



**Universidade Federal de Uberlândia  
Faculdade de Matemática**

**Bacharelado em Estatística**

**MODELOS DE PREVISÃO VIA ANÁLISE  
DE SÉRIES TEMPORAIS DA  
PRODUÇÃO DE CAFÉ NO BRASIL NO  
PERÍODO DE 1960 A 2021**

**Nágela de Sousa Machado**

**Uberlândia-MG**

**2022**



**Nágela de Sousa Machado**

**MODELOS DE PREVISÃO VIA ANÁLISE  
DE SÉRIES TEMPORAIS DA  
PRODUÇÃO DE CAFÉ NO BRASIL NO  
PERÍODO DE 1960 A 2021**

Trabalho de conclusão de curso apresentado à Coordenação do Curso de Bacharelado em Estatística como requisito parcial para obtenção do grau de Bacharel em Estatística.

Orientador: Dr. Quintiliano Siqueira Schroden Nomelini

**Uberlândia-MG  
2022**





**Universidade Federal de Uberlândia  
Faculdade de Matemática**

**Coordenação do Curso de Bacharelado em Estatística**

A banca examinadora, conforme abaixo assinado, certifica a adequação deste trabalho de conclusão de curso para obtenção do grau de Bacharel em Estatística.

Uberlândia, 10 de Março de 2022

**BANCA EXAMINADORA**

---

Dr. Quintiliano Siqueira Schrodin Nomelini

---

Dr. José Waldemar da Silva

---

Dr. Lúcio Borges de Araújo

**Uberlândia-MG  
2022**



# AGRADECIMENTOS

Agradeço primeiramente à Deus, que está acima de todas as coisas.

Aos meus professores, que durante todo o processo da minha graduação se mostraram dispostos a ajudarem tanto dentro quanto fora das disciplinas ofertadas.

À minha tia Jesuína, sem ela nem sei se eu teria chegado até aqui, sempre me incentivando a estudar, ser independente.

Ao meu amado Heráclito, que sempre esteve comigo em todos os momentos, bons e ruins, e sempre me incentivando a não desistir dos meus sonhos.

Enfim, agradeço a todos os meus familiares e amigos que de alguma forma contribuiu pra eu chegar até aqui, e a todos que torceram e torcem pelo meu sucesso e minha felicidade.





# RESUMO

O Brasil ocupa a primeira posição como exportador de café no mundo, além de ocupar a segunda posição em relação ao consumo de café. O volume de produção dos Cafés do Brasil para a safra de 2020, correspondeu a um aumento de 25% em relação à safra de 2019. Objetivou-se neste estudo analisar a série histórica da produção de café no Brasil no período das safras de 1960/1961 e 2021/2022 e realizar previsões para os próximos 5 anos / safra, com o auxílio dos modelos de séries temporais, ou seja, até a safra de 2026/2027. Para seleção do melhor modelo utilizado, foram realizadas as análises dos critérios de adequação de modelos. A construção dos modelos de séries temporais foi feita pela teoria de Box Jenkins. Após a definição do modelo ARIMA (1,1,1) para a série temporal dos dados, as previsões se mostraram adequadas em relação aos valores da série analisada.

**Palavras-chave:** Produção, Café, Safra, Séries Temporais, Previsão.



# ABSTRACT

Brazil occupies the first position as a coffee exporter in the world, in addition to occupying the second position in relation to coffee consumption. The production volume of Brazilian Coffee for the 2020 harvest corresponded to a 25% increase in comparison with the 2019 harvest. The objective of this study was to analyze the historical series of coffee production in Brazil in the period of the 1960/1961 and 2021/2022 and make forecasts for the next 5 years/harvest, with the help of time series models, that is, until the 2026/2027 harvest. In order to select the best model used, analyzes of the model adequacy criteria were performed. The construction of time series models was made by the theory of Box Jenkins. After defining the ARIMA model (1,1,1) for the time series of the data, the predictions proved to be adequate in relation to the values of the analyzed series.

**Keywords:** Production, Coffee, Crop, Time Series, Forecast.



# SUMÁRIO

<b>Lista de Figuras</b>	<b>I</b>
<b>Lista de Tabelas</b>	<b>III</b>
<b>1 Introdução</b>	<b>1</b>
<b>2 Materiais e Métodos</b>	<b>3</b>
2.1 Processos Estocásticos . . . . .	3
2.2 Séries Temporais . . . . .	3
2.3 Tendência . . . . .	4
2.4 Sazonalidade . . . . .	5
2.5 Ruído Branco . . . . .	5
2.6 Função Autocorrelação e Correlograma . . . . .	6
2.7 Modelo Arima . . . . .	7
2.8 Suavização Exponencial de Holt (SEH) . . . . .	9
2.9 Seleção dos modelos . . . . .	9
2.9.1 Critério de Akaike (AIC) . . . . .	9
2.9.2 Erro percentual médio absoluto (MAPE) . . . . .	9
2.9.3 Raiz do erro quadrático médio . . . . .	10
2.10 Análise dos resíduos . . . . .	10
<b>3 Resultados</b>	<b>11</b>
<b>4 Conclusões</b>	<b>21</b>
<b>Referências Bibliográficas</b>	<b>23</b>
<b>Apêndice A Scripts R</b>	<b>25</b>
<b>Apêndice B P-valores Modelos propostos</b>	<b>29</b>



---

# LISTA DE FIGURAS

2.1	Exemplo de um processo ruído branco . . . . .	6
3.1	Série histórica da produção de café entre 1960/1961 e 2021/2022 (a) e entre 1960/1961 e 2016/2017 (b). . . . .	11
3.2	Função de Autocorrelação e Autocorrelação parcial da série histórica da produção de café no Brasil - safra até 2016/2017. . . . .	12
3.3	Série com uma diferença – dados parciais. . . . .	12
3.4	Função de autocorrelação e autocorrelação parcial da série diferenciada – dados parciais. . . . .	13
3.5	Série estimada e série histórica da produção de café no Brasil - dados parciais. . . . .	16
3.6	Análise dos resíduos do modelo Arima (1,1,1) – dados parciais. . . . .	17
3.7	Periodograma acumulado – dados parciais. . . . .	17
3.8	Função de Autocorrelação e Autocorrelação parcial da série completa. . . . .	18
3.9	Função de autocorrelação e autocorrelação parcial da série completa diferenciada. . . . .	18
3.10	Série estimada e série histórica da produção de café no Brasil – dados completos. . . . .	19
3.11	Representação gráfica da previsão da produção de café no Brasil entre 2022/2023 até 2026/2027. . . . .	20





---

# LISTA DE TABELAS

3.1	Resultados teste Cox-Stuart . . . . .	13
3.2	Modelos Arima propostos – dados parciais . . . . .	14
3.3	Estimativas dos critérios de seleção do modelo – até a safra 2016/2017 . . . . .	14
3.4	Valores previstos pelos modelos e observados para as safras entre 2017/2018 a 2021/2022, para a série histórica da produção de café no Brasil – dados parciais . . . . .	15
3.5	Parâmetros do modelo selecionado – dados parciais . . . . .	15
3.6	Estimativas dos critérios de seleção do modelo – dados completos . . . . .	19
3.7	Parâmetros do modelo selecionado – dados completos . . . . .	19
3.8	Previsão da Produção de café no Brasil entre 2022/2023 a 2026/2027 . . . . .	20



# 1. INTRODUÇÃO

O Brasil, sendo o maior exportador de café no mercado mundial e ocupando a segunda posição em relação aos países mais consumidores da bebida, exerce um papel fundamental na economia tanto nacional quanto mundial. Conforme dados do Ministério da Agricultura, Pecuária e Abastecimento (Mapa), a cafeicultura brasileira é uma das mais exigentes do mundo, em relação às questões sociais e ambientais, e há uma preocupação em se garantir a produção de um café sustentável[1].

O volume de produção dos Cafés do Brasil para a safra de 2020, somadas as espécies arábica e conilon, foi estimado em 61,62 milhões de sacas de 60kg, o que representou um aumento de 25% em relação à safra de 2019. A produtividade também apresentou uma evolução de 20,2% em 2020, alcançando 32,7 sacas/ha, contra 27,2 sacas/ha em 2019[5].

Entre os séculos XIX e início do século XX, o café foi um dos principais produtos de exportação Brasileira, tendo seu início no século XVIII, sendo que suas primeiras plantações foram realizadas em meados da década de 1720 no Estado do Pará.

As regiões produtoras de café são: Minas Gerais, Espírito Santo, São Paulo, Bahia, Paraná e Rondônia, sendo o mapa representativo das regiões produtoras de café, segundo informações coletadas no site da CONAB[2]:

- Minas Gerais: localizado na região Sudeste, Minas Gerais é o maior estado produtor de café do Brasil, responde por cerca de 50% da produção nacional e é uma das principais fontes de cafés especiais do país. Praticamente 100% das plantações são de café Arábica, cultivado em quatro regiões produtoras: Sul de Minas, Cerrado de Minas, Chapada de Minas e Matas de Minas, que exportam seus cafés pelos portos de Santos, Rio de Janeiro e Vitória.

- Espírito Santo: o Espírito Santo é o segundo maior estado produtor de café do país e o principal produtor de Conilon (Robusta). Com plantações de café Conilon nas áreas mais quentes ao norte, região chamada de Conilon Capixaba, e de Arábica ao sul, região conhecida como Montanhas do Espírito Santo, o estado é grande fornecedor do mercado brasileiro e escoar seus cafés especiais pelo porto de Vitória.

- São Paulo: o estado de São Paulo é um dos mais tradicionais no cultivo de café. Sua produção é exclusivamente de Arábica, distribuída em duas regiões: Mogiana e Centro-Oeste Paulista, que alternam fazendas com pequenas propriedades e produzem cafés especiais em áreas específicas. São Paulo abriga o porto de Santos, que escoar cerca de 2/3 das exportações de café do Brasil.

- Bahia: o estado da Bahia está localizado na região nordeste do Brasil, de clima quente e

temperaturas mais altas. São duas as regiões produtoras de café no estado: Planalto da Bahia e Cerrado da Bahia, onde se cultiva Arábica. Ao sul do estado também há áreas onde se produz café Conilon (Robusta).

- Paraná: o Paraná é o estado produtor de café localizado mais ao sul do país. Apenas café Arábica é cultivado em plantações adensadas, que usam variedades adequadas ao clima mais frio da região. Outrora o maior estado produtor do país, vem recuperando sua produção com forte ênfase no processo de cereja descascado.

- Rondônia: o estado de Rondônia se localiza na região norte do país. Com uma produção anual de aproximadamente 2 milhões de sacas, o estado produz exclusivamente café Conilon (Robusta). A cafeicultura é tradicional e familiar, com pequenas propriedades.

Tendo em vista a grande importância da produção e consumo de café no Brasil, objetivou-se aplicar-se a técnica de análise de séries temporais na série histórica da produção brasileira de café, e posteriormente realizar previsões para os próximos 5 (cinco) anos.

A escolha do tema para o presente estudo é importante para mostrar o quão pode ser amplo a aplicabilidade dos métodos estatísticos no cotidiano independentemente da área de pesquisa, em especial neste caso aplicando a metodologia de séries temporais. Com os resultados desse estudo, teremos uma perspectiva de como será a produção de café no Brasil nos próximos 5 anos, sendo um importante resultado para um reflexo da economia futura do país.

Uma série temporal é uma descrição do passado, um procedimento lógico para realizar previsões e fazer uso desses dados históricos. Se os dados passados são indicativos do que se esperar no futuro, pode-se então postular um modelo matemático que é representativo do processo. O modelo pode então ser usado para gerar previsões.[10].

Esse trabalho teve como objetivo realizar a previsão para a produção de café no Brasil para os próximos 5 (cinco) anos sendo o intervalo entre safras de 2022/2023 até a safra de 2026/2027, podendo dessa forma ter uma estimativa do comportamento futuro da produção do café no país. Objetivou-se ajustar modelos de séries temporais através do método Box e Jenkins, bem como verificar todos os pressupostos necessários para a realização das análises com o auxílio do software R v.4.1.1[12].

## 2. MATERIAIS E MÉTODOS

A série histórica da safra de 1960/1961 até 2021/2022 da produção de café no Brasil foi coletada através do site da USDA (United States Department of Agriculture Foreign Agricultural Service, 2020). As análises foram realizadas a partir do método de Box Jenkins[9] com o auxílio do software R v.4.1.1[12].

### 2.1 PROCESSOS ESTOCÁDICOS

Segundo Morettin e Toloi[9], seja  $T$  um conjunto arbitrário. Um processo estocástico é uma família  $Z = \{Z(t), t \in T\}$  tal que, para cada  $t \in T$  é uma variável aleatória.

Ainda conforme Morettin e Toloi[9], os modelos utilizados para descrever séries temporais são processos estocásticos, isto é, processos controlados por leis probabilísticas. Um processo estocástico é uma família de variáveis aleatórias que suporemos definidas num espaço de probabilidade.

Um exemplo de processo estocástico é um conjunto infinito de séries formadas pelos resultados de 100 lançamentos consecutivos de um dado[7].

Os processos estocásticos podem ser divididos em processos estacionários e processos não estacionários. Um processo estocástico  $Z_t$  é estacionário quando as propriedades estatísticas de qualquer sequência finita  $Z_1, Z_2, \dots, Z_k$  de componentes de  $Z_t$  são semelhantes às da sequência  $Z_{1+h}, Z_{2+h}, \dots, Z_{k+h}$  para qualquer número inteiro  $h$ [3]. Em contrapartida, um processo estocástico  $Z_t$  é não estacionário quando as propriedades estatísticas de ao menos uma sequência finita  $Z_1, Z_2, \dots, Z_k$  de componentes de  $Z_t$  são diferentes das de sequência  $Z_{1+h}, Z_{2+h}, \dots, Z_{k+h}$  para ao menos um número inteiro  $h$ .

### 2.2 SÉRIES TEMPORAIS

Segundo Nogueira[10], uma série de tempo é uma série de observações de alguma quantidade de interesse (uma variável aleatória) em relação ao tempo. Assim, se  $X_i$  é uma variável aleatória de interesse no tempo  $i$ , e se observações são tomadas nos tempos  $i = 1, 2, \dots, t$ , então os valores observados  $\{X_1 = x_1, X_2 = x_2, X_t = x_t\}$  são uma série de tempo.

A partir da possibilidade de mensuração do tempo, pode-se estabelecer algumas relações entre a passagem do tempo e a ocorrência de fenômenos. Na observação de um fenômeno na natureza, pode-se perceber que este é estimulado por uma infinidade de outros fenômenos,

com diferentes graus de interferência. Uma das variáveis que podem estar interferindo em um fenômeno é o transcurso do tempo. Dessa forma, o tempo pode ser utilizado como uma variável auxiliar para o seu entendimento[6].

Há basicamente, dois enfoques usados na análise de séries temporais. Em ambos, o objetivo é construir modelos para as séries, com propósitos determinados. No primeiro enfoque, a análise é feita no domínio temporal e os modelos propostos são modelos paramétricos (com um número finito de parâmetros). No segundo, a análise é conduzida no domínio de frequências e os modelos propostos são os modelos não paramétricos[9]

Segundo FERRAZ (1999)[6], a possível correlação introduzida pela amostragem de observações em instantes de tempo adjacentes pode severamente restringir a aplicabilidade de alguns métodos estatísticos convencionais, os quais dependem da suposição que observações adjacentes são independentes e identicamente distribuídas. A abordagem sistemática pela qual responderemos à questão de termos dados correlacionados no tempo é chamada comumente como análise de séries temporais.

As séries temporais podem ser divididas entre discretas, contínuas, univariadas e multivariadas. Nas séries temporais discretas, as observações são realizadas em tempos específicos, enquanto que as séries temporais contínuas são feitas a partir de observações realizadas continuamente no tempo. Uma série temporal é dita univariada quando são realizadas observações únicas ao longo do tempo, ou seja, uma única sequência de dados; já as séries multivariadas, são várias sequências de dados para o mesmo período de tempo.

Conforme EHLERS (2005)[4], muitas das propriedades observadas em uma série temporal  $Z_t$  podem ser captadas assumindo-se a seguinte forma de decomposição:

$$Z_t = T_t + C_t + R_t \quad (2.1)$$

onde  $T_t$  é uma componente de tendência,  $C_t$  é uma componente cíclica ou sazonal e  $R_t$  é uma componente aleatória ou ruído (a parte não explicada, que espera-se ser puramente aleatória).

Com o uso das séries temporais, objetiva-se a modelagem dos dados bem como definir tendências e periodicidades nas séries analisadas, além de realizar previsão, sendo este um elemento determinante na tomada de decisão do pesquisador[11].

## 2.3 TENDÊNCIA

Atualmente, existem vários métodos estatísticos para a determinação do componente tendência em uma série temporal.

O estudo da tendência tem por objetivo atender a duas finalidades básicas: uma, é identificar a tendência e usá-la, por exemplo, em previsões, e a outra, é remover a tendência, de modo a permitir o estudo de outros componentes da série[6].

A forma mais simples da tendência é[4]:

$$X_t = \alpha + \beta t + \epsilon_t \quad (2.2)$$

onde  $\alpha$  e  $\beta$  são constantes a serem estimadas e  $\epsilon_t$  denota um erro aleatório com média zero e variância constante. O nível médio da série no tempo  $t$  é dado por

$$m_t = \alpha + \beta_t \quad (2.3)$$

que é algumas vezes chamado de termo de tendência.

Segundo Morettin e Toloi[9], é possível o uso de testes estatísticos de hipóteses para verificar se existe tendência na série, como por exemplo o teste do sinal (Cox-Stuart). Vários outros testes podem ser adotados, mas no presente trabalho será utilizado apenas o teste do sinal de Cox-Stuart. Esse teste baseia-se em agrupar as observações em pares, sendo  $c = \frac{n}{2}$  caso  $n$  seja par, ou  $c = \frac{n+1}{2}$  caso  $n$  seja ímpar, sendo  $n$  o total de pares de observações em que foi atribuído o sinal de “+”. Caso a probabilidade de sinais positivos for idêntica à probabilidade de sinais negativos, conclui-se que não há existência de tendência na série. Caso isso não aconteça, então conclui-se que há tendência na série estudada.

Um procedimento para tornar uma série estacionária é fazermos as diferenças:

$$W_t = Z_t - Z_{t-1} = (1 - B)Z_t = \Delta Z_t \quad (2.4)$$

sendo  $B$  a defasagem de  $k$  períodos de tempo passado.

## 2.4 SAZONALIDADE

Uma série temporal é dita como sazonal quando os dados observados se repetem em determinado período  $t$  de tempo, podendo ser determinística (que apresenta padrão regular e estável) ou estocástica (que varia ao longo do tempo).

Utiliza-se testes estatísticos com o objetivo de identificar a presença de sazonalidade determinística, podendo ser paramétricos ou não paramétricos. Neste estudo, utilizou-se o teste não paramétrico de Kruskal-Wallis. Considera-se que a série apresenta sazonalidade se a hipótese nula do teste de ausência de sazonalidade for rejeitada, do acordo com a significância  $\alpha$  adotada.

## 2.5 RUÍDO BRANCO

Um ruído branco é um processo estocástico onde todas as variáveis aleatórias seguem uma distribuição normal de média zero, variância constante e as covariâncias são nulas. Além disso, um processo ruído branco é estacionário em média e variância. (Davila, 2020)[3]:

$$Z_t = a_t \quad (2.5)$$

$$E(Z_t) = 0 \quad (2.6)$$

$$\text{Var}(Z_t) = \sigma^2 \quad (2.7)$$

$$\text{Cov}(Z_t, Z_{t-k}) = 0 \quad (2.8)$$

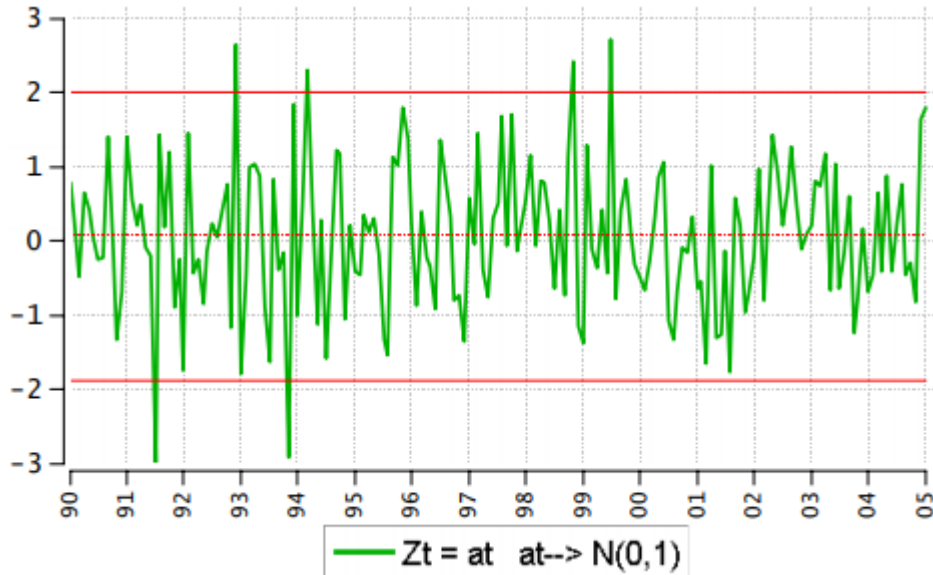


Figura 2.1: Exemplo de um processo ruído branco

## 2.6 FUNÇÃO AUTOCORRELAÇÃO E CORRELOGRAMA

Segundo Ehlers[4], uma importante ferramenta para se identificar as propriedades de uma série temporal consiste de uma série de quantidades chamadas coeficientes de autocorrelação amostral. A ideia equivale-se ao coeficiente de correlação usual.

A função de autocovariância para  $X_t$  é dada por:

$$Y_k = \text{Cov}(X_t, X_{t-k}) = E([X_t - E(X_t)][X_{t-k} - E(X_{t-k})]) \quad (2.9)$$

com  $E(X_t) = E(X_{t-k}) = 0$ , tem-se que

$$Y_k = E([X_t][X_{t-k}]) \quad (2.10)$$

A função de autocorrelação (fac) é dada por:

$$\rho_k = \frac{Y_k}{Y_0}, k = 0, 1, 2, \dots \quad (2.11)$$

com  $Y_0 = \text{Var}(X_t)$

Segundo FERRAZ[6], A função que associa cada valor de "k" com o seu respectivo coeficiente de autocorrelação é chamado de função de autocorrelação (fac). A representação gráfica da função de autocorrelação é chamada de correlograma, sendo obtida colocando-se em um gráfico



os valores das correlações em função dos "lags" $k$ , ou seja, a associação de cada valor de uma distância de tempo ao seu respectivo coeficiente de correlação. Através do correlograma, verifica-se diversas características da série, sendo utilizada para medir o número de defasagens e a intensidade da correlação existente entre as observações da série.

A função de autocorrelação também é importante para verificar se a série é estacionária. Se ao construir o correlograma, o coeficiente de autocorrelação não diminui rapidamente para zero à medida que "k" cresce, então está indicada a não-estacionaridade da série. Em caso contrário, caso se o coeficiente de correlação diminui rapidamente, estar-se-á em presença de uma série estacionária[6].

Com o auxílio da função de autocorrelação, também é possível analisar a presença de sazonalidade conforme o padrão dos lags, bem como definir a quantidade de parâmetros a serem definidos nos parâmetros auto regressivos do modelo AR(p) e de médias móveis MA(q) do modelo Arima.

Outra importante medida utilizada é a função de autocorrelação parcial (facp), onde são eliminadas as correlações implícitas entre a variável e suas respectivas defasagens, sendo possível estimar o parâmetro de defasagem. A Facp auxilia na determinação do parâmetro auto regressivo, ou seja, AR(p).

## 2.7 MODELO ARIMA

Segundo Morettin e Tolo[9], a estratégia para a construção do modelo será baseada em um ciclo iterativo, no qual a escolha da estrutura do modelo é baseada nos próprios dados. Os estágios do ciclo iterativo são:

- a) Uma classe geral de modelos é considerada para a análise (especificação);
- b) Há identificação de um modelo, com base na análise de autocorrelações, autocorrelações parciais e outros critérios;
- c) A seguir vem a fase de estimação, na qual os parâmetros do modelo identificado são estimados;
- d) Finalmente, há a verificação ou diagnóstico do modelo ajustado, através de uma análise de resíduos, para se saber se este é adequado para os fins em vista.

Caso o modelo não seja adequado, o ciclo deverá ser repetido, voltando-se à fase de identificação.

Nos modelos auto regressivos, temos que  $Y_t$  no instante  $t$  é função dos  $Y$ 's nos instantes anteriores a  $t$ . O modelo AR(p) pode ser escrito por:

$$\Phi(B)\tilde{z} = a_t \quad (2.12)$$

onde  $\Phi(B)$  é o operador auto regressivo:

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (2.13)$$

sendo  $B$  o operador translação para o passado:

$$BZ_t = Z_{t-1} \rightarrow B^m Z_t = Z_{t-m} \quad (2.14)$$

Logo, AR(p):

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \mu_t \quad (2.15)$$

Nos modelos de médias móveis,  $Y_t$  é uma função soma algébrica ponderada de  $\mu_t$  que se move no tempo. O modelo MA(q) pode ser escrito como:

$$\tilde{z} = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t = \Theta(B) a_t \quad (2.16)$$

onde  $\Theta(B)$  é o operador de médias móveis, sendo:

$$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (2.17)$$

Logo, MA(q):

$$Y_t = \mu_t - \theta_1 \mu_{t-1} - \theta_2 \mu_{t-2} - \dots - \theta_q \mu_{t-q} \quad (2.18)$$

Para muitas séries encontradas na prática, se quisermos um modelo com um número não muito grande de parâmetros, a inclusão de termos auto regressivos e de médias móveis é a solução adequada[9]. Dessa forma, surgem os modelos ARMA(p,q), de forma:

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \dots + \phi_p \tilde{Z}_{t-p} + a_t + \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (2.19)$$

Entretanto, os modelos AR(p), MA(q) e ARMA(p,q) têm o pressuposto de estacionaridade, que nem sempre acontecem com as séries históricas.

Box e Jenkins propuseram um método iterativo para a identificação do modelo de uma série temporal, o modelo ARIMA – Modelos auto regressivos integrados de médias móveis. Este método envolve investigações sobre os dados da série, sem a necessidade de se ter informações prévias sobre ela. Os termos auto regressivos se referem às defasagens da série diferenciada, e as médias móveis, às defasagens dos erros aleatórios. O termo integrado faz referência ao processo de diferenciação em que a série original é levada para torná-la estacionária. O modelo ARIMA é representado pela notação ARIMA(p,d,q), onde (p,d,q) representa a ordem do modelo, sendo p a ordem para o componente auto regressivo, d a ordem do componente de integração ou diferenciação, e q para a ordem de médias móveis.

Um procedimento para tornar uma série estacionária é tirar as diferenças, conforme a equação 2.4. Ou tirando d diferenças,

$$W_t = \Delta^d Z_t \quad (2.20)$$

Ou

$$\Phi(B) \Delta^d Z_t = \theta(B) a_t \quad (2.21)$$

Sendo este o modelo ARIMA(p,d,q).

Em geral, o modelo ARIMA é parcimonioso, sendo necessário  $d = 0, 1$  ou  $2$  para se obter a estacionaridade e identificação dos modelos.

## 2.8 SUAVIZAÇÃO EXPONENCIAL DE HOLT (SEH)

O método de Suavização exponencial de Holt (SEH) é aplicado em séries que apresentam tendência linear, sendo similar ao método da suavização linear simples, com a diferença de que além de suavizar o nível, também é incluído uma nova constante de suavização afim de modelar a tendência da série analisada.

Os valores do nível e tendência da série no instante  $t$  serão estimadas por:

$$\bar{Z}_t = AZ_t + (1 - A)(\bar{Z}_{t-1} + \hat{T}_{t-1}), 0 < A < 1, t = 2, \dots, N \quad (2.22)$$

$$\hat{T}_t = C(\bar{Z}_t - \bar{Z}_{t-1}) + (1 - C)\hat{T}_{t-1}, 0 < C < 1, t = 2, \dots, N \quad (2.23)$$

Sendo  $A$  e  $C$  constantes de suavização.

## 2.9 SELEÇÃO DOS MODELOS

Como critério de seleção do melhor modelo, utilizou-se neste estudo o Critério de Akaike (AIC), Erro percentual médio absoluto (MAPE) e Raiz do erro quadrático médio (RMSE). Na bibliografia, existem diversos outros critérios para seleção de modelos. Como regra, seleciona-se o modelo que apresentar os menores valores para as métricas dos critérios escolhidos.

### 2.9.1 CRITÉRIO DE AKAIKE (AIC)

O critério de informação de Akaike é dado por:  $AIC = 2\log(L) + 2m$ , onde  $m$  é o número de parâmetros do modelo e  $L$  a verossimilhança maximizada dos dados.

Vale destacar que o critério de Akaike pode assumir qualquer valor, sendo positivos ou negativos, sendo que deverá ser escolhido aquele modelo que apresentar o menor valor de AIC.

### 2.9.2 ERRO PERCENTUAL MÉDIO ABSOLUTO (MAPE)

O erro médio absoluto percentual (MAPE) expressa precisão como uma porcentagem do erro. Como o MAPE é uma porcentagem, pode ser mais fácil de entender do que as outras estatísticas de medida de precisão.[8]. O seu cálculo é dado por:

$$\frac{\sum \left| \frac{y_i - \hat{y}_i}{y_i} \right|}{n} \quad (2.24)$$

onde  $y_i$  representa os valores reais,  $\hat{y}_i$  representa os valores previstos, e  $n$  é total de amostras. O modelo a ser escolhido deverá apresentar o menor valor de MAPE.

### 2.9.3 RAIZ DO ERRO QUADRÁTICO MÉDIO

A raiz quadrática do erro médio (RMSE) apresenta os valores do erro nas mesmas dimensões da variável estudada, sendo comumente utilizada para avaliar a precisão do modelo. Seu cálculo é dado por:

$$\sqrt{\frac{\sum(O_i - P_i)^2}{n}} \quad (2.25)$$

onde  $O_i$  representa os valores observados, e  $P_i$ , os valores previstos pelo modelo. Da mesma forma que os demais critérios de seleção do modelo, escolhe-se o modelo com o menor valor para RMSE.

## 2.10 ANÁLISE DOS RESÍDUOS

Para a análise dos resíduos, deve-se observar a suposição de normalidade, sendo possível observá-la por meio do histograma e teste de normalidade. Para este estudo, utilizou-se o teste de Lilliefors, que é uma adaptação do teste de Kolmogorof-Smirnoff.

Além disso, os resíduos devem ser i.i.d (independente e identicamente distribuído), ou seja, devem ser ruído branco. Um teste utilizado para verificar se o resíduo é ruído branco é o de Box-Pierce.

Box e Pierce (1970) sugeriram um teste para as autocorrelações dos resíduos estimados, que, apesar de não detectar quebras específicas no comportamento de ruído branco, pode indicar se esses valores são muito altos. Uma variação desse teste foi sugerido por Ljung e Box (1978)[9].

Se o modelo for apropriado, a estatística:

$$Q(K) = n(n + 2) \sum_{j=1}^K \frac{\hat{r}_j^2}{(n - j)} \quad (2.26)$$

terá aproximadamente uma distribuição qui-quadrado com  $K - p - q$  graus de liberdade.

Ainda segundo Morettin e Tolo[9], para verificar se uma série é ruído branco, constrói-se o gráfico da função de autocorrelação com o seu respectivo intervalo de confiança. Caso as correlações residuais estejam dentro do intervalo, afirma-se que o processo é ruído branco.

Também é utilizado para análise dos resíduos, o teste do espectograma. Este estimador foi um dos primeiros a serem propostos nos estágios iniciais da Análise Espectral de séries temporais, com a finalidade de detectar periodicidades nos dados.

### 3. RESULTADOS

O gráfico da série histórica da produção de café entre os anos 1960/1961 e 2021/2022 é apresentado na Figura 3.1a). Visualmente, observa-se que os dados são não estacionários, uma vez que apresentam uma tendência crescente após uma queda ocorrida entre os anos 1960 e 1970, além de não apresentar sazonalidade e variação cíclica. De uma forma geral, após os anos 70, a produção de café no Brasil cresceu de maneira considerável.

Com a finalidade de averiguar a adequabilidade do modelo, foram retiradas as 5 últimas observações da série de dados para posterior comparação com os valores observados vs preditos pelo modelo referente a estes 5 últimos anos / safra (Figura 3.1b)). Posteriormente, realizou-se a mesma análise da série temporal com os dados completos, afim de realizar a previsão da produção de café para os próximos 5 anos desconhecidos, além de verificar se o modelo selecionado para os dois casos seria o mesmo.

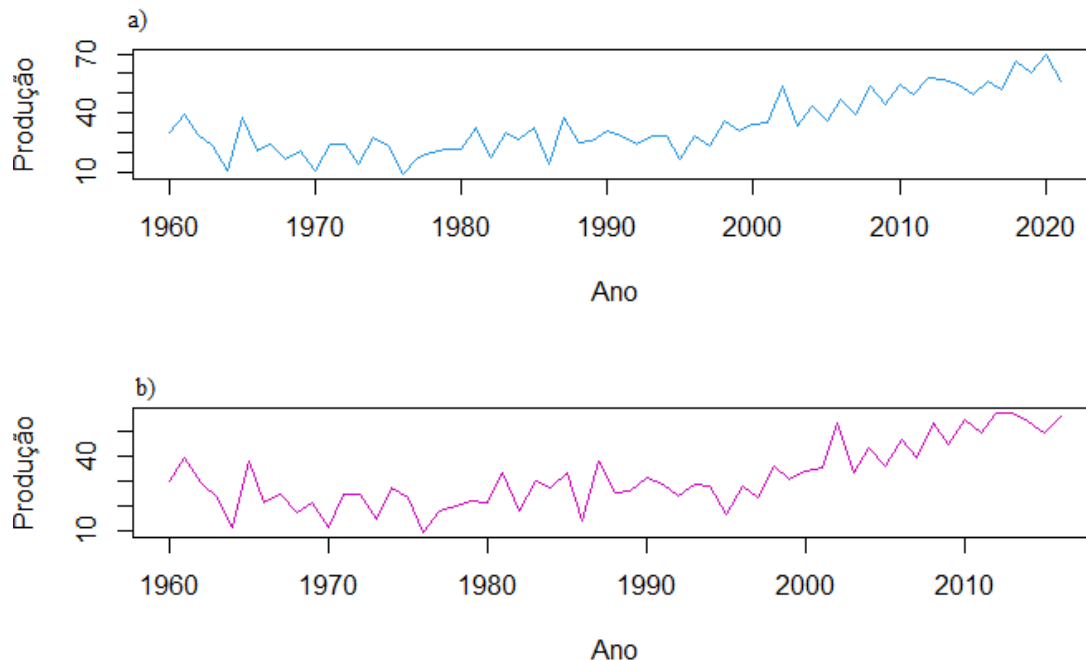


Figura 3.1: Série histórica da produção de café entre 1960/1961 e 2021/2022 (a) e entre 1960/1961 e 2016/2017 (b).

A função de autocorrelação (FAC) e autocorrelação parcial (FACp) da série até a safra de 2016/2017 está apresentada na Figura 3.2. Observa-se no correlograma que a série não decai

rapidamente para zero, caracterizando a não estacionaridade, além de ter decrescimento lento, acusando também indício de tendência.

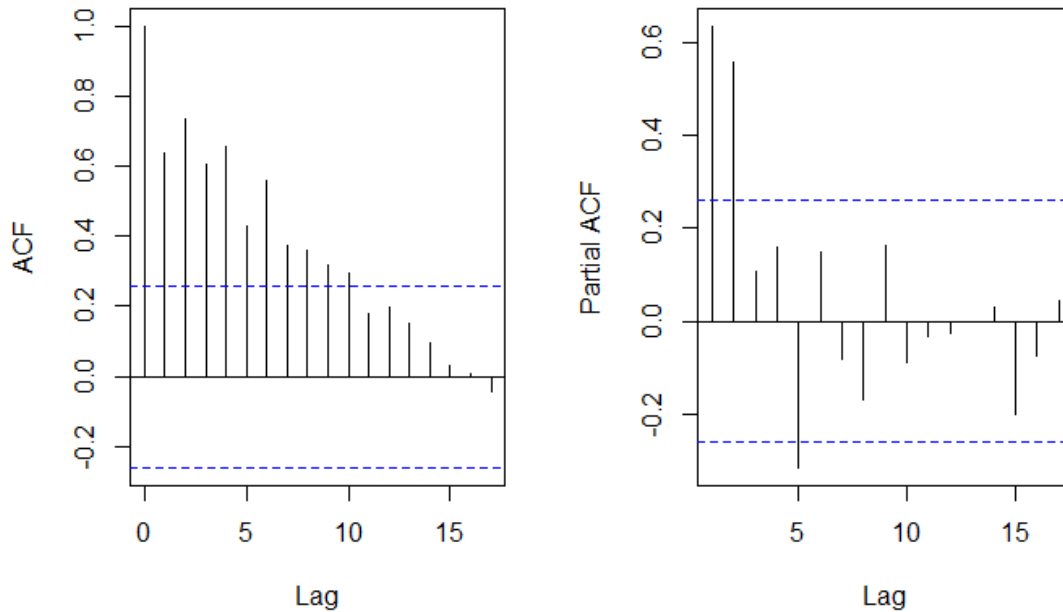


Figura 3.2: Função de Autocorrelação e Autocorrelação parcial da série histórica da produção de café no Brasil - safra até 2016/2017.

Segundo Morettin e Toloi[9], para que a tendência seja eliminada, é necessário realizar a diferenciação da série, até que a mesma seja extinta. Tendo em vista que a tendência caracteriza um processo não-estacionário, e os modelos ajustados devem seguir o pressuposto de estacionaridade, foi verificado com o auxílio da função *ndiffs* do pacote *forecast*, que apenas uma diferença seria necessária, ou seja,  $d(1)$ , para que a componente tendência fosse eliminada. Conforme a Figura 3.3, observa-se que a série diferenciada apresenta comportamento aleatório e média constante sendo então confirmado a exclusão da componente tendência.

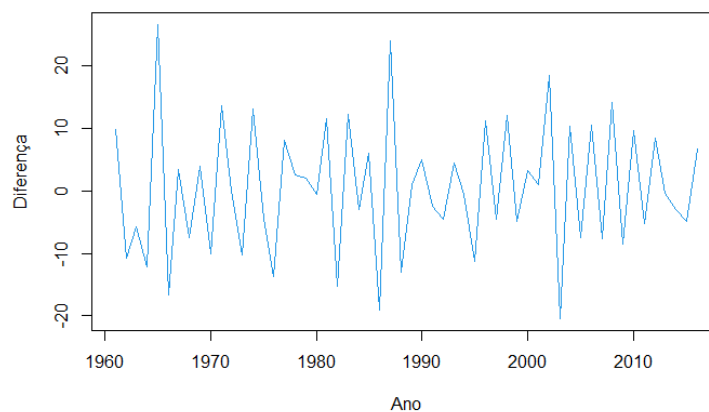


Figura 3.3: Série com uma diferença – dados parciais.

Além disso, foi realizado o teste de Cox-Stuart afim de verificar se a série histórica possuía

média constante, tanto para os dados originais, quanto para a série diferenciada, onde constatou-se presença de tendência na série original, e sua ausência na série com uma diferença, conforme resultados na Tabela 3.1. Vale ressaltar que a ausência de sazonalidade também foi verificada pelo teste de Kruskal-Wallis (p-valor = 0.48)

Tabela 3.1: Resultados teste Cox-Stuart

Série	Estatística	P-valor	Conclusão
Original	23	<0,01	Série com tendência
Com 1 diferença	16	0,57	Série sem tendência

Na Figura 3.5, observa-se o comportamento das funções FAC e FACp da série diferenciada. De acordo com Morettin e Tolo (2006), a função de auto correlação parcial indica a ordem da parte auto regressiva e a função de auto correlação, a ordem de médias móveis. Analisando graficamente, a FACp sugere um modelo AR(1) e a FAC sugere um modelo MA(6). Tendo em vista essas observações, foram averiguados 12 modelos, conforme consta na Tabela 3.2, sendo que foram considerados até 1 parâmetro auto regressivo e até 6 parâmetros de médias móveis na série diferenciada uma única vez.

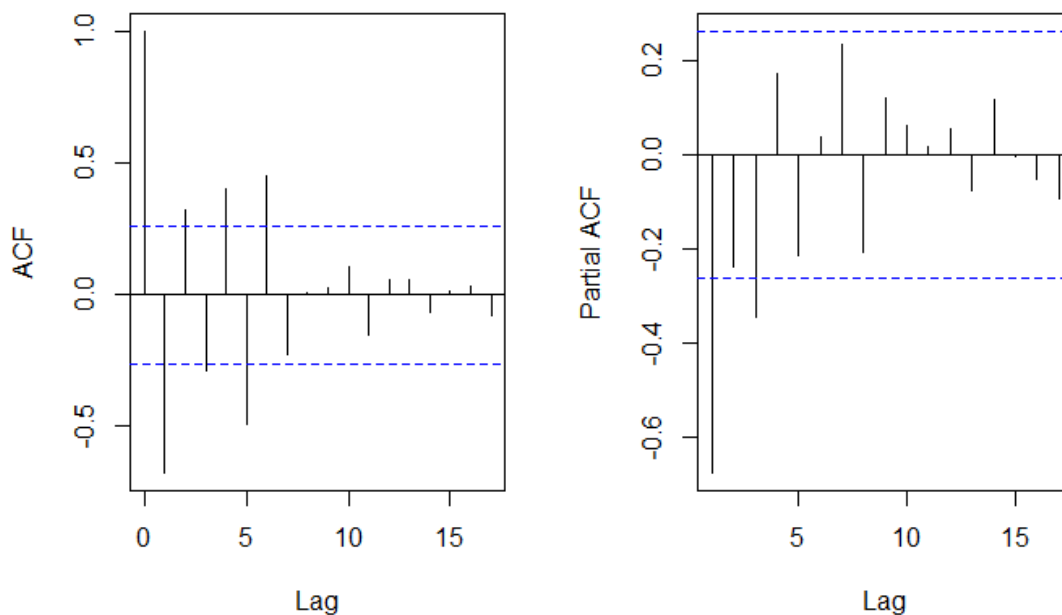


Figura 3.4: Função de autocorrelação e autocorrelação parcial da série diferenciada – dados parciais.

Na Tabela 3.2, tem-se o resumo dos modelos propostos. Foram considerados para posteriores análises os modelos que continham todos os parâmetros significativos, a partir do resultado do teste T-Student para significância dos parâmetros. A tabela contendo os respectivos p-valores dos parâmetros estão contidos no apêndice B. Tendo isso, foram selecionados apenas 3 modelos, sendo eles os modelos M9 - ARIMA (0,1,2), M11 - ARIMA (0,1,1) e M12 - ARIMA (1,1,1).

Para encontrar o melhor modelo ajustado, foram utilizados como critério de seleção o RMSE,

Tabela 3.2: Modelos Arima propostos – dados parciais

Modelo	Descrição	Todos os Parâmetros significativos?
M1	ARIMA (0,1,6)	Não
M2	ARIMA (1,1,6)	Não
M3	ARIMA (0,1,5)	Não
M4	ARIMA (1,1,5)	Não
M5	ARIMA (0,1,4)	Não
M6	ARIMA (1,0,4)	Não
M7	ARIMA (0,1,3)	Não
M8	ARIMA (1,1,3)	Não
<b>M9</b>	<b>ARIMA (0,1,2)</b>	<b>Sim</b>
M10	ARIMA (1,1,2)	Não
<b>M11</b>	<b>ARIMA (0,1,1)</b>	<b>Sim</b>
<b>M12</b>	<b>ARIMA (1,1,1)</b>	<b>Sim</b>

MAPE e AIC (Tabela 3.3). Além dos modelos de Box e Jenkins, também foi analisado o ajuste do modelo a partir da suavização exponencial de Holt (SEH).

Para o modelo SEH, obteve-se como resultados para as constantes de suavização, os valores  $A = 63,36$  e  $C = 0,91$ .

Tabela 3.3: Estimativas dos critérios de seleção do modelo – até a safra 2016/2017

Modelo	RMSE	MAPE	AIC
<b>M9 - ARIMA (0,1,2)</b>	<b>7,3</b>	<b>24,32</b>	<b>389,33</b>
M11 - ARIMA (0,1,1)	7,79	25,55	394,44
<b>M12 - ARIMA (1,1,1)</b>	<b>7,3</b>	<b>25,02</b>	<b>389,33</b>
SEH	7,79	25,46	-

Os melhores modelos selecionados foram os modelos ARIMA (0,1,2) e ARIMA (1,1,1). Observa-se que ambos tiveram valores para RMSE e AIC iguais e MAPE muito próximos, sendo então observado o comportamento dos erros de previsão, onde identificou-se que, em geral, o modelo ARIMA(1,1,1) apresentou menores erros em relação aos últimos 5 valores analisados, além de apresentar o menor valor de EQM (Tabela 3.4). O Erro quadrado médio (EQM) é calculado a partir da soma dos quadrados dos erros de previsão dividido pelo total de observações, ou seja, calcula-se a média do quadrado do erro, sendo útil para avaliar a distância entre um valor estimado e o valor real.

Outra observação a ser considerada é que o modelo ARIMA(0,1,2) apresentou o mesmo valor previsto para as safras entre 2018/2019 a 2021/2022, o que nos mostra que o parâmetro auto regressivo contido no modelo 12 apresentou melhor ajuste aos dados.



Tabela 3.4: Valores previstos pelos modelos e observados para as safras entre 2017/2018 a 2021/2022, para a série histórica da produção de café no Brasil – dados parciais

Modelo	Safra	Valor previsto	Valor real	Erro de previsão	Erro quadrado médio (EQM)
ARIMA(0,1,2)	2017/2018	53,2727	52,10	1,1727	100,30
	2018/2019	53,4322	66,50	13,0678	
	2019/2020	53,4322	60,50	7,0678	
	2020/2021	53,4322	69,90	16,4678	
	2021/2022	53,4322	56,30	2,8678	
ARIMA(1,1,1)	2017/2018	52,2822	52,10	0,1822	96,18
	2018/2019	54,0392	66,50	12,4608	
	2019/2020	53,3060	60,50	7,1940	
	2020/2021	53,6027	69,90	16,2973	
	2021/2022	53,4315	56,30	2,8685	

Na Tabela 3.5, são apresentados os parâmetros do modelo selecionado, bem como os respectivos erros padrões e p-valor, onde verifica-se que todos os parâmetros são significativos com  $\alpha = 0,05$ . Na Figura 3.5 tem-se a comparação dos valores preditos pelo modelo ARIMA (1,1,1) com a série de dados parcial; observa-se que o modelo foi capaz de acompanhar o comportamento da série.

Tabela 3.5: Parâmetros do modelo selecionado – dados parciais

Modelo	Parâmetro	Estimativa	Erro padrão	P-valor
ARIMA (1,1,1)	$\phi$	-0,4602	0,1539	0,0058
	$\theta$	-0,4244	0,1515	0,0094

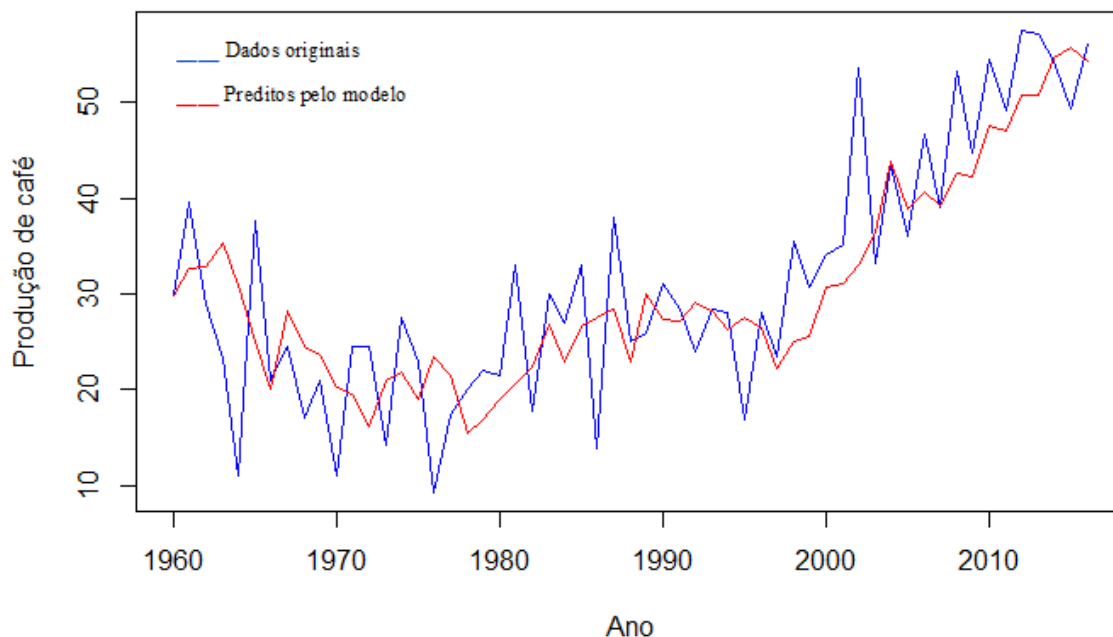


Figura 3.5: Série estimada e série histórica da produção de café no Brasil - dados parciais.

Posterior à seleção do modelo, prosseguiu-se com a análise dos resíduos afim de averiguar sua adequabilidade, sendo averiguado os pressupostos de normalidade, média zero e variância constante e resíduos não correlacionados. Na Figura 3.6a), tem-se o gráfico dos resíduos, onde observa-se comportamento aleatório, com média zero e variação constante ao longo do tempo. Na Figura 3.6b), tem-se o gráfico da autocorrelação dos resíduos, onde verifica-se que todos os lags, com exceção do lag 6, estão dentro dos limites, indicando que os resíduos se comportam como ruído branco. Na Figura 3.6c), tem-se o histograma dos resíduos, onde observa-se graficamente que os resíduos estão normalmente distribuídos.

Além disso, também foi realizado o teste de Box-Pierce, onde foi confirmado que os resíduos são ruído branco ( $p\text{-valor} = 0,86$ ), bem como também o teste de normalidade de Lilliefors ( $p\text{-valor} = 0,07$ ), confirmando a normalidade dos resíduos.

Na Figura 3.7, tem-se o periodograma acumulado do modelo ARIMA(1,1,1), onde todos os valores estão dentro do limite de confiança, também confirmando que os resíduos são processo ruído branco.

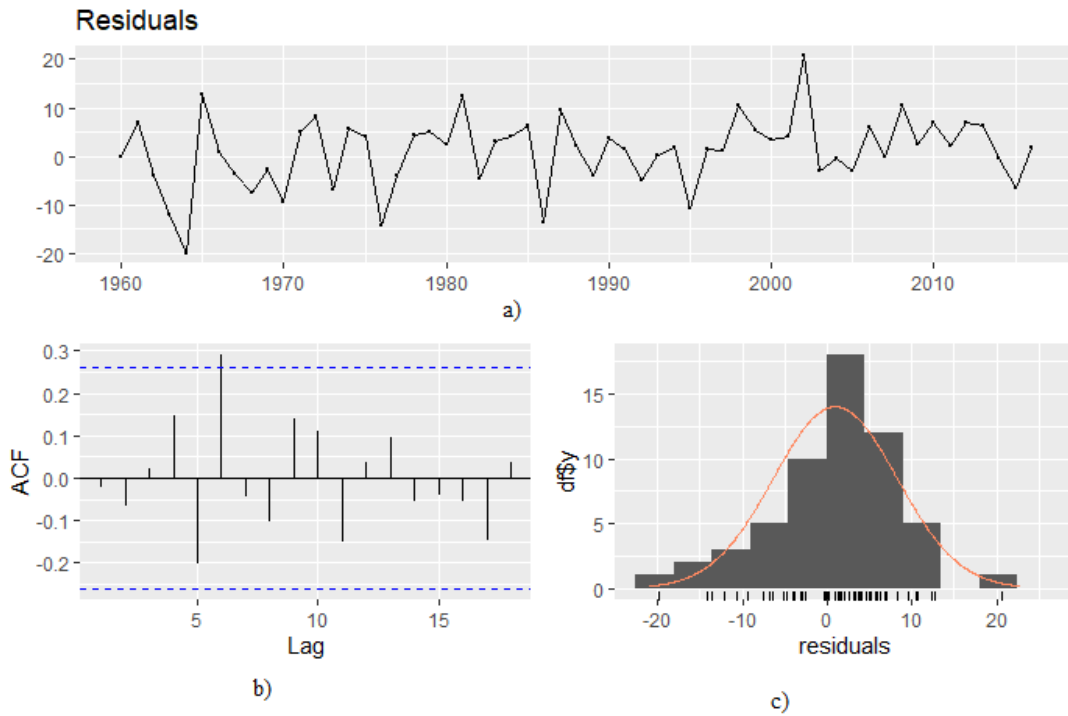


Figura 3.6: Análise dos resíduos do modelo Arima (1,1,1) – dados parciais.

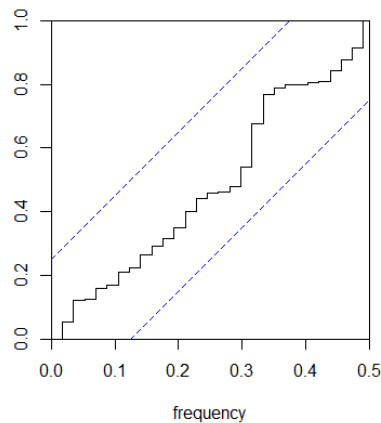


Figura 3.7: Periodograma acumulado – dados parciais.

Identificado o melhor modelo ajustado para os dados da safra até 2016/2017, prosseguiu-se com o ajuste do modelo com a série completa, ou seja, até a safra de 2021/2022 com o objetivo de prever os 5 valores desconhecidos para as safras compreendidas entre os anos 2022/2023 até 2026/2027.

Da mesma maneira, analisou-se o comportamento da FAC e FACp da série, tendo o mesmo comportamento da série inicial sem os 5 últimos valores (Figura 3.8).

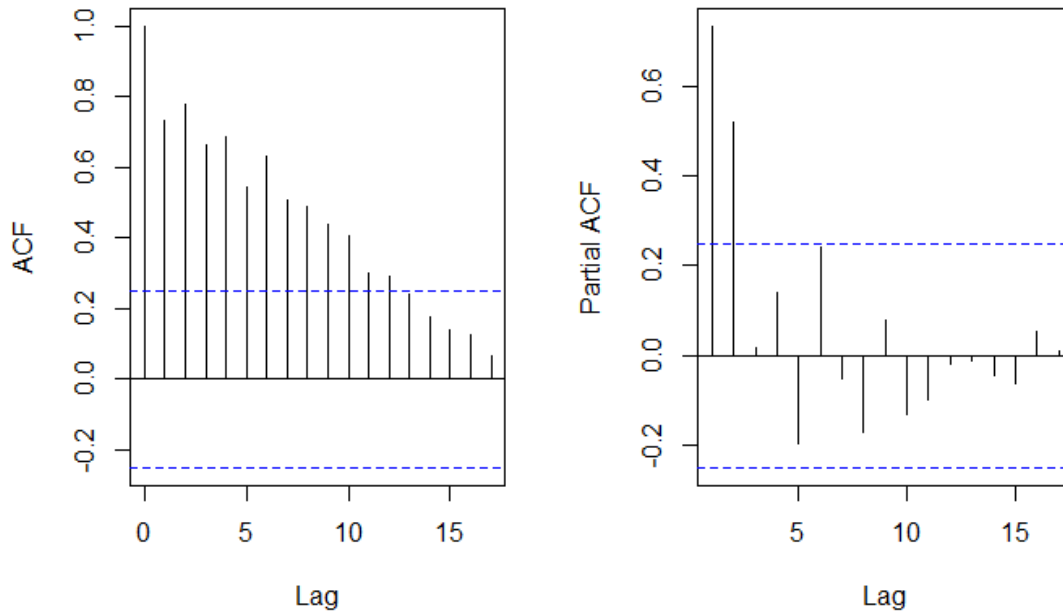


Figura 3.8: Função de Autocorrelação e Autocorrelação parcial da série completa.

Foi realizado uma única diferenciação necessária na série, retirando-se a tendência, e confirmando-a pelo teste de Cox-Stuart ( $p$ -valor = 0,36). Na Figura 3.9, observa-se o comportamento dos lags na FAC e FACp, onde igualmente à série de dados até a safra 2016/2017, foram propostos modelos com parte auto regressiva de ordem até 1 AR(1) e de médias móveis até ordem 6 MA(6).

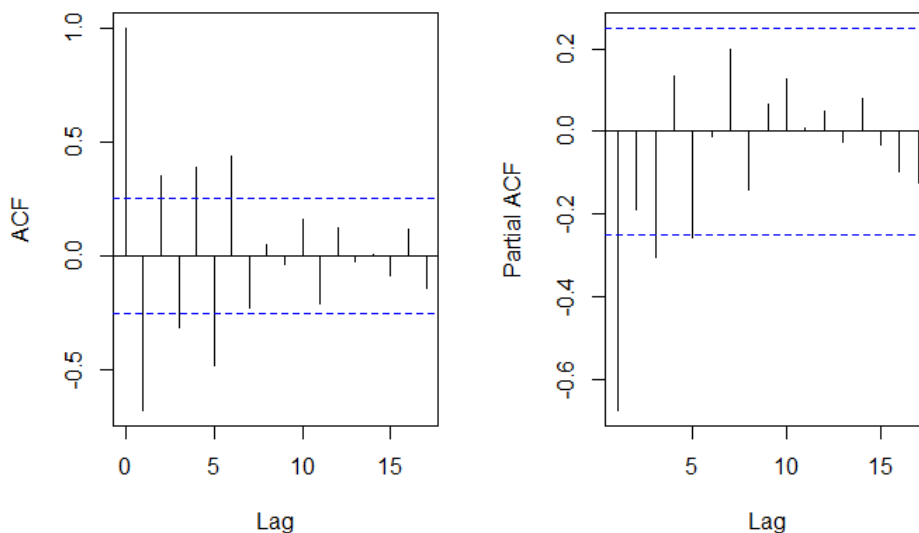


Figura 3.9: Função de autocorrelação e autocorrelação parcial da série completa diferenciada.

Dentre os modelos selecionados, apenas os modelos M9 - ARIMA (0,1,2), M11 - ARIMA (0,1,1) e M12 - ARIMA (1,1,1) apresentaram significância em todos os parâmetros, sendo o mesmo ocorrido com a série até 2016/2017. Na Tabela 3.6 tem-se o resumo dos modelos

selecionados.

Tabela 3.6: Estimativas dos critérios de seleção do modelo – dados completos

Modelo	RMSE	MAPE	AIC
<b>M9 - ARIMA (0,1,2)</b>	<b>7,4</b>	<b>23,25</b>	<b>425,11</b>
M11 - ARIMA (0,1,1)	7,85	24,38	430,01
<b>M12 - ARIMA (1,1,1)</b>	<b>7,34</b>	<b>24,02</b>	<b>424,01</b>
SEH	7,85	24,3	-

Para o modelo SEH, obteve-se como resultados para as constantes de suavização, os valores  $A = 54,11$  e  $C = -1,97$ .

Observa-se que o modelo M12 – Arima (1,1,1) foi o melhor modelo selecionado, sendo o mesmo ocorrido com a série parcial. O valor de RMSE e AIC foram os menores entre os três modelos, sendo então adotado para realizar as previsões para os próximos 5 valores desconhecidos para a produção de café no Brasil.

Na Tabela 3.7, têm-se os parâmetros do modelo selecionado, bem como os respectivos erros padrões e p-valor, sendo que todos os parâmetros são significativos com  $\alpha = 0,05$ . Na Figura 3.10, também é apresentado o gráfico com a série original e a série ajustada pelo modelo, onde verifica-se que os valores preditos acompanharam o comportamento da série original.

Tabela 3.7: Parâmetros do modelo selecionado – dados completos

Modelo	Parâmetro	Estimativa	Erro padrão	P-valor
ARIMA (1,1,1)	$\phi$	-0,4899	0,1547	0,0035
	$\theta$	-0,3809	0,1633	0,0281

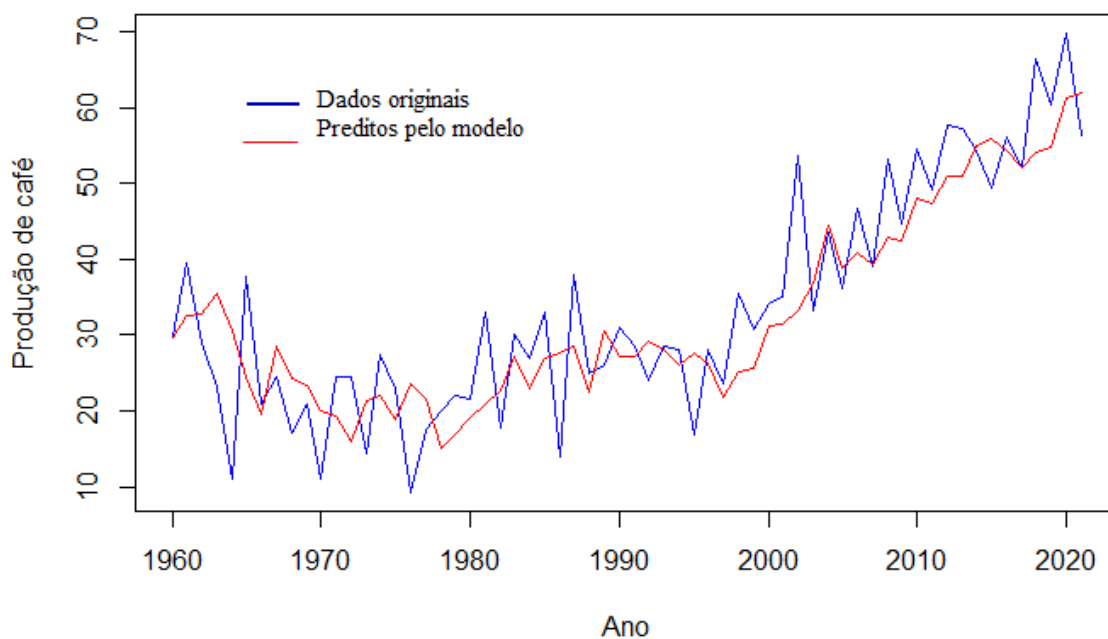


Figura 3.10: Série estimada e série histórica da produção de café no Brasil – dados completos.

Analisando os modelos ajustados de um modo geral, os valores previstos pelos modelos não acompanharam as safras que tiveram produções muito altas ou muito baixas. Todavia, estes foram os melhores modelos selecionados, dentro da técnica utilizada para este estudo.

Após a validação dos pressupostos do modelo, foi realizado a previsão para os próximos 5 anos da série histórica da produção de café, sendo os períodos compreendidos entre 2022/2023 até 2026/2027. Estas previsões, bem como seus respectivos intervalos de confiança estão apresentados na Figura 3.11. É importante ressaltar que quanto maior o tempo predito, maior será o intervalo de confiança.

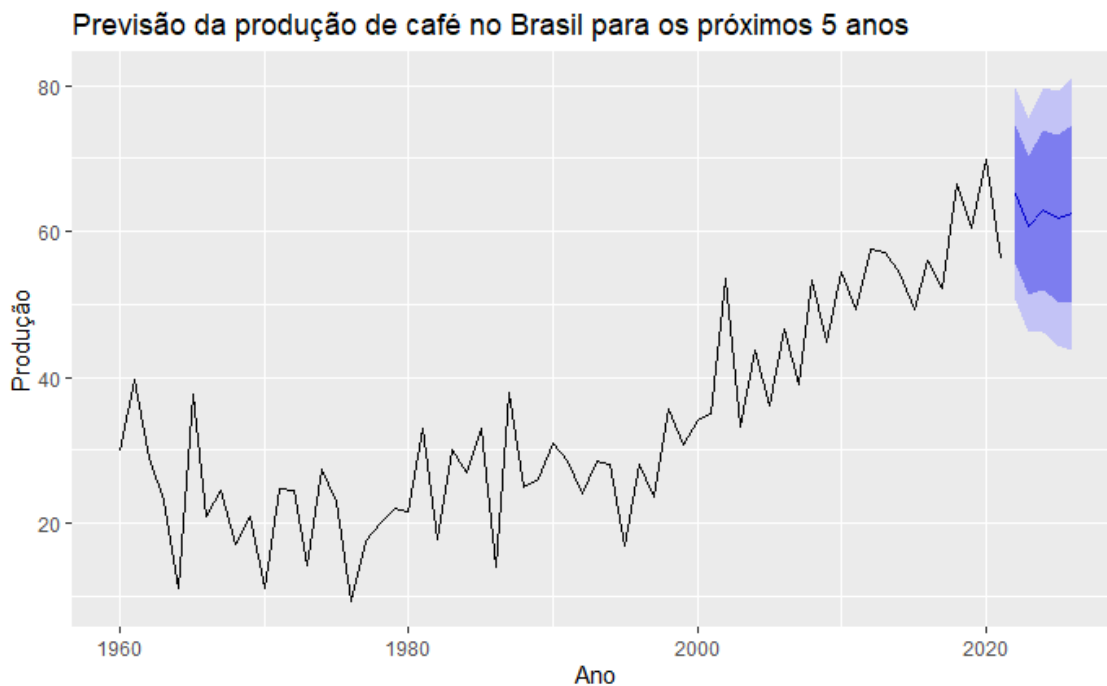


Figura 3.11: Representação gráfica da previsão da produção de café no Brasil entre 2022/2023 até 2026/2027.

Na Tabela 3.8 são apresentados os valores preditos e o intervalo de confiança (com 80 e 95% de confiança) para as previsões dos anos 2026/2027. Para a próxima safra, a previsão para a produção de café no Brasil é de 65,13 milhões de toneladas, com intervalo entre 50,63 e 79,63, com 95% de confiança.

Tabela 3.8: Previsão da Produção de café no Brasil entre 2022/2023 a 2026/2027

Safra	Previsão	LI 80%	LS 80%	LI 95%	LS 95%
2022/2023	65,13	55,65	74,61	50,63	79,63
2023/2024	60,81	51,25	70,36	46,19	75,42
2024/2025	62,92	52,01	73,84	46,23	79,62
2025/2026	61,89	50,49	73,28	44,45	79,32
2026/2027	62,39	50,23	74,56	43,79	81,00

## 4. CONCLUSÕES

A partir do modelo ajustado, foi possível realizar previsões para a produção de café no Brasil, para os próximos 5 anos, compreendidos entre 2022 e 2026. A previsão para a safra de 2022/2023 foi de 65,13 milhões de toneladas, representando um aumento de cerca de 16% em relação à safra de 2021/2022. Isso nos mostra que a produção de café no Brasil tem previsão de crescimento ao longo dos próximos anos.

Para estudos futuros, sugere-se analisar possíveis interferências que podem ter ocorrido durante os anos estudados, tendo em vista que tais interferências podem influenciar diretamente na produção, bem como analisar a série temporal com uma variável regressora afim de inserir o fator bienalidade do setor cafeeiro, ou seja, incluir uma variável *dummy*. Outra forma de estudar possíveis melhores modelos, seria tentar ajustar mais parâmetros para a parte auto regressiva.

Conclui-se que com este trabalho, foi possível visualizar a importância e eficácia das séries temporais no dia a dia em diversas áreas.





## REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Abic: *Associação Brasileira da indústria de café*, 2020. <https://www.abic.com.br/o-cafe/historia/o-cafe-brasileiro-na-atualidade-2>, acessado em 20/12/2020.
- [2] Conab: *Companhia Nacional de Abastecimento*, 2020. <https://www.conab.gov.br/info-agro/safras/cafe>, acessado em 15/12/2020.
- [3] Davila, V. H. L.: *Introdução às Séries Temporais*. Notas de aula utilizados na disciplina Séries Temporais, 2020.
- [4] Ehlers, R. S.: *Análise de Séries Temporais*. Material didático e instrucional, 2007.
- [5] Embrapa: *Empresa Brasileira de Pesquisa Agropecuária – Ministério da Agricultura, Pecuária e Abastecimento*, 2020. <https://www.embrapa.br/busca-de-noticias/-/noticia/56084554/producao-dos-cafes-do-brasil-atinge-6162-milhoes-de-sacas-de-60kg-em-2020-volume-25-maior-que-2019>, acessado em 15/12/2020.
- [6] Ferraz, M. I. F.: *Uso de modelos de séries temporais na previsão da série de precipitações pluviais mensais no município de Lavras - MG*. Dissertação de Mestrado, 1999.
- [7] Hoffmann, R.: *Análise de regressão: uma introdução à econometria*. O Autor, 4<sup>a</sup> ed., 2016.
- [8] Minitab, M. S. ao: *Interpretar todas as estatísticas e gráficos para Média móvel*, 2022. <https://support.minitab.com/pt-br/minitab/18/help-and-how-to/modeling-statistics/time-series/how-to/moving-average/interpret-the-results/all-statistics-and-graphs/#mape>, acessado em 20/02/2022.
- [9] Morettin, P. A. e C. Tolo, C. M. de: *Análise de series temporais*. Egard Blucher, 2<sup>a</sup> ed., 2006.
- [10] Nogueira, F.: *Modelagem e Simulação - Modelos de Previsão*. Notas de aula, 2009.
- [11] Reis, M. M.: *Análise de Séries Temporais*. Notas de aula, 2020.
- [12] Team, R. C.: *R: A Language and Environment for Statistical Computing*, 2021. <https://www.R-project.org/>.



## A. SCRIPTS R

```

require(forecast)
require(tseries)
require(stats)
require(GeneCycle)
require(randtests)
require(nortest)
cafe=read.table("cafe.txt", h=T)–chamando os dados
summary(cafe) –resumo dos dados
head(cafe)-visualizando os dados
cafets = ts(cafe, start = 1960, frequency = 1) –gerando a série temporal ts.plot(cafets, xlab
= 'Ano', ylab = 'Produção', col = '4') – gráfico da série temporal
summary(cafets)
par(mfrow=c(1,2)) acf(cafets, main = "") -fac pacf(cafets, main = "") -facp
nsdiffs(cafets) –dados não são sazonais ndiffs(cafets, alpha=0.05, test=c("adf")) –necessário
apenas 1 diferença cafedif=diff(cafets) - realizando a diferença plot(cafedif, ylab = "Diferença",
xlab = 'Ano', col = '4')
cox.stuart.test(cafets) – teste de cox stuart para tendência cox.stuart.test(cafedif)–série di-
ferenciada não apresenta tendência
par(mfrow=c(1,2)) – fac e facp série diferenciada acf(cafedif, main = "") pacf(cafedif, main
= "")
AEH=HoltWinters(cafets, alpha = NULL, beta = TRUE, gamma = FALSE) –Modelo
Holt plot(AEH, lwd=2, col="blue", xlab="Ano",ylab=NA) summary(AEH) previsaoholt <-
ses(cafets, alpha = NULL, beta = TRUE, gamma = FALSE, h=5)
–previsão para os próximos h anos
summary(previsaoholt) accuracy(previsaoholt) -critérios de seleção dos modelos
—Estimando modelos Arima—
m1 = arima(cafedif, order = c(0, 0, 6)) summary(m1) t1<-dt((-0.7218)/ 0.1239,
length(cafedif)) t2<-dt((0.0972)/ 0.1349, length(cafedif)) t3<-dt((-0.0878)/ 0.1357, length(cafedif))
t4<-dt((0.3770)/ 0.1467, length(cafedif)) t5<-dt((-0.5128)/ 0.1608, length(cafedif)) t6<-
dt((0.4443)/ 0.1119, length(cafedif)) tab1= matrix = list(t1,t2, t3, t4, t5, t6) tab1
m2 = arima(cafedif, order = c(1, 0, 6)) summary(m2) t7<-dt((-0.5541)/ 0.1805,
length(cafedif)) t8<-dt((-0.1772)/ 0.1881, length(cafedif)) t9<-dt((-0.2126)/ 0.1485, length(cafedif))

```

```

t10<-dt((-0.1182)/ 0.1364, length(cafedif)) t11<-dt((0.2065)/ 0.1645, length(cafedif)) t12<-
dt((-0.2414)/ 0.1356, length(cafedif)) t13<-dt((0.4229)/ 0.1717, length(cafedif)) tab2= matrix
= list(t7,t8,t9,t10,t11,t12,t13) tab2
  m3 = arima(cafedif, order = c(0, 0, 5)) summary(m3) t14<-dt((-0.732)/ 0.2005,
length(cafedif)) t15<-dt((0.0505)/ 0.2892, length(cafedif)) t16<-dt((-0.0203)/ 0.1341, length(cafedif))
t17<-dt((0.3629)/ 0.2674, length(cafedif)) t18<-dt((-0.3021)/ 0.2528, length(cafedif)) tab3=
matrix = list(t14, t15, t16, t17, t18) tab3
  m4 = arima(cafedif, order = c(1, 0, 5)) summary(m4) t19<-dt((-0.7496)/ 0.1378,
length(cafedif)) t20<-dt((0.0446)/ 0.1744, length(cafedif)) t21<-dt((-0.4041)/ 0.1513, length(cafedif))
t22<-dt((-0.0562)/ 0.1261, length(cafedif)) t23<-dt((0.3716)/ 0.1417, length(cafedif)) t24<-
dt((-0.2195)/ 0.14, length(cafedif)) tab4= matrix(list(t19, t20, t21, t22, t23, t24)) tab4
  m5 = arima(cafedif, order = c(0, 0, 4)) summary(m5) t25<-dt((-0.9136)/ 0.1311,
length(cafedif)) t26<-dt((0.3407)/ 0.1915, length(cafedif)) t27<-dt((-0.1318)/ 0.1433, length(cafedif))
t28<-dt((0.0986)/ 0.1137, length(cafedif)) tab5= matrix(list(t25, t26, t27, t28)) tab5
  m6 = arima(cafedif, order = c(1, 0, 4)) summary(m6) t29<-dt((-0.825)/ 0.1021,
length(cafedif)) t30<-dt((0.126)/ 0.1473, length(cafedif)) t31<-dt((-0.5413)/ 0.1327, length(cafedif))
t32<-dt((-0.1536)/ 0.1408, length(cafedif)) t33<-dt((0.4003)/ 0.1453, length(cafedif)) tab6 =
matrix(list(t29, t30, t31, t32, t33)) tab6
  m7 = arima(cafedif, order = c(0, 0, 3)) summary(m7) t34<-dt((-0.9033)/ 0.1281,
length(cafedif)) t35<-dt((0.3547)/ 0.1589, length(cafedif)) t36<-dt((-0.0827)/ 0.1313, length(cafedif))
tab7 = matrix(list(t34, t35, t36)) tab
  m8 = arima(cafedif, order = c(1, 0, 3)) summary(m8) t37<-dt((-0.9427)/ 0.0726,
length(cafedif)) t38<-dt((0.1019)/ 0.1579, length(cafedif)) t39<-dt((-0.4556)/ 0.1285, length(cafedif))
t40<-dt((0.0904)/ 0.1483, length(cafedif)) tab8 = matrix(list(t37, t38, t39, t40)) tab8
  m9 = arima(cafets, order = c(0, 1, 2)) summary(m9) t41<-dt((-0.8713)/ 0.1234,
length(cafets)) t42<-dt((0.3084)/ 0.1119, length(cafedif)) tab9 = matrix(list(t41, t42)) tab9
  m10 = arima(cafets, order = c(1, 1, 2)) summary(m10) t43<-dt((-0.9316)/ 0.0766,
length(cafedif)) t44<-dt((0.1907)/ 0.1451, length(cafedif)) t45<-dt((-0.4138)/ 0.1208, length(cafedif))
tab10 = matrix(list(t43, t44, t45)) tab10
  m11 = arima(cafets, order = c(0, 1, 1)) summary(m11) tab11 = matrix(list(t46)) tab11
  m12 = arima(cafets, order = c(1, 1, 1)) summary(m12) accuracy(m12) t47<-dt((-0.4899)/
0.1547, length(cafedif)) t48<-dt((-0.3809)/ 0.1633, length(cafedif)) tab12 = matrix(list(t47,
t48)) tab12
  -Analisando o modelo selecionado
  r = residuals(m12) lillie.test(r) - teste de normalidade Box.test(r, type=c("Box-Pierce"))
  -teste para média 0 e var const cpgram(r, main = ") -periodograma checkresiduals(r, main =
'Análise dos resíduos')
  -Realizando a previsão
  prev = fitted(m12) seqplot.ts(cafets,fitted(m12), ylab = "Produção de café", xlab = "Ano",
colx = "blue", coly = "red") autoplot(forecast(m12, h = 5), main = "Previsão da produção de

```

café no Brasil para os próximos 5 anos", xlab = "Ano", ylab = "Produção")

–Análise sem os últimos 5 valores

novo = cafets[1:57] -chamando os novos valores novots = ts(novo, start = 1960, frequency = 1) -criando a série temporal head(novots) - verificando

par(mfrow=c(1,2)) acf(novots, main = ") pacf(novots, main = ") ts.plot(novots)

cox.stuart.test(novots) novotsdif = diff(novots) -diferenciação cox.stuart.test(novotsdif)

par(mfrow=c(1,2)) acf(novotsdif, main = ") pacf(novotsdif, main = ") ts.plot(novotsdif)

–alisamento exponencial

AEH=HoltWinters(novots, alpha = NULL, beta = TRUE, gamma = FALSE) plot(AEH, lwd=2, col="blue", xlab="Ano", ylab=NA) previsaoholt1 <- ses(novots, alpha = NULL, beta = TRUE, gamma = FALSE, h=5) - previsão para os próximos h anos summary(previsaoholt1)

—Estimando modelos Arima—

n1 = arima(novots, order = c(0, 1, 6)) summary(n1)

a1<-dt((-0.7182)/ 0.1329, length(novots)) a2<-dt((0.0792)/ 0.1444, length(novots)) a3<-dt((0.0007)/ 0.1364, length(novots)) a4<-dt((0.3756)/ 0.1440, length(novots)) a5<-dt((-0.5799)/ 0.1681, length(novots)) a6<-dt((0.44)/ 0.1306, length(novots)) taba = matrix(list(a1, a2, a3, a4, a5, a6)) taba

n2 = arima(novots, order = c(1, 1, 6)) summary(n2) b1<-dt((-0.4552)/ 0.3599, length(novots)) b2<-dt((-0.2699)/ 0.3683, length(novots)) b3<-dt((-0.1733)/ 0.2366, length(novots)) b4<-dt((-0.0359)/ 0.1604, length(novots)) b5<-dt((0.3026)/ 0.1966, length(novots)) b6<-dt((-0.3156)/ 0.232, length(novots)) b7<-dt((0.3086)/ 0.1795, length(novots)) tabb = matrix(list(b1, b2, b3, b4, b5, b6, b7)) tabb

n3 = arima(novots, order = c(0, 1, 5)) summary(n3) c1<-dt((-0.6287)/ 0.1691, length(novots)) c2<-dt((-0.0941)/ 0.2257, length(novots)) c3<-dt((0.0539)/ 0.1112, length(novots)) c4<-dt((0.4925)/ 0.219, length(novots)) c5<-dt((-0.4072)/ 0.1825, length(novots)) tabc = matrix(list(c1, c2, c3, c4, c5)) tabc

n4 = arima(novots, order = c(1, 1, 5)) summary(n4) d1<-dt((-0.7309)/ 0.1802, length(novots)) d2<-dt((0.0408)/ 0.2116, length(novots)) d3<-dt((-0.3954)/ 0.1716, length(novots)) d4<-dt((0.0130)/ 0.1364, length(novots)) d5<-dt((0.3945)/ 0.1414, length(novots)) d6<-dt((-0.2173)/ 0.1399, length(novots)) tabd = matrix(list(d1, d2, d3, d4, d5, d6)) tabd

n5 = arima(novots, order = c(0, 1, 4)) summary(n5) e1<-dt((-0.9027)/ 0.1354, length(novots)) e2<-dt((0.3548)/ 0.1864, length(novots)) e3<-dt((-0.0741)/ 0.1454, length(novots)) e4<-dt((0.074)/ 0.1108, length(novots)) tabe = matrix(list(e1, e2, e3, e4)) tabe

n6 = arima(novots, order = c(1, 1, 4)) summary(n6) f1<-dt((-0.8169)/ 0.1024, length(novots)) f2<-dt((0.1628)/ 0.1475, length(novots)) f3<-dt((-0.5174)/ 0.143, length(novots)) f4<-dt((-0.1045)/ 0.1361, length(novots)) f5<-dt((0.4306)/ 0.1214, length(novots)) tabf = matrix(list(f1, f2, f3, f4, f5)) tabf

n7 = arima(novots, order = c(0, 1, 3)) summary(n7) g1<-dt((-0.9064)/ 0.1317, length(novots)) g2<-dt((0.3647)/ 0.1631, length(novots)) g3<-dt((-0.0287)/ 0.1268, length(novots)) tabg = matrix(list(g1, g2, g3)) tabg

```
n8 = arima(novots, order = c(1, 1, 3)) summary(n8) h1<-dt((-0.9242)/ 0.0917,
length(novots)) h2<-dt((0.0584)/ 0.1729, length(novots)) h3<-dt((-0.4048)/ 0.1377, length(novots))
h4<-dt((0.1627)/ 0.1506, length(novots)) tabh = matrix(list(h1, h2, h3, h4)) tabh
n9 = arima(novots, order = c(0, 1, 2)) summary(n9) i1<-dt((-0.9023)/ 0.1281,
length(novots)) i2<-dt((0.3403)/ 0.1195, length(novots)) tabi = matrix(list(i1, i2)) tabi
n10 = arima(novots, order = c(1, 1, 2)) summary(n10) j1<-dt((-0.9088)/ 0.0924,
length(novots)) j2<-dt((0.1475)/ 0.1569, length(novots)) j3<-dt((-0.4140)/ 0.1263, length(novots))
tabj = matrix(list(j1, j2, j3)) tabj
n11 = arima(novots, order = c(0, 1, 1)) summary(n11) k1<-dt((-0.6657)/ 0.076,
length(novots)) tabk = matrix(list(k1)) tabk
n12 = arima(novots, order = c(1, 1, 1)) summary(n12) l1<-dt((-0.4602)/ 0.1539,
length(novots)) l2<-dt((-0.4244)/ 0.1515, length(novots)) tabl = matrix(list(l1, l2)) tabl
-Realizando a previsão
prev12 = fitted(n12) seqplot.ts(novots,fitted(n12), ylab = "Produção de café", xlab =
"Ano", colx = "blue", coly = "red") forecast(n12, h=5)
-Verificando os resíduos res12 = residuals(n12) lillie.test(res12) Box.test(res12, type=c("Box-
Pierce")) cpgram(res12, main = ") checkresiduals(res12, main = 'Análise dos resíduos')
```

## B. P-VALORES MODELOS PROPOSTOS

Modelo	Parâmetro	P-valor	Modelo	Parâmetro	P-valor
M1-ARIMA (0,1,6)	$\theta_1$	<0,01	M5-ARIMA (0,1,4)	$\theta_1$	<0,01
	$\theta_2$	0,34		$\theta_2$	0,07
	$\theta_3$	0,4		$\theta_3$	0,35
	$\theta_4$	0,02		$\theta_4$	0,32
	$\theta_5$	<0,01	M6-ARIMA (1,1,4)	$\theta_1$	<0,01
	$\theta_6$	<0,01		$\theta_1$	0,22
$\theta_1$	0,18	$\theta_2$		<0,01	
M2-ARIMA (1,1,6)	$\theta_1$	0,3	$\theta_3$	0,29	
	$\theta_2$	0,3	$\theta_4$	<0,01	
	$\theta_3$	0,39	M7-ARIMA (0,1,3)	$\theta_1$	<0,01
	$\theta_4$	0,12		$\theta_2$	0,03
	$\theta_5$	0,16		$\theta_3$	0,39
	$\theta_6$	0,09	M8-ARIMA (1,1,3)	$\theta_1$	<0,01
$\theta_1$	<0,01	$\theta_1$		0,37	
$\theta_2$	0,36	$\theta_2$		<0,01	
M3-ARIMA (0,1,5)	$\theta_3$	0,35	$\theta_3$	0,22	
	$\theta_4$	0,03	M9-ARIMA (0,1,2)	$\theta_1$	<0,01
	$\theta_5$	0,04		$\theta_2$	<0,01
	$\theta_1$	<0,01	M10-ARIMA (1,1,2)	$\theta_1$	<0,01
$\theta_1$	0,39	$\theta_1$		0,25	
$\theta_2$	0,03	$\theta_2$		<0,01	
M4-ARIMA (1,1,5)	$\theta_3$	0,4	M11-ARIMA (0,1,1)	$\theta_1$	<0,01
	$\theta_4$	<0,01	M12-ARIMA (1,1,1)	$\theta_1$	<0,01
	$\theta_5$	0,12		$\theta_1$	<0,01