVOS E DELIMITAÇÃO

Buscando responder às questões de pesquisa, o objetivo geral deste trabalho é identificar como se dá a relação da variação dos dados em séries históricas da Conab e como se adequam os principais modelos de previsão no seu tratamento. Por sua vez, os objetivos específicos são:

- Investigar, com o auxílio de *softwares*, os diferentes comportamentos das séries históricas visando seu uso em modelos de previsão;
- Identificar situações em que é recomendado o uso de um determinado modelo de previsão;
- Determinar um intervalo de variação dos dados em que cada método de previsão analisado funcione e ofereça valores previstos compatíveis com a realidade modelada.

O trabalho se delimita, portanto, a estudar a relação entre dados vizinhos nas séries históricas da produção de *commodities* disponibilizadas pela Conab e apresentar uma

metodologia que permita, dentre modelos conhecidos encontrar aquele que garanta a maior precisão de previsão, utilizando como ferramentas de análise o *software* Excel, para construção de gráficos e modelagem de alisamento exponencial simples e média móvel e o Gretl, para modelagem ARIMA E SARIMA.

1.3 JUSTIFICATIVA

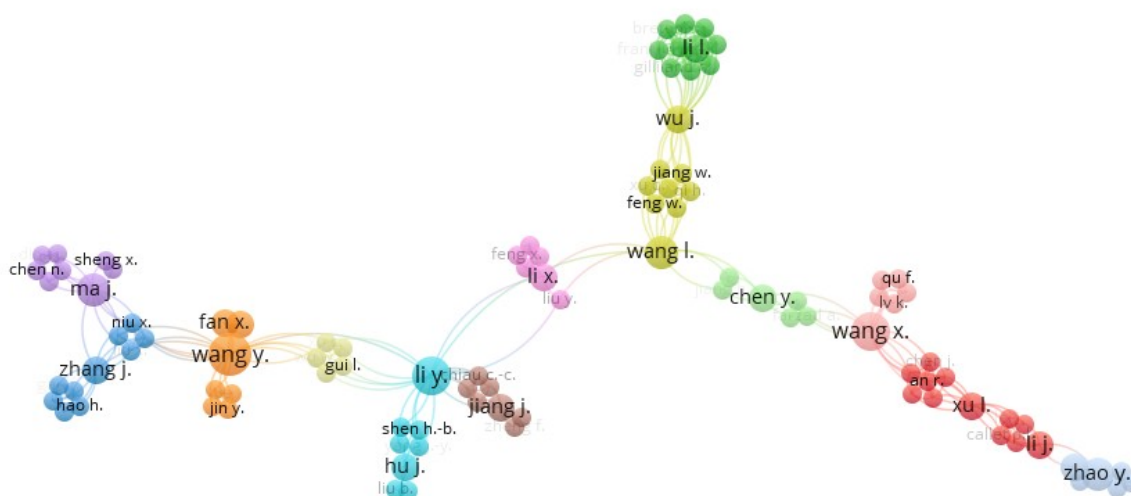
A disponibilidade de uma grande quantidade de séries históricas, que possibilitam o desenvolvimento do estudo de métodos de previsão, e a importância estratégica da CONAB foram os principais fatores para a escolha de sua base de dados como base para o trabalho. Além disso, a escolha dos dados da Conab e do tema se apoiou também no resultado de uma análise bibliométrica preliminar feita sobre o assunto.

Com o uso da base de dados Scopus (2020) e o auxílio do *Software* VOSview foi feita uma análise preliminar no assunto. Utilizando-se como palavras-chave “*forecasting, techniques, data e base*”, levando-se em conta o período de 2016 a 2020, identificou-se 214 artigos e um total de 1.540 citações, o que indica que o assunto tem relevância científica e vem sendo discutido no mundo.

Dentre os artigos identificados, o artigo “*A review on time series forecasting techniques for building energy consumption*” de Deb et al. (2017), citado por 172 artigos, garante embasamento teórico sobre séries temporais, tendência, sazonalidade, resíduos e técnicas de modelagem, fornece auxílio na identificação de casos onde é aplicável o modelo autorregressivo integrado de médias móveis (ARIMA), que é um dos principais modelos utilizados nesse trabalho, além de apresentar uma revisão teórica que auxilia na busca por novas fontes de informação.

De maneira geral, os principais autores citados (indicados pelos círculos maiores na Figura 1) apontam em seus artigos os benefícios do auxílio de softwares como ferramentas de previsão, como, por exemplo, Yang (2017) em “*Integrated Forecasting & Planning for Strategic & Operational Telecom Field Service Delivery*”, sendo um ponto importante para as escolhas dos softwares Excel e GRETL, que serão discutidos posteriormente, para a realização do trabalho.

Figura 1 – Principais redes de citação de acordo com os filtros indicados.



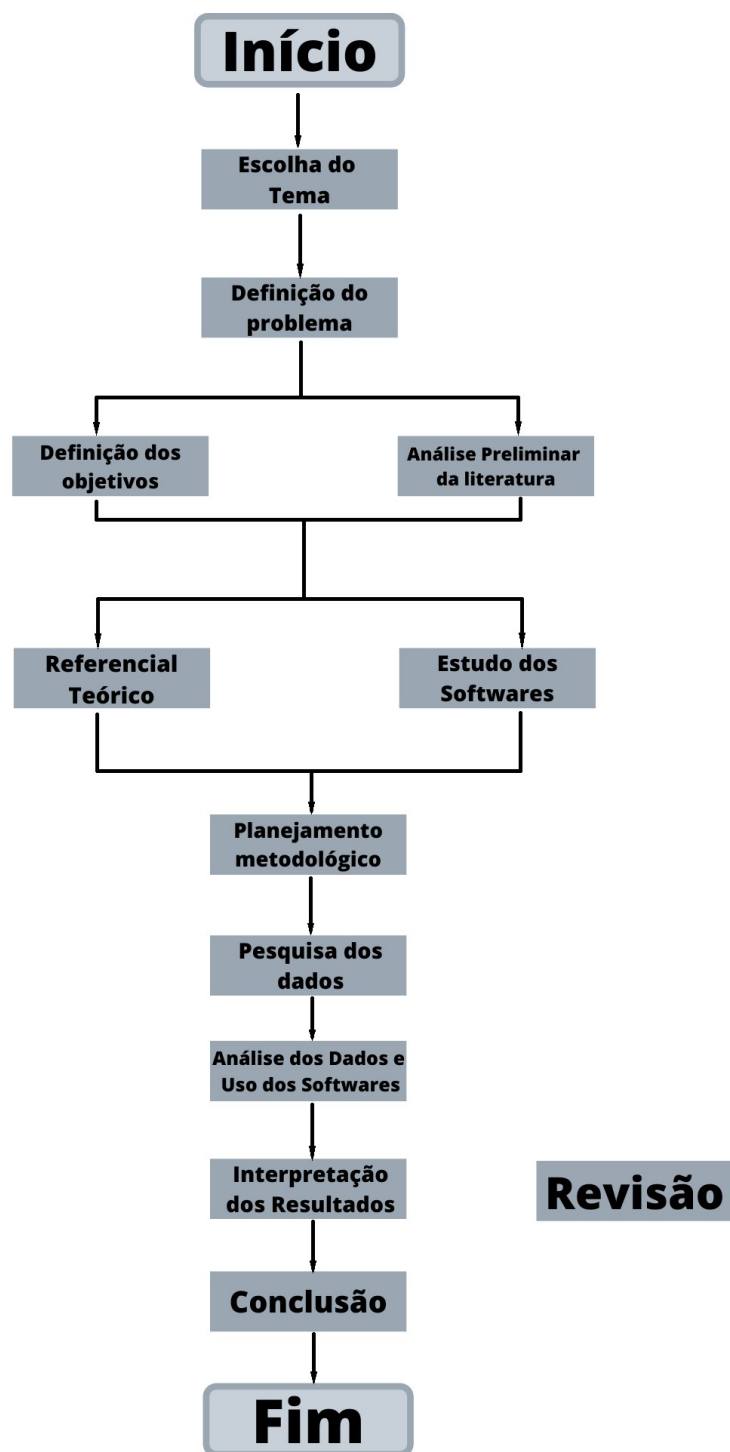
Fonte: Scopus (2020) e VOSview (2020).

Em outra análise, a partir das mesmas palavras-chave e levando em conta também a palavra “*agriculture*” foram encontrados apenas 15 artigos, quando se reduz a filiação de autores à região “*Brazil*” não foram encontrados artigos publicados, o que indica uma possível lacuna de pesquisa na relação entre técnicas de previsão e produção agrícola no país, que estabelece onde este trabalho se delimita. Esta lacuna que, também, poderá servir de motivação também para trabalhos futuros.

1.4 ETAPAS E CLASSIFICAÇÃO DA PESQUISA

A pesquisa a ser realizada se enquadra como uma pesquisa aplicada. A pesquisa aplicada pode ser definida como “conjunto de atividades nas quais conhecimentos previamente adquiridos são utilizados para coletar, selecionar e processar fatos e dados, a fim de se obter e confirmar resultados, e se gerar impacto” (FLEURY et al, 2017). Neste trabalho o tratamento dos dados propõe criar uma visão geral sobre determinado fenômeno e conhecer os fatos básicos que circundam a situação, proposta que, segundo Fleury (2017), se enquadra como um dos objetivos da pesquisa aplicada. Para que se possa planejar e organizar o trabalho foi desenvolvido o fluxo de atividades na Figura 2 e o cronograma da Figura 3.

Figura 2 – Fluxo de atividades para organização do trabalho



Fonte: Autor (2020).

Figura 3 – Cronograma das etapas.

Etapa / atividade	Mês 1				Mês 2				Mês 3				Mês 4				Mês 5				Mês 6				Mês 7				Mês 8				Mês 9			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4				
1																																				
2																																				
3																																				
4																																				
5																																				
6																																				
7																																				
8																																				
9																																				
10																																				
11																																				
12																																				

Fonte: Autor (2020).

As atividades a serem desenvolvidas estão descritas sucintamente na sequência:

- 1. Escolha do Tema:** Foi pesquisado e selecionado o tema. Foi escolhido um tema dentro da área de Pesquisa Operacional devido à importância de se ter familiaridade com os métodos de previsão no ambiente industrial.
- 2. Definição do problema:** Foram selecionadas as questões a serem respondidas no trabalho.
- 3. Definição dos objetivos:** Foram definidos os objetivos do trabalho de modo a responder as questões levantadas pelo problema.
- 4. Análise Preliminar da literatura:** Etapa onde serão pesquisados livros, artigos e monografias para auxiliar no embasamento do trabalho e delimitação do tema.
- 5. Referencial Teórico:** Serão apresentadas as técnicas estatísticas de previsão, e os conceitos a serem utilizados.
- 6. Estudo dos Softwares:** Os softwares serão estudados de modo a se identificar ferramentas adequadas para se cumprir os objetivos do trabalho.
- 7. Planejamento metodológico:** Serão selecionados os métodos estatísticos de previsão a serem utilizados, e como se dará o desenvolvimento do trabalho.
- 8. Pesquisa dos dados:** As séries históricas serão obtidas no site da Conab possibilitando a aplicação dos conceitos.
- 9. Análise dos Dados e Uso dos Softwares:** As séries históricas obtidas serão submetidas à aplicação de métodos de previsão de interesse por meio do uso dos softwares.
- 10. Interpretação dos Resultados:** Os dados gerados pelos softwares serão interpretados e relacionados com os conceitos utilizados.
- 11. Conclusão:** Após a interpretação dos resultados serão discutidas as respostas para as

questões de pesquisa que motivaram o trabalho.

12. Revisão: Revisão sistemática da organização do trabalho para aprimorar o desenvolvido nas etapas do trabalho.

1.5 ORGANIZAÇÃO DO TRABALHO

Este trabalho está estruturado em mais quatro capítulos. No Capítulo 2 está a fundamentação teórica, onde estão descritos, de forma sucinta, os conceitos que dão embasamento teórico para o trabalho. No capítulo 3 está descrita a metodologia do trabalho, onde são explicados os instrumentos de pesquisa, a unidade de pesquisa e como os dados serão tratados. No Capítulo 4 serão obtidos e tratados os dados de forma a se gerar resultados para que haja discussão do que foi obtido. Finalmente no Capítulo 5 estão as conclusões, respostas às questões de pesquisa e verificação dos objetivos, bem como sugestões para continuidade do trabalho, sendo seguidas das referências bibliográficas consultadas.

2 FUNDAMENTAÇÃO TEÓRICA

Nesse capítulo serão introduzidos os conceitos e ferramentas provenientes da revisão da literatura feita para o desenvolvimento do trabalho. O capítulo tem como objetivo explicar de forma sucinta tudo que dá embasamento teórico para a perspectiva do estudo realizado.

2.1 SÉRIES TEMPORAIS

Uma série temporal é uma sequência de observações ordenadas em intervalos iguais de tempo, tendo como principal característica a dependência entre os dados vizinhos. Yang et al. (2017), divide a análise de séries temporais em duas partes. Na primeira, deve-se obter a estrutura e o padrão subjacente dos dados observados, e na segunda, ajustar um modelo à série para fazer previsões futuras. Yang (2017) também destaca que o tempo pode ser substituído por outras variáveis, como espaço ou profundidade, desde que se leve em consideração a dependência entre as observações vizinhas.

Segundo Morettin e Toloi (2018), séries temporais podem ser contínuas ou discretas. Contínuas, quando as observações são feitas continuamente no tempo, e discretas por meio da amostragem de uma série temporal contínua em intervalo de tempos iguais. Deve ser observado que, estes termos se referem à série, e não a variável observada, esta pode assumir valores contínuos ou discretos.

No que se diz respeito aos possíveis objetivos do estudo de séries temporais, pode-se dividi-los em quatro áreas:

- **Descrever o comportamento da série:** construção do gráfico, diagramas de dispersão, verificação das propriedades da série, como sazonalidade, tendência alterações estruturais, *outliers* etc;
- **Explicação:** investigar o mecanismo gerador da série temporal;
- **Predição:** Predizer valores futuros com base no histórico de valores;
- **Controle de processos:** como, por exemplo, controle estatístico de qualidade.

O estudo realizado neste trabalho tem os objetivos dentro das categorias de descrição, ao analisar o comportamento da série e a dependência entre os dados para fazer a

modelagem correta, e predição, ao utilizar a modelagem para prever futuros valores nas séries estudadas.

2.2 DECOMPOSIÇÃO CLÁSSICA E SÉRIES ESTACIONÁRIAS.

Bastos (2016) afirma que a previsão baseada em modelos clássicos de decomposição de séries temporais é constituída pela filosofia de que um padrão existente em uma série histórica pode ser distinguido da aleatoriedade e decomposto em sub padrões que permitam identificar cada componente responsável pela variação dos dados.

A decomposição clássica seria então uma maneira de dividir estes sub padrões de acordo com a existência de três características: tendência, sazonalidade e ruído. Sendo a variação dos dados decorrente do somatório desses três efeitos. Yang et al. (2017) descreve estes efeitos da seguinte maneira:

- Efeito de tendência: Efeito que confere à série tendência de crescimento ou decrescimento ao longo do tempo, diz respeito ao movimento geral que a variável exibe durante o período de observação. Para a venda de um determinado produto, por exemplo, uma tendência de crescimento pode ser associada a aceitação do produto no mercado;
- Efeito sazonal (ou cíclico): Representa o fato de que a série tem comportamentos semelhantes em certos ciclos de tempo. Para a venda de sorvete, por exemplo, as vendas podem apresentar um crescimento no verão de cada ano;
- Variações ao acaso (ou ruído): Varáveis devidas a causas não identificadas que ocorrem em curtíssimo prazo e não são passíveis de previsão. Este efeito, por muitas vezes, pode mascarar a existência de tendência ou sazonalidade.

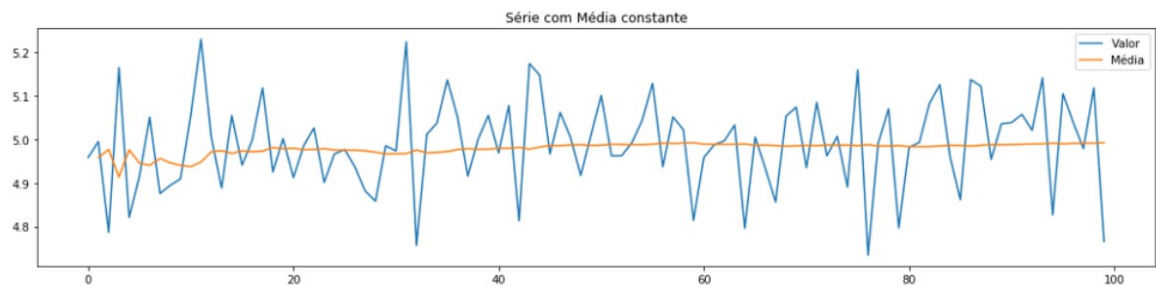
De acordo com esta divisão, a fórmula para a decomposição clássica então pode ser estabelecida na equação (1):

$$X_t = T_t + C_t + R_t \quad (1)$$

sendo que X_t é o valor obtido na série; T_t a componente de tendência; C_t a componente de sazonalidade; e R_t uma componente aleatória correspondente aos resíduos.

Quando não há a existência de tendência ou sazonalidade atribui-se toda a variação dos dados à componente aleatória, isso significa que a longo prazo a média e a variância dos dados podem ser consideradas constantes. Séries estacionárias seriam aquelas onde a variação dos dados se dá apenas devido à componente aleatória, conforme ilustrado na Figura 4.

Figura 4 - Valor da média (laranja) de acordo com a variação dos dados (azul).

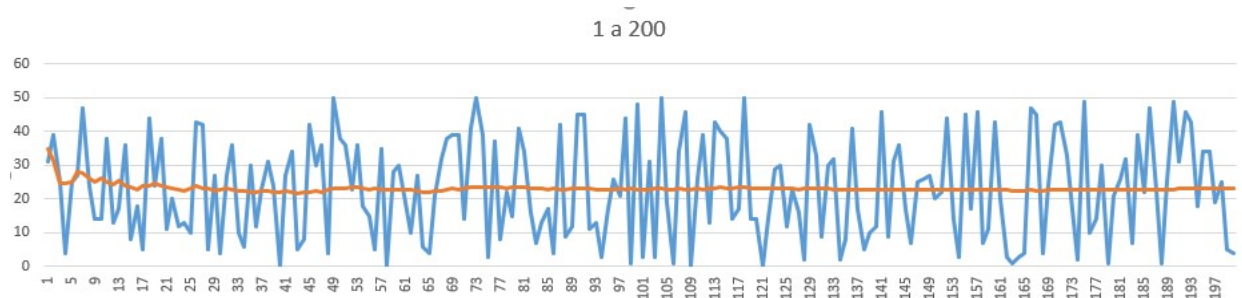


Fonte: Rabelo (2019).

2.3 TENDÊNCIA

Segundo Xavier (2016), as duas principais fontes de variação de dados em séries temporais são a sazonalidade e a tendência. A tendência está presente quando os dados apresentam crescimentos ou declínios em períodos sucessivos, isso acarreta o crescimento ou declínio da variância e da média, não podendo-se assumir mais a série como estacionária. Um dos principais problemas ao se trabalhar com séries tendenciosas está na escolha do intervalo de dados a serem utilizados, como está ilustrado na Figura 5.

Figura 5 - Valor da média (laranja) de acordo com a variação dos dados (azul).



Fonte: Autor (2020).

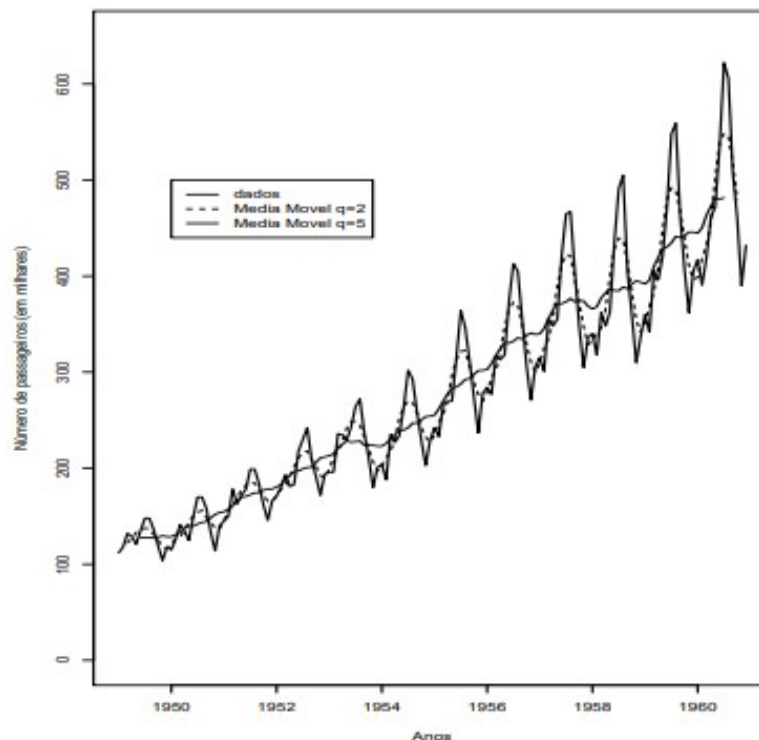
Segundo Antunes e Cardoso (2015) séries temporais podem apresentar trechos com diferentes tendências. Na Figura 5 quando são consideradas as 200 observações, o processo é visualmente estacionário, entretanto observe que caso fossem selecionados os dados de 180 a 189, por exemplo, os valores apresentariam uma tendência de crescimento, e caso de 190 a 200, os valores apresentariam declínio. Nota-se então que a identificação de tendência dentro de uma série depende do intervalo de dados a serem avaliados.

Antunes e Cardoso (2015) evidenciam que o comportamento observado na série pode variar de acordo com o intervalo de dados considerado. Nesse raciocínio, séries com dados muito antigos podem conter informações que não mais descrevem o sistema atual a ser analisado, assim como ao se considerar um intervalo muito curto de dados pode-se acabar por atribuir uma tendência a uma variação de caráter residual, esse processo de escolha dos dados está incluído dentro dos objetivos de estudo da descrição.

2.4 SAZONALIDADE

Yang et al. (2017) definem sazonalidade como a flutuação periódica do objeto de análise, são variações decorrentes de efeitos regulares e estáveis ao longo do tempo, sendo, em geral, significativa quando o intervalo de tempo entre as observações é menor que um ano. Visualmente essas flutuações apresentam um padrão de repetição facilmente reconhecível, como está exemplificado na Figura 6.

Figura 6 – Número de passageiros ao longo dos anos



Fonte: Ehlers (2005).

Pode-se identificar na Figura 6 que, além da notável tendência de crescimento, o gráfico apresenta um padrão de repetições ao longo dos anos, esse padrão, segundo Antunes e Cardoso (2015), seria a sazonalidade que, no caso, poderia ser descrita como a variação da demanda por viagens de avião de acordo com a época do ano. Nota-se também que a existência de um efeito não anula a variação decorrente de outros efeitos, o que é previsto pela decomposição clássica.

2.5 MÉTODO GRÁFICO E TESTES DE RAIZ UNITÁRIA

Ainda na Figura 5, pode-se observar que é possível se observar a existência de tendência na variação dos dados, visto que, entre o primeiro valor observado e o último houve uma quantidade de variações positivas consideravelmente maior (o valor inicial é cerca de 300% o valor inicial). É possível também, por meio deste gráfico, identificar um padrão, já discutido, de sazonalidade.

Na prática, esse tipo de gráfico auxilia, por meio da identificação visual das

componentes da decomposição clássica, na escolha dos modelos de previsão a serem implementados, isso porque modelos estatísticos de previsão tem como por objetivo se adequar a existência dessas componentes. Zuculoto et al. (2018) afirmam que a construção de um gráfico antes de se modelar uma série temporal possibilita se ter uma visão aproximada de tendências, movimentos cíclicos, sazonalidade e observações atípicas, o que possibilita uma melhor descrição da série e, conseqüentemente, selecionar um modelo de previsão mais adequado.

Além do método gráfico, existem testes, mais detalhados, que permitem uma conclusão sobre a existência de estacionaridade, são os chamados “testes de raiz unitária”. Como o próprio nome diz, esses testes se baseiam, na maioria dos casos, em se confirmar a hipóteses da existência de uma raiz unitária para série estudada. Xavier (2016) afirma que se deve utilizar mais de um teste para se certificar a estacionaridade da série.

Para o trabalho realizado serão levados em conta o método gráfico e os testes da raiz unitária presentes no *software* GRETl. Segundo Mattos (2018) ao se realizar um teste de raiz unitária em um *software* deve-se observar o p-valor, o qual se abaixo do nível de significância proposto acarretará a rejeição da hipótese de não-estacionariedade da série. Neste trabalho, serão utilizados os testes Dickey-Fuller aumentado, ADF-GLS (modificação do teste Dickey-Fuller aumentado) e o KPSS (sigla devida ao nome dos autores), para o KPSS deve-se comparar se o valor da estatística de teste está abaixo dos valores críticos. Após aplicação dos modelos estudados, será discutido se o intervalo observado no gráfico corresponde ao resultado da adequação dos modelos com a realidade.

2.6 MÉTODOS ESTATÍSTICOS DE PREVISÃO

A análise de séries temporais normalmente pode ser dividida em dois grupos, a análise univariada e a análise multivariada. Segundo Yang et al. (2017), séries univariadas são aquelas que contêm uma única observação registrada ao longo do tempo, como, por exemplo, volume de produção de grãos de café por ano. Séries multivariadas, por sua vez, são aquelas em que se considera que um grupo de variáveis da série temporal está envolvido na variação dos dados e a interação entre essas variáveis deve ser considerada.

Yang et al (2017) também destacam que a parte de ajuste de um modelo de série temporal é um processo complexo que demanda cálculos matemáticos detalhados, neste trabalho serão utilizados softwares para a realização desses cálculos. Nos tópicos a seguir serão explicados de maneira conceitual os modelos de séries temporais, univariadas, que

serão utilizados no trabalho.

2.6.1 Média móvel

De acordo com Almeida (2019) a média móvel é definida como a média aritmética de um determinado intervalo de observações mais recentes, como está descrito na equação (2):

$$M_t = \frac{X_{t-1} + X_{t-2} + \dots + X_{t-r+1}}{r} \quad (2)$$

sendo M_t é a média das últimas r observações dos valores das variáveis X_t .

Nesse método, M_t é a própria previsão para o próximo valor, e é sempre recalculada com a entrada de uma observação mais recente. Note que neste modelo todas as observações incluídas no intervalo r tem o mesmo peso.

De acordo com Xavier (2016), dado que séries temporais tem comportamento diferentes, na aplicação do método da média móvel deve-se:

- Utilizar uma quantidade maior de dados, para séries que apresentam muita aleatoriedade, modo que o cálculo da média móvel fique imune a ruídos e movimentos curtos;
- Utilizar uma quantidade menor de dados, para séries que apresentam pouca variação aleatória, de modo a reagir mais rápido a mudanças nas características da série.

Em geral, modelos de média móvel tem maior aplicabilidade em séries estacionárias.

2.6.2 Alisamento exponencial simples

O modelo de alisamento exponencial simples, assim como o modelo da média móvel, utiliza das observações registradas para determinar o comportamento da série. Entretanto, a diferença entre os dois modelos está, principalmente, na presença de uma constante de alisamento que atribui um peso para a observação mais recente. Segundo Xavier (2016) a fórmula do cálculo do alisamento exponencial pode ser dada pela equação (3):

$$F_{t+1} = \alpha X_t + (1 - \alpha)F_t \quad (3)$$

com F_{t+1} representando a previsão para o tempo $t+1$; X_t é a observação mais recente; α ($0 < \alpha < 1$) é a constante de alisamento (peso atribuído à observação); e F_t é o valor da previsão anterior.

Percebe-se que quanto maior o valor de α , mais importância se dá ao valor da última observação, ou seja, o modelo estará mais propenso a detectar rapidamente mudanças nas características da série. Entretanto corre-se o risco de, ao assumir um alto valor para α , ter uma previsão imprecisa devido à existência de um alto valor da componente aleatória para a última observação. Martin et al (2015) destaca que quanto menor o valor de α mais estáveis se mantêm os valores de previsão. Assim como modelos de média móvel, modelos de alisamento exponencial tem maior aplicabilidade em séries estacionárias.

2.6.3 ARIMA

Os Modelos Autorregressivos Integrados de Médias Móveis (ARIMA), segundo Yang et al. (2017), são a forma mais geral de técnicas de previsão. Eles se baseiam em transformar uma série temporal em estacionária, por meio do processo de diferenciação. Estes modelos são utilizados quando é detectada a presença de tendência na série.

Xavier (2016) descreve o processo de diferenciação em modelos ARIMA como a subtração de um dado ao dado anterior. É realizada uma diferenciação de 1ª ordem caso a série esteja crescer a uma taxa constante e de 2ª ordem se a série crescer com uma taxa crescente. De tal forma a equação ARIMA para uma série temporal é uma equação linear que tem como entrada defasagens da variável dependente associadas com defasagens do erro de previsão.

De acordo com o apresentado por Yang (2017) a saída (resultado previsto) do modelo pode ser descrita como uma constante mais a soma ponderada de uma ou mais previsões anteriores mais a soma ponderada de um ou mais valores anteriores estimados de erro. As defasagens do termo de erro não podem ser consideradas como variáveis independentes, pois não são funções lineares dos coeficientes. Os coeficientes dos modelos ARIMA com erros defasados devem ser calculados por outras técnicas como a *otimização não linear*.

Yang et al (2017) também destacam os termos do modelo da forma a seguir. Os termos de defasagem da série temporal estacionária são referidos como "autorregressivos" (AR), e os atrasos dos termos de erro previstos são referidos como "média móvel"

(MA). Uma série temporal que necessita de uma diferenciação para que possa ser analisada como estacionária é considerada uma versão "integrada" (I) de uma série estacionária.

Araújo e Souza (2016) denotam estes modelos como ARIMA (p, d, q) , em que p é a ordem (número de defasagens) do modelo autorregressivo, d é o grau de diferenciação (o número de vezes em que os dados tiveram valores passados subtraídos) e q é a ordem do modelo de média móvel. A parte autorregressiva p do modelo pode ser representada de acordo com a equação (4):

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + e_t \quad (4)$$

O modelo da média móvel q pode ser representado pela equação (5):

$$Y_t = c + \emptyset_1 e_{t-1} + \emptyset_2 e_{t-2} + \dots + \emptyset_p e_{t-p} + e_t \quad (5)$$

sendo Y é a saída da série temporal φ e $\emptyset$ são pesos atribuídos e e_t é a série de erro.

2.6.4 SARIMA

A fim de englobar o comportamento de sazonalidade presente em algumas séries temporais têm-se os chamados modelos ARIMA sazonais multiplicativos (SARIMA). Xavier (2016), descreve que os modelos SARIMA apresentam uma parte não sazonal, assim como no ARIMA, correspondentes aos parâmetros p , d e q , e uma parte sazonal que, de maneira alusiva ao visto anteriormente, corresponde aos parâmetros P , D e Q . Tendo-se assim uma componente sazonal autorregressiva SAR (P) e média móvel SMA (Q), em que D se refere à ordem de diferenciação sazonal. Pode-se também apresentar esse modelo como SARIMA $(p,d,q)(P,D,Q)$.

2.7 ERROS DE PREVISÃO

No que se diz respeito a modelos de previsão, uma das principais medidas para validação do modelo está no cálculo do erro associado. Segundo Freire (2017) a diferença entre os métodos estatísticos e outros métodos de previsão reside no fato de que os métodos estatísticos permitem calcular a probabilidade de erro. Serão detalhados alguns tipos de erro

a serem calculados.

2.7.1 Erros simples e erro absoluto de previsão

Azevedo (2016) afirma variabilidade de um conjunto de dados como sendo a maior ou menor diversificação dos valores em torno de uma medida de tendência central de previsão. Seguindo este raciocínio, erros simples de previsão são distância entre o valor real e o valor previsto. O cálculo deste erro se dá de acordo com a equação (6):

$$E_i = D_i - P_i \quad (6)$$

Onde E_i é o erro simples de previsão no período i ; D_i é a demanda observada no período i ; e P_i é a previsão estimada para o período i .

O erro absoluto, por sua vez, é dado pelo módulo do erro simples calculado.

2.7.2 Desvio padrão da previsão

O desvio padrão é uma medida de dispersão, ou seja, é uma medida que indica o quanto o conjunto de dados é uniforme. É uma importante ferramenta para se interpretar os erros de previsão. Segundo Azevedo (2016) o desvio padrão pode ser calculado de acordo com a equação (7).

$$S = \sqrt{\frac{\sum_{i=1}^n (e_i - \bar{e})^2}{n-1}} \quad (7)$$

Sendo S o desvio padrão de n períodos; e_i o erro simples no período i ; e $\bar{e}$ a média dos erros simples de n períodos.

2.7.3 Desvio médio absoluto

Morais et al. (2015) definem o desvio médio absoluto como sendo a média acumulada dos erros absolutos dos últimos períodos. Seu cálculo se dá de acordo com a equação (8):

$$DMA_n = \frac{1}{n} \sum_{i=1}^n EA_i \quad (8)$$

sendo DMA_n é o desvio médio absoluto do período n ; e EA_i o erro absoluto cometido no período i .

2.7.4 Intervalo de confiança

O Intervalo de Confiança é um dos conceitos mais importantes dentro dos testes de hipóteses na estatística por ser utilizado como uma medida de incerteza, seu cálculo depende diretamente das observações obtidas, também do desvio padrão. Segundo Azevedo (2016) ele é importante para indicar a imprecisão frente a um cálculo efetuado. E pode ser interpretado como a confiança (em porcentagem) de que o intervalo entre o limite inferior e limite superior contém o verdadeiro valor do parâmetro populacional.

Neste trabalho, o intervalo de confiança da previsão será gerado automaticamente por meio do Gretl para os modelos ARIMA e SARIMA, de acordo com o nível de confiança preestabelecido.

2.8 *SOFTWARES* PARA PREVISÃO E ANÁLISE DE SÉRIES TEMPORAIS

Morettin e Tollói (2018) afirmam que a operacionalização de modelos de previsão é complexa na prática, e para que haja rapidez na execução da análise, garantindo viabilidade prática do estudo, se faz necessário o uso *softwares* apropriados aos modelos de previsão escolhidos. No caso deste trabalho, serão utilizados os *softwares* Excel e Gretl para aplicação dos modelos.

2.8.1 Excel

O Excel é um Software comercializado pela empresa Microsoft. Segundo o site oficial da empresa (2020) o Excel é um software de planilhas de análise. O site destaca que o software dispõe de inúmeras ferramentas para tratamento, simulação e análise de dados. Como ferramenta de cálculo, o Excel possibilita a utilização de fórmulas e o recurso funções predefinidas, além da criação de fórmulas e lógicas dentro de suas planilhas.

Com base nos dados das folhas de cálculo é possível elaborar os mais variados tipos

de gráficos, ordenar listas de dados e criar listas de sub -totais, assim como recorrer a diversas ferramentas de simulação. No estudo a ser realizado o Excel servirá como ferramenta de produção de gráficos, e de cálculos principalmente para alisamento exponencial simples e média móvel.

2.8.2 Gretl

Segundo a *homepage* do site do Gretl (2020), o Gretl é um *software* multi-plataforma para análise econométrica, escrito em linguagem C. É livre e de código aberto. De acordo com o site, as bases de dados existentes no formato “planilha de Excel” podem ser importadas e utilizadas para a modelagem por meio do Gretl. Além disso, dentro das suas competências estão as análises de previsão por intermédio dos métodos ARIMA e SARIMA. No estudo a ser realizado no trabalho, o *software* será a ferramenta para a análise das séries da Conab utilizando esses métodos.

3 DESENVOLVIMENTO DA PESQUISA

Neste capítulo estão descritos com mais detalhes o problema a ser tratado, a unidade de pesquisa, os instrumentos de pesquisa e como será realizado o tratamento dos dados.

3.1 DESCRIÇÃO DO PROBLEMA A SER TRATADO

Segundo Martin et al (2015) a previsão de demanda é um processo para estimar condições futuras de vendas, mercado ou fatores que possam influenciar na estratégia e planejamento de uma organização. Séries temporais são sequências ordenadas obtidas em intervalos iguais de tempo que tem na dependência entre seus dados informações importantes sobre o objeto de estudo.

Para Xavier (2016) espera-se que ao se analisar uma série temporal seja observada uma causa relacionada com o tempo que influenciou o comportamento dos dados e pode influenciá-los no futuro. Dessa forma, pode-se, segundo Yang (2017), incorporar esta causa em um modelo matemático para que sejam realizadas previsões. Entretanto a escolha do modelo requer cuidados, para que seja selecionado um modelo que interprete de maneira precisa a variação dos dados da série.

Utilizar de um modelo adequado acarreta num melhor planejamento e, conseqüentemente, numa economia de recursos. Para Martin et al. (2015) a grande competitividade existente no ambiente empresarial leva as empresas a buscarem métodos de produzir de maneira mais eficiente e eficaz, fazer previsões precisas é uma ferramenta poderosa para uma empresa que busca se líder de mercado.

Yang et al. (2017) afirma que o ajuste de um modelo a uma série temporal é um processo desafiador. Ainda neste capítulo será apresentada uma metodologia para que seja possível descrever o comportamento das séries históricas e escolher um modelo de previsão adequado.

3.2 UNIDADE DE PESQUISA

A unidade de pesquisa deste trabalho será a Conab, especificamente suas séries históricas de produção de *commodities*. A Conab (Companhia Nacional de Abastecimento) é uma empresa pública, com sede em Brasília. A companhia foi criada em 1990 (tendo suas

atividades iniciadas no dia 1º de janeiro de 1991) após a fusão de três empresas públicas: a Companhia de Financiamento da Produção (CFP), a Companhia Brasileira de Alimentos (Cobal) e a Companhia Brasileira de Armazenamento (Cibrazem), segundo a *homepage* da página institucional da Conab (2017).

A empresa está presente em todas as regiões do Brasil e suas atividades visam a inclusão social, com ênfase em geração de emprego e renda, e a viabilização de análises do quadro de oferta e demanda sobre o cenário agropecuário nacional, em geral, a empresa busca criar inteligência para que o governo possa tomar decisões nesse âmbito.

A Conab possui importância estratégica ao oferecer ao Governo Federal informações técnicas para embasar a sua tomada de decisão quanto à elaboração de políticas voltadas à agricultura. Para isso, fornece informações detalhadas e atualizadas sobre a produção agropecuária nacional, por meio de levantamentos de previsão de safras [...] (CONAB, 2017, Homepage).

As áreas de atuação da empresa estão divididas em:

Abastecimento Social: que promove a segurança alimentar, garantindo o direito de todos ao acesso regular e permanente a alimentos de qualidade.

Agricultura Familiar: que dá apoio a projetos que contribuam para o fortalecimento social e econômico de organizações produtivas rurais de base familiar.

Armazenagem: cujo a principal finalidade é garantir o fluxo e abastecimento constante, de forma a proporcionar estabilidade e preços de mercado.

Comercialização: que reúne informações relativas a operações de comercialização realizadas pela Conab.

Estoques: que tem como objetivo executar a política governamental de intervenção no mercado para garantir o preço e a renda do produtor, e tornar regular o abastecimento interno.

Informações Agropecuárias: cujo objetivo é munir os diversos agentes envolvidos no desenvolvimento da agropecuária brasileira de dados e informações.

Preços Mínimos: área que atua como balizadora de oferta de alimentos para diminuir oscilações na renda dos produtores rurais e assegurar uma remuneração mínima.

Dentre as informações geradas pela criação de inteligência da Conab, têm-se as séries históricas de produção de *commodities* no Brasil, que servirão de base de dados para o trabalho.

3.3 INSTRUMENTOS DE PESQUISA

Os instrumentos de pesquisa deste trabalho, são as ferramentas de modelagem e análise de dados disponíveis nos *softwares* Excel e Gretl.

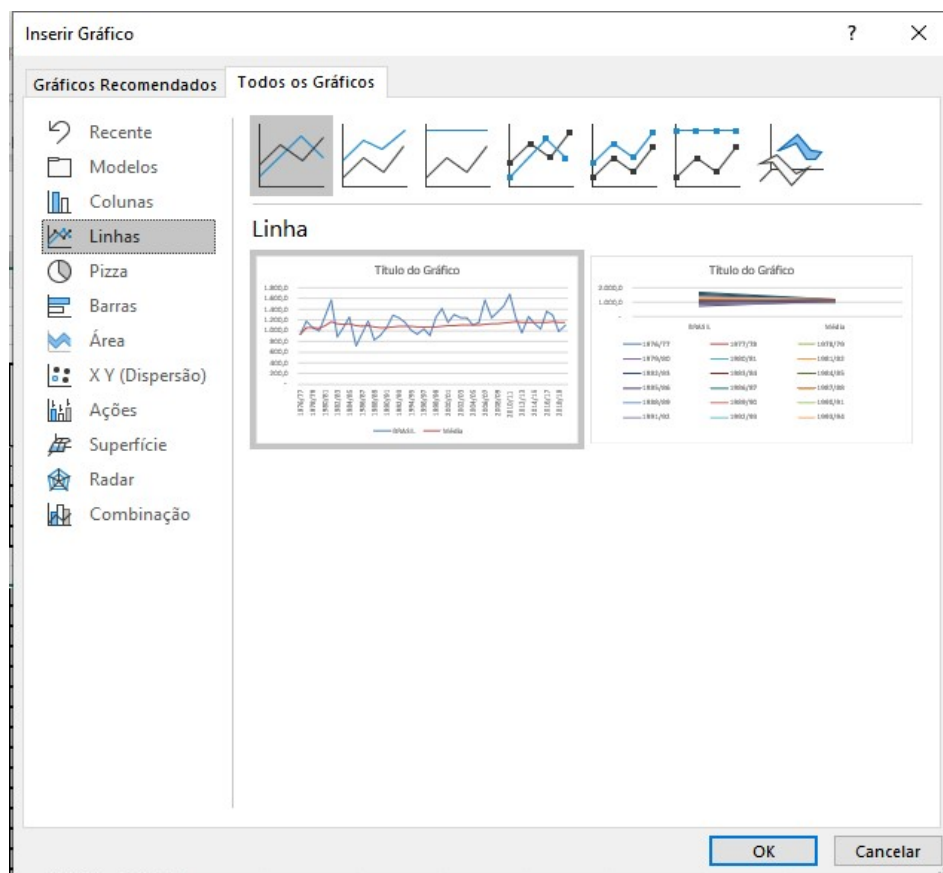
3.3.1 Ferramentas do Excel

Serão apresentadas as ferramentas do Excel utilizadas neste trabalho.

3.3.1.1 Plotagem de gráficos

Um dos principais recursos utilizados neste trabalho é a construção de gráficos para descrição das séries históricas. Entre os recursos do Excel está a plotagem de gráficos, vide a interface mostrada pela Figura 7.

Figura 7 – Interface para construção de gráficos no Excel.



Fonte: Autor (2020).

Neste trabalho serão utilizados gráficos de linha, que são exemplificados na Figura 8 por meio da plotagem do gráfico da variação da produção anual de feijão no Brasil (em mil toneladas) e a variação da média dessa produção ao longo do tempo.

Figura 8 - variação da produção anual e da média da produção anual de feijão no Brasil.



Fonte: Autor (2020).

3.3.1.2 Tabelas, funções e contas

As séries históricas da Conab são fornecidas por meio de dados organizados em planilhas do Excel. A disposição de células do Excel permite a construção de tabelas e planilhas de contas, além de existirem funções dentro do programa que auxiliam em cálculos conhecidos, há um trecho de uma tabela na Figura 9.

Figura 9 – Trecho de tabela feita no Excel.

ANO	BRASIL	MÉDIA	MÉDIA MOVEL	ERRO	SIMPLES
1976/77	935,0	935,0	-		
1977/78	1.179,6	1.057,3	-		
1978/79	1.054,5	1.056,4	-		
1979/80	1.000,0	1.042,3	-		
1980/81	1.274,0	1.088,6	-		
1981/82	1.566,0	1.168,2	-		
1982/83	890,3	1.128,5	1.128,5		
1983/84	1.058,7	1.119,8	1.146,2	-	69,8
1984/85	1.257,3	1.135,0	1.157,3	-	111,1
1985/86	714,0	1.092,9	1.108,6	-	443,3
1986/87	947,2	1.079,7	1.101,1	-	161,4
1987/88	1.180,4	1.088,1	1.087,7	-	79,3
1988/89	823,0	1.067,7	981,6	-	264,7
1989/90	911,0	1.056,5	984,5	-	70,6
1990/91	1.069,8	1.057,4	986,1	-	85,3
1991/92	1.292,3	1.072,1	991,1	-	306,2
1992/93	1.234,1	1.081,6	1.065,4	-	243,0

Fonte: Autor (2020).

Na Figura 9 foi criada uma tabela em que as colunas “média” e “média móvel” foram construídas por meio da função “=média” presente no Excel, que permite o cálculo da média de um determinado intervalo de dados. Entretanto, a coluna “erro simples” foi construída utilizando-se uma fórmula inserida manualmente. Há também comandos dentro do Excel, como o “atingir meta”, que altera o valor de uma variável de interesse até que seja obtido o resultado desejado em uma equação.

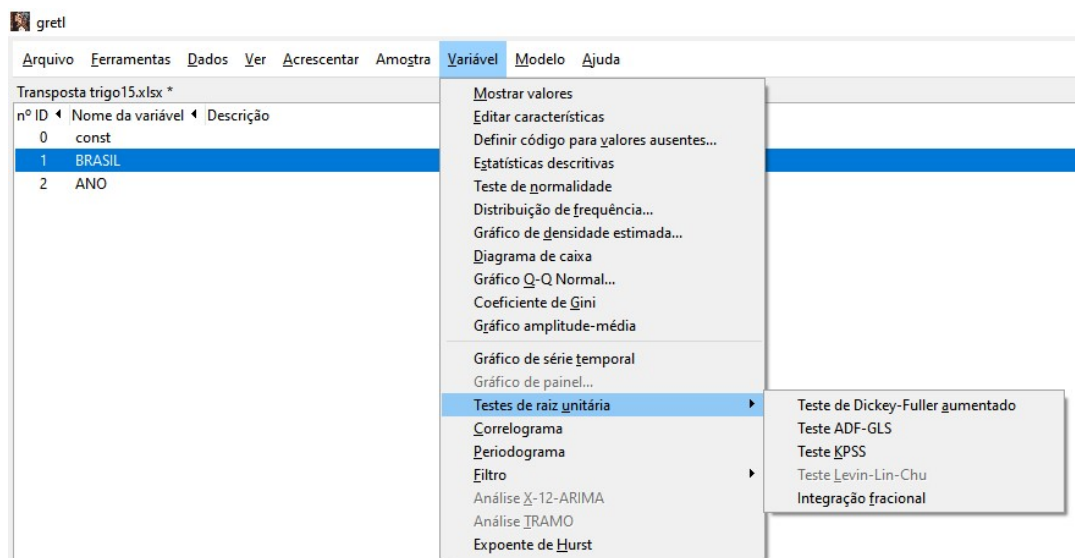
3.3.2 Ferramentas do Gretl

Serão apresentadas as ferramentas do Gretl utilizadas neste trabalho. Vale ressaltar que os dados de séries históricas utilizados no Gretl podem ser importados das planilhas do Excel mantendo a disposição.

3.3.2.1 Testes de raiz unitária

Entre as ferramentas de análise de dados do GRETL existem três testes de raiz unitária que serão utilizados para a verificação de estacionariedade em uma série. Sendo eles o Dickey-Fuller aumentado, o ADF-GLS e o KPSS. Para se aplicar um destes testes, após importados os dados para o GRETL, basta ir na aba “Testes de Raiz Unitária” dentro da aba “Variável” e selecionar um dos testes, como indicado na Figura 10.

Figura 10 – Seleção de abas para utilização dos testes de raiz unitária.



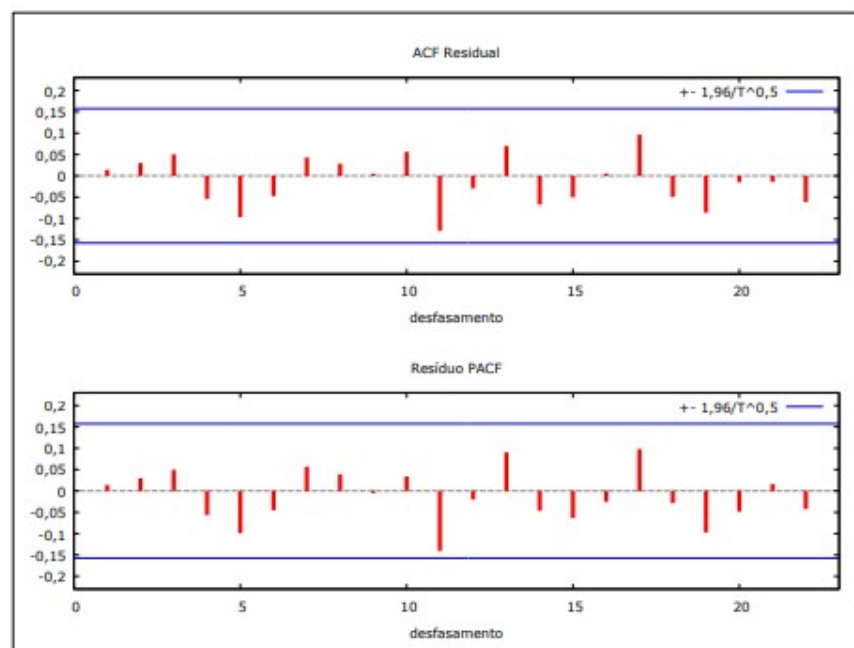
Fonte: Autor (2021).

3.3.2.2 Identificação dos parâmetros ARIMA

A identificação ARIMA por meio do software Gretl requer um passo a passo para que se possam ser identificados de maneira coerente os coeficientes p , d e q . Após analisados os gráficos por meio do Excel e importados os dados para o GRETTL deve-se primeiro construir um correlograma (utilizando a opção “correlograma” do GREL), uma ferramenta usada para verificar a autocorrelação entre os dados.

Se os dados do correlograma obtido forem próximos de zero, pode-se considerar a série estacionária. Entretanto caso isso não ocorra, deve-se diferenciar as variáveis (utilizando a função “primeiras diferenças” do GRETTL) até que isso ocorra, sendo d o número de vezes que necessárias até que o correlograma indique dados próximos de zero. Na Figura 11 há um exemplo de correlograma.

Figura 11 – Exemplo de correlograma.



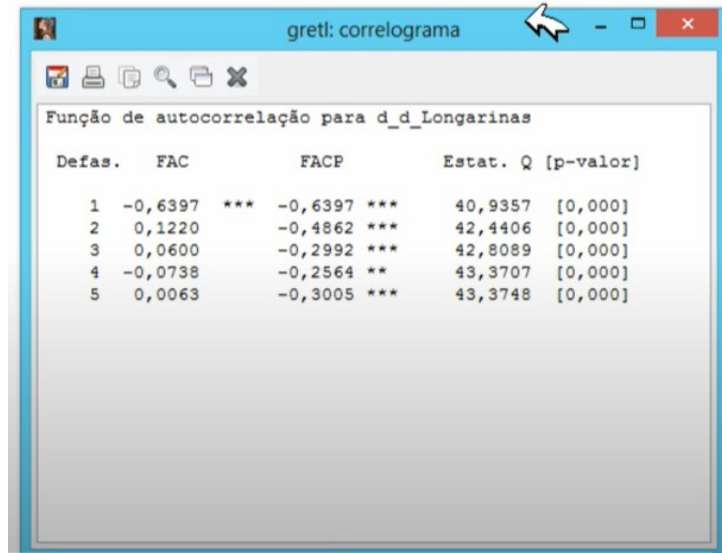
Fonte: Xavier (2016).

Nota-se que nesta Figura 10 o valor dos dados exibidos pelo correlograma está próximo de zero, podendo então assumir a ordem de diferenciação como o parâmetro d da modelagem ARIMA. Caso houvesse muitos valores acima da linha azul, seria necessária outra defasagem.

Após obtido o parâmetro d deve-se construir outro correlograma, em relação aos

dados da última defasagem e observar os valores de FAC (função autocorrelação) e FACP (função autocorrelação parcial), conforme ilustrado na Figura 12.

Figura 12 – Valores para FAC e FACP.



Fonte: Silva (2015).

É calculado então um ponto de referência para escolha dos parâmetros de acordo com a estimativa dada pela equação (9):

$$O = \frac{2}{\sqrt{n}} \quad (9)$$

sendo O o ponto de referência; e n o número de dados utilizados.

Para se obter o parâmetro p deve-se observar o último valor, em módulo, acima do ponto de referência para a coluna de FACP e escolher o valor correspondente na coluna “defasagem”, e para o parâmetro q deve-se observar o último valor em módulo acima do ponto de referência para FAC e escolher o valor correspondente para a coluna “defasagem”. Vale ressaltar que serão sempre utilizados valores dentro do intervalo de confiança pré-estabelecido.

3.3.2.3 Identificação dos parâmetros SARIMA

A identificação dos parâmetros SARIMA ocorre de maneira semelhante à da

modelagem ARIMA. A obtenção dos parâmetros d , p e q se dão da forma vista anteriormente para o modelo ARIMA. A maneira de se obter os parâmetros se dá de forma semelhante.

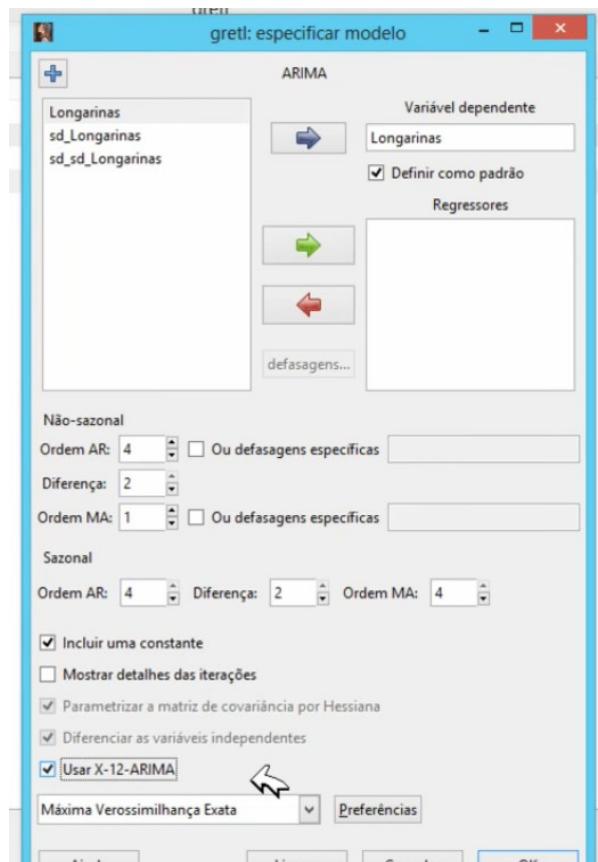
Após obtidos os parâmetros para o modelo ARIMA deve-se voltar aos valores originais da série e realizar as defasagens utilizando a função “diferenças de sazonalidade”. Da mesma forma como acontece para d , quando os valores do correlograma estão próximos a zero se obtém o parâmetro D .

Constrói-se então um outro correlograma em relação aos dados da última defasagem e se observa os valores de FAC (função autocorrelação) e FACP (função autocorrelação parcial). Para se obter o parâmetro P deve-se observar o último valor, em módulo, acima do ponto de referência, já calculado, para a coluna de FACP e escolher o valor correspondente na coluna “defasagem”, e para o parâmetro Q deve-se observar o último valor em módulo acima do ponto de referência para FAC e escolher o valor correspondente para a coluna “defasagem”.

3.3.2.4 Modelagem

A modelagem ocorre após obtidos os parâmetros, escolhendo-se a opção “ARIMA” na aba de modelos. Após selecionada a opção entra-se com os parâmetros do modelo, conforme mostrado na Figura 13.

Figura 13 – Interface para aplicação do modelo ARIMA.



Fonte: Silva (2015)

Aplica-se então o modelo e se obtém os coeficientes. Então, avaliando-se o *p-value* se adapta o modelo, para que não haja coeficientes com *p*-valores acima do estabelecido para confiança que foi determinada, alterando os parâmetros identificados anteriormente. A Figura 14 mostra como os coeficientes são apresentados no *software*.

Figura 14 – coeficientes obtidos na modelagem ARIMA.

Variável dependente: $(1-L)^2(1-L_s)^2$ Longarinas

	coeficiente	erro padrão	z	p-valor
const	-1,67969	5,55603	-0,302	0,7624
phi_1	-0,578314	0,0809342	-7,145	8,97e-013 ***
phi_2	-0,247086	0,0953146	-2,592	0,0095 ***
phi_3	0,0209074	0,0906926	0,230	0,8177
phi_4	-0,0419878	0,0790441	-0,531	0,5953
Phi_1	-1,15530	0,0774545	-14,92	2,60e-050 ***
Phi_2	-0,424621	0,0880914	-4,820	1,43e-06 ***
theta_1	-0,999998	0,0351539	-28,45	5,41e-178 ***
Theta_1	-0,208319	0,132068	-1,577	0,1147
Theta_2	-0,304698	0,129600	-2,351	0,0187 **
Theta_3	0,896577	0,155818	5,754	8,72e-09 ***
Média var. dependente	211,9178	D.P. var. dep.		
Média de inovações	134,9000	D.P. das inovações		3218,522
Log da verossimilhança	-723,6562	Critério de Akaike		1471,312
Critério de Schwarz	1498,798	Critério Hannan-Quinn		1482,266

Fonte: Silva (2015).

3.4 TRATAMENTO DOS DADOS

Para se atender aos objetivos propostos, os dados serão tratados da seguinte forma:

- **Obtenção da série:** obtenção da série histórica da *comodity* desejada, por meio da base de dados da Conab;
- **Análise da série:** Análise do comportamento da série de dados, com auxílio do Excel e do teste da raiz unitária do GRET, para que posteriormente se possa discutir o que era esperado e o que se obteve;
- **Aplicação do modelo:** Aplicação dos modelos levando em conta o comportamento observado e o intervalo selecionado;
- **Cálculo dos erros:** Cálculo dos erros para que se possa ser discutida a precisão de cada modelo;
- **Resultados:** Análise e comparação dos resultados.

4 RESULTADOS

Neste capítulo os dados serão obtidos e tratados, trazendo a discussão e a análise dos resultados obtidos.

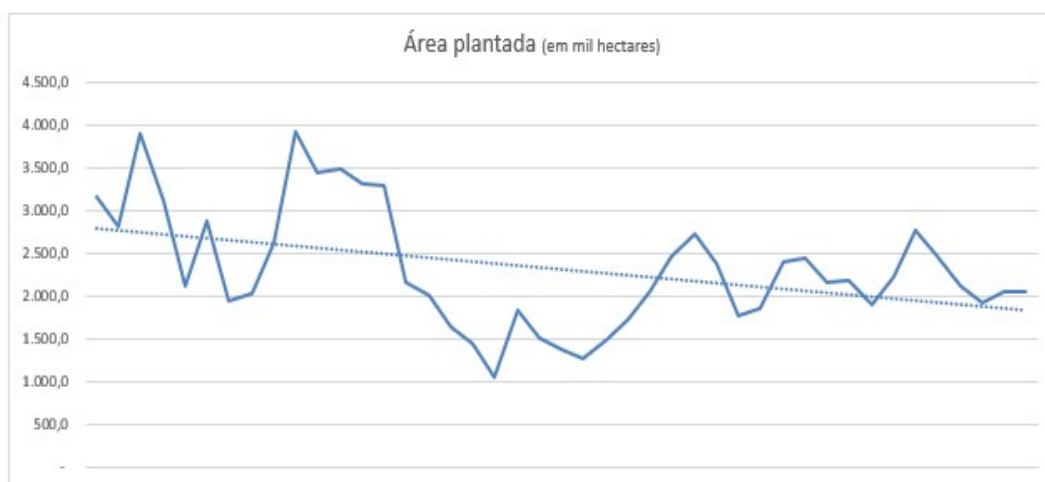
4.1 TRIGO

Serão analisadas as séries históricas da Conab da área de plantação, da produtividade e produção de trigo. Os dados das séries históricas anuais referente ao trigo disponibilizadas pela Conab tem início em 1977. As séries serão analisadas levando em conta todo o período, entretanto serão analisados trechos de interesse para aplicação de alguns modelos.

4.1.1 Série histórica da área de plantação de trigo.

A plotagem do gráfico da série de área plantada de trigo, referente ao período de 1977 a 2019, está apresentada na Figura 15.

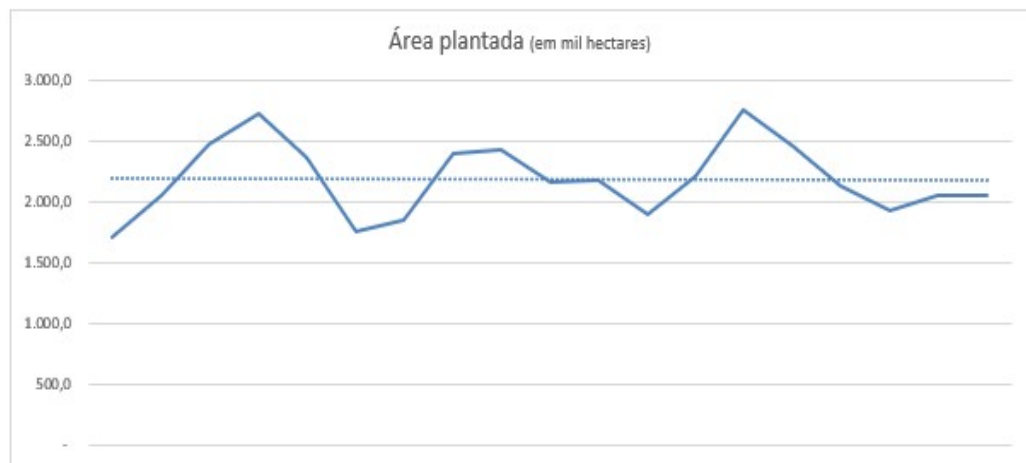
Figura 15 – Variação da área plantada de trigo 1977 a 2019.



Fonte: Autor (2021).

Foi adicionada no gráfico uma linha de tendência para o auxílio da visualização. É observado que o gráfico apresenta uma tendência de decréscimo, entretanto ao se selecionar um período mais recente (2001 – 2019) se observa uma possível estacionariedade, como pode ser observado na Figura 16.

Figura 16 - Variação da área plantada de trigo de 2001 a 2019.



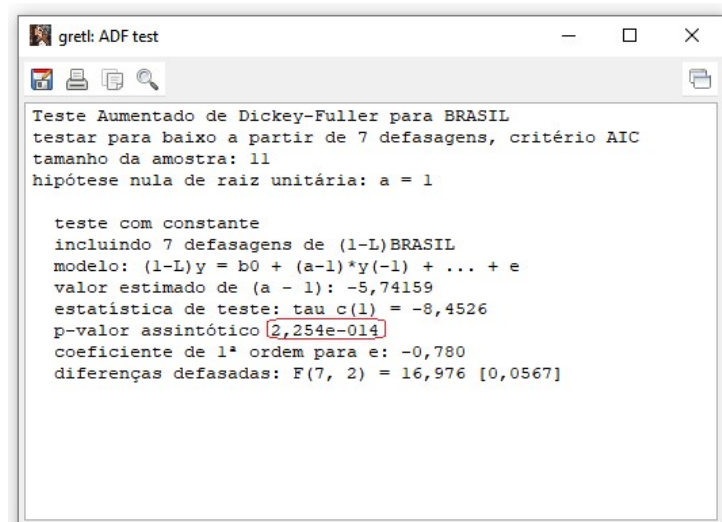
Fonte: Autor (2021).

Vale ressaltar que, como o período entre as observações é de um ano, não se mostra interessante considerar a presença de sazonalidade.

4.1.1.1 Teste da estacionariedade

Como visto na Figura 15 ao se considerar todo o intervalo de dados há uma forte presença de tendência, entretanto para o intervalo considerado na Figura 16 tem-se uma possível estacionariedade que pode ser confirmada pela realização dos testes. O p-valor obtido para o teste Dickey-Fuller aumentado está indicado na Figura 17.

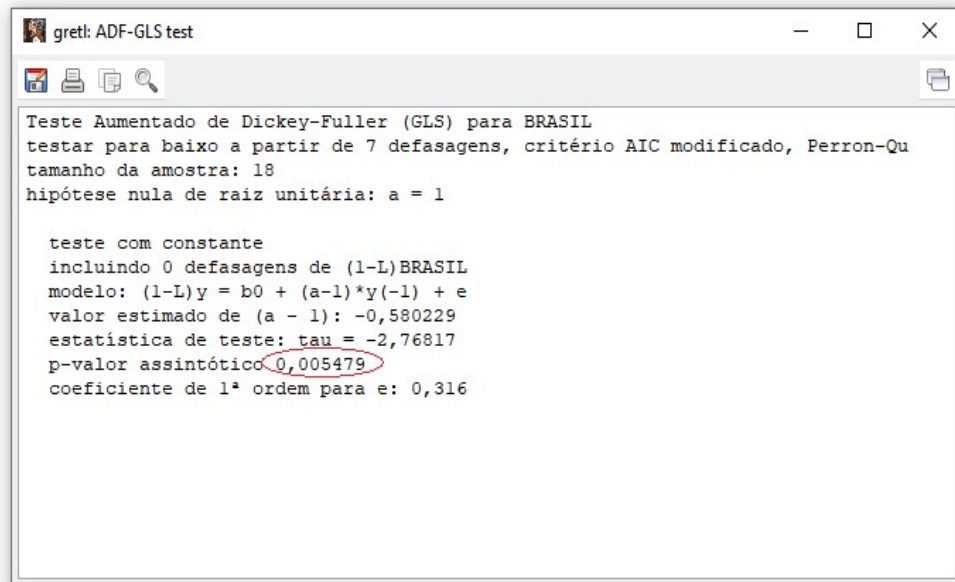
Figura 17 – p-valor para o teste Dickey-Fuller aumentado.



Fonte: Autor (2021).

É possível observar que o p-valor está abaixo 0,05 ou até mesmo 0,01. O p-valor obtido para o teste ADF-GLS está indicado na Figura 18.

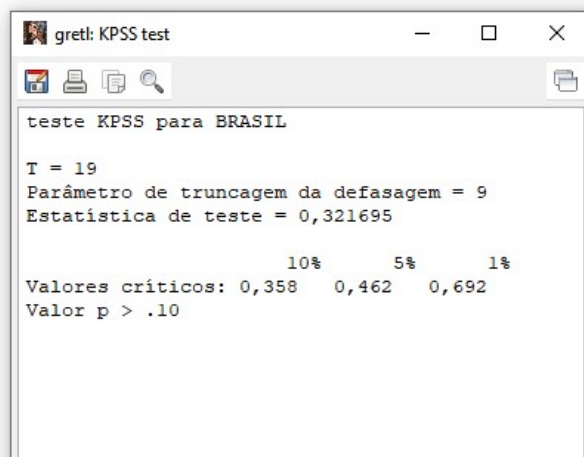
Figura 18 – p-valor para o teste ADF-GLS.



Fonte: Autor (2021).

Nota-se que, assim como no teste anterior, o p-valor obtido está abaixo de 0,05 ou até mesmo 0,01. A estatística de teste e os valores críticos obtidos para o teste KPSS estão indicados na Figura 19.

Figura 19 – Estatística de teste e valores críticos para o teste KPSS.



Fonte: Autor (2021).

É possível observar que a estatística de teste está abaixo dos valores críticos obtidos. Conclui-se então que para o intervalo de dados selecionado a série é estacionária.

4.1.1.2 Modelo ARIMA

Para se fazer a previsão da área para o ano seguinte para o intervalo de dados de 1977 a 2019 será utilizado o modelo ARIMA, devido ao intervalo entre as observações ser maior que um ano e devido a forte presença de tendência. A princípio será identificado se a presença de tendência realmente existe. A Figura 20 apresenta o correlograma da série.

Figura 20 – correlograma da série referente a área plantada de trigo.

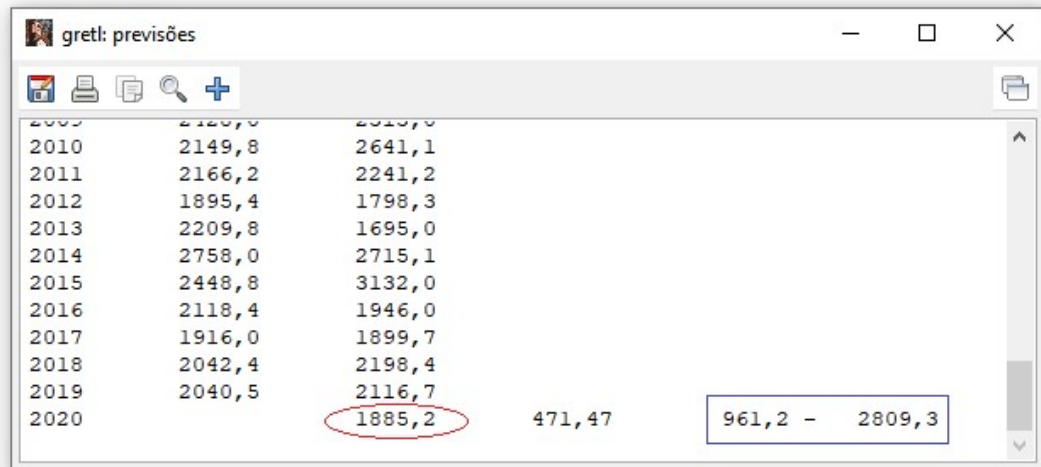
Função de autocorrelação para BRASIL
 ***, **, * indicam significância aos níveis de 1%, 5% e 10%
 usando erro padrão $1/T^{0,5}$

Defas.	FAC		FACP		Estat. Q [p-valor]
1	0,7063 ***		0,7063 ***		22,9839 [0,000]
2	0,4971 ***		-0,0035		34,6472 [0,000]
3	0,3013 **		-0,0970		39,0378 [0,000]
4	0,1410		-0,0733		40,0240 [0,000]
5	0,0061		-0,0804		40,0259 [0,000]
6	0,0194		0,1689		40,0456 [0,000]
7	0,0343		0,0257		40,1089 [0,000]
8	0,0005		-0,1135		40,1089 [0,000]
9	-0,0308		-0,0525		40,1628 [0,000]
10	0,0007		0,0973		40,1629 [0,000]
11	-0,0233		-0,0326		40,1957 [0,000]
12	-0,1076		-0,1716		40,9182 [0,000]
13	-0,1594		-0,0813		42,5571 [0,000]
14	-0,1599		0,0506		44,2637 [0,000]
15	-0,1629		0,0343		46,0971 [0,000]
16	-0,1543		-0,0545		47,8025 [0,000]
17	-0,0977		-0,0180		48,5135 [0,000]
18	-0,1138		-0,1155		49,5160 [0,000]
19	-0,1566		-0,0450		51,4935 [0,000]
20	-0,2175		-0,1108		55,4728 [0,000]
21	-0,1741		0,0853		58,1398 [0,000]
--	--		--		--

Fonte: Autor (2021).

É observado que na décima linha o valor de FAC quase zero, entretanto nas linhas seguintes ele volta a subir, isto indica que há presença de tendência na série. Confirmada a existência de tendência o modelo ARIMA é aplicado. A previsão da área plantada para o ano de 2020 resulta de um modelo ARIMA com $d=2$, $p=5$ e $q=2$ e está indicada dentro da elipse vermelha na Figura 21.

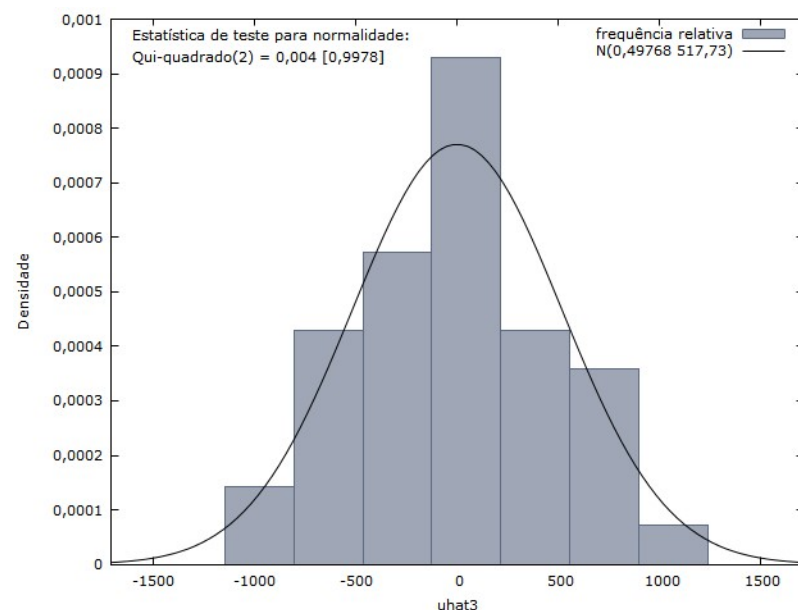
Figura 21 – Previsão ARIMA da área plantada de trigo para 2020.



Fonte: Autor (2021)

Nota-se que os outros valores previstos pelo modelo ARIMA (terceira coluna) estão próximos dos valores reais (segunda coluna). Note também que foi gerado um intervalo de confiança da previsão a 95%, que está dentro do retângulo azul. Além do observado nas previsões feitas pelo ARIMA uma ferramenta de auxílio para confirmar a aderência dos modelos é verificar a normalidade dos resíduos, a Figura 22 apresenta a distribuição dos resíduos.

Figura 22 – Distribuição dos resíduos ARIMA para a área plantada de trigo.



Fonte: Autor (2021).

É possível observar que os resíduos representados pelas colunas apresentam uma distribuição semelhante à da curva normal, o que indica uma boa aderência do modelo à série.

4.1.1.3 Média móvel e Alisamento exponencial

Para a aplicação da média móvel e do alisamento exponencial utilizou-se o intervalo de 2001 a 2019. A Tabela 1 mostra, utilizando $r=10$, as médias móveis de 2010 a 2019.

Tabela 1 – Médias móveis para a área plantada de trigo.

Ano	Área	Média Móvel
2010	2.149,8	2.189,8
2011	2.166,2	2.235,4
2012	1.895,4	2.219,8
2013	2.209,8	2.194,4
2014	2.758,0	2.197,5
2015	2.448,8	2.206,2
2016	2.118,4	2.242,2
2017	1.916,0	2.248,7
2018	2.042,4	2.213,3
2019	2.040,5	2.174,5

Fonte: Autor (2021).

A previsão para 2020 será então o último valor da coluna das médias móveis. Nota-se que os valores das outras previsões realizadas estão próximos aos valores reais, apresentados na segunda coluna.

Para o cálculo do alisamento exponencial foi estimado um valor para α utilizando o comando atingir meta do Excel de forma a alterar o valor de α na equação do alisamento exponencial até que a previsão para 2018 fosse exatamente o valor real para 2018. As previsões obtidas para o alisamento exponencial estão apresentadas na Tabela 2.

Tabela 2 – Previsão da área plantada de trigo utilizando alisamento exponencial.

Ano	Área	Alisamento exponencial
2010	2.149,8	2.224,1
2011	2.166,2	2.187,9
2012	1.895,4	2.005,2
2013	2.209,8	2.133,0
2014	2.758,0	2.523,5
2015	2.448,8	2.476,8
2016	2.118,4	2.252,9
2017	1.916,0	2.042,4
2018	2.042,4	2.042,4
2019	2.040,5	2.041,2
alfa =	0,625	

Fonte: Autor (2021).

A previsão para 2020 será então o valor da última linha da terceira coluna. É possível observar que os valores de previsão obtidos se mostram próximos aos valores reais.

4.1.1.4 Cálculo dos erros.

O valor real obtido pela Conab para a área plantada de trigo em 2020 foi 2341,5 (em mil hectares) e será utilizado para o cálculo dos erros dos modelos. Por sua vez, os valores obtidos para o erro absoluto e o desvio médio absoluto estão apresentados na Tabela 3.

Tabela 3 – Erro absoluto e desvio médio absoluto para área plantada de trigo.

	Valor Real 2020	Valor obtido 2020	Erro Absoluto	Desvio médio absoluto	% de Erro (Erro Absoluto)
ARIMA	2341,5	1885,2	456,3	374,8	19,5%
Média Móvel	2341,5	2174,5	167	214,9	7,1%
Alisamento Exponencial	2341,5	2041,2	300,3	290,65	12,8%

Fonte: Autor (2021).

4.1.1.5 Discussão dos resultados

A partir dos valores obtidos de erro pode-se identificar a precisão dos modelos. Para a série avaliada o modelo do alisamento exponencial se mostra preciso, isso indica que o comando “atingir meta” do Excel se mostrou útil e estimar o valor de α com base no valor real de 2018, uma observação recente, garantiu um valor coerente.

Em outro aspecto a previsão utilizando a média móvel foi o mais preciso, o que indica que para séries estacionárias a média móvel, assim como o alisamento exponencial, se mostra um modelo de previsão coerente. Entretanto, vale-se notar que para aplicação desses dois modelos não se levou em conta todo o intervalo de dados disponíveis, pois a série apresentaria outro comportamento, sendo então uma limitação para esses modelos, o que pode ser corrigido com a escolha de um intervalo adequado.

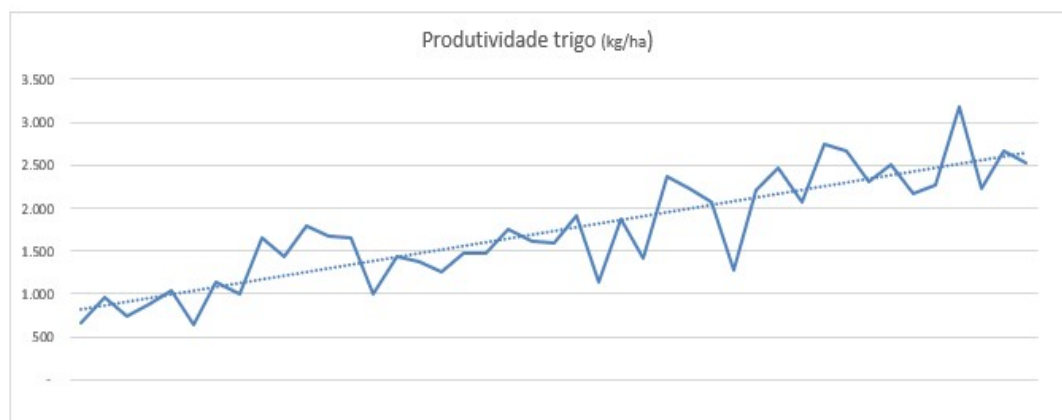
Apesar de na previsão para 2020 o modelo ARIMA apresentar o maior erro, ao se analisar as previsões realizadas o modelo apresentou valores bem próximos à realidade para outros anos. Vale ressaltar que o valor obtido está garantido dentro do intervalo de confiança do modelo.

Deve-se lembrar que dados muito antigos podem não ser muito eficientes para auxiliar na descrição do cenário atual ao que se remete uma série histórica, considerá-los ou não influencia na verificação das componentes da decomposição clássica. É necessário levar em conta qual é o contexto da análise da série.

4.1.2 Série histórica da produtividade em relação ao trigo

A plotagem do gráfico da série da produtividade de trigo, referente ao período de 1977 a 2019, está apresentada na Figura 23.

Figura 23 - Variação da produtividade em relação ao trigo de 1977 a 2019.



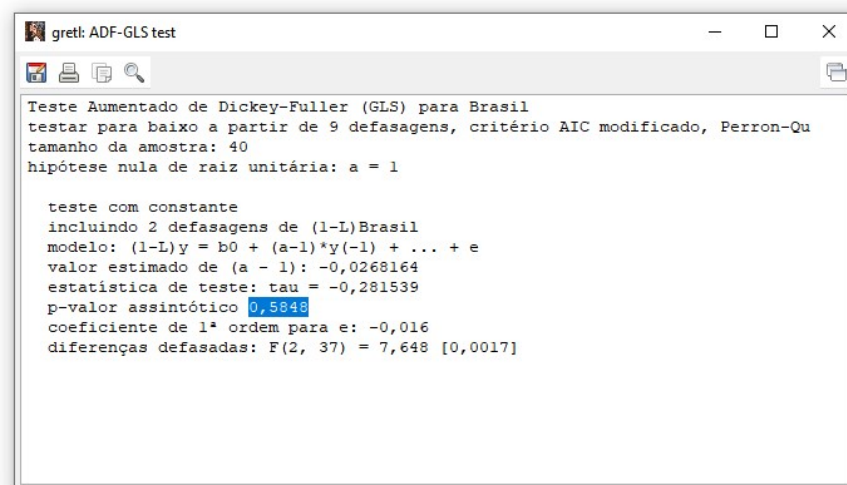
Fonte: Autor (2021).

Nota-se que aparentemente não há a estacionariedade na série mesmo ao se selecionar trechos do gráfico. A linha de tendência adicionada indica uma forte presença de uma tendência de crescimento. Por meio da visualização do gráfico o modelo mais indicado para a previsão é o modelo ARIMA, devido a não estacionariedade.

4.1.2.1 Modelo ARIMA

A princípio será identificado se a presença de tendência realmente existe para a série de produtividade de trigo referente ao período de 1977 a 2019. A Figura 24 apresenta o teste de raiz unitária ADF-GLS para a série

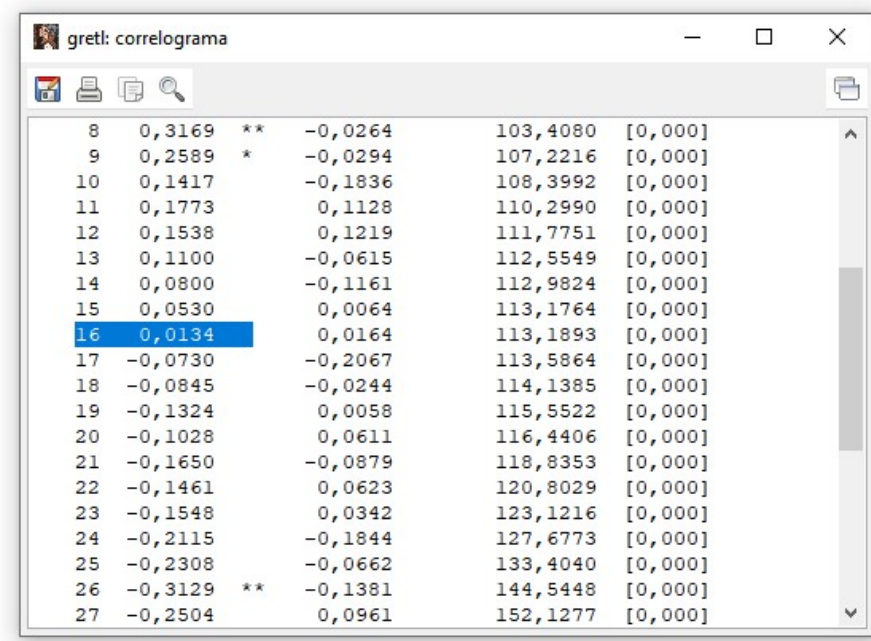
Figura 24 – Teste ADF-GLS para a série de produtividade em relação ao trigo.



Fonte: Autor (2021).

É possível observar que o p-valor, destacado em azul, está bem acima de até mesmo 0,1, o que indica que a série não é estacionária. A Figura 25 confirma a presença de tendência mostrando o correlograma da série.

Figura 25 – Correlograma da série de produtividade em relação ao trigo.



Fonte: Autor (2021).

Nota-se que na linha 16 do correlograma, destacada em azul, o valor de FAC quase zero, porém nas linhas seguintes ele volta a subir, indicando que há presença de tendência na série. É aplicado então o modelo ARIMA. A previsão da produtividade para o ano de 2020 resulta de um modelo ARIMA com $d=2$, $p=8$ e $q=3$, o resultado da aplicação do modelo está apresentado na Figura 26.

Figura 26 – Aplicação do modelo ARIMA à série de produtividade em relação ao trigo.

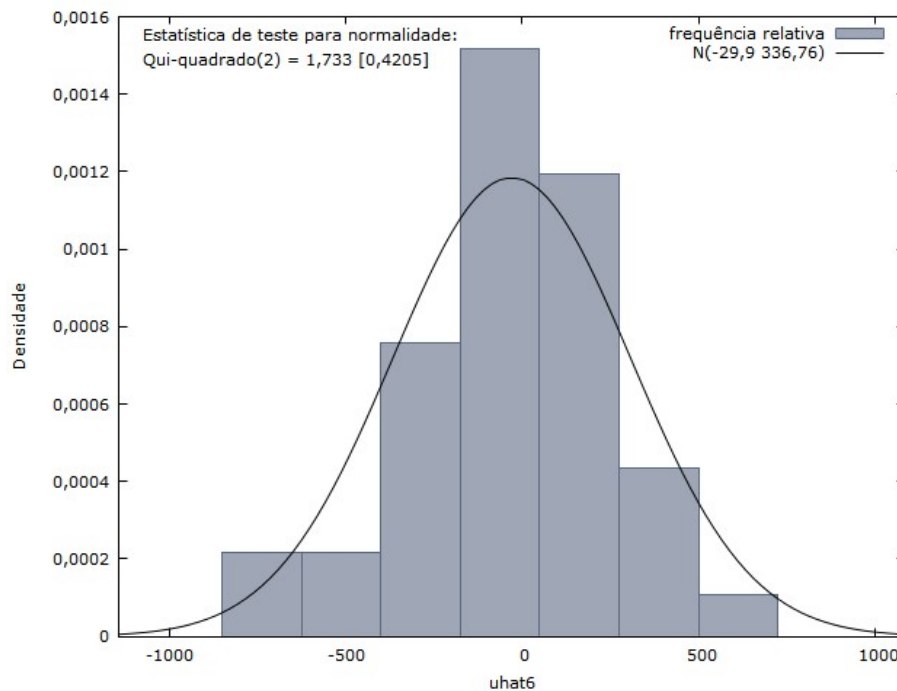
2005	2063,00	2119,56			
2006	1271,00	1841,79			
2007	2212,00	2060,02			
2008	2456,00	1844,27			
2009	2070,00	2118,96			
2010	2736,00	2468,83			
2011	2672,00	2415,22			
2012	2311,00	2575,47			
2013	2502,00	2518,74			
2014	2165,00	2219,97			
2015	2260,00	2463,00			
2016	3175,00	2845,40			
2017	2225,00	2594,63			
2018	2657,00	2667,17			
2019	2526,00	2781,76			
2020		2779,86	289,609	2212,24 -	3347,48

Fonte: Autor (2021).

Observa-se que os valores anteriores previstos pelo modelo ARIMA (terceira coluna) estão próximos dos valores reais (segunda coluna). Para se confirmar a aderência dos

modelos foi verificada a normalidade dos resíduos, a Figura 27 apresenta a distribuição dos resíduos.

Figura 27 - Distribuição dos resíduos ARIMA para a produtividade em relação ao trigo.



Fonte: Autor (2021).

É observado que os valores dos resíduos se aproximam da distribuição normal, afirmando a aderência do modelo.

4.1.2.2 Média móvel e alisamento exponencial

A série histórica não apresenta estacionariedade ao analisar trechos que se relacionem a um período mais recente, entretanto para auxiliar na discussão sobre a importância da verificação da estacionariedade ao se aplicar um modelo serão aplicados os modelos de média móvel e alisamento exponencial. A Tabela 4 mostra os resultados da aplicação do modelo de média móvel.

Tabela 4 – Média móvel aplicada à produtividade em relação ao trigo.

Ano	Produtividade	Média móvel
2003	2.375	1.661
2004	2.227	1.736
2005	2.063	1.794
2006	1.271	1.747
2007	2.212	1.808
2008	2.456	1.894
2009	2.070	1.909
2010	2.736	2.070
2011	2.672	2.150
2012	2.311	2.239
2013	2.502	2.252
2014	2.165	2.246
2015	2.260	2.266
2016	3.175	2.456
2017	2.225	2.457
2018	2.657	2.477
2019	2.526	2.523

Fonte: Autor (2021).

É possível notar uma maior parcela das previsões se mostra afastada do valor real (observado na linha abaixo, na coluna “produtividade”), em alguns momentos o modelo tem dificuldade de acompanhar a variação dos dados, resultando em alguns pontos de grande distância entre o previsto e o real, além disso as previsões estão frequentemente abaixo do valor real, o que indica a não aleatoriedade dos erros. A Tabela 5 mostra a aplicação do alisamento exponencial.

Tabela 5 – Alisamento exponencial aplicado à produtividade em relação ao trigo.

Ano	Produtividade	Alisamento exponencial
2003	2.375	1.841
2004	2.227	1.976
2005	2.063	2.007
2006	1.271	1.749
2007	2.212	1.911
2008	2.456	2.102
2009	2.070	2.091
2010	2.736	2.317
2011	2.672	2.441
2012	2.311	2.395
2013	2.502	2.433
2014	2.165	2.339
2015	2.260	2.311
2016	3.175	2.614
2017	2.225	2.478
2018	2.657	2.540
2019	2.526	2.535
Alfa	0,35	

Fonte: Autor (2021).

O α escolhido foi de 0,35 por ser um valor que aproxima a previsão para 2018 para o valor real, entretanto utilizando o comando atingir meta foi percebido que não há valor para α que possibilite que a previsão para 2018 fosse exata, o que evidencia uma possível limitação do modelo quando se há tendência.

4.1.2.3 Cálculo dos erros

O valor real obtido para a produtividade em 2020 foi de 2663 (em kg/há). Os resultados dos cálculos dos erros para os modelos aplicados estão apresentados na Tabela 6.

Tabela 6 – Erros dos modelos para a produtividade em relação ao trigo.

	Valor Real 2020	Valor obtido 2020	Erro Absoluto	Desvio médio absoluto	% de Erro (Erro Absoluto)
ARIMA	2663	2779,86	116,86	247,4	4,4%
Média Móvel	2663	2523	140	331,8	5,3%
Alisamento Exponencial	2663	2535	128	286,45	4,8%

Fonte: Autor (2021).

É possível notar que o modelo ARIMA tem os menores erros associados.

4.1.2.4 Discussão dos resultados

Ao se analisar os resultados da aplicação dos modelos foi possível perceber que o modelo ARIMA apresentou o menor erro, tanto para o desvio médio absoluto quanto para o erro absoluto, o que diz que o modelo foi o mais preciso na descrição da série não só para a previsão de 2020, mas também, de maneira geral, para as previsões dos outros anos.

Ao se aplicar o modelo de média móvel, foi possível perceber que houve momentos em que o modelo previu um valor distante do valor real, o que indica que nesses períodos houve dificuldade da aderência do modelo à série. O modelo apresentou, para a série de produtividade do trigo, a maior dificuldade da adesão do modelo ao comportamento de tendência, o que pode ser visto pelos erros obtidos, tendo o modelo os maiores erros entre os três. Entretanto o erro absoluto obtido foi baixo, isso se deve ao fato de a tendência ser relativamente baixa no último trecho da série.

O modelo de alisamento exponencial se mostrou intermediário, o que indica que, a

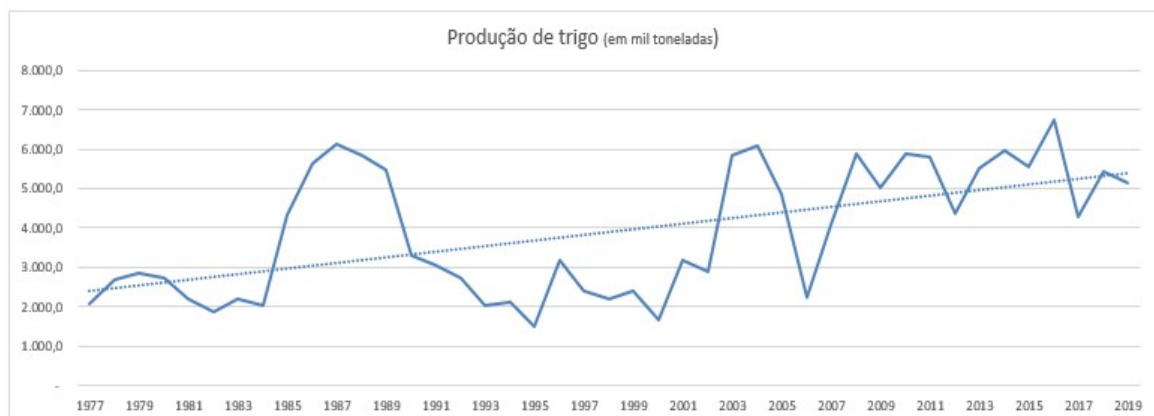
escolha de um α adequado pode reduzir de maneira significativa os erros de previsão. Entretanto foi evidenciado que há situações em que não havia valor para α que garantisse uma previsão exata, o que indica uma limitação do modelo ao se trabalhar com tendência.

Ao se comparar a Figura 23 com a Figura 15 é possível notar que a tendência tem um comportamento mais constante na série da produtividade em relação ao trigo se comparada a série da área plantada de trigo, o que explica valores menores de erros ao se aplicar os modelos para a série de produtividade.

4.1.3 Série histórica da produção de trigo

A plotagem do gráfico da série da produção de trigo, referente ao período de 1977 a 2019, está apresentada na Figura 28.

Figura 28 - Variação da produção de relação ao trigo de 1977 a 2019.



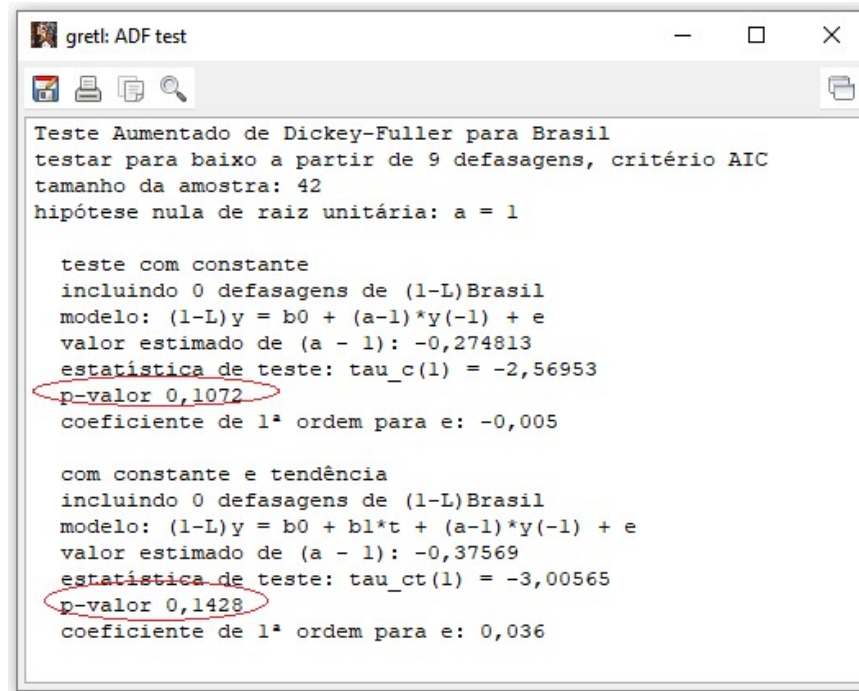
Fonte: Autor (2021).

Aparente não há a estacionariedade na série. Pode -se avaliar o período de 2009 a 2019 para a aplicação dos modelos de média móvel e alisamento exponencial. Entretanto a seleção do período pode estar abdicando de dados relevantes em relação ao contexto ao qual a série se refere e, devido ao fato de o número de observações para o período ser baixo, há a impossibilidade de se aplicar o teste ADF-GLS e o Dickey-Fuller aumentado. Percebe-se que há uma limitação de análise que influencia na aplicação dos modelos.

A linha de tendência adicionada indica a presença de uma possível tendência de crescimento. A visualização do gráfico e a limitação da aplicação dos testes de estacionariedade para o período de 2009 a 2019 sugerem o uso modelo ARIMA.

Para verificar a existência de tendência na série foram aplicados testes de estacionariedade. A Figura 29 apresenta os resultados obtidos para o teste Dickey-Fuller aumentado.

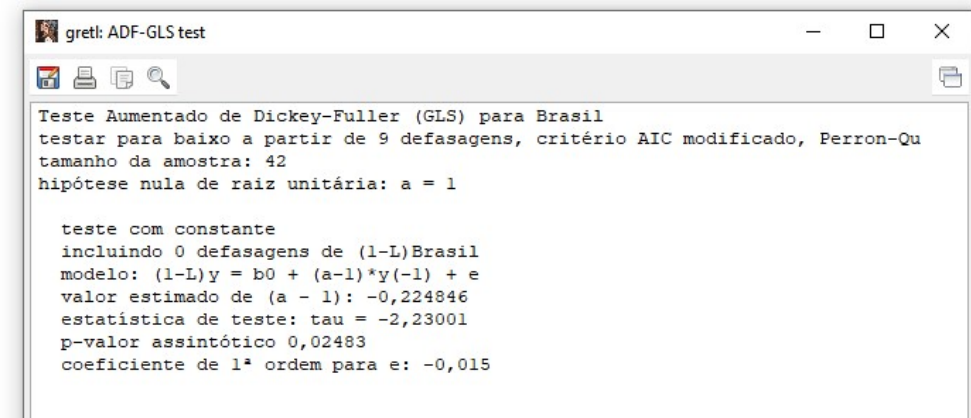
Figura 29 – Teste Dickey-Fuller para a produção de trigo.



Fonte: Autor (2021).

Percebe-se que o p-valor obtido para o teste, tanto para o teste com constante quanto para o teste com constante e tendência está acima de 0,1, o que indica a presença de tendência, entretanto o valor se apresenta bem próximo 0,1. A Figura 30 apresenta o resultado da aplicação do teste ADF-GLS.

Figura 30 – Aplicação do teste ADF-GLS para a produção de trigo.



Fonte: Autor (2021).

Nota-se que o p-valor se encontra abaixo de 0,05 o que indica estacionariedade. Percebe-se que existe um movimento de tendência pouco significativo. A Figura 31 mostra o correlograma para a série.

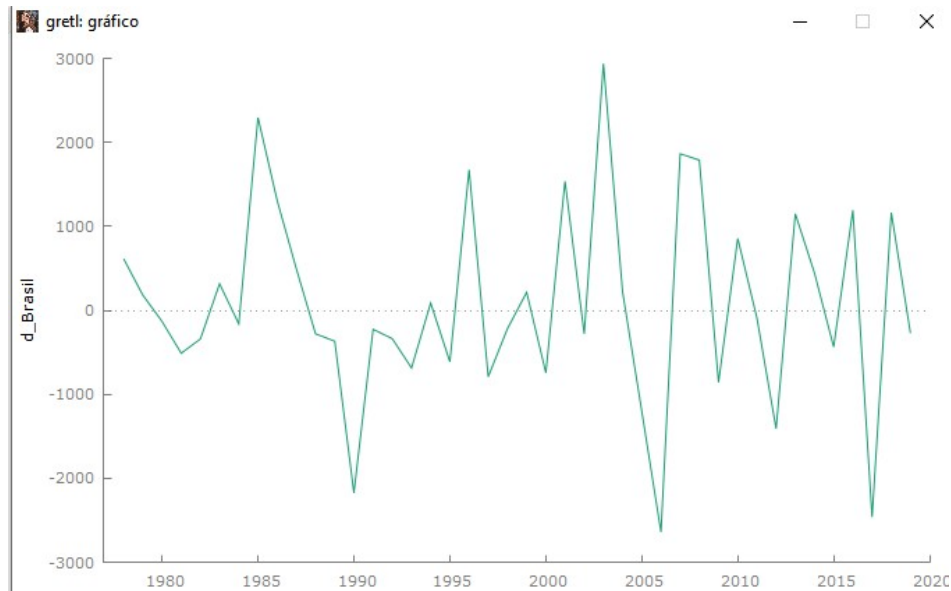
Figura 31 – Correlograma para a série de produção de trigo.

Defas.	FAC	FACP	Estat. Q	[p-valor]
1	0,7141 ***	0,7141 ***	23,4930	[0,000]
2	0,5080 ***	-0,0039	35,6733	[0,000]
3	0,3339 **	-0,0561	41,0671	[0,000]
4	0,1905	-0,0558	42,8667	[0,000]
5	0,1490	0,0992	43,9974	[0,000]
6	0,0827	-0,0664	44,3547	[0,000]
7	0,0273	-0,0400	44,3948	[0,000]
8	-0,0719	-0,1440	44,6805	[0,000]
9	-0,1350	-0,0196	45,7173	[0,000]
10	-0,1492	0,0100	47,0233	[0,000]
11	-0,0757	0,1483	47,3700	[0,000]
12	-0,0848	-0,1488	47,8189	[0,000]
13	-0,0798	0,0055	48,2292	[0,000]
14	-0,0396	0,0750	48,3339	[0,000]
15	0,0334	0,1586	48,4110	[0,000]
16	0,0331	-0,1619	48,4896	[0,000]
17	0,0100	-0,0624	48,4970	[0,000]
18	-0,0319	-0,0836	48,5756	[0,000]
19	-0,1082	-0,0357	49,5194	[0,000]
20	-0,1125	0,0231	50,5848	[0,000]
21	-0,1069	0,0021	51,5902	[0,000]
22	-0,0165	0,1024	51,6152	[0,000]
23	-0,0246	-0,0754	51,6738	[0,001]
24	-0,0446	0,0152	51,8766	[0,001]
25	-0,0375	0,0281	52,0279	[0,001]
26	-0,0850	-0,1505	52,8511	[0,001]
27	-0,0344	0,1069	52,9941	[0,002]
28	-0,0019	0,0122	52,9946	[0,003]
29	0,0075	-0,0565	53,0024	[0,004]
30	-0,0512	-0,1767	53,3919	[0,005]
31	-0,1106	-0,0107	55,3652	[0,005]
32	-0,1479	-0,0100	59,2114	[0,002]
33	-0,1606	-0,0390	64,2026	[0,001]
34	-0,1766	-0,0555	70,9072	[0,000]
35	-0,1568	0,0633	76,8540	[0,000]
36	-0,1476	-0,0590	82,8773	[0,000]
37	-0,1286	0,0431	88,2149	[0,000]
38	-0,0967	-0,0413	91,8347	[0,000]

Fonte: Autor (2021).

É possível perceber que na linha 7 o valor de FAC quase zero, porém, a partir da linha 31 o valor volta a subir. A Figura 32 mostra o gráfico das primeiras diferenças para auxílio da análise.

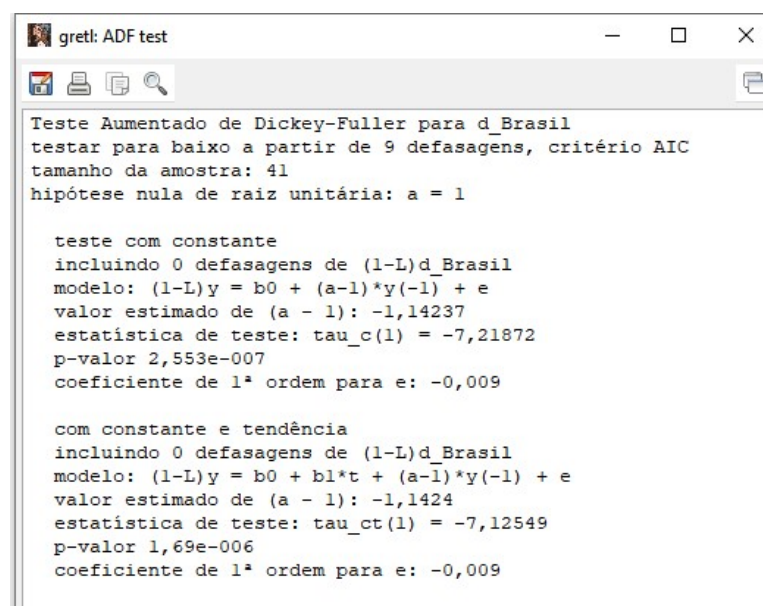
Figura 32 – Gráfico das primeiras diferenças em relação à série de produção de trigo.



Fonte: Autor (2021).

Nota-se que a série indica uma possível estacionariedade para os dados. Realiza-se então os testes de estacionariedade para a série das primeiras diferenças. A Figura 33 mostra o resultado do teste Dickey-Fuller aumentado.

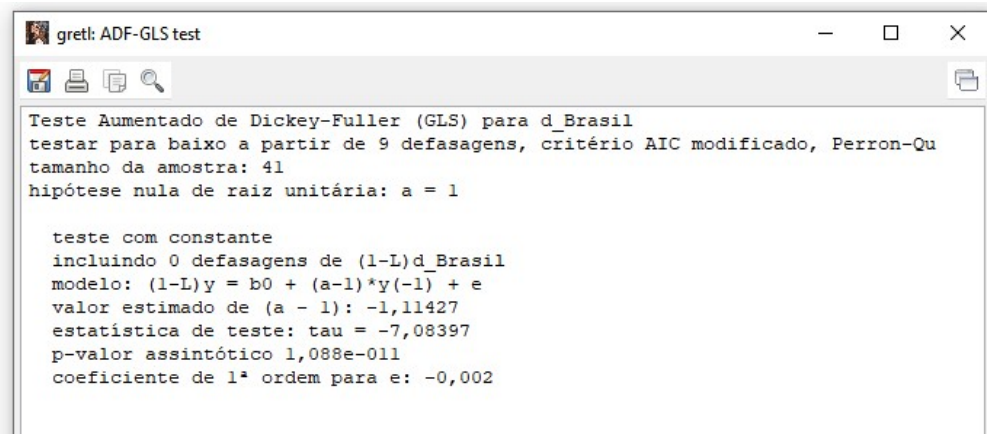
Figura 33- Teste Dickey-Fuller aumentado para as primeiras diferenças.



Fonte: Autor (2021).

Percebe-se que o p-valor agora encontra-se consideravelmente baixo. A Figura 34 mostra o resultado da aplicação do teste ADF-GLS.

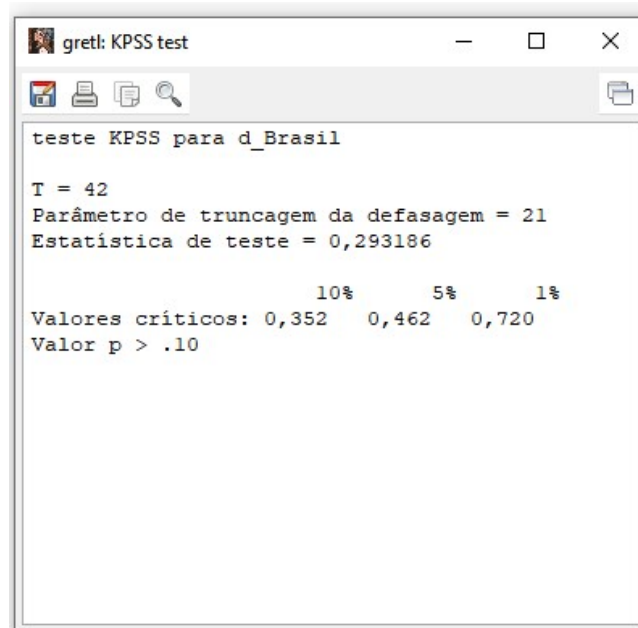
Figura 34 – Teste ADF-GLS para as primeiras diferenças.



Fonte: Autor (2021).

Nota-se que o p-valor obtido está consideravelmente baixo. A Figura 35 mostra os resultados obtidos para a aplicação do teste KPSS.

Figura 35 – Teste KPSS para as primeiras diferenças.



Fonte: Autor (2021).

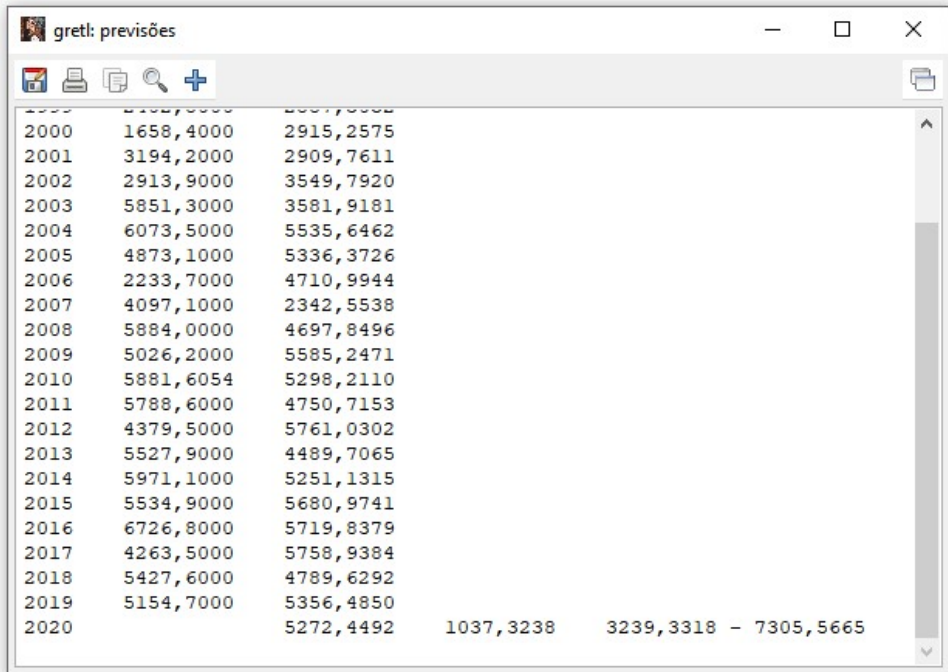
Observe que a estatística de teste está abaixo de todos os valores críticos. Conclui-se

que a série das primeiras diferenças pode ser considerada estacionária.

4.1.3.1 Modelo ARIMA

O modelo ARIMA foi aplicado após confirmada a estacionariedade da série das primeiras diferenças em relação a produção de trigo. O resultado obtido foi um modelo com $d=1$, $p=1$ e $q=4$. A previsão obtida após a aplicação do modelo está apresentada na Figura 36.

Figura 36 – Aplicação do Modelo ARIMA à série de produção de trigo.



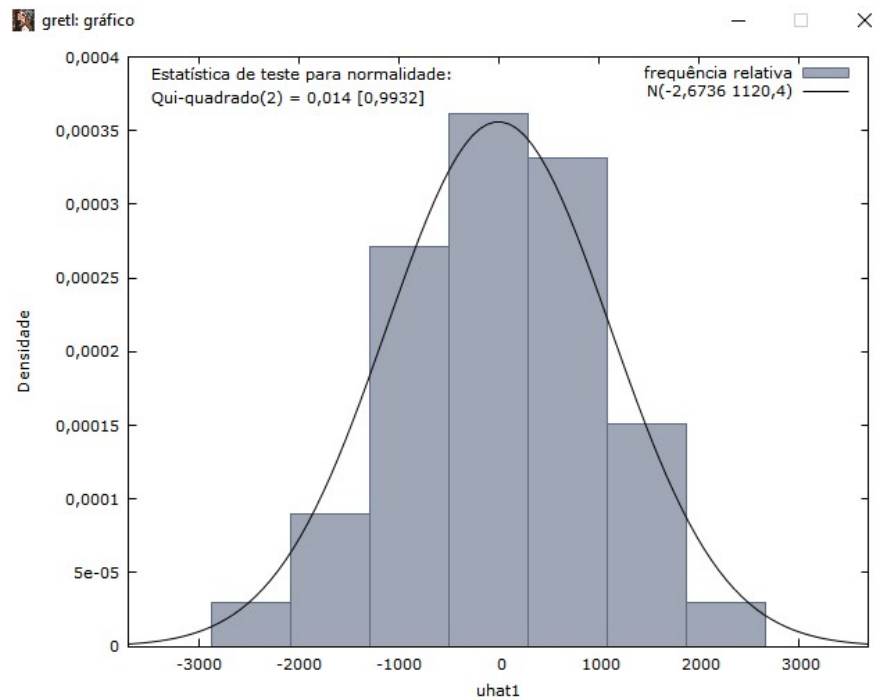
The screenshot shows the 'gretl: previsões' window. It contains a table with three columns: Year, Observed Production, and Forecasted Production. The data spans from 2000 to 2020. The forecast for 2020 includes a point estimate and a confidence interval.

Ano	Produção (Observed)	Previsão (Forecast)
2000	1658,4000	2915,2575
2001	3194,2000	2909,7611
2002	2913,9000	3549,7920
2003	5851,3000	3581,9181
2004	6073,5000	5535,6462
2005	4873,1000	5336,3726
2006	2233,7000	4710,9944
2007	4097,1000	2342,5538
2008	5884,0000	4697,8496
2009	5026,2000	5585,2471
2010	5881,6054	5298,2110
2011	5788,6000	4750,7153
2012	4379,5000	5761,0302
2013	5527,9000	4489,7065
2014	5971,1000	5251,1315
2015	5534,9000	5680,9741
2016	6726,8000	5719,8379
2017	4263,5000	5758,9384
2018	5427,6000	4789,6292
2019	5154,7000	5356,4850
2020		5272,4492
		1037,3238 3239,3318 - 7305,5665

Fonte: Autor (2021).

Para verificar a aderência do modelo foi feito o gráfico da distribuição dos resíduos apresentado na Figura 37.

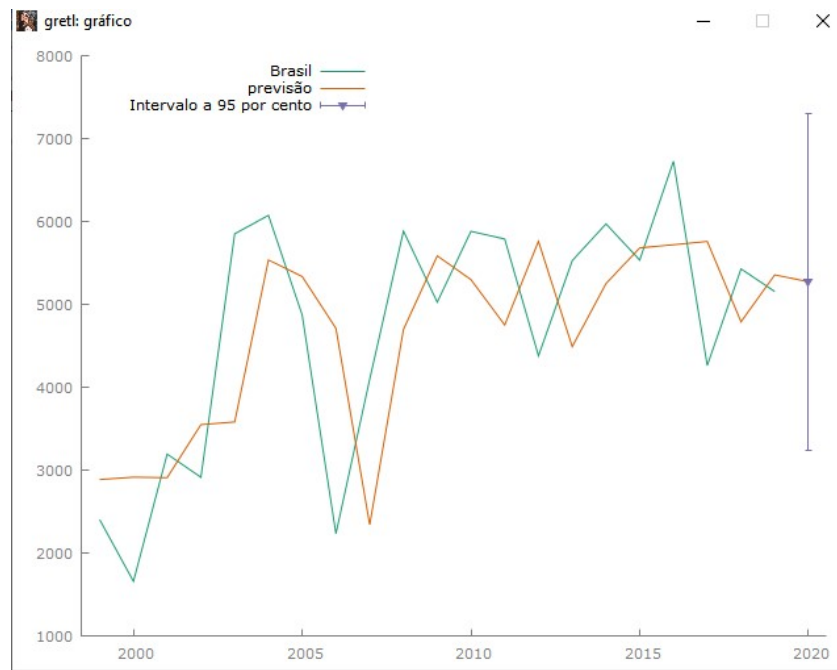
Figura 37 – Distribuição dos resíduos ARIMA para a produção de trigo.



Fonte: Autor (2021).

Ainda verificando a aderência do modelo foi plotado o gráfico comparando os valores previstos com os valores reais, apresentado na Figura 38.

Figura 38 – Previsão do modelo ARIMA em relação a produção de trigo.



Fonte: Autor (2021).

Percebe-se que o modelo reage de forma coerente à variação dos dados. Para que se possa discutir a precisão do modelo foram calculados os erros, apresentados na Tabela 7.

Tabela 7 – Erros para o modelo ARIMA aplicado à série de produção de trigo.

	Valor Real 2020	Valor obtido 2020	Erro Absoluto	Desvio médio absoluto	% de Erro (Erro Absoluto)
ARIMA	6234,6	5272,449	962,151	960,0	15,4%

Fonte: Autor (2021).

4.1.3.2 Discussão dos resultados

Ao se analisar o gráfico da série temporal da produção de trigo foi notado que havia tendência nos dados, entretanto, por ser uma tendência fraca, a análise dos dados exigiu novas etapas.

Devido ao fato de os testes de estacionariedade apontarem comportamentos diferentes, a análise do gráfico foi crucial para o entendimento do comportamento da série, além do uso do correlograma.

A análise gráfica das primeiras diferenças se mostrou importante para situações em que apenas a análise do correlograma das primeiras diferenças não apresenta resultados tão sólidos. Além disso, os testes de raiz unitária se mostraram bastante úteis para analisar qualquer etapa do processo de diferenciação da série.

A normalidade da distribuição dos erros, assim como a análise gráfica das previsões, indicou que os resultados obtidos por meio do modelo ARIMA foram coerentes com a realidade. Apesar de serem utilizadas mais etapas para a detecção das componentes da decomposição clássica no estudo da série da produção de trigo, o modelo ARIMA apresentou melhores resultados se comparada a aplicação do modelo para a série de área plantada de trigo, isso mostra que a utilização adequada das ferramentas de auxílio garante análises precisas mesmo em situações em que há mais dificuldade para a descrição da série.

Em outro aspecto percebeu-se que uma baixa quantidade de observações impacta na descrição da série, pois impossibilita a aplicação de alguns testes de estacionariedade, consequentemente a aplicação do modelo de média móvel e de alisamento exponencial simples é afetada.

4.2 SOJA

Será analisada a série histórica da Conab produção de soja. Os dados da série de produção de soja anual são referentes ao período de 1977 a 2019.

4.2.1 Série histórica de produção de soja

A plotagem do gráfico da série da produção de soja, referente ao período de 1976 a 2019, está apresentada na Figura 39.

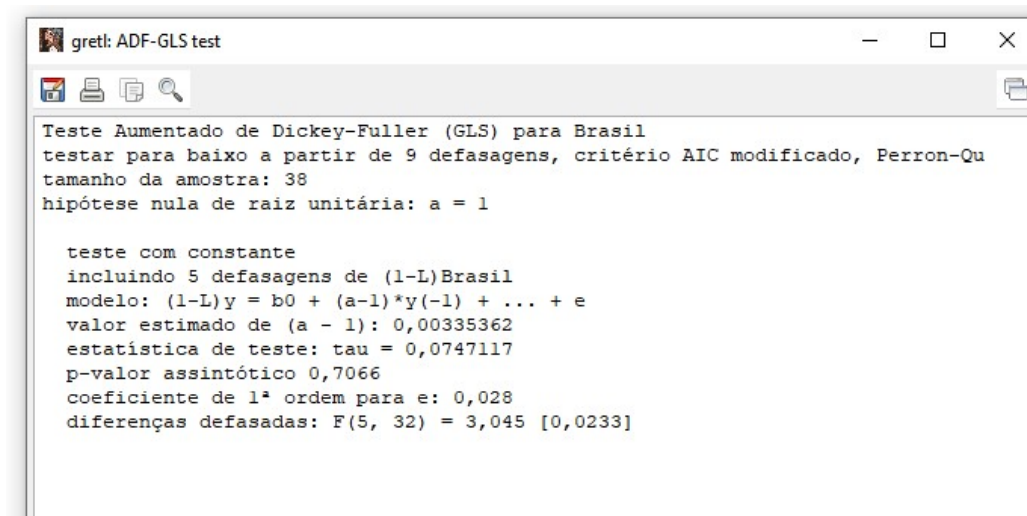
Figura 39 – Variação da produção de soja de 1976 a 2019.



Fonte: Autor (2021).

Observa-se que a série apresenta visualmente uma forte presença de tendência de crescimento. Para se atestar a não estacionariedade da série foi aplicado o teste ADF-GLS mostrado na Figura 40.

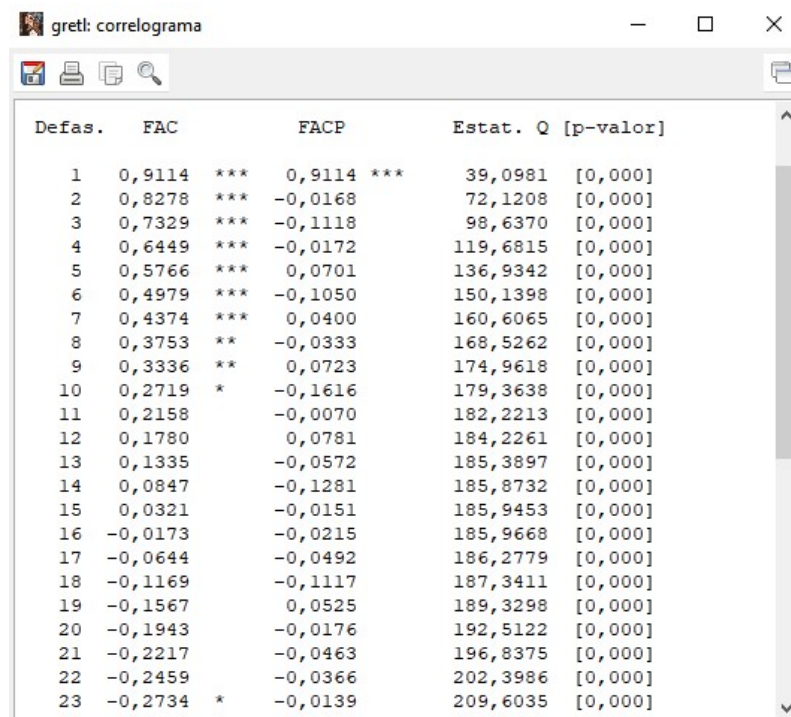
Figura 40 – Teste ADF-GLS para a série de produção de trigo.



Fonte: Autor (2021)

Nota-se que há um valor alto obtido para p-valor, o que indica que a série não é estacionária. A Figura 41 mostra o correlograma da série.

Figura 41 – Correlograma para a série de produção de trigo.



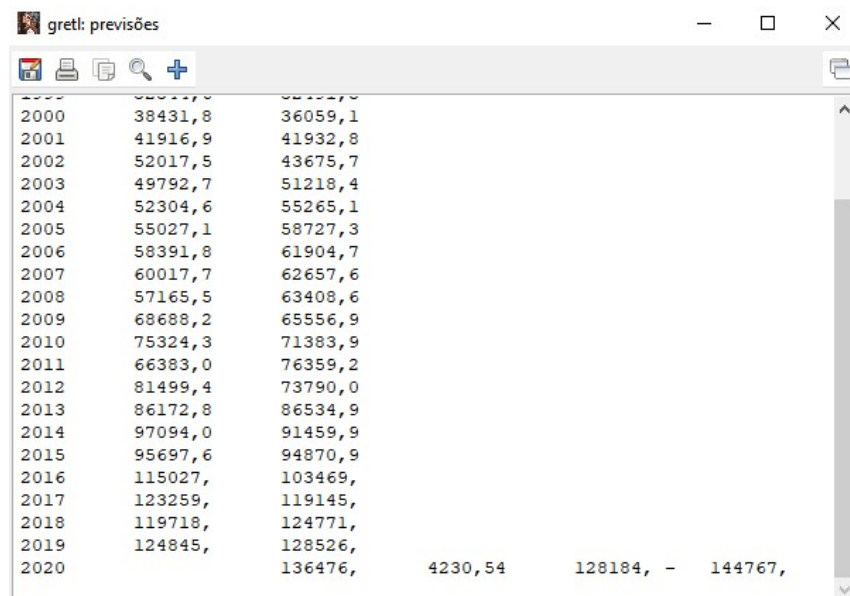
Fonte: Autor (2021).

Observa-se que o valor FAC está próximo de zero na linha 15, entretanto a partir da linha 18 ele volta a crescer. Confirma-se então a presença de tendência, é indicado o uso do modelo ARIMA.

4.2.1.1 Modelo ARIMA

O ajuste dos parâmetros na aplicação do modelo resultou em um ARIMA de $d=2$, $p=5$ e $q=2$. A previsão para 2020, assim como as previsões para anos anteriores podem ser observadas na Figura 42.

Figura 42 – Aplicação do modelo ARIMA para a série de produção de soja.



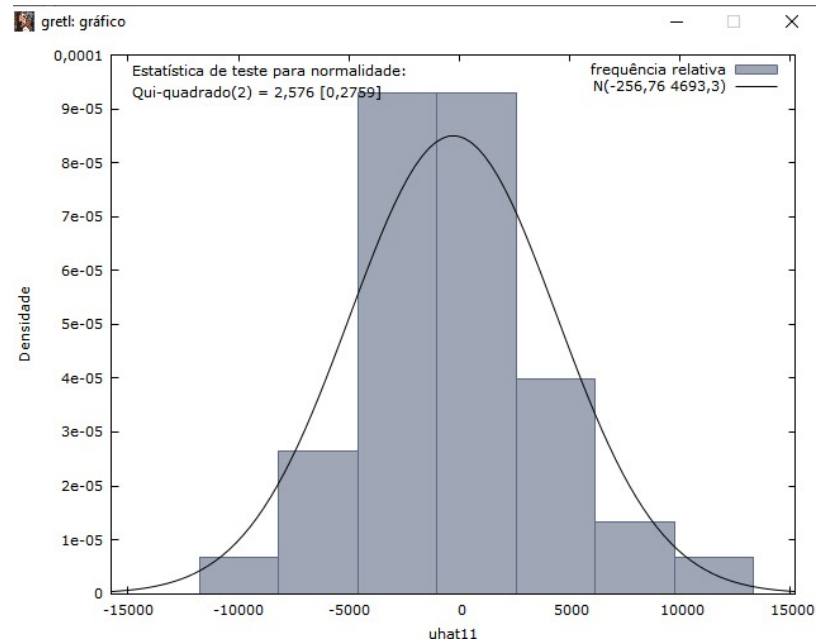
Ano	Produção Observada	Previsão
2000	38431,8	36059,1
2001	41916,9	41932,8
2002	52017,5	43675,7
2003	49792,7	51218,4
2004	52304,6	55265,1
2005	55027,1	58727,3
2006	58391,8	61904,7
2007	60017,7	62657,6
2008	57165,5	63408,6
2009	68688,2	65556,9
2010	75324,3	71383,9
2011	66383,0	76359,2
2012	81499,4	73790,0
2013	86172,8	86534,9
2014	97094,0	91459,9
2015	95697,6	94870,9
2016	115027,	103469,
2017	123259,	119145,
2018	119718,	124771,
2019	124845,	128526,
2020		136476,

Intervalo de 95% para 2020: 4230,54 a 144767,

Fonte: Autor (2021).

A fim de verificar a aderência do modelo foi construído o gráfico de dispersão dos erros, apresentado na Figura 43.

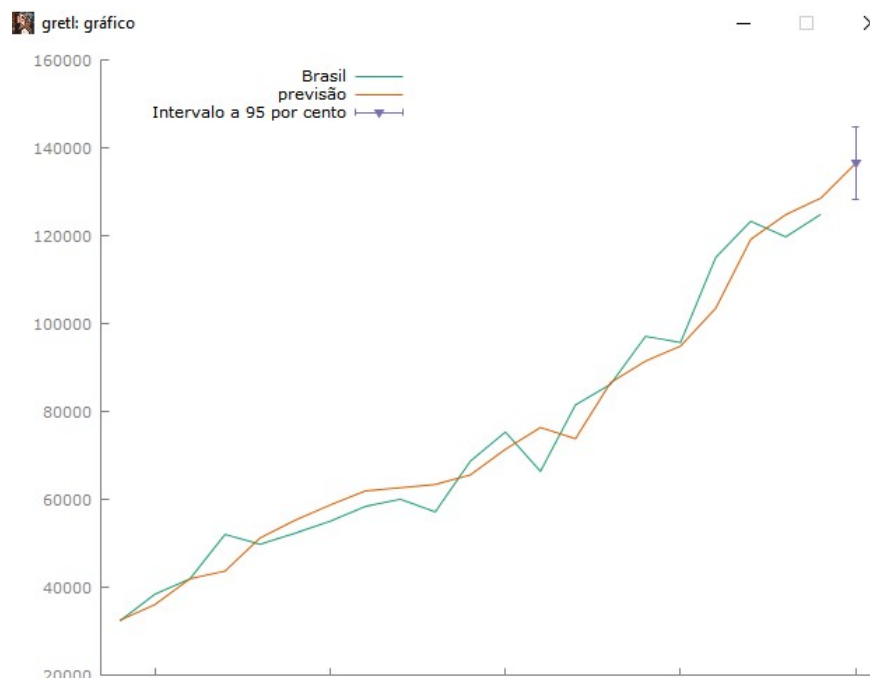
Figura 43 – Dispersão dos erros para o modelo ARIMA da produção de soja.



Fonte: Autor (2021).

Nota-se que a distribuição dos erros se aproxima da curva normal. Verificando-se ainda a aderência do modelo foi plotado o gráfico das previsões em comparação aos valores reais obtidos pela série, como está apresentado na Figura 44.

Figura 44 - Previsão do modelo ARIMA em relação a produção de soja.



Fonte: Autor (2021).

Percebe-se que a curva das previsões é compatível a curva dos valores reais. Os erros calculados para as previsões do modelo ARIMA estão apresentados na Tabela 8.

Tabela 8 – Erros para as previsões do modelo ARIMA para a produção de soja.

	Valor Real 2020	Valor obtido 2020	Erro Absoluto	Desvio médio absoluto	% de Erro (Erro Absoluto)
ARIMA	133692,30	136476	2783,7	4096,76	2,1%

Fonte: Autor (2021).

Nota-se que os valores de desvio médio absoluto e erro absoluto são consideravelmente baixos em relação aos valores dos dados ao longo da série.

4.2.1.2 Média móvel.

Para efeito de discussão sobre o efeito do valor da variável r na aplicação em modelos de média móvel, será estudada a aplicação dos modelos à série de produção de trigo. A Figura 45 mostra como o gráfico da média móvel se relaciona com o comportamento da série ao se escolher um $r=10$.

Figura 45 – Média móvel para $r=10$ aplicada a produção de soja.



Fonte: Autor (2021).

Nota-se que o modelo não seria interessante, pois a distribuição dos erros simples se concentraria em valores positivos, tendo em vista que a curva do modelo (em vermelho) está abaixo dos valores da série durante quase todo o intervalo observado. A Figura 46 apresenta o gráfico da média móvel se relaciona com o comportamento da série, utilizando

$r=3$.

Figura 46 - Média móvel para $r=3$ aplicada a produção de soja.



Fonte: Autor (2021).

Percebe-se que a aderência do modelo aparenta estar mais próxima à realidade, entretanto o comportamento dos erros simples não se daria de forma aleatória, sendo estes concentrados em valores positivos, o que enviesaria a aplicação do modelo.

4.2.1.3 Alisamento exponencial

O modelo de alisamento exponencial foi aplicado à série para que seja possível a análise dos resíduos. Utilizando o comando atingir meta para que a previsão para 2018 seja o valor exato obtido na série para esse ano foi obtido o valor de 0,7 para α . O gráfico relacionando os valores previstos para o modelo com os valores obtidos na série estão apresentados na Figura 47.

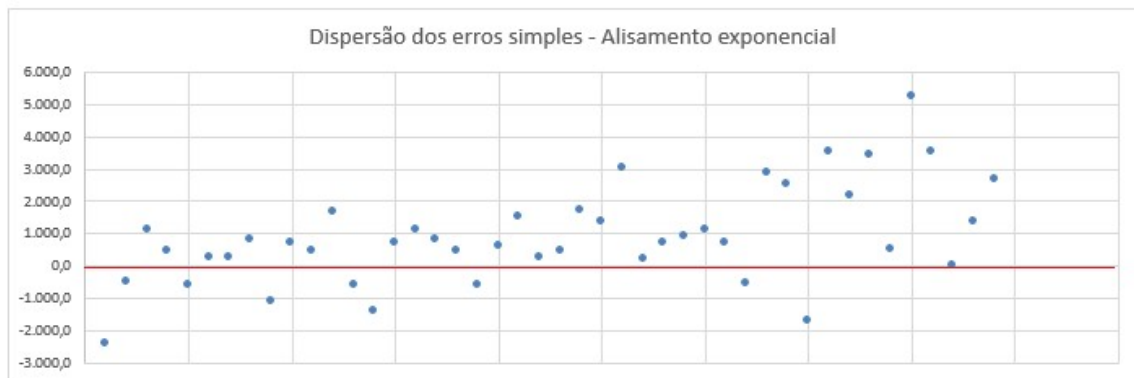
Figura 47 – Alisamento exponencial em relação à série de produção de soja.



Fonte: Autor (2021).

Percebe-se que a aderência do modelo aparenta estar próxima à realidade, entretanto é necessário avaliar o comportamento dos erros simples para que se possa ser observado se o modelo descreve de maneira coerente a série, garantindo a aleatoriedade dos erros. A Figura 48 mostra a dispersão dos erros ao se aplicar o alisamento exponencial.

Figura 48 – Dispersão dos erros para o alisamento exponencial.



Fonte: Autor (2021).

Ao se observar a distribuição dos erros percebe-se que 34 das 42 previsões estão acima de zero, sendo então 80,9% dos erros valores positivos, o que indica que os erros não se dão de forma aleatória, o que enviesaria a aplicação do modelo.

4.2.1.4 Discussão dos resultados

O modelo ARIMA se mostrou consideravelmente preciso ao se analisar a série de produção de soja, e a presença de uma tendência forte impactou diretamente na agilidade da aplicação do modelo, por requerer menos etapas para identificação do comportamento da série. O fato de a tendência apresentada ser significativamente constante impactou diretamente na precisão do modelo, que apresentou um erro bastante baixo ao se comparar com o erro do modelo para as séries de produção, produtividade e área plantada do trigo.

Percebeu-se que utilizar um menor valor para r no cálculo do modelo de média móvel impacta na rapidez de absorção do comportamento de tendência na série, entretanto o modelo se mostrou limitado ao não incorporar com precisão o comportamento da tendência exibindo erros que não se apresentam como ruído.

O alisamento exponencial, assim como a média móvel, apresentou uma aparente aderência à série, porém, da mesma forma, não foi eficiente em incorporar o comportamento da tendência, apresentando erros que não podem ser considerados aleatórios.

A análise feita da aplicação dos modelos em relação à série de produção de soja sugere que os modelos de média móvel e alisamento exponencial não se mostram eficientes na descrição de séries com tendência, entretanto o modelo ARIMA se mostra de fácil aplicabilidade e bastante precisão, principalmente em séries onde a presença de tendência é forte e constante.

4.3 MILHO

Será analisada a série histórica da Conab produção de milho. Os dados da série de produção de milho anual são referentes ao período de 1976 a 2019.

4.3.1 Série histórica da produção de milho

A plotagem do gráfico da série da produção de milho, referente ao período de 1976 a 2019, está apresentada na Figura 49.

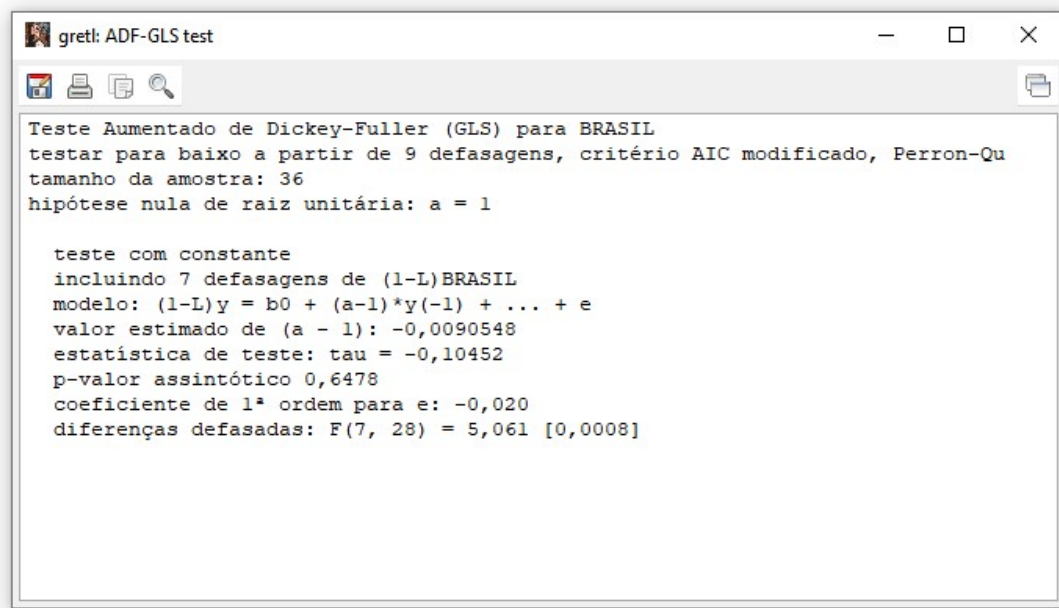
Figura 49 – Variação da produção de milho de 1976 a 2019.



Fonte: Autor (2021).

A análise do gráfico aponta a existência de uma tendência de crescimento. A fim de se confirmar a não estacionariedade da série foi realizado o teste ADF-GLS, os resultados estão apresentados na Figura 50.

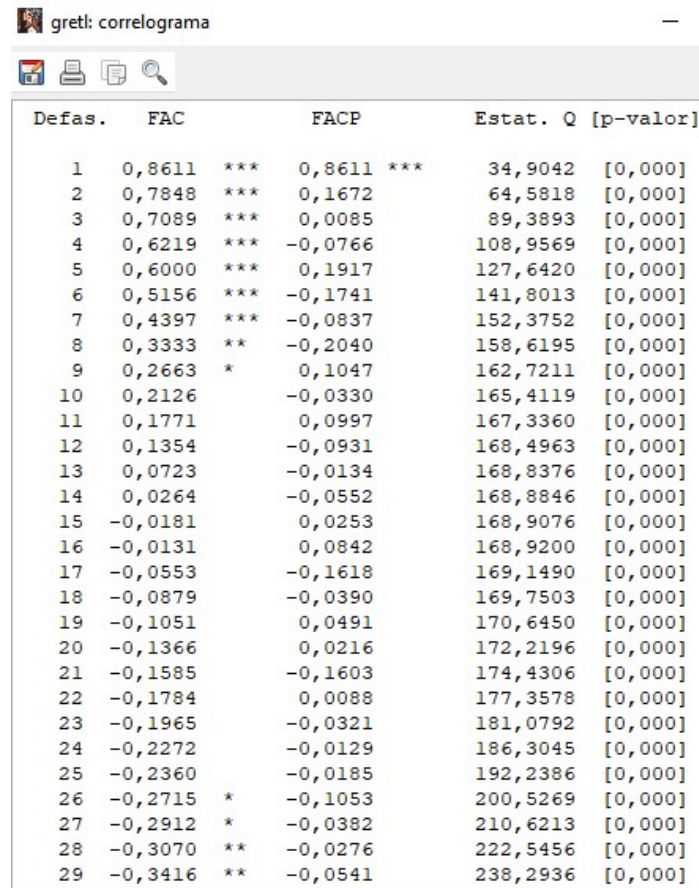
Figura 50 – Teste ADF-GLS para a série de produção de milho.



Fonte: Autor (2021).

Percebe-se que há um valor alto obtido para p-valor, o que indica que a série não é estacionária. A Figura 51 mostra o correlograma da série.

Figura 51 – Correlograma da série de produção de milho.



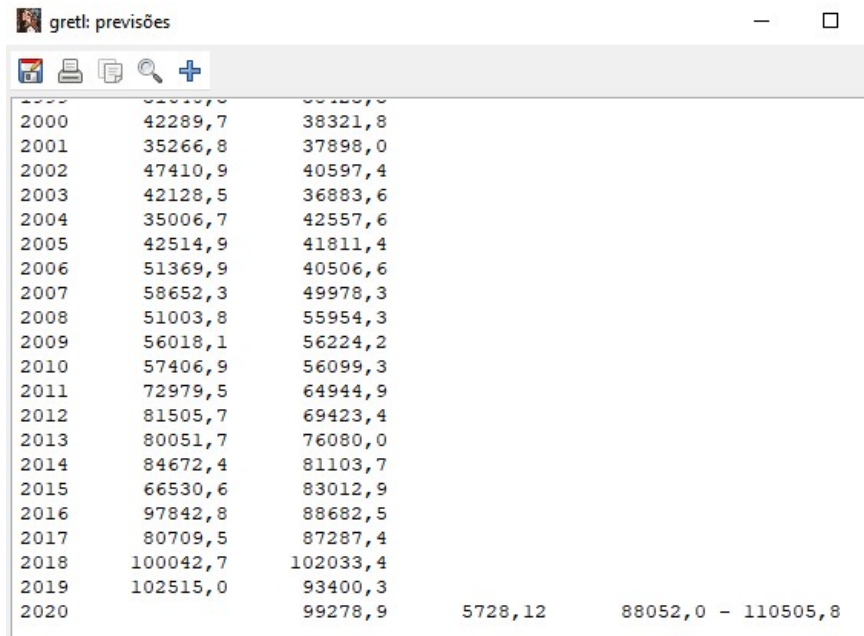
Fonte: Autor (2021).

Observa-se chega próximo de zero na linha 16, porém o valor volta a subir nas linhas abaixo. Atesta-se então a presença de tendência, indica-se o uso do modelo ARIMA.

4.3.1.1 Modelo ARIMA

O ajuste dos parâmetros na aplicação do modelo resultou em um ARIMA de $d=1$, $p=1$ e $q=4$. A previsão para 2020, assim como as previsões para anos anteriores podem ser observadas na Figura 52.

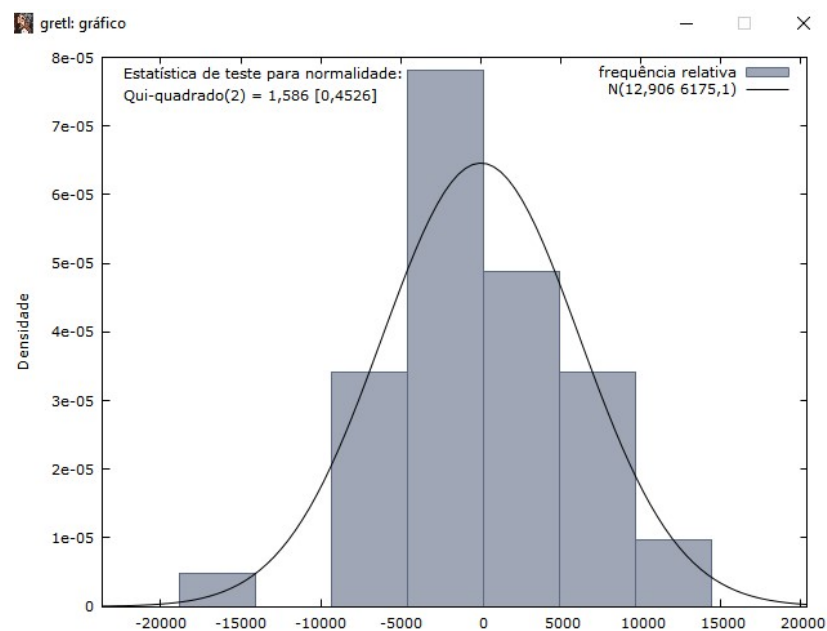
Figura 52- Aplicação do modelo ARIMA para a série de produção de milho.



Fonte: Autor (2021).

Com o objetivo de verificar a aderência do modelo foi construído o gráfico de dispersão dos erros, apresentado na Figura 53.

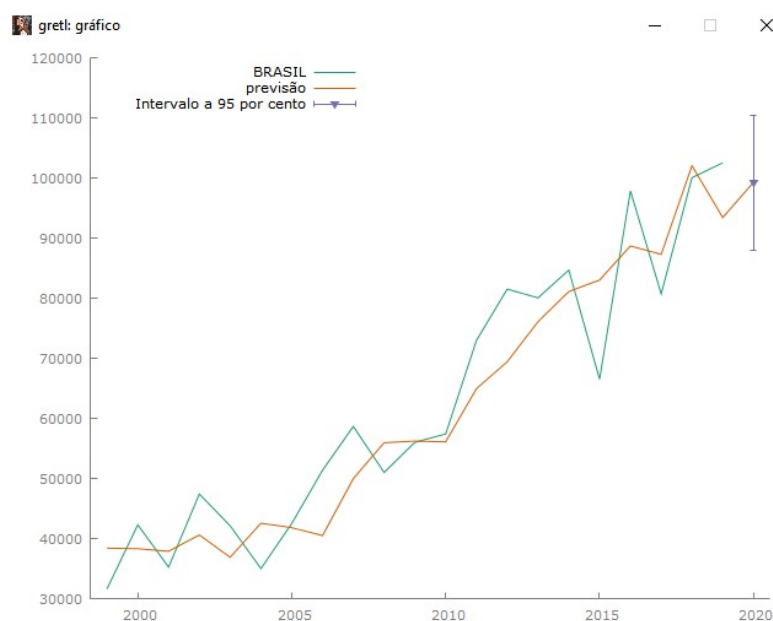
Figura 53 - Dispersão dos erros para o modelo ARIMA da produção de milho.



Fonte: Autor (2021).

É possível perceber que a distribuição dos erros se aproxima da curva normal. Verificando-se ainda a aderência do modelo foi plotado o gráfico das previsões em comparação aos valores reais obtidos pela série, como está apresentado na Figura 54.

Figura 54 - Previsão do modelo ARIMA em relação a produção de milho.



Fonte: Autor (2021).

Percebe-se que a curva das previsões é compatível a curva dos valores reais. Os erros calculados para as previsões do modelo ARIMA estão apresentados na Tabela 9.

Tabela 9 - Erros para as previsões do modelo ARIMA para a produção de milho.

	Valor Real 2020	Valor obtido 2020	Erro Absoluto	Desvio médio absoluto	% de Erro (Erro Absoluto)
ARIMA	102313,2	99278,9	3034,3	6.077,9	3,0%

Fonte: Autor (2021).

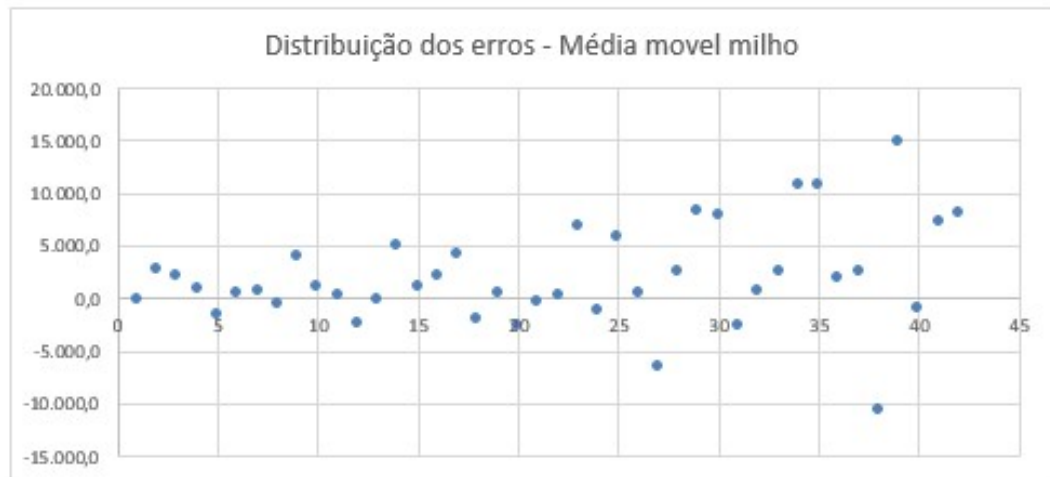
Percebe-se que os valores de desvio médio absoluto e erro absoluto são consideravelmente baixos em relação aos valores dos dados ao longo da série.

4.3.1.2 Média Móvel

Para efeito de discussão sobre a efetividade do modelo na descrição de séries com

forte presença de tendência foi plotado o gráfico da distribuição dos erros para um modelo de média móvel com $r=3$, apresentado na Figura 55.

Figura 55 – Distribuição dos erros para o modelo de médias móveis.



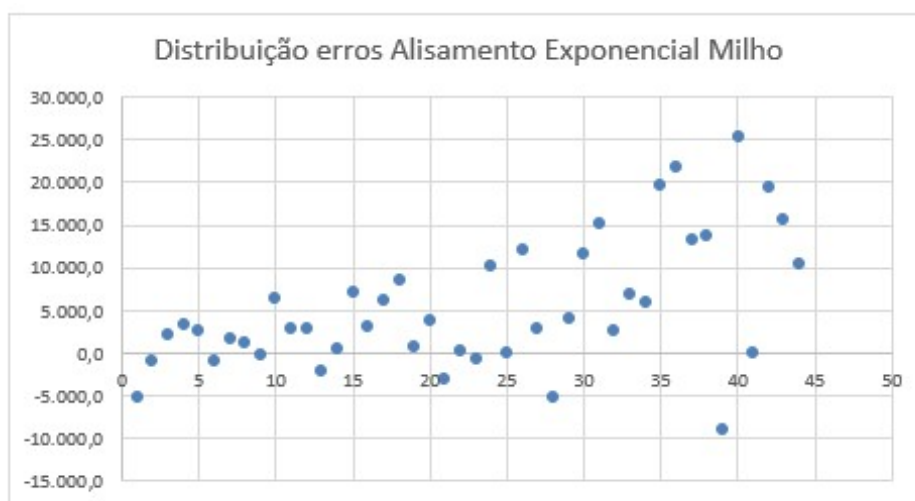
Fonte: Autor (2021).

Nota-se que o modelo não seria interessante devido a distribuição dos erros simples estar concentrada em valores positivos. O comportamento dos erros simples não se dá de forma aleatória, o que enviesaria a aplicação do modelo.

4.3.1.3 Alisamento exponencial

Para a análise da efetividade do modelo na descrição de séries com forte presença de tendência, foi plotado o gráfico da distribuição dos erros para um modelo com $\alpha=0,3$, obtido usando o comando “atingir meta”. O gráfico está apresentado na Figura 56.

Figura 56 – Distribuição dos erros para o modelo de alisamento exponencial.



Fonte: Autor (2021).

Observa-se que o modelo não seria descreveria a série adequadamente devido a distribuição dos erros simples estar concentrada em valores positivos. O comportamento dos erros simples não se dá de forma aleatória, o que enviesaria a aplicação deste modelo.

4.3.1.4 Discussão dos resultados

O modelo ARIMA se mostrou novamente preciso ao se analisar uma série com presença de uma tendência forte. Além disso, a aplicação aconteceu de forma ágil por requerer menos etapas para identificação do comportamento da série. Nota-se que, igualmente ao que foi avaliado para a aplicação do modelo em relação à série da produção de trio, a presença de uma tendência significativa e relativamente constante impactou diretamente na precisão do modelo, acarretando um erro absoluto de apenas 3,0%.

Ao se analisar o modelo de média móvel e o modelo de alisamento exponencial foi verificado novamente que os modelos não tiveram êxito em se adequar à série, por não apresentar erros de caráter residual. Percebe-se então que os modelos não são indicados caso haja uma forte presença de tendência.

4.4 FEIJÃO

Será analisada a série histórica da Conab produção de feijão. Os dados da série de

produção de feijão anual são referentes ao período de 1976 a 2019.

4.4.1 Série histórica da produção de feijão

A plotagem do gráfico da série da produção de feijão, referente ao período de 1976 a 2019, está apresentada na Figura 57.

Figura 57 – Variação da produção de feijão de 1976 a 2019.



Fonte: Autor (2021).

Foi adicionada no gráfico uma linha de tendência para o auxílio da visualização. É observado que o gráfico apresenta uma possível tendência de crescimento, porém ao se selecionar um período mais recente (1993 – 2019) se observa uma possível estacionariedade, como pode ser observado na Figura 58.

Figura 58 – Variação da produção de feijão de 1993 a 2019.

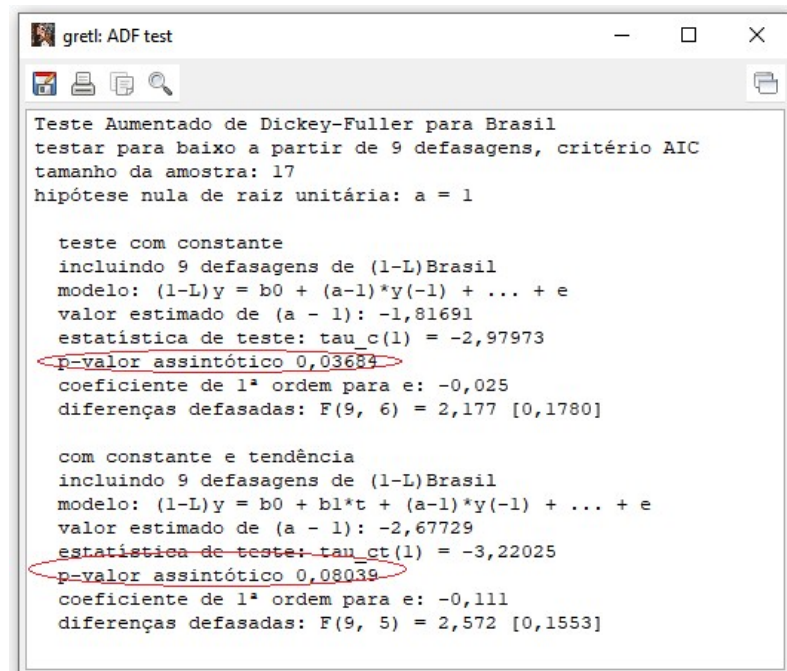


Fonte: (Autor) 2021.

4.4.1.1 Testes de estacionariedade

Para verificar a possível estacionariedade identificada na Figura 58, foram realizados os testes de estacionariedade. O p-valor obtido para o teste Dickey-Fuller aumentado está indicado na Figura 59.

Figura 59 – Teste Dickey-Fuller aumentado para a produção de feijão de 1993 a 2019.



```

gretl: ADF test

Teste Aumentado de Dickey-Fuller para Brasil
testar para baixo a partir de 9 defasagens, critério AIC
tamanho da amostra: 17
hipótese nula de raiz unitária: a = 1

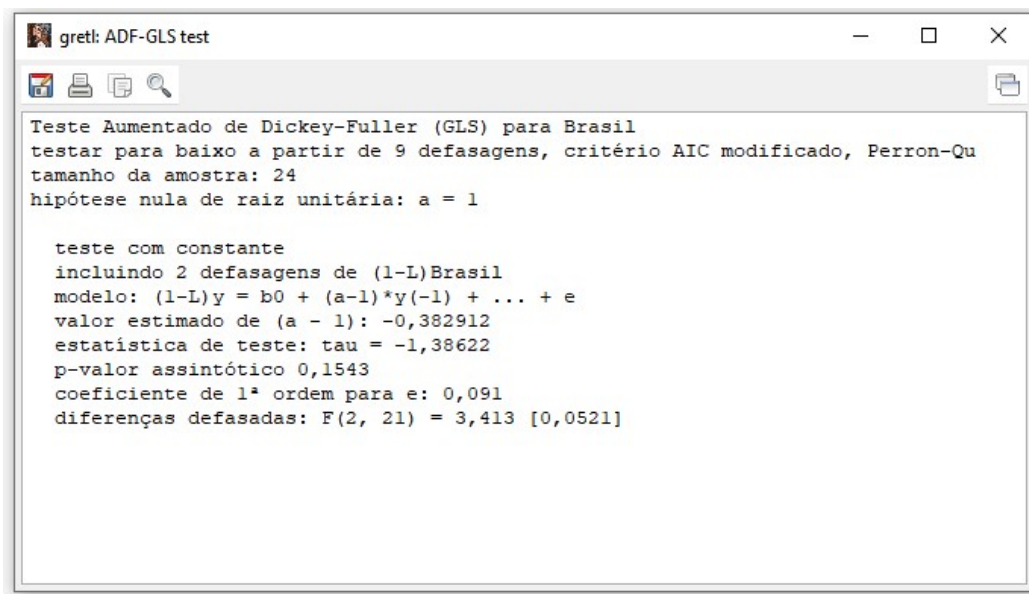
teste com constante
incluindo 9 defasagens de (1-L)Brasil
modelo: (1-L)y = b0 + (a-1)*y(-1) + ... + e
valor estimado de (a - 1): -1,81691
estatística de teste: tau_c(1) = -2,97973
p-valor assintótico 0,03684
coeficiente de 1ª ordem para e: -0,025
diferenças defasadas: F(9, 6) = 2,177 [0,1780]

com constante e tendência
incluindo 9 defasagens de (1-L)Brasil
modelo: (1-L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
valor estimado de (a - 1): -2,67729
estatística de teste: tau_ct(1) = -3,22025
p-valor assintótico 0,08039
coeficiente de 1ª ordem para e: -0,111
diferenças defasadas: F(9, 5) = 2,572 [0,1553]
  
```

Fonte: Autor (2021).

Observa-se que o p-valor obtido para o teste foi abaixo, o que indica que se pode rejeitar a hipótese de não estacionariedade. Para validar obtido foi aplicado o teste ADF-GLS, o resultado do teste está apresentado na Figura 60.

Figura 60 – Teste ADF-GLS para a produção de feijão de 1993 a 2019.



```

gretl: ADF-GLS test

Teste Aumentado de Dickey-Fuller (GLS) para Brasil
testar para baixo a partir de 9 defasagens, critério AIC modificado, Perron-Qu
tamanho da amostra: 24
hipótese nula de raiz unitária: a = 1

teste com constante
incluindo 2 defasagens de (1-L)Brasil
modelo: (1-L)y = b0 + (a-1)*y(-1) + ... + e
valor estimado de (a - 1): -0,382912
estatística de teste: tau = -1,38622
p-valor assintótico 0,1543
coeficiente de 1ª ordem para e: 0,091
diferenças defasadas: F(2, 21) = 3,413 [0,0521]
  
```

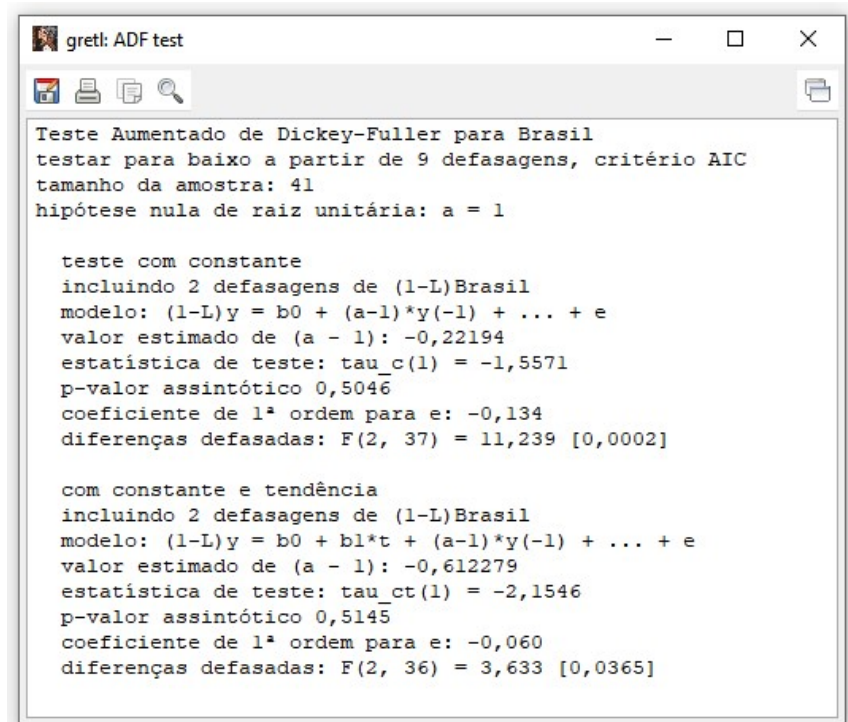
Fonte: Autor (2021).

Percebe-se que o valor obtido foi acima de 0,1, o que indica que a série é não estacionária. A seleção de intervalos para a obtenção de estacionariedade e aplicação dos modelos da média móvel e do alisamento exponencial se mostra suscetível a falhas, devido à possível contradição de valores obtidos, a limitação da escolha do intervalo e impossibilidade da realização dos testes para intervalos muito pequenos.

4.4.1.2 Modelo ARIMA

Para a aplicação do modelo ARIMA foi realizado o teste de estacionariedade Dikey-Fuller aumentado para a série histórica de produção de feijão de 1976 a 2019. A Figura 61 mostra o resultado obtido.

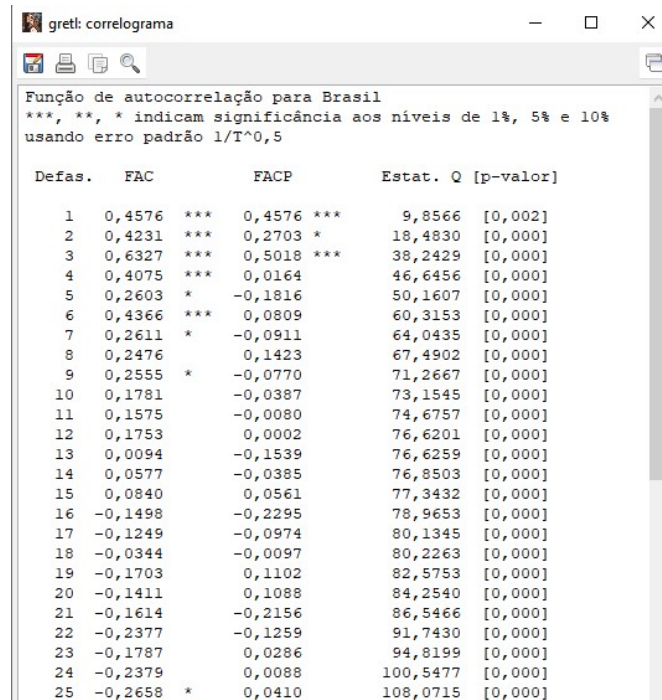
Figura 61 - Teste Dickey-Fuller aumentado para a produção de feijão de 1976 a 2019.



Fonte: Autor (2021).

Observa-se que os resultados para o p-valor estão altos, o que indica que a série não é estacionária. A Figura 62 mostra o correlograma da série.

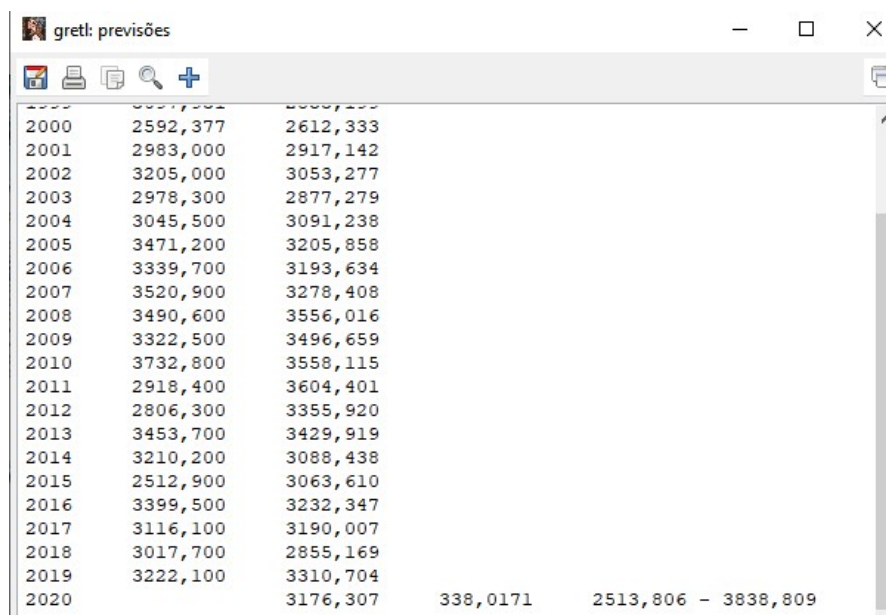
Figura 62 – Correlograma para a série de produção de feijão de 1976 a 2019.



Fonte: Autor (2021).

Observe que o valor de FAC quase zero na linha 14 e logo após volta a crescer. A análise das diferenças para a série resultou em um modelo ARIMA com $d=2$, $p=2$ e $q=1$. A Figura 63 mostra os resultados obtidos para as previsões do modelo.

Figura 63 – Aplicação do modelo ARIMA à série de produção de feijão.



Fonte: Autor (2021).

Observa-se que os valores previstos estão próximos aos reais, os resultados para o cálculo dos erros obtidos para as previsões do modelo estão apresentados na Tabela 10.

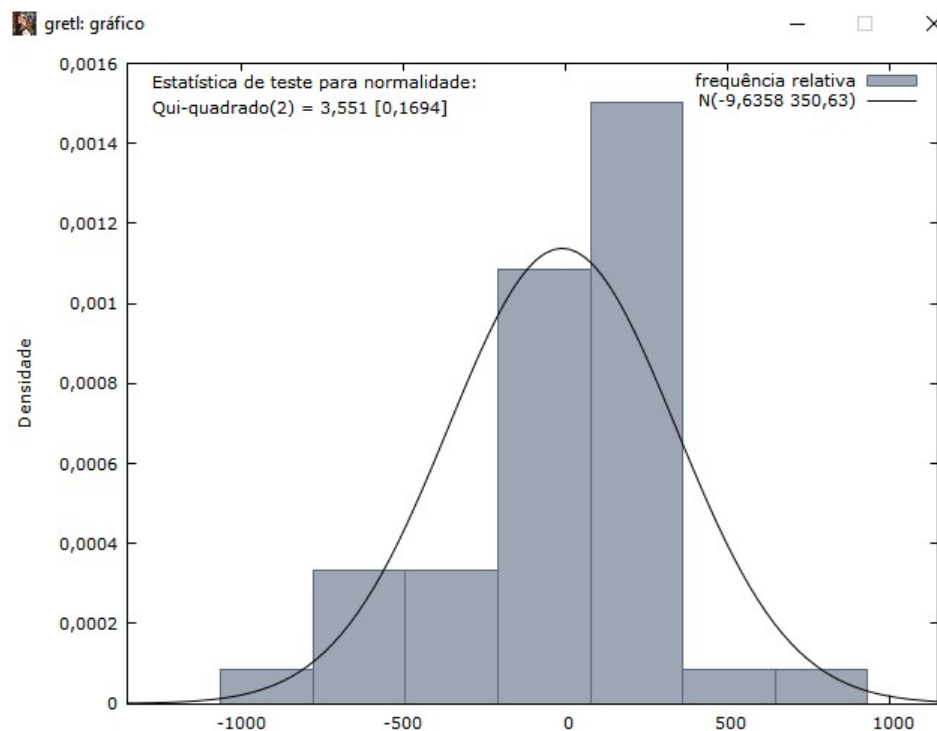
Tabela 10 – Erros para aplicação do modelo ARIMA à série de produção de feijão.

	Valor Real 2020	Valor obtido 2020	Erro Absoluto	Desvio médio absoluto	% de Erro (Erro Absoluto)
ARIMA	3144,2	3176,307	32,107	188,8	1,0%

Fonte: Autor (2021).

Observa-se que o erro obtido para a previsão realizada para 2020 foi baixo e o desvio médio absoluto se mostrou baixo se comparado aos valores da produção ao longo dos anos. O teste da normalidade dos resíduos está apresentado na Figura 64.

Figura 64 – Teste da normalidade dos resíduos para a previsão da produção de feijão.



Fonte: Autor (2021).

Nota-se que a distribuição dos resíduos se assemelha à distribuição normal, entretanto não há tanta aderência como nas séries anteriores.

4.4.1.3 Discussão dos resultados

Percebe-se pelos resultados obtidos que a seleção de intervalos para a obtenção de estacionariedade e aplicação dos modelos da média móvel e do alisamento exponencial se mostra suscetível a falhas, devido à possível contradição de valores obtidos, a limitação da escolha do intervalo e impossibilidade da realização dos testes para intervalos muito pequenos, além de requerer mais etapas devido a seleção de novos intervalos, caso o primeiro intervalo selecionado não se mostre interessante.

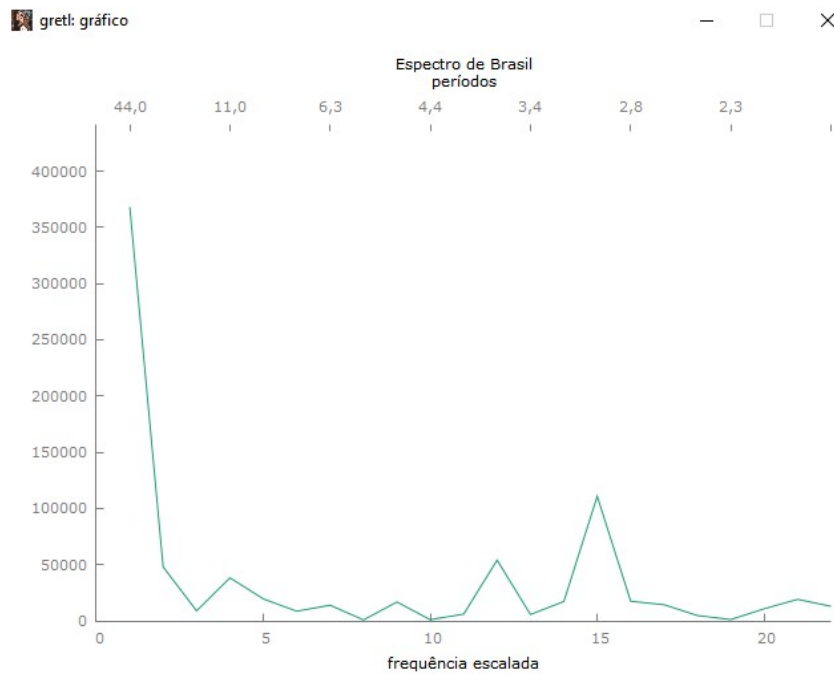
O modelo ARIMA novamente se mostra bastante adequado para descrever séries com tendência, garantindo, na maioria das vezes, uma aplicação ágil e precisa. A distribuição dos erros mostra que por mais que o modelo se mostre coerente ele pode ser afetado por um possível caráter sazonal dos dados, mesmo que as observações ocorram de maneira anual.

4.5 MODELO SARIMA

A aplicação do modelo SARIMA se mostrou prejudicada devido ao fato de o intervalo entre as observações se dar de maneira anual. Mesmo ao se analisar *commodities* em que se tem mais de uma safra durante o ano a aplicação do modelo não poderia ocorrer, pois as safras têm períodos diferentes.

A análise do periodograma para algumas das séries estudadas indica uma possível presença de sazonalidade, como pode ser visto na Figura 65 para a série de produção de feijão.

Figura 65 – Períodograma da produção de feijão.



Fonte: Autor (2021).

A presença dessa possível sazonalidade, indicada por um ponto destoante no primeiro ponto do gráfico, pode explicar a diferença da aderência dos erros a uma distribuição normal, entretanto mesmo para esses casos o modelo ARIMA se mostrou bem eficaz e requiriu menos etapas. Vale ressaltar que a sazonalidade para observações em intervalos de um ano para o contexto de produção de *commodities* no brasil não teria muito significado.

5 CONSIDERAÇÕES FINAIS

Serão verificados os objetivos e as respostas as questões de pesquisa e apresentadas sugestões para a continuidade do trabalho.

5.1 VERIFICAÇÃO DAS RESPOSTAS AS QUESTÕES DE PESQUISA

A pesquisa realizada, sendo classificada como pesquisa aplicada, cumpriu com a sua definição de se gerar conhecimentos para a aplicação prática dirigida à solução de problemas específicos. A verificação da resposta a questão de pesquisas será realizada para cada questão de forma individual.

a) Como determinar, em uma série histórica, um modelo mais aderente à realidade?

Para se determinar um modelo mais aderente a realidade, deve-se analisar o intervalo entre as observações que constituem a série, analisar se é possível considerar a série como estacionária, analisar os resíduos da série para verificar se houve uma boa descrição do modelo utilizado, calcular os erros para verificação da precisão do modelo e entender qual o contexto daquilo que é analisado.

b) Quais são os principais fatores que influenciam na precisão de um modelo de previsão?

Os principais fatores que influenciam na precisão de um modelo de previsão são o intervalo de dados utilizado e a capacidade do modelo de incorporar as componentes da decomposição clássica. Foi observado que modelos que absorvem melhor as componentes da decomposição clássica conseguem ser aplicados obtendo resultados precisos e obtidos de forma ágil.

c) Quais modelos de previsão descrevem adequadamente séries históricas da Conab?

Devido ao intervalo entre as observações para as séries históricas da Conab ser de um ano e para commodities que tem mais de uma safra no ano o intervalo entre as observações variar, os modelos ARIMA, média móvel e alisamento exponencial se mostraram como

úteis para a descrição das séries. O modelo ARIMA se mostrou o mais eficiente, devido à grande maioria das séries apresentar um comportamento tendencioso, os modelos de média móvel e alisamento exponencial se mostraram adequados em apenas algumas ocasiões.

Verifica-se então que foram obtidas respostas para todas as questões de pesquisa.

5.2 VERIFICAÇÃO DO ATENDIMENTO AOS OBJETIVOS

Assim como a verificação das respostas às questões de pesquisa, a verificação do atendimento aos objetivos será realizada de forma individual e está apresentada a seguir.

- a) Investigar, com o auxílio de *softwares*, os diferentes comportamentos das séries históricas visando seu uso em modelos de previsão.

Os *softwares* utilizados se mostraram bastante úteis para investigar o comportamento das séries históricas, disponibilizando ferramentas complementares que possibilitaram o entendimento da série antes da aplicação dos modelos.

- b) Identificar situações em que é recomendado o uso de um determinado modelo de previsão.

Por meio da comparação entre o que se era esperado ao observar o gráfico da série e o que se obteve depois da aplicação dos modelos, assim como o estudo dos resíduos, foi possível perceber que os modelos de alisamento exponencial simples e média móvel se mostram interessantes apenas para séries estacionárias, e que deve se ter dados suficientes para que seja possível atestar a estacionariedade da série. O modelo ARIMA se mostrou bastante eficiente para a descrição das séries, pois a maioria das séries apresentou comportamento de tendência. O modelo SARIMA teve pouca aplicação devido ao fato das observações ou serem anuais ou serem feitas em intervalos diferentes de tempo, tendo então a sua aplicação acontecendo de forma mais coerente em situações em que as observações são realizadas com mais frequência.

- c) Determinar um intervalo de variação dos dados em que cada método de previsão analisado funcione e ofereça valores previstos compatíveis com a realidade modelada.

Foi observado que para a aplicação do modelo ARIMA, caso não haja sazonalidade nos dados, pode ocorrer desde que se tenha uma base de dados suficientemente grande. Para os modelos de alisamento exponencial simples e média móvel foi observado que devem ser levados em conta trechos mais recentes da série, de forma a tornar o comportamento analisado o mais estacionário possível.

Foi percebido que estimar o parâmetro α , para o modelo de alisamento exponencial, com base em valores mais recentes obtidos se mostrou bem útil para aumentar a precisão do modelo. Para o modelo de média móvel se mostrou interessante reduzir o valor de r quando se identificar variações no comportamento estacionário da série.

O modelo SARIMA mostrou que sua aplicação requer observações feitas mais de uma vez ao longo do ano. Por ser uma variação do modelo ARIMA, o modelo SARIMA se mostra uma opção para casos que o uso do modelo ARIMA não se mostre eficiente, levando em consideração o contexto que está sendo analisado.

O objetivo geral do trabalho, por sua vez, é “identificar como se dá a relação da variação dos dados em séries históricas da Conab e como se adequam os principais modelos de previsão no seu tratamento “. Foi verificado que a variação dos dados se dá, de maneira geral, com a presença de tendência e é necessário alterar os parâmetros para cada modelo e/ou o período analisado para se garantir uma maior precisão.

Na realização do trabalho foi verificado que os objetivos do estudo estão bem relacionados. Por meio dos usos das ferramentas dos *softwares* as séries foram devidamente descritas, o que possibilitou identificar com agilidade padrões para situações em que é mais indicado determinado modelo de previsão. Além disso, foi estudado como variar o trecho estudado de modo a possibilitar a aplicação de um determinado modelo com uma maior precisão. Verifica-se então o atendimento aos objetivos.

5.3 SUGESTÕES PARA A CONTINUIDADE DO TRABALHO

Como sugestão para trabalhos futuros, tem-se primeiramente a utilização de uma outra base de dados, em que as observações ocorram em intervalos menores de tempo. De forma a estudar de maneira mais precisa a aplicação do modelo SARIMA.

Outra sugestão para a continuidade do trabalho seria o estudo de outros modelos de previsão, para que sejam possíveis novas discussões ao se estudar uma série histórica e melhor selecionar modelos existentes para se obter mais precisão.

Como última sugestão, seria interessante um estudo de maneira conjunta sobre sobre as políticas do país, a variação do mercado e o comportamento das séries de uma determinada *commodity*, de forma a atribuir outras explicações para a variação dos dados.

REFERÊNCIAS

- ANTUNES, J. L. F.; CARDOSO, M. R. A. Uso da análise de séries temporais em estudos epidemiológicos. **Epidemiologia e Serviços de Saúde**, São Paulo, v. 20 p. 565-576, 2015. Disponível em: <https://www.scielo.org/article/ress/2015.v24n3/565-576/pt/>. Acesso em: 04 dez. 2020.
- ARAÚJO, J. C. C. DE; SOUZA, R. C. Estatística clássica e inteligência computacional na previsão de séries temporais. In: CONGRESSO BRASILEIRO DE INTELIGÊNCIA COMPUTACIONAL, 12. 2015, Curitiba. **Anais [...]**. Curitiba: Abricom, 2015. Disponível em: http://abricom.org.br/wp-content/uploads/2016/01/cbic2015_submission_99.pdf. Acesso em: 27 set. 2020.
- AZEVEDO, P. R. M. DE, Testes de hipótese: primeiras ideias. In: AZEVEDO, P. R. M. de **Introdução à estatística**. Natal: EDUFN, 2016. cap. 9, p. 166-178. Disponível em: <https://repositorio.ufrn.br/jspui/bitstream/123456789/21298/2/Introduc%cc%a7a%cc%83o%20a%cc%80%20Estati%cc%81stica%20%28digital%29.pdf>. Acesso em: 28 set. 2020
- BASTOS, R. F. Avaliação de desempenho de modelos de séries temporais para previsões da produção de óleo diesel nacional. In: SIMPÓSIO DE PESQUISA OPERACIONAL & LOGÍSTICA DA MARINHA, Rio de Janeiro, v. 18, 2016. **Anais [...]**. Rio de Janeiro: Blucher, 2016. Disponível em: <http://pdf.blucher.com.br.s3-sa-east-1.amazonaws.com/marineengineeringproceedings/spolm2015/140494.pdf>. Acesso em: 25 set. 2020.
- CONAB. Companhia Nacional de Abastecimento. Institucional: **A Conab**. Disponível em: <https://www.conab.gov.br/institucional>. Acesso em: 01 jun. 2020.
- DEB, C.; ZHANG, F.; YANG, J.; LEE, S. E.; SHAH, W. K. A review on time series forecasting techniques for building energy consumption. **Renewable and Sustainable Energy Reviews**, Singapura, v. 74, p. 902-924, 2017. Disponível em: <https://www-sciencedirect.ez87.periodicos.capes.gov.br/science/article/pii/S1364032117303155?via%3Dihub%20Artigo>. Acesso em: 23 jul. 2020.
- FLEURY, M. T. L.; COSTA, W. S. R. DA, Pesquisa aplicada: conceitos e abordagens. **Anuário de pesquisa GVPesquisa**, São Paulo, p. 10-15, 2017. Disponível em: <http://bibliotecadigital.fgv.br/ojs/index.php/apgvpesquisa/article/download/72796/69984>. Acesso em: 27 jul. 2020.
- GRETL. **Gnu regression, econometrics and time-series library**, 2017. Disponível em: <http://gretl.sourceforge.net/pt.html>. Acesso em: 08 set. 2020.
- LI, Y.; SHAKYA, S.; OWUSU, G. Integrated forecasting and planning for strategic and operational telecom field service delivery. In: 2017 COMPUTING CONFERENCE, Londres, p. 981-989, jul. 2017. **Anais [...]**. Londres: IEEE, 2017. Disponível em: <https://ieeexplore-ieee-org.ez87.periodicos.capes.gov.br/stamp/stamp.jsp?tp=&arnumber=8252212&tag=1>. Acesso em: 27 jul. 2020.

MARTIN, A. C.; HENNING, E.; WALTER, O. M.; KONRATH, A. C. Análise de séries temporais para previsão da evolução do número de automóveis no município de Joinville. **Espacios**, Joinville, v. 37, n. 1, p. 29, nov. 2015. Disponível em: <https://www.revistaespacios.com/a16v37n06/16370629.html>. Acesso em: 28 set. 2020.

MATTOS, R. S. **Tendências e raízes unitárias**, 2018. Apresenta uma publicação com material didático do departamento de economia da UFJF. Disponível em: https://www.ufjf.br/wilson_rotatori/files/2011/05/Tendencias-e-Raizes-Unitarias-2018.pdf. Acesso em: 21 jan. 2021.

MICROSOFT. **Microsoft Excel**. Versão 365. São Paulo: Microsoft, 2020. Disponível em: <https://www.microsoft.com/pt-br/microsoft-365/excel?rtc=1>. Acesso em: 28 set. 2020

MORAIS, J.; FERREIRA, P. B.; JACOME, I. M. T.; MELLO, R.; BREDAS, F. C.; RORATO, P. R. Curva de crescimento de diferentes linhagens de frango de corte caipira. **Ciência Rural**, Santa Maria, v. 45, n. 10, p. 1872-1878, out. 2015. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-84782015001001872&lng=en&nrm=iso. Acesso em: 02 out. 2020.

MORETTIN, A.; TOLOI, C. M. C. **Análise de séries temporais**. 3. ed. São Paulo: BLUSHER, 2018. 455 p.

OLIVEIRA, E. S. de; DOURADO, A. D. J.; MELLO, J. V. B. Aplicação de modelos de previsão de demanda em uma fábrica de embalagens plásticas. **Revista da Universidade Vale do Rio Verde**, Três Corações, v. 15, n. 1, dez. 2017. Disponível em: http://periodicos.unincor.br/index.php/revistaunincor/article/view/2911/pdf_703. Acesso em: 29 set. 2020.

RABELO, L. **Princípios básicos para criar previsões de séries temporais**, 2019. Apresenta uma discussão sobre as características de séries temporais. Disponível em: <https://medium.com/ensina-ai/princ%C3%ADpios-b%C3%A1sicos-para-criar-previs%C3%B5es-de-s%C3%A9ries-temporais-e58c451a25b#:~:text=S%C3%A9rie%20estacion%C3%A1ria&text=Uma%20s%C3%A9rie%20temporal%20estacion%C3%A1ria%20%C3%A9,estat%C3%ADsticas%20mudam%20com%20o%20tempo>. Acesso em: 08 set. 2020.

SCOPUS. **Author details Wang Xiangyu**, 2020. Apresenta *links* para leitura dos *abstracts* dos artigos do autor de acordo com o filtro solicitado. Disponível em: <https://www-scopus-com.ez87.periodicos.capes.gov.br/authid/detail.uri?authorId=8945580300&eid=2-s2.0-84944511801>. Acesso em: 23 jul. 2020.

SCOPUS. **Citation overview**, 2020. Apresenta um infográfico com informações sobre o número de publicações de artigos e citações. Disponível em: https://www-scopus-com.ez87.periodicos.capes.gov.br/cto2/main.uri?origin=resultslist&stateKey=CTOF_1224242308. Acesso em: 22 jul. 2020.

SILVA, A. **Modelo SARIMA (aplicação real)**, 2020. Apresenta um vídeo sobre a utilização do Gretl. Disponível em: <https://www.youtube.com/watch?v=PyUIttebgGA>. Acesso em: 03 out. 2020.

XAVIER, J. M. N. **Análise e previsão de séries temporais com modelos ARIMA e análise espectral singular**. 2016. 92 f. Dissertação (Mestrado em Bioestatística e Biometria) – Universidade Aberta, 2016. Disponível em: https://repositorioaberto.uab.pt/bitstream/10400.2/5873/1/Disserta%C3%A7%C3%A3o_Jorge.Xavier_%201104916.pdf. Acesso em: 22 jun. 2020.

ZUCOLOTO, A. C.; GIAROLA, L. T. P.; ROCHA, R. C. DA, Modelagem da exportação brasileira de automóveis. **Revista eletrônica Matemática e Estatística em Foco**, Uberlândia, v. 6, n. 1, p. 12-23, maio. 2017. Disponível em: <http://www.seer.ufu.br/index.php/matematicaeestatisticaemfoco/article/view/39080>. Acesso em: 28 set. 2020.