

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE MATEMÁTICA
GRADUAÇÃO EM ESTATÍSTICA

Luiz Fernando Silva Resende

Modelagem via séries temporais da Produção de Milho no Brasil

Uberlândia – MG

Dezembro - 2016

Luiz Fernando Silva Resende

Modelagem via séries temporais da Produção de Milho no Brasil

Trabalho de Conclusão de Curso apresentado à Coordenação do Curso de Estatística, da Universidade Federal de Uberlândia, para a obtenção do grau de Bacharel em Estatística.

Orientador:

Dr. Quintiliano Siqueira Schroden Nomelini

Uberlândia – MG

Dezembro – 2016

Luiz Fernando Silva Resende

Modelagem via séries temporais da Produção de Milho no Brasil

Trabalho de Conclusão de Curso apresentado à Coordenação do Curso de Estatística, da Universidade Federal de Uberlândia, para a obtenção do grau de Bacharel em Estatística.

Aprovado em: ____ de ____ de ____.

Prof. Dr. Quintiliano Siqueira Schroden Nomelini

Orientador

Prof. Dr. Janser Moura Pereira

Convidado

Prof. Dr. José Waldemar da Silva

Convidado

Uberlândia - MG

Dezembro - 2016

Resumo

O milho é o cereal de maior volume de produção no mundo, com mais de 1 bilhão de toneladas na safra de 2015/2016. O Brasil, com uma produção de 67 milhões de toneladas, é hoje o terceiro maior produtor e o segundo maior exportador mundial de milho. Objetiva-se com esse trabalho analisar e fazer previsões, utilizando modelos de séries temporais da produção brasileira de milho para os próximos anos, até a safra de 2025/2026. Para a escolha do modelo que melhor se ajusta aos dados, os modelos serão testados por testes de hipóteses e comparados por 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 (1,1,2) para a logarítmica da série temporal original como a melhor, as previsões se mostraram realistas quanto aos valores da série.

Palavras-chaves: Produção, Milho, Séries Temporais, Previsão.

Abstract

Corn is the world's largest cereal crop with more than 1 billion tones in the 2015/2016 harvest. Brazil, with a production of 67 million tons, is now the third largest producer and the second largest exporter of maize in the world. The objective of this work is to analyze and make forecasts using time-series models of Brazilian maize production for the next years, up to the 2025/2026 crop. In order to choose the model that best fits the data, the models will be tested by hypothesis tests and compared by model adequacy criteria. The construction of the time series models was done by the Box & Jenkins theory. After the ARIMA (1,1,2) model was defined for the logarithm of the original time series as the best, the predictions were realistic about the values of the series.

Key-words: Production, Corn, Time Series, Forecasting.

Lista de Ilustrações

Figura 1 – Comparação entre Variações Cíclicas (a) e Variações Sazonais (b)	8
Figura 2 – Processo não estacionário (a) que passa a ser um processo estacionário (b) ao tomarmos sua diferença.....	10
Figura 3 – Gráficos de linha da série histórica de produção de milho no Brasil, diferença dessa série, logaritmo e diferença do logaritmo dos dados originais	16
Figura 4 – Função de Autocorrelação e Função de Autocorrelação Parcial, para Diferença da Produção de Milho no Brasil e Diferença do Logaritmo da Produção de Milho no Brasil	18
Figura 5 – Série Estimada e Série da Produção de Milho no Brasil das safras de 1960/1961 à 2015/2016	20
Figura 6 – Histograma dos Resíduos do modelo ARIMA(1,1,2)	21
Figura 7 – Gráfico de Linha dos Resíduos do modelo ARIMA(1,1,2)	21
Figura 8 – Teste do Espectrograma para o modelo ARIMA(1,1,2)	22
Figura 9 – Previsão da Produção Brasileira de Milho para as safras de 2016/2017 a 2025/2026	23

Lista de tabelas

Tabela 1. Testes do Sinal de Cox-Stuart para tendência.....	17
Tabela 2. Teste F para comparar a variância da primeira com a segunda metade das séries	17
Tabela 3. Teste T-student para significância dos Parâmetros nos modelos	19
Tabela 4. Critério de Seleção do Modelo	20
Tabela 5. Teste para normalidade dos resíduos do modelo ARIMA(1,1,2).....	21
Tabela 6. Teste para igualdade da variância da primeira metade e variância da segunda metade dos resíduos do modelo ARIMA(1,1,2)	22
Tabela 7. Previsão da Produção Brasileira de Milho de 2016/2017 a 2025/2026	23

Sumário

1. Introdução	1
2. Objetivos	4
2.1 Geral	4
2.2 Específicos	4
3. Materiais e Métodos	5
3.1 Processos Estocásticos	5
3.2 Séries Temporais	6
3.2.1 Modelo Clássico	7
3.2.2 Tendência	7
3.2.3 Sazonalidade	8
3.2.4 Ruído Branco	8
3.3 Autocorrelograma	9
3.4 Modelo ARIMA	10
3.5 Significância dos Parâmetros	12
3.6 Seleção do Modelo	12
3.6.1 Critério de Informação de AKAIKE	13
3.6.2 Erro Percentual Médio Absoluto	13
3.6.3 Erro Quadrático Médio de Previsão	13
3.7 Análise de Resíduos	14
3.7.1 Normalidade	14
3.7.2 Variação	14
3.7.3 Autocorrelação	15
3.7.4 Teste do Espectrograma	15
4. Resultados	16
5. Conclusão	24
6. Referências	25
7. Anexo I	28
8. Anexo II	29

1. Introdução

O agronegócio brasileiro é uma atividade próspera, segura e rentável. Com um clima diversificado, chuvas regulares, energia solar abundante e quase 13% de toda a água doce disponível no planeta, o Brasil tem 388 milhões de hectares de terras agricultáveis férteis e de alta produtividade. Esses fatores fazem do país um lugar de vocação natural para a agropecuária e todos os negócios relacionados às suas cadeias produtivas, fazendo assim do agronegócio a principal locomotiva da economia brasileira (BRASIL A, 2004). O Brasil situa-se, no contexto mundial atual, como celeiro mundial em termos de agronegócio. Segundo RODRIGUES (2006), o país possui 22% das terras agricultáveis do mundo, além de elevada tecnologia utilizada no campo, dados estes que fazem do agronegócio brasileiro um setor moderno, eficiente e competitivo no cenário internacional.

Na década de 60, a maioria dos brasileiros vivia no campo. A atividade agropecuária ainda não tinha atingido os níveis de excelência de hoje. O Brasil exportava café – que respondia por mais de 60% das vendas externas – cacau, açúcar e algodão. Soja, milho e carne in natura eram mínimos. A agricultura não deixou de evoluir, mas isso aconteceu mais lentamente. Em 30 anos, de uma produção quase rudimentar para produção agropecuária brasileira em agronegócio, com cotação em bolsas internacionais, como a de Chicago. Soja, milho e carnes são itens importantes da nossa pauta de exportações. A agricultura brasileira chegou ao século 21 como nosso principal pilar econômico. Enquanto a indústria e os serviços sofreram perda e queda significativas na produção, os produtores rurais e nossa agroindústria foram responsáveis por 49% de tudo que o Brasil exportou no primeiro semestre deste ano. Com venda de US\$ 45 bilhões e um superávit de US\$ 38,9 bilhões (MAGGI, 2016).

O faturamento da agropecuária alcançou R\$ 498,49 bilhões em 2015, praticamente estável em relação a 2014 (R\$ 494,32 bilhões), de acordo com levantamentos do Ministério da Agricultura, Pecuária e Abastecimento (Mapa). O faturamento do setor é medido pelo Valor Bruto da Produção (VBP), que mostra o desempenho das lavouras e da pecuária ao longo do ano e corresponde ao faturamento bruto. O cálculo do indicador envolve duas variáveis: a produção das culturas agrícolas e da pecuária, e os preços recebidos pelos produtores nas principais praças de produção e comercialização, considerando os 26 maiores produtos agropecuários produzidos no país. Dos R\$ 498,49 bilhões gerados pelo setor em 2015, 64,4% (R\$ 320,96 bilhões)

tiveram origem na agricultura, na comparação com 2014, o VBP das principais lavouras cresceu 0,29%. O resultado de 2015 deve-se principalmente ao bom desempenho da safra de grãos, 208,54 milhões de toneladas, segundo levantamentos da Companhia Nacional de Abastecimento (Conab), que compensou a queda dos preços agrícolas (BRASIL B, 2016).

Cultivado em diferentes sistemas produtivos, no Brasil o milho é plantado principalmente nas regiões Centro-Oeste, Sudeste e Sul. O grão é transformado em óleo, farinha, amido, margarina, xarope de glicose e flocos para cereais matinais. A primeira ideia é o cultivo do grão para atender ao consumo na mesa dos brasileiros, mas essa é a parte menor da produção. O principal destino da safra são as indústrias de rações para animais. O estudo das projeções de produção do cereal, realizado pela Assessoria de Gestão Estratégica do Mapa, indica que o Brasil está entre os países que terão aumento significativo das exportações de milho, ao lado da Argentina. O crescimento será obtido por meio de ganhos de produtividade. Enquanto a produção de milho está projetada para crescer 2,67% ao ano nos próximos anos, a área plantada deverá aumentar 0,73% (BRASIL C, 2016).

O milho é o cereal de maior volume de produção no mundo, com aproximadamente 960 milhões de toneladas. Estados Unidos, China, Brasil e Argentina são os maiores produtores, representando 70% da produção mundial. Com uma área agrícola de 60 milhões de hectares, ocupando 7% do total de terras, estimado em 851 milhões de hectares, aproximadamente 5,5 milhões de imóveis rurais e uma produção ao redor de 190 milhões de toneladas, o Brasil é um país de grande importância dentro do cenário agrícola mundial. Dentro deste cenário, o Brasil, com uma área cultivada com milho de 15,12 milhões de hectares e produção de 82 milhões de toneladas, é hoje um país estratégico, pois, é o terceiro maior produtor e o segundo maior exportador mundial de milho (PEIXOTO, 2014).

A produção nacional de milho, em 2015/16, está distribuída nos estados de Mato Grosso, 24,9%, Paraná, 21,4%, Mato Grosso do Sul, 10,6%, Goiás, 9,8%, Minas Gerais, 7,9%, Rio Grande do Sul, 7,7% e São Paulo, 5,1%. Estes 7 estados têm produção estimada em 66,7 milhões de toneladas, devem contribuir com 87,5% da produção nacional esperada em 2015/16. A previsão de produção de milho no Brasil na safra de 2015/16 está estimada em 76,2 milhões de toneladas. (CONAB, 2016). Desse total, cerca de 50,0 milhões correspondem ao milho de segunda safra. A área plantada de milho deve ter um acréscimo de 4,4% entre 2015/16 e 2025/26, passando de 15,7

milhões de hectares em 2015/16 para 16,4 milhões no final do período das projeções. Não haverá necessidade de novas áreas para expansão dessa atividade, pois as áreas de soja liberam a maior parte das áreas requeridas pelo milho. O aumento de área projetado de 4,4% está bem abaixo do crescimento ocorrido nos últimos 10 anos, que foi de 11,4%. Mas o milho teve nos últimos anos elevados ganho de produtividade resultando em menor necessidade adicional de áreas. O consumo interno de milho que em 2015/16 representa 76,6% da produção deve reduzir-se nos próximos anos para 74,5%. As exportações devem passar de 30,4 milhões de toneladas em 2016 para 46,3 milhões de toneladas em 2025/26. Segundo técnicos que trabalham com essa cultura a área deve aumentar mais do que está sendo projetado e talvez se aproximar mais do seu limite superior de crescimento que é de 22,0 milhões de hectares (BRASIL D, 2016).

Devido a importância do agronegócio na economia do país justifica-se a aplicação de técnicas estatísticas para o estudo das series históricas da produção brasileira do agronegócio. Assim, técnicas de séries temporais podem ser aplicadas ao banco de dados de produção de milho, soja, algodão, cana, suínos, frango, bovinos dentre outros. 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 (NOGUEIRA, 2009). Assim essa pesquisa busca realizar uma previsão para a produção de milho no Brasil para os próximos dez anos.

2. Objetivos

2.1 Geral

Esse estudo pretende obter a previsão para a produção de milho no Brasil para os próximos dez anos, safras de 2016/2017 até a safra de 2025/2026, podendo então saber o quanto podemos esperar do crescimento dessa área para a próxima década.

2.2 Específicos

Identificar características inerentes às séries em estudo, como tendência, sazonalidade e ruído branco por meio de testes específicos;

Observar a presença e os efeitos de possíveis intervenções;

Ajustar modelos de séries temporais que representem os dados adequadamente por meio do método Box e Jenkins;

Analisar os parâmetros por meio de testes de significância;

Escolher o melhor modelo por métodos específicos;

Compreender o mecanismo gerador da série para que através desses resultados possamos prever o comportamento futuro da série.

Implementar as análises com o auxílio do software R.

3. Materiais e Métodos

A série histórica da produção brasileira de milho da safra de 1960/1961 até a safra de 2015/2016 retirada do site PSD online (PRODUCTION, SUPPLY AND DISTRIBUTION, 2017), utiliza como fonte a *Foreign Agricultural Service, Official USDA Estimates* (Serviço Agrícola Estrangeiro, USDA Estimativas Oficiais). As análises foram realizadas pela teoria de Box & Jenkins (MORETTIN & TOLOI, 1981) utilizando o ambiente R: A Language and Environment for Statistical Computing (2017).

Para estudar séries temporais, é importante entender sobre processos estocásticos, visto que um processo estocástico é uma coleção de variáveis aleatórias que, em geral, são utilizadas para estudar a evolução de fenômenos que são observados ao longo do tempo. Assim, ao invés de descrevermos o sistema através de equações determinísticas, que dado uma condição inicial, conhecemos toda a evolução do sistema, utiliza-se processos estocásticos, as quais dada uma condição inicial, ainda se tem diversas trajetórias possíveis para a evolução do sistema (FARIA, 2015).

3.1 Processos Estocásticos

Žitković (2010) diz que, com T sendo um subconjunto de $[0, \infty)$, uma família de variáveis aleatórias $\{X_t\}_{t \in T}$, indexada por T , é chamada de processo estocástico (ou aleatório). Quando $T = \mathbb{N}$ (ou $T = \mathbb{N}_0$), $\{X_t\}_{t \in T}$ é dito ser um processo de tempo discreto, e quando $T = [0, \infty)$, é chamado de processo de tempo contínuo.

Ainda de acordo com Žitković (2010), quando T é único (diz-se $T = \{1\}$), o processo $\{X_t\}_{t \in T} = X_1$ é realmente apenas uma variável aleatória. Quando T é finito (ou seja, $T = \{1, 2, \dots, n\}$), tem-se um vetor aleatório. Portanto, processos estocásticos são generalizações de vetores aleatórios. A interpretação é, entretanto, diferente. Enquanto os componentes de um vetor aleatório geralmente (nem sempre) representam coordenadas espaciais, o índice $t \in T$ é frequentemente interpretado como tempo. Processos estocásticos usualmente modelam a evolução de um método aleatório no tempo. Quando $T = [0, \infty)$ (processo de tempo contínuo), o valor do processo pode mudar a todo momento. Quando $T = \mathbb{N}$ (processo de tempo discreto), as mudanças ocorrem discretamente.

3.2 Séries Temporais

Uma série temporal é qualquer conjunto de observações geradas sequencialmente no tempo, tendo como característica a dependência entre as observações. A variável é observada em pontos temporais discretos, usualmente equidistantes, e a análise de tal comportamento temporal envolve a descrição do processo ou fenômeno que gera a sequência, sendo assim uma descrição do passado. Um procedimento lógico para realizar previsões é fazer uso desses dados históricos (NOGUEIRA, 2009).

A suposição básica que norteia a análise de séries temporais é que há um sistema causal relativamente constante, relacionado com o tempo, que exerceu influência sobre os dados no passado e pode continuar a fazê-lo no futuro. Este sistema causal costuma atuar criando padrões não aleatórios que podem ser detectados em um gráfico da série temporal, ou mediante algum outro processo estatístico. O objetivo da análise de séries temporais é identificar padrões não aleatórios na série temporal de uma variável de interesse, e a observação deste comportamento passado pode permitir fazer previsões sobre o futuro, orientando a tomada de decisões (REIS, 2015).

De acordo com EHLERS (2009), uma série é dita ser contínua quando as observações são feitas continuamente no tempo. Definindo o conjunto $T = \{t: t_1 < t < t_2\}$, a série temporal será denotada por $\{X(t): t \in T\}$. Uma série temporal será discreta quando as observações são feitas em tempos específicos (diariamente, mensalmente, anualmente, etc). Definindo o conjunto $T = \{t_1, \dots, t_2\}$, a série será denotada por $\{X(t): t \in T\}$ ou, para facilitar $T = \{1, 2, \dots, n\}$.

Para EHLERS (2009), de um modo geral, os principais objetivos em se estudar séries temporais são descrição, explicação, predição e controle. Descrever propriedades da série como padrão de tendência, existência de variação sazonal ou cíclica, observações discrepantes (outliers) e alterações estruturais, usar a variação da série para explicar o comportamento, prever valores futuros com base em valores passados e por fim medir a “qualidade” de um processo através dos valores obtidos no estudo da série.

3.2.1 Modelo Clássico

Em geral no estudo de uma série temporal existe o interesse na análise e modelagem, que visa descrever, verificar suas características mais relevantes, suas possíveis relações com outras séries e a previsão a partir de valores históricos para estimar previsões. O número de instantes à frente para o qual é feita a previsão é chamado de horizonte de previsão, chamaremos de s . As propriedades de uma série temporal X_t podem ser captadas, conforme EHLERS (2009), assumindo-se a seguinte forma de decomposição:

$$X_t = T_t + C_t + R_t \text{ ou } X_t = T_t \cdot C_t \cdot R_t$$

Onde t é o tempo definido em $t=1,2, \dots, n$; T_t é uma componente de tendência; C_t é uma componente cíclica (ou sazonal) e R_t é uma componentes aleatória ou ruído. A questão crucial do modelo clássico é decidir como será a equação que relaciona as componentes com a variável, modelo aditivo ou multiplicativo. O modelo aditivo é a soma das componentes T_t, C_t e R_t e o modelo multiplicativo o produto.

3.2.2 Tendência

Segundo MORETTIN & TOLOI (1981), é possível o uso de testes estatísticos de hipóteses para verificar se existe tendência na série pelo o teste do sinal (Cox-Stuart). Considerando um conjunto de observações $X_1, X_2, \dots, X_n$ o teste baseia-se em agrupar as observações em pares $(X_1, X_{1+c}), \dots, (X_{n-c}, X_n)$, para $c = \frac{n}{2}$ se n for par e $c = \frac{(n+1)}{2}$ se n for ímpar, em que n é o número de observações e X_i a observação i da série. A cada par de observações se associa o sinal '+', se $X_i < X_{i+c}$ e o sinal '-', se $X_i > X_{i+c}$, eliminando os empates. Considerando n_t o numero de pares onde $X_i \neq X_{i+c}$.

A estatística do teste é dada por T_{Est} = numero de pares com sinal positivo. Para $T_{Est} > n - T_{Est}$ rejeita-se a hipótese H_0 . O valor de T_{Tab} é encontrado numa tabela de distribuição binomial, com parâmetros $p = 0,5$ e n para um dado nível α , se $n \leq 20$ e para $n > 20$, pode-se usar distribuição normal em que o n é o número de pares.

3.2.3 Sazonalidade

Os ciclos são caracterizados pelas oscilações de subida e de queda nas séries, de forma suave e repetida ao longo da componente de tendência, flutuações nos valores da variável e que se repetem com certa periodicidade. A sazonalidade em uma série corresponde às oscilações de subida e de queda que sempre ocorrem em um determinado período do ano, do mês, da semana ou do dia, a sazonalidade pode ser verificada pela análise espectral do correlograma. A diferença essencial entre as componentes sazonal e cíclica é que a primeira possui movimentos facilmente previsíveis, ocorrendo em intervalos regulares de tempo, enquanto que movimentos cíclicos tendem a ser irregulares (GUTIERREZ, 2003), tais diferenças são mostradas na Figura 1.

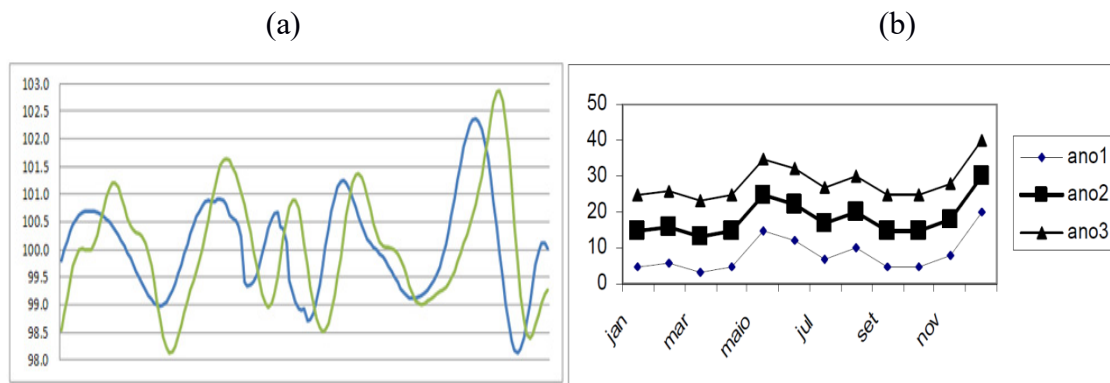


Figura 1 – Comparação entre Variações Cíclicas (a) e Variações Sazonais (b)

3.2.4 Ruído Branco

As variações irregulares são as flutuações inexplicáveis, resultado de fatos que ocorre sem que haja previsão ou planejamento, inesperados. A decomposição da série permitirá identificar quais componentes estão atuando na série em estudo, além de possibilitar obter índices e equações para realizar previsões para períodos futuros da série. Se o modelo for adequado, espera-se que os resíduos não apresentem nenhum padrão bem definido, para que possam ser classificados como ruído branco. De acordo com MORETTIN & TOLOI (1981) uma sequência R_t é dita ruído branco se cada valor da série dos erros de estimação tiver distribuição normal, média zero, variância constante e não-correlacionados.

3.3 Autocorrelograma

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 é similar ao coeficiente de correlação usual, para n pares de observações das variáveis x e y . A função autocorrelação é uma função que associa a cada valor uma distância entre tempo, e sua representação gráfica é chamada de autocorrelograma ou simplesmente correlograma. Espera-se um decréscimo à medida que aumentamos a distância entre as observações e, no caso das séries temporais, à medida que o tempo passa. Podemos dizer que valores vizinhos guardam mais semelhanças entre si do que valores muito distantes. A função autocorrelação é, portanto, uma medida de interdependência entre as observações. A função de autocorrelação será dada por:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\sum_{t=1}^{n-k} (y_t - \bar{Y})(y_{t+k} - \bar{Y})}{\sum_{t=1}^n (y_t - \bar{Y})^2}$$

Segundo FRANCO (2009) a ideia de autocorrelação pode ser estendida. Se medirmos a correlação entre duas observações seriais, y_t e y_{t+k} , eliminando a dependência dos termos intermediários $y_{t+1}, y_{t+2}, \dots, y_{t+k-1}$, temos o que se denomina autocorrelação parcial, representada por:

$$\begin{bmatrix} 1 & \rho_1 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \dots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k_1} \\ \phi_{k_2} \\ \vdots \\ \phi_{k_k} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix}$$

A autocorelação parcial de um processo é definida como a sequência dos ϕ_{k_k} 's obtidos pela resolução das equações acima. Por Cramer temos:

$$\phi_{k_1} = \rho_1; \phi_{k_2} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}; \phi_{k_3} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}, \dots; \phi_{k_k} = \frac{|P_k^*|}{|P_k|}$$

Onde $|\cdot|$ é o determinante da matriz, P_k é a matriz de autocorrelação e P_k^* é a matriz P_k com a última coluna substituída pelo vetor de autocorrelações. Para definir a quantidade de parâmetros de Média Móveis, faz-se a análise do autocorrelograma, sendo os possíveis valores de q as “lags” acima do intervalo de confiança, o mesmo vale para os parâmetros Auto-Regressivos, analisando a FACP.

3.4 Modelo ARIMA

Segundo GUTIERREZ (2003) os modelos ARIMA são modelos estatísticos lineares para análise de séries temporais. A abreviação em língua inglesa refere-se a “Auto-Regressive Integrated Moving Average model”, ou seja, um modelo auto-regressivo integrado de médias móveis. Os termos auto-regressivos correspondem a defasagens da série transformada (isto é, série estacionária obtida por diferenciação) e as médias móveis a defasagens dos erros aleatórios. O termo "integrado" refere-se ao processo de diferenciação da série original para torná-la estacionária.

De acordo com MORETTIN & TOLOI (1981), para a modelagem ARIMA acontecer a série a ser estudada deve ser estacionária, isto quer dizer, a série deve se distribuir aleatoriamente em torno de uma média constante, admitindo certa estabilidade ao longo do tempo. Ao construir o correlograma se o coeficiente diminui rapidamente para zero, temos uma série estacionária. Será necessário transformar os dados originais se estes não formam uma série estacionária como mostra a Figura 2. A transformação mais comum consiste em tomar diferenças sucessivas da série original, até se obter uma série estacionária.

$$\Delta Z(t) = Z(t) - Z(t - 1).$$

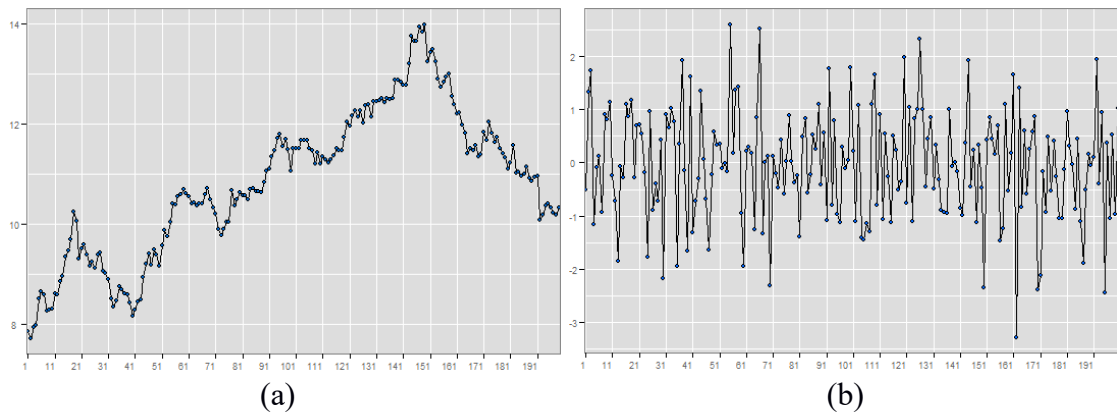


Figura 2 – Processo não estacionário (a) que passa a ser um processo estacionário (b) ao tomarmos sua diferença

Seja a estimativa da variável S para um tempo t depende de uma combinação linear de p termos da série observada, incluindo o termo aleatório $\varepsilon(t)$ de ruído branco. Os coeficientes φ_i são parâmetros que ponderam os valores de $S(t)$, do instante imediatamente anterior $t-1$ até o mais distante $t-p$, sendo determinados através de técnicas de minimização do erro.

De acordo com PINDYCK & RUBINFELD (2004) o modelo Auto-Regressivo AR(p) é definido por:

$$S(t) = \sum_{i=1}^p \varphi_i S(t-i) + \varepsilon(t)$$

O modelo AR(p) é limitado, pois assume a existência de uma relação linear entre os elementos da sequência e baseia-se na hipótese de que a série é estacionária, isto é, a média e o desvio padrão das observações medidas não variam com o tempo.

O modelo de médias móveis MA(q) assume que a série modelada é gerada através de uma combinação linear de q sinais e o termo aleatório $\varepsilon(t)$ de ruído branco. PINDYCK & RUBINFELD (2004, p. 586) afirmam que “no processo de médias móveis de ordem q , cada observação de Y_t é gerada por uma média ponderada de perturbações aleatórias que vão para trás q períodos”. MAKRIDAKIS (1998), de maneira análoga, apontam que a relação de dependência entre os erros e a equação é denominada MA(q). Nesse sentido, o termo MA pode ser definido pela média móvel dos erros ε . O processo MA (q) é representado:

$$S(t) = - \sum_{i=1}^q \theta_i \varepsilon(t-i) + \varepsilon(t)$$

Segundo MAKRIDAKIS et al (1998), os parâmetros $\theta_1, \theta_2, \dots, \theta_i$ são determinadores das ponderações da média móvel, os quais podem ser positivos ou negativos. A combinação dos modelos AR(p) e MA(q) dá então origem ao modelo ARMA(p,q). As limitações previamente mencionadas no modelo AR(p), concernentes à linearidade e estacionariedade do fenômeno modelado, são também aplicáveis aos modelos MA(q) e ARMA(p,q). A aplicação dos modelos de Box & Jenkins para fins de previsão de séries temporais segue as seguintes etapas: identificação, estimação e verificação, após realizar os três passos, se o modelo for satisfatório, faz-se a estimação.

Para estimação dos parâmetros este estudo utilizará o método de mínimos quadrados, em que o vetor autorregressivo e o vetor dos parâmetros médias móveis que minimizem a soma das diferenças quadradas entre os pontos observados na amostra e o esperado pela estimativa obtida com estes parâmetros estimados (SOUZA, 1991). Esta etapa foi executada através de “software” específico para análise de séries temporais. Existem diferentes métodos para a estimação dos parâmetros do modelo, e todos produzem resultados semelhantes, embora possam existir diferenças quanto à eficiência na implementação computacional (MORETTIN & TOLOI, 1981).

3.5 Significância dos Parâmetros

Para sabermos se um parâmetro possui influência suficiente em um modelo realizamos um teste de hipótese para cada parâmetro estimado. Para realizarmos esse teste é feita a suposição de que os erros são independentes e identicamente distribuídos. Utilizamos o teste t-student com $n-1-d$ graus de liberdade, seja $\hat{\varphi}_i$ um parâmetro estimado de Médias Móveis, a Estatística do teste será dada por:

$$T_{\hat{\varphi}_i} = \frac{\hat{\varphi}_i - 0}{\sqrt{\sigma_{\hat{\varphi}_i}^2}}$$

O mesmo vale para $\hat{\theta}_i$, os parâmetros estimados Auto-Regressivos. Logo, rejeitamos H_0 a um nível de confiança de $(1 - \alpha)$ se $|T_0| > t_{(1-\frac{\alpha}{2}, n-1-d)}$ (ACTION, 2012).

Uma vez identificado o processo estocástico que gera a série temporal de interesse e os parâmetros do modelo, passa-se então à etapa de previsão de valores futuros, nesse estudo o software R será utilizado para os processos de identificação e previsão dos valores.

3.6 Seleção do Modelo

Os critérios de seleção do modelo realizados nesse estudo foram o Critério de Akaike, Erro Percentual Absoluto Médio (MAPE) e Erro Quadrático Médiode Previsão (EQM_p). Vale ressaltar que estas medidas são comparativas, não tem significado quando olhada individualmente (considerando apenas um modelo), assim a seleção do melhor modelo escolhido será sempre aquele que apresentar o menor valor nos três critérios. Neste estudo os três métodos convergiram, mas em caso de divergência o modelo escolhido seria aquele que obtivesse o menor valor no Erro Quadrático Médiode Previsão.

3.6.1 Critério de Informação de AKAIKE

O Critério de Informação de Akaike (AIC) é definido como

$$AIC_p = -2\log(L_p) + 2[(\rho + 1) + 1]$$

em que L_p é a função de máxima verossimilhança do modelo e ρ é o número de variáveis explicativas consideradas no modelo (GUTIERREZ, 2003). O critério de informação de Akaike (AIC) pode assumir qualquer valor, inclusive valor negativo, já que ele depende da forma da função de verossimilhança. Para se comparar o AIC, é necessário que os modelos ajustados sejam feitos no mesmo período amostral.

3.6.2 Erro Percentual Médio Absoluto

O Erro percentual absoluto médio (MAPE) expressa a precisão como porcentagem do erro. Como esse número é uma porcentagem, ele pode ser mais fácil de compreender do que outras estatísticas. A equação é:

$$\frac{\sum \left| \frac{(y_i - \hat{y}_i)}{y_i} \right|}{n}, (y_i \neq 0)$$

em que y_i é igual ao valor real, $\hat{y}_i$ é igual ao valor previsto, e n é igual ao número de observações (MINITAB, 2016).

3.6.3 Erro Quadrático Médio de Previsão

O Erro Quadrático Médio de Previsão (EQM_p) é definido por:

$$\frac{\sum_{i=1}^s (y_{t-s+i} - \hat{y}_{t-s+i})^2}{s}$$

em que s é o horizonte de previsão, y_{t-s+i} é o valor real das s últimas observações, $\hat{y}_{t-s+i}$ é o valor previsto em um modelo feito a partir dos dados sem as s últimas observações, e n é igual ao número de observações (MORETTIN & TOLOI, 1981). Assim compara-se os s valores retirados com os valores previstos a partir dos dados sem as s últimas observações.

3.7 Análise de Resíduos

De acordo com MORETTIN & TOLOI (1981) ao ajustarmos um modelo ARIMA, os resíduos de um modelo bem ajustado devem ser uma sequência de ruídos brancos, ou seja, erros de estimação com distribuição normal, média zero, variação constante ao longo do tempo e não autocorrelacionados. Para comprovar tais pressuposições realiza-se análises gráficas e testes de hipóteses.

3.7.1 Normalidade

De acordo com ABDI & MOLIN (2007) o teste de Lilliefors é uma adaptação do teste de Kolmogorof-Smirnoff, usado para verificação de normalidade de um conjunto de dados. A diferença básica entre os testes é que o teste de Kolmogorof-Smirnoff testa a normalidade a partir de média e desvio fornecidos, enquanto Lilliefors usa a média e o desvio-padrão, calculados no próprio conjunto de dados, testando a hipótese de normalidade dos dados. Para a decisão do teste utiliza-se a tabela de Lilliefors. Para análise informal de verificação de normalidade faz-se uma análise gráfica do histograma dos resíduos.

3.7.2 Variabilidade

Para verificação da variação ao longo do tempo dos resíduos faz-se o teste F para comparar duas variâncias, teste esse realizado para comparar as variâncias de duas amostras de populações normalmente distribuídas. Nesse estudo irá realizar a comparação da variância da primeira metade dos resíduos com a variância da segunda metade, testando a hipótese de que as variâncias são iguais. A regra de decisão será dada por n_1-1 e n_2-1 graus de liberdade, ou simplesmente $F_{(n_1-1; n_2-1)}$ (SILVA, 2014). A análise informal do gráfico de linhas dos resíduos do modelo também será realizada nesse estudo.

3.7.3 Autocorrelação

Um teste utilizado para verificar se o resíduo é independente e identicamente distribuído, isto é, se o resíduo não é autocorrelacionado, é o teste de Box e Pierce, (PRIESTLEY, 1989), o qual é baseado nas k primeiras autocorrelações, $\hat{r}_k$, dos resíduos. Para um processo ARIMA (p, d, q), se o modelo ajustado é apropriado, então faz-se a estatística do teste com distribuição aproximadamente qui-quadrado (χ^2). Testando a hipótese de que os resíduos são i.i.d. Rejeita H_0 se $Q > \chi^2$ com $(k - p - q)$ graus de liberdade, em que k é o número de "lags", p é a ordem da parte auto-regressiva e q a ordem da parte de médias móveis. Uma análise informal sugerida por MORETTIN & TOLOI (1981) é que para testar se uma série é constituída de observações independentes identicamente distribuídas, basta construir o correlograma a um intervalo de confiança $(1 - \alpha)$, $(1 - \alpha)\%$ das correlações deverão estar dentro deste intervalo de confiança.

3.7.4 Teste do Espectrograma

O espectrograma foi proposto no início dos estudos em análise espectral de série temporais para detectar periodicidades nos dados. O teste é realizado pela análise do estimador não tendencioso do espectro acumulado, seja:

$$I_a(f_1) = \frac{2}{n} \left[\left(\sum_{t=1}^n a_t \cos(2\pi f_i t) \right) + \left(\sum_{t=1}^n a_t \sin(2\pi f_i t) \right) \right]$$

O espectro acumulado e o periodograma acumulado, que é uma estimativa do espectro acumulado, são respectivamente:

$$C(f_j) = \frac{\sum_{i=1}^j I_a(f_i)}{n\sigma_a^2}; P_a(f) = \int_0^f P_a(g)dg = \begin{cases} 0, & f < 0 \\ 2\sigma_a^2 f, & 0 \leq f < \frac{1}{2} \\ \sigma_a^2, & f \geq \frac{1}{2} \end{cases};$$

Em que um pico na frequência f_i indica uma periodicidade no período $\frac{n}{f_i}$. Aplicando o teste Smirnov-Kolmogorov, podemos obter limites inferiores e superiores. Se os pontos estiverem dentro dos limites de confiança será ruído branco, caso contrário não será (SOUZA, 1991).

4. Resultados

Na Figura 3a, é apresentado o gráfico da série histórica da Produção de milho no Brasil, pela análise gráfica nota-se que os dados não possuem sazonalidade nem ciclicidade, pode-se dizer também que a série é não estacionária, uma vez que apresenta tendência. Segundo MORETTIN & TOLOI (1981) para retirarmos a tendência faz-se a diferenciada série, diferença apresentada na Figura 3b. Na Figura 3b foi feita a primeira diferença e nota-se média constante porém alta variabilidade nas ultimas observações, foi feito a segunda e terceira diferença, porém a variabilidade nas ultimas observações se manteve. COUTO (2008) em um experimento com culturas olerícolas com variabilidade heterogêneas, testou transformações para redução de variabilidade, e a transformação logarítmica mostrou-se mais eficaz. Ao realizar a transformação nos dados deste estudo, como mostra a Figura 3c, percebe-se maior linearidade quando comparado com a produção de milho no Brasil. Na figura 3d ao se realizar a diferença do logaritmo dos dados originais já se observa variabilidade homogênea em torno da média zero, dando indícios de estacionariedade, portanto não se faz necessário realizar outra diferença.

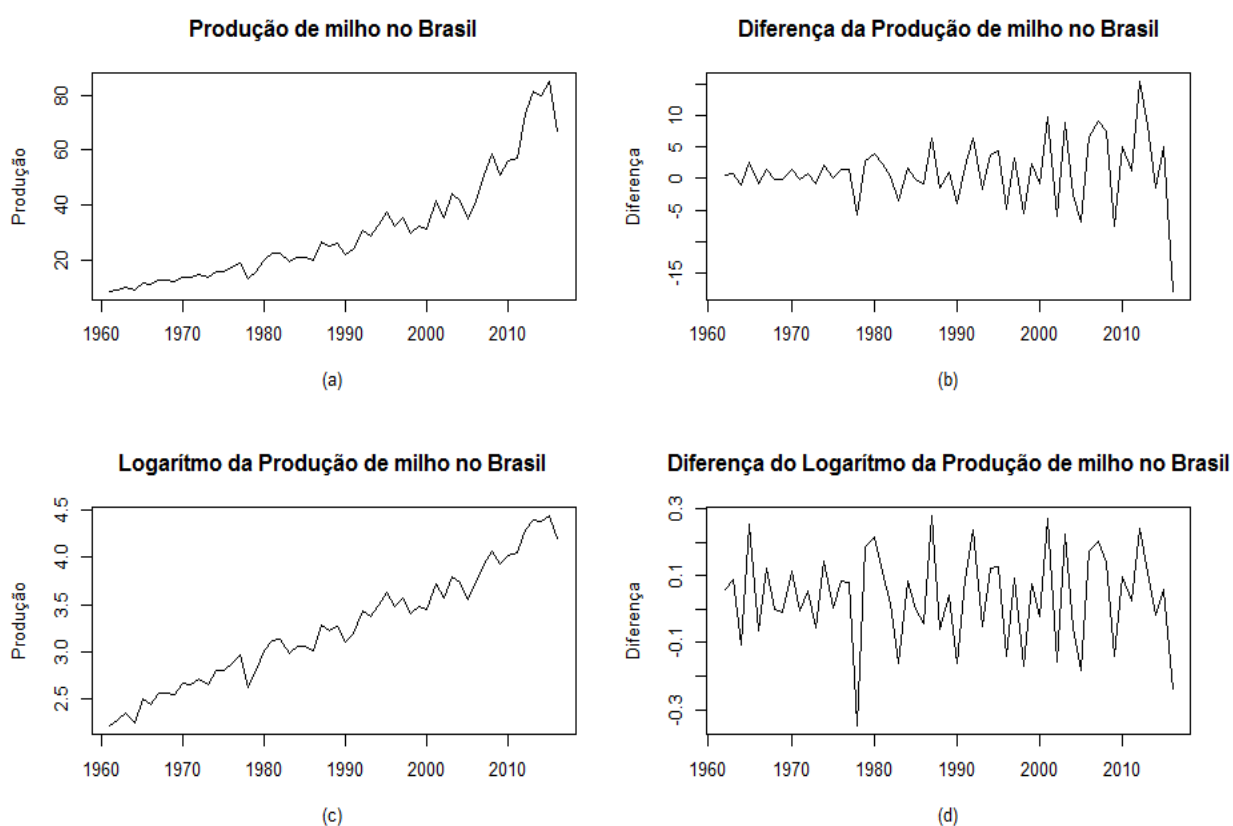


Figura 3 – Gráficos de linha da série histórica de produção de milho no Brasil, diferença dessa série, logaritmo e diferença do logaritmo dos dados originais

Na Tabela 1 o teste de Cox-Stuart foi realizado para testar a hipótese de que a série histórica possui média constante, ou seja, não possui tendência, essa hipótese foi rejeitada para as séries (a) e (c), produção de milho do Brasil e o logaritmo da mesma. As séries que a hipótese de nulidade não foi rejeitada, ou seja, não apresentaram tendência foram as séries (b) e (d), primeira diferença da serie original e a primeira diferença do logaritmo. Justificando o uso do método da diferença para transformar séries com tendência em séries sem tendência.

Tabela 1. Testes do Sinal de Cox-Stuart para tendência

Série Temporal	Estatística	P-valor
Produção de milho no Brasil (a)	28	0
Diferença da Produção de milho no Brasil (b)	15	0,7
Logaritmo da Produção de milho no Brasil (c)	28	0
Diferença do Logaritmo da Produção de milho no Brasil (d)	14	1

As séries temporais (a), (b), (c) e (d) foram divididas em duas partes, uma com a primeira metade das observações e a segunda parte com observações restantes, para a realização do teste F comparando a variância dos dois grupos, o resultado é apresentado na Tabela 2. Rejeita-se a hipótese de igualdade de variâncias para as séries (a) e (b), já as variâncias dos dados transformados mostraram melhores resultados, justificando a transformação logarítmica.

Tabela 2. Teste F para comparar a variância da primeira com a segunda metade das séries

Série Temporal	Estatística	P-valor
Produção de Milho no Brasil (a)	13,6687	0,0000
Diferença da Produção de Milho no Brasil (b)	0,1027	0,0000
Logaritmo da Produção de Milho no Brasil (c)	0,6519	0,2725
Diferença do Logaritmo da Produção de Milho no Brasil (d)	1,0000	1,0000

Pela análise da Figura 3, Tabela 1 e Tabela 2 a série (a), série com os dados originais, possui tendência e alta variabilidade, ou seja, não é estacionária. Ao realizar a primeira diferença, observa-se em (b) que foi retirada a tendência dos dados, porém alta variação nas ultimas observações. Nos dados de (c) as variâncias se mostraram estatisticamente iguais, mas apresentam tendência. Após transformação logarítmica e diferença da Produção de Milho no Brasil, a série histórica (d) atende as pressuposições de estacionariedade com variação homogênea e sem tendência.

Na Figura 4a, o gráfico de autocorrelação para a diferença da produção de milho no Brasil não nos dá pistas da ordem de Médias Móveis e na Figura 4b o mesmo para a ordem Auto-Regressiva. Para a diferença do logaritmo da produção de milho no Brasil, a função de autocorrelação, na Figura 4c, observa-se também o “lag” 1 significativo, o que pode indicar a ordem da parte de Médias Móveis, e na Figura 4d os “lag” 1 e 4 podendo indicar a ordem da parte Auto-Regressiva, a partir disso, por meio do software R realizou-se ajustes dos modelos com até dois parâmetros de Média Moveis e quatro Auto-Regressivos. Também serão feitos ajustes para a diferença da série original para comparar com os ajustes da diferença da série transformada.

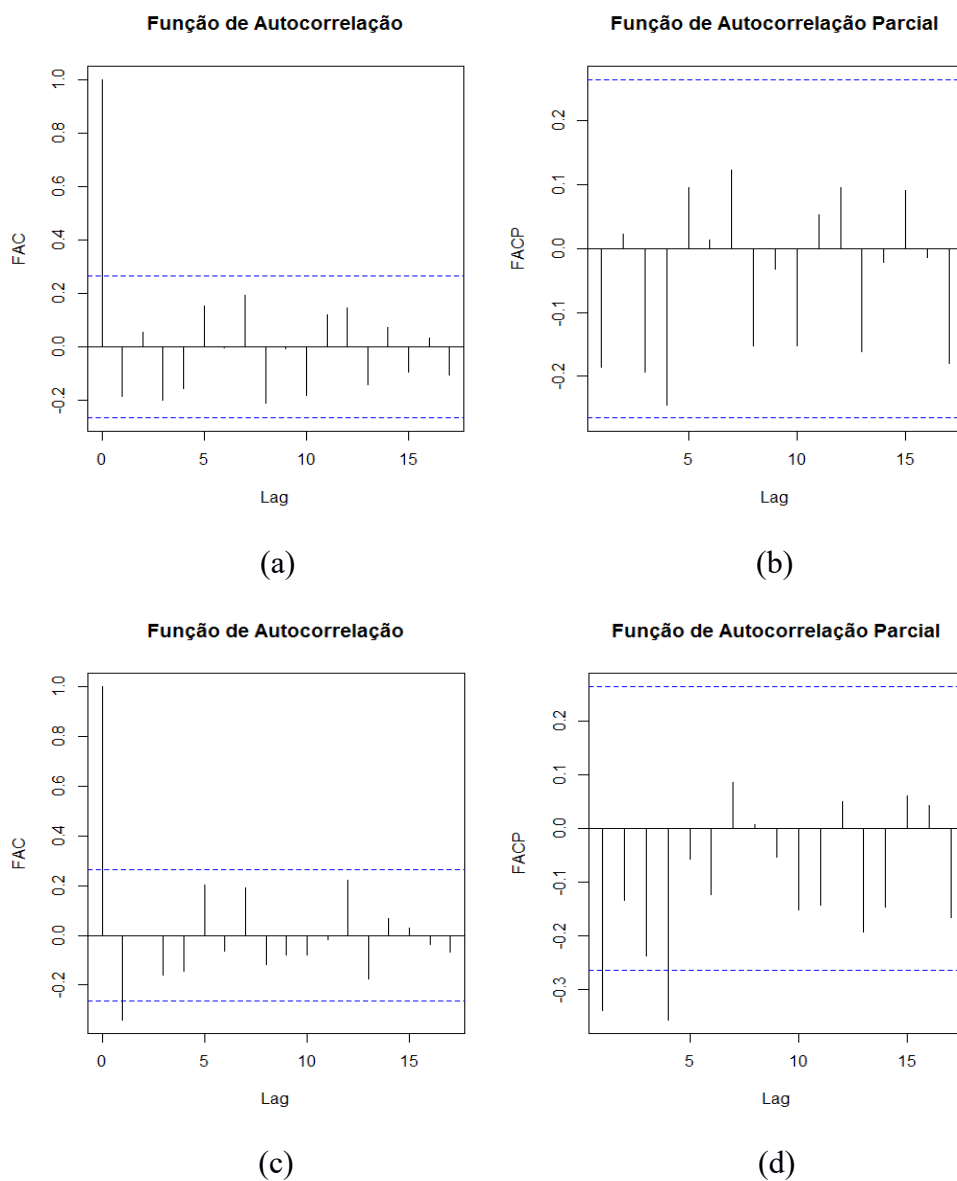


Figura 4 – Função de Autocorrelação e Função de Autocorrelação Parcial, para Diferença da Produção de Milho no Brasil e Diferença do Logaritmo da Produção de Milho no Brasil

Testando a hipótese em que os parâmetros são não significativos, foi feito o teste T-student ao nível de significância de 5%, à esquerda os modelos a partir da diferença dos dados originais e à direita a partir da diferença do logaritmo. Os modelos que tiveram a hipótese nula rejeitada, ou seja, que possuem todos os parâmetros significativos foram os modelos da primeira diferença da série original ARIMA: (0,1,1); (1,1,2); e (2,1,2); da série com a primeira diferença da transformação ARIMA(1,1,2) e (2,1,2).

Tabela3. Teste T-student para significância dos Parâmetros nos modelos ajustados

ARIMA	Parâmetro	Diferença da Produção de Milho no Brasil		Diferença do Logaritmo da Produção de Milho no Brasil	
		p-valor	Estimativa	p-valor	Estimativa
(0,1,1)	θ_1	0,00000	-0,1311	0,06313	0,2483
(1,1,0)	φ_1	0,22835	-0,1557	0,06509	-0,2542
(1,1,1)	φ_1	0,10718	-0,5521	0,36179	-0,2812
	θ_1	0,21713	0,3852	0,39678	0,0289
(0,1,2)	θ_1	0,00000	-0,0957	0,07197	-0,3527
	θ_2	0,33978	0,1128	0,38588	0,0281
(2,1,0)	φ_1	0,17799	-0,1453	0,07412	-0,2531
	φ_2	0,29280	0,1149	0,39691	0,0045
(1,1,2)	φ_1	0,22835	-0,4574	0,00000	0,9998
	θ_1	0,21179	0,3528	0,00000	-1,6751
	θ_2	0,25485	0,1834	0,00000	0,6817
	φ_1	0,28382	-0,3730	0,00012	-1,0785
(2,1,1)	φ_2	0,33860	0,0948	0,06984	-0,2573
	θ_1	0,34422	0,2329	0,00083	0,8160
	φ_1	0,00000	-0,8900	0,00000	-1,6390
(2,1,2)	φ_2	0,00000	-0,9693	0,00005	-0,8714
	θ_1	0,00000	0,8615	0,00000	1,4682
	θ_2	0,00000	1,0000	0,02916	0,6940
	φ_1	0,00015	-0,8355	0,00275	-0,8805
(4,1,1)	φ_2	0,38891	0,0374	0,26427	-0,1694
	φ_3	0,34754	-0,0925	0,33175	-0,1055
	φ_4	0,03773	-0,3077	0,12785	-0,2062
	θ_1	0,00000	0,7279	0,01546	0,6435

Os critérios de seleção do modelo realizados nesse estudo foram o critério de Akaike, Erro percentual absoluto médio (MAPE) e Erro Quadrático Médio de previsão (EQM_p). O modelo escolhido será sempre aquele que apresentar o menor valor nos três testes. O modelo que obteve os menores resultados foi da série da diferença do logaritmo da produção de milho no Brasil, com um parâmetro auto-regressivo e dois médias-móveis.

Tabela4. Critério de Seleção do Modelo

Série	Modelo	AIC	EQM_p	MAPE
Diferença da Série Original	(0,1,1)	342,0692	829,9954	0,3727
	(1,1,2)	344,6067	859,7281	0,3836
	(2,1,2)	344,9590	844,5083	0,3775
Diferença do Logaritmo	(1,1,2)	-63,9802	343,8588	0,2229
	(2,1,2)	-54,3783	849,5205	0,3774

O modelo escolhido foi ARIMA (1,1,2): $\widehat{W}_t = w_{t-1} \cdot \hat{\varphi}_1 + w_{\alpha_t-1} \cdot \hat{\theta}_1 + w_{\alpha_t-2} \cdot \hat{\theta}_2$, ao substituir os valores dos coeficientes encontrado teremos $\widehat{W}_t = w_{t-1} \cdot 0,998 - w_{\alpha_t-1} \cdot 1,6751 + w_{\alpha_t-2} \cdot 0,6817$ na Figura 5 temos a comparação do modelo ARIMA (1,1,2) com a série histórica da produção de milho no Brasil, apresentando um bom ajuste com valores próximos aos reais.

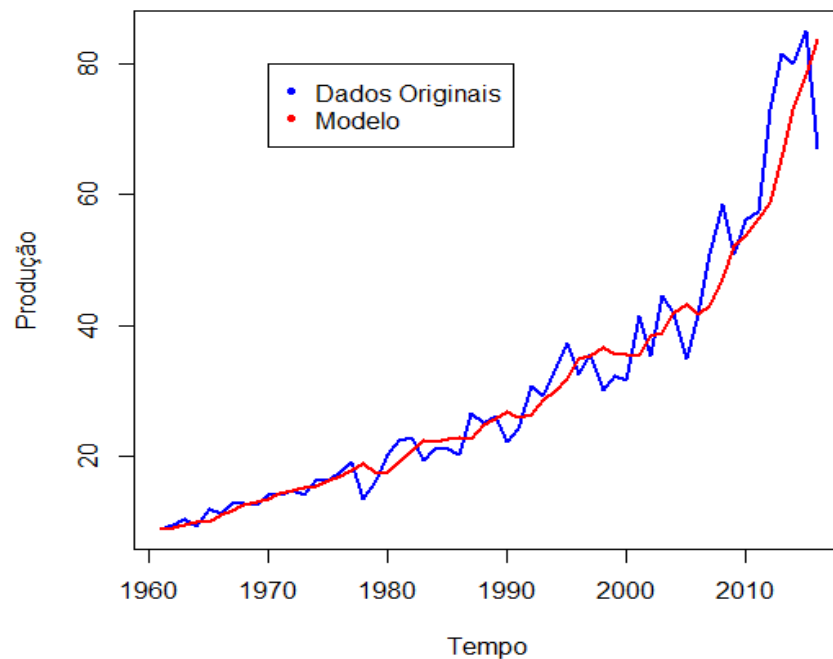


Figura 5. Série Estimada e Série da Produção de Milho no Brasil das safras de 1960/1961 à 2015/2016

Após seleção do modelo os resíduos devem estar normalmente distribuídos, ter média zero com variação constante ao longo do tempo e ser não autocorrelacionados. Pela análise gráfica da Figura 6 percebe-se que os resíduos estão normalmente distribuídos.

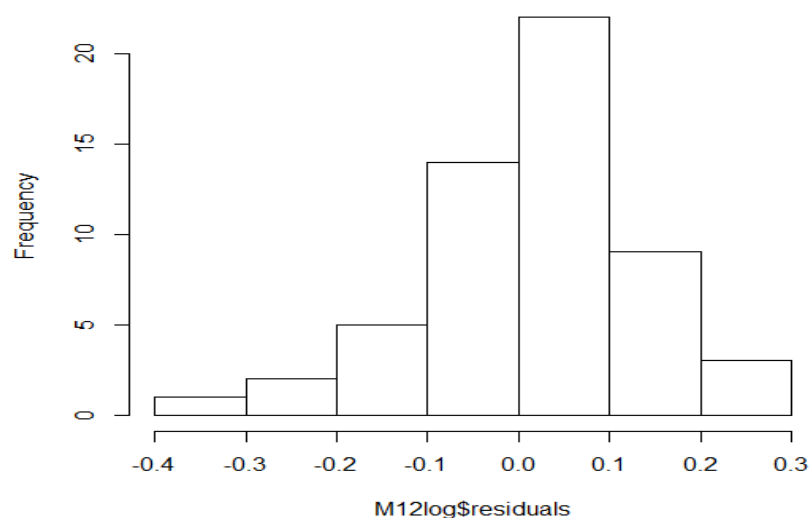


Figura 6. Histograma dos Resíduos do modelo ARIMA(1,1,2)

Na Tabela 5, o teste de Lilliefors para normalidade foi realizado, testando a hipótese de que os resíduos do modelo ARIMA(1,1,2) seguem distribuição normal. Não rejeita-se a hipótese nula, ou seja, à 95% de confiança os resíduos seguem distribuição normal.

Tabela 5. Teste para normalidade dos resíduos do modelo ARIMA(1,1,2)

Teste de Hipótese	Estatística	P-valor
Testes de Lilliefors	0,0913	0,2916

Pela Figura 7 observa-se que os resíduos seguem média zero e possuem homocedasticidade na variação ao longo do tempo.

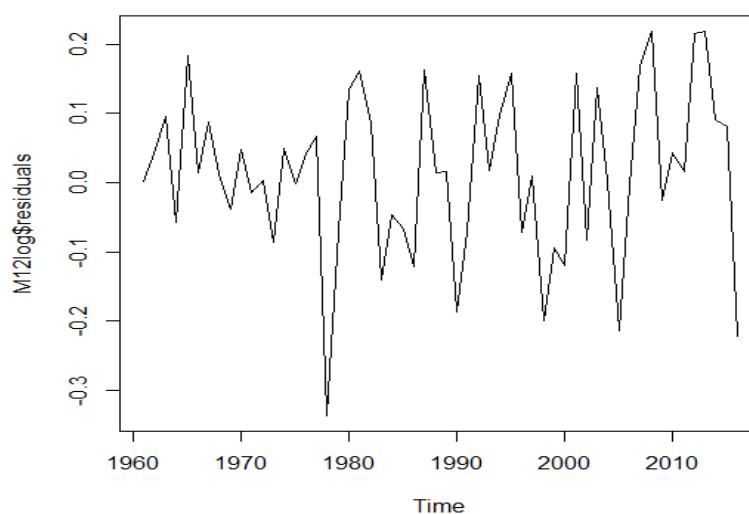


Figura 7. Gráfico de linha dos Resíduos do modelo ARIMA(1,1,2)

Na Tabela 6, o teste F para igualdade de variâncias foi realizado comparando a primeira metade dos resíduos com a segunda metade, testando a hipótese em que as variâncias são iguais. Não rejeita a hipótese nula, ou seja, à 95% de confiança as duas metades possuem variâncias estatisticamente iguais.

Tabela 6. Teste para igualdade da variância da primeira metade e variância da segunda metade dos resíduos do modelo ARIMA(1,1,2)

Teste de Hipótese	Estatística	P-valor
Teste F	0,8595	0,6968

Na Figura 8, o modelo ARIMA(1,1,2) foi testado pelo Espectrograma. Segundo SOUZA (1991) como o teste baseado no estudo do espectro acumulado mostra os valores dentro dos limites de confiança obtidos pelo teste de Smirnov-Kolmogorov, afirma-se que os resíduos são ruído branco.

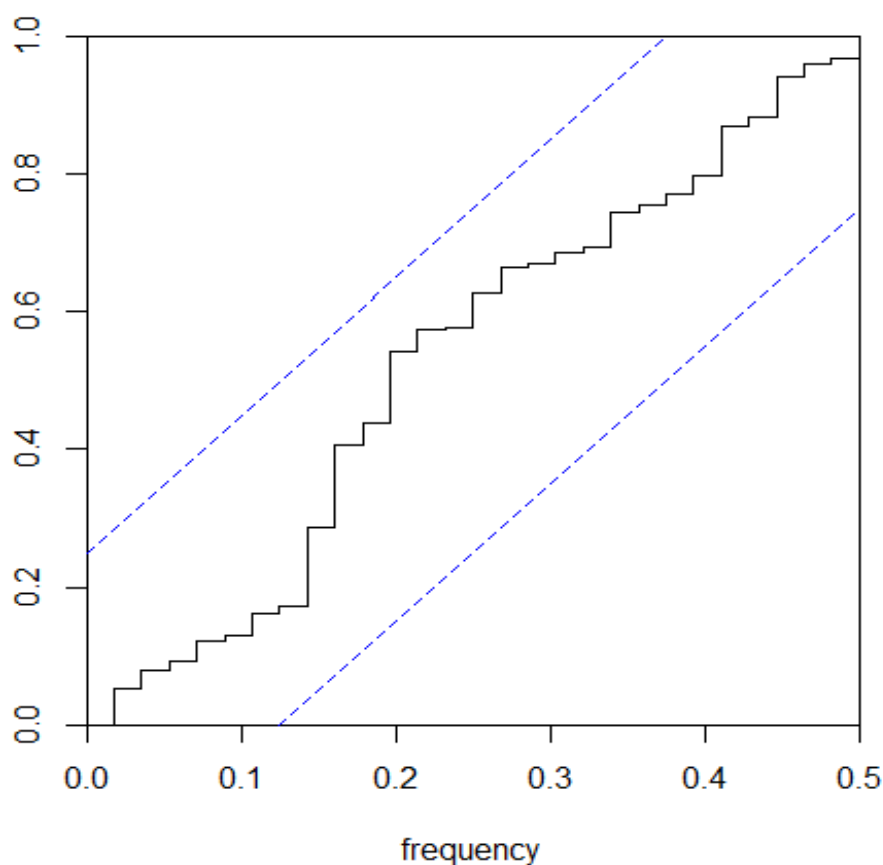


Figura 8. Teste do Espectrograma para o modelo ARIMA(1,1,2)

Após as análises informais e os testes de hipóteses pode-se afirmar que os erros do modelo estimados estão normalmente distribuídos, tem média zero com variação constante ao longo do tempo e não são autocorrelacionados, ou seja, é ruído branco, concluindo então que o modelo foi bem ajustado.

Na figura 9, o modelo $\widehat{W}_t = w_{t-1} \cdot 0,998 - w_{\alpha_t-1} \cdot 1,6751 + w_{\alpha_t-2} \cdot 0,6817$ é apresentado com a previsão da produção brasileira de milho em azul escuro para os próximos 10 anos, e em azul claro o intervalo de confiança para cada safra prevista.

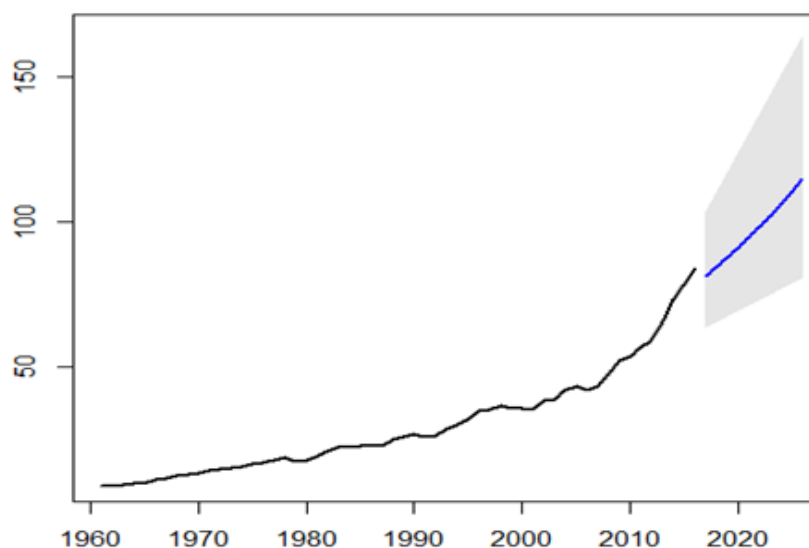


Figura 9. Previsão da Produção Brasileira de Milho para as safras de 2016/2017 a 2025/2026

Para a próxima safra a previsão é de aproximadamente 81 milhões de toneladas, vale ressaltar que em pesquisa no site PSD online no segundo semestre de 2016 a produção de milho no Brasil foi de 83,5 milhões de toneladas, valor que deve se manter dentro dos limites do valor previsto. Esse modelo faz uma previsão superior ao obtido pelo Ministério de Agricultura em sua revista projeções do agronegócio (BRASIL D, 2016) que para a safra 2025/26, a produção projetada pelo Ministério da Agricultura é de 94,7 milhões de toneladas, e seu limite superior pode chegar a 140,1 milhões de toneladas enquanto o modelo ARIMA(1,1,2) estima 114,77 milhões de toneladas com limite superior de 163,20 milhões de toneladas. Tais valores são apresentados na Tabela 6.

Tabela7.Previsão da Produção Brasileira de Milho de 2016/2017 a 2025/2026

Safra	Produção	Lim. Inf.	Lim. Sup.
2016/2017	81,0066	63,7845	102,8797
2017/2018	84,2072	65,4738	108,3016
2018/2019	87,5334	67,2092	114,0036
2019/2020	90,9900	68,9920	120,0022
2020/2021	94,5823	70,8227	126,3139
2021/2022	98,3153	72,7014	132,9549
2022/2023	102,1957	74,6283	139,9451
2023/2024	106,2272	76,6056	147,3028
2024/2025	110,4177	78,6329	155,0488
2025/2026	114,7711	80,7114	163,2038

5. Conclusão

Pelo modelo ajustado foi possível fazer previsões para a produção brasileira de milho no Brasil, para a próxima safra (2016/2017), o valor previsto foi uma produção de 81 milhões de toneladas, com limite inferior de 63,78 milhões de tonelada e limite superior de 102,87 milhões de tonelada, para a safra de 2025/2026 estima 114,77 milhões de toneladas com limite superior de 163,20 milhões de toneladas e limite inferior de 80,7114.

6. Referências

ABDI, HervéAbdi & MOLIN,Paul. **Lilliefors/Van Soest's test of normality**, 2007. Disponível em: <http://www.utdallas.edu/~herve/Abdi-Lillie2007-pretty.pdf>. Acesso em 23 de dezembro de 2016.

ACTION, Portal. **Teste e Intervalo de Confiança para Parâmetros**. Disponível em <<http://www.portalaction.com.br/analise-de-regressao/14-testes-e-intervalos-de-confianca-para-os-parametros>>. Acesso em 30 de novembro de 2016.

BRASIL A. **Balanço do Exercício 2004**. Ministério da Agricultura, Pecuária e Abastecimento. Disponível em: <http://www.agricultura.gov.br/arq_editor/image/RELATORIO_GESTAO/GM/2004.pdf>. Acesso em 30 de novembro de 2016.

BRASIL B. **Informe Econômico da Política Agrícola nº1/2016**. Ministério da Agricultura, Pecuária e Abastecimento. Disponível em: <http://www.agricultura.gov.br/arq_editor/Informe_economico_politica_agricola_Edicao_n01-2016.pdf>. Acesso em 30 de novembro de 2016.

BRASIL C. **Ministério da Agricultura, Pecuária e Abastecimento**, 2016. Disponível em: <<http://www.agricultura.gov.br/vegetal/culturas/milho>>. Acesso em 30 de novembro de 2016.

BRASIL D. **Projeções do Agronegócio: Brasil 2015/16 a 2025/26**. Ministério da Agricultura, Pecuária e Abastecimento. Secretaria de Política Agrícola, 2016. Disponível em: <http://www.agricultura.gov.br/arq_editor/Proj_Agronegocio2016.pdf>. Acesso em 30 de novembro de 2016.

COUTO,Mara Rubia Machado. **Transformação de dados com excesso de zero em experimentos com culturas olerícolas**, 2008. Disponível em: <http://cascavel.ufsm.br/tede/tde_arquivos/4/TDE-2008-08-25T160952Z-1685/Publico/MARARUBIACOUTO.pdf>. Acesso em 21 de dezembro de 2016.

CONAB. **Companhia Nacional de Abastecimento**, 2016. Disponível em: <<http://www.conab.gov.br/conteudos.php?a=1252&t=2>>. Acesso em 30 de novembro de 2016.

EHLERS, Ricardo Sandes. **Análise de séries temporais: Material didático e instrucional**, 2003. Laboratório de Estatística e Geoinformação. UFP. Disponível em: <<http://www.each.usp.br/rvicente/AnaliseDeSeriesTemporais.pdf> >.Acesso em: 5 de dezembro de 2016.

FARIA,Katon Oliveira Santos. **Estudo da série temporal do preço da arroba do boi gordo da BM&F para o estado de Goiás**. 2015. 39 f. Trabalho de Conclusão de Curso

(Graduação em Estatística) – Universidade Federal de Uberlândia, Uberlândia, 2015. Disponível em: <<https://repositorio.ufu.br/bitstream/123456789/17827/1/EstudoS%C3%A9rieTempora.pdf>>. Acesso em: 5 de dezembro de 2016.

FRANCO, Glauro da Conceição. **Apostila sobre Modelagem Box e Jenkins**, Departamento de Estatística, 2009, UFMG. Disponível em <<ftp://est.ufmg.br/pub/glauro/Met%20Previsao/>>. Acesso em: 6 de dezembro de 2016.

GUTIERREZ, José Luis Carrasco. **Monitoring of the Corumbá - i Dam Instrumentation Neural Networks and the Box & Jenkins null Models**, 2003. Disponível em: <http://www.maxwell.vrac.puc-rio.br/4244/4244_5.PDF>. Acesso em 30 de novembro de 2016.

MAGGI, Blairo. **Carta da Agricultura: Crescer e Simplificar**. Revista de Política Agrícola nº 2/2016. Disponível em: <http://www.agricultura.gov.br/arq_editor/RPA%202%202016.pdf>. Acesso em 30 de novembro de 2016.

MAKRIDAKIS, Spyros; WHEELWRIGHT, Seven; HYNDMAN, Rob. **Forecasting methods and applications**. 3.ed. New York: John Wiley & Sons, 1998.

MINITAB. **Suporte ao Minitab**. Disponível em <<http://support.minitab.com/pt-br/minitab/17/topic-library/modeling-statistics/time-series/time-series-models/what-are-map-mad-and-msd/>>. Acesso em 30 de novembro de 2016.

MORETTIN, Pedro Alberto; TOLOI, Clélia Maria de Castro. **Previsão de series temporais**, Rio de Janeiro – RJ, 1981. Disponível em: <http://www.impa.br/opencms/pt/biblioteca/cbm/13CBM/13_CBM_81_01.pdf>.

NOGUEIRA, Fernando. **Modelagem e Simulação - Modelos de Previsão**, 2009. Disponível em: <<http://www.ufjf.br/epd042/files/2009/02/previsao1.pdf>>. Acesso em: Acesso em: 5 de dezembro de 2016.

PEIXOTO, Claudio de Miranda. **O milho no Brasil, sua importância e evolução**. Disponível em <<http://www.pioneersementes.com.br/media-center/artigos/165/o-milho-no-brasil-sua-importancia-e-evolucao>>. Acesso em 30 de novembro de 2016.

PINDYCK, Robert Stephen. RUBINFELD, Daniel. **Econometria Modelos & Previsões**. Trad. da 4ª Edição. Rio de Janeiro: Elsevier/Campus, 2004.

PRIESTLEY, M. B. **Spectral analysis and time series**. 6 ed. New York: Academic Press, 1989.

PRODUCTION, SUPPLY AND DISTRIBUTION. **The official United States Department of Agriculture Supply**. Disponível em: <<https://apps.fas.usda.gov/psdonline/>>. Acesso em 5 de Janeiro de 2017.

SILVA, Tais Medeiros. **Teste t-student, igualdade de variâncias**, 2014. Disponível em <http://www.ufpa.br/heliton/arquivos/aplicada/seminarios/M1_01_Testes_t_Tais.pdf>. Acesso em 24 de dezembro de 2016.

SOUZA, Adriano Mendonça. **Aplicação e performance da análise de intervenção em séries macroeconômicas brasileiras**, 1991. Disponível em <<http://w3.ufsm.br/adriano/mdta/mono.pdf>>. Acesso em 24 de dezembro de 2016.

RODRIGUES, Roberto. **O céu é o limite para o agronegócio brasileiro**. Conjuntura Econômica, Rio de Janeiro, V.60, n.11, p.14-15, 2006.

REIS, Marcelo Menezes. **Análise de Séries Temporais**, 2015. Disponível em: <<http://www.inf.ufsc.br/~marcelo/Cap4.pdf>>. Acesso em 30 de Novembro de 2015.

R CORE TEAM (2017). **R: A language and environment for statistical computing. R Foundation for Statistical Computing**. Vienna, Austria. Disponível em: <<https://www.r-project.org>>. Acesso em 8 de Janeiro de 2017.

ŽITKOVIC, Gordan. **Introduction to stochastic processes**. December 24, 2010. 106 p.

7. Anexo I

Série Histórica da Produção de milho no Brasil

Safr				Safr			
		Produção				Produção	
1960	/	1961	9,036	1988	/	1989	26,270
1961	/	1962	9,587	1989	/	1990	22,300
1962	/	1963	10,478	1990	/	1991	24,330
1963	/	1964	9,408	1991	/	1992	30,800
1964	/	1965	12,112	1992	/	1993	29,200
1965	/	1966	11,371	1993	/	1994	32,934
1966	/	1967	12,824	1994	/	1995	37,440
1967	/	1968	12,814	1995	/	1996	32,480
1968	/	1969	12,693	1996	/	1997	35,700
1969	/	1970	14,216	1997	/	1998	30,100
1970	/	1971	14,130	1998	/	1999	32,393
1971	/	1972	14,891	1999	/	2000	31,641
1972	/	1973	14,109	2000	/	2001	41,536
1973	/	1974	16,284	2001	/	2002	35,501
1974	/	1975	16,354	2002	/	2003	44,500
1975	/	1976	17,751	2003	/	2004	42,000
1976	/	1977	19,256	2004	/	2005	35,000
1977	/	1978	13,569	2005	/	2006	41,700
1978	/	1979	16,310	2006	/	2007	51,000
1979	/	1980	20,214	2007	/	2008	58,600
1980	/	1981	22,555	2008	/	2009	51,000
1981	/	1982	22,932	2009	/	2010	56,100
1982	/	1983	19,500	2010	/	2011	57,400
1983	/	1984	21,180	2011	/	2012	73,000
1984	/	1985	21,170	2012	/	2013	81,500
1985	/	1986	20,264	2013	/	2014	80,000
1986	/	1987	26,760	2014	/	2015	85,000
1987	/	1988	25,220	2015	/	2016	67,000

8. Anexo II

Script.R

```
#biblioteca
require(nortest)
require(Rcpp)
require(forecast)
require(graphics)
require(TSA)
require(car)
require(randtests)
library(nlme)

#Entrada dos dados
dados<-read.table("milho.txt", header=T)
dados=ts(dados,start=1961,frequency=1)#classificando como série temporal,
colocando data inicial e frequencia

#Figura 3 - Gráficos de linhas
par(mfrow=c(1,1))
ts.plot(dados,main = "Produção de milho no
Brasil",ylab="Produção",xlab="(a)")#apresenta tendencia, sem sazonalidade, não
estacionária
difdados<-diff(dados)
ts.plot(difdados,main = "Diferença da Produção de milho no
Brasil",ylab="Diferença",xlab="(b)")#apresenta tendencia, sem sazonalidade, não
estacionária, alta variabilidade nas utimas diferenças
milho<-log(dados)
ts.plot(milho,main = "Logarítmo da Produção de milho no
Brasil",ylab="Produção",xlab="(c)")#apresenta tendencia, sem sazonalidade, maior
linearidade
dif<-diff(milho)
ts.plot(dif,main = "Diferença do Logarítmo da Produção de milho no
Brasil",ylab="Diferença",xlab="(d)")

#Teste para verificar homocedasticidade das variancias

dadosa<-dados[1:28,1]
dadosa
dadosb<-dados[29:56,1]
dadosb

difdadosa<-diff(dadosa)
difdadosa
```

```
difdadosb<-diff(dadosb)
difdadosb
```

```
milhoa<-milho[1:28,1]
milhoa
milhob<-milho[29:56,1]
milhob
```

```
difa<-diff(milhoa)
difa
difb<-diff(milhob)
difb
```

#Tabela 1 - Teste F para Variância

```
var.test(dadosb,dadosa)
var.test(difdadosa,difdadosb)
var.test(milhoa,milhob)
var.test(difa,difb)
```

#Tabela 2 - Teste de Box-Stuart

```
cox.stuart.test(dados)
cox.stuart.test(difdados)
cox.stuart.test(milho)
cox.stuart.test(dif)
```

#Figura 4. FAC e FACP

```
par(mfrow=c(1,2))
acf(dif,lag.max=NULL,main="Função de Autocorrelação",ylab="FAC")#serie não
estacionária, correlação apresenta tendencia
pacf(dif,lag.max=NULL,main="Função de Autocorrelação Parcial",ylab="FACP")#serie
não estacionária, correlação apresenta tendencia
```

```
acf(difdados,lag.max=NULL,main="Função de Autocorrelação",ylab="FAC")#serie não
estacionária, correlação apresenta tendencia
pacf(difdados,lag.max=NULL,main="Função de Autocorrelação
Parcial",ylab="FACP")#serie não estacionária, correlação apresenta tendencia
```

```
acf(dados,lag.max=NULL,main="Função de Autocorrelação dos Dados
Originais",ylab="FAC")#serie não estacionária, correlação apresenta tendencia
pacf(dados,lag.max=NULL,main="Função de Autocorrelação
Parcial",ylab="FACP")#serie não estacionária, correlação apresenta tendencia
```

#Modelos e Teste de Significancia dos parâmetros

```

M01log=stats::arima(dados,order=c(0,1,1))
AIC(M01log)
M01log
mm1<--0.1311/ 0.1384
mm1
dt(mm1, length(difdados))

```

```

M02log=stats::arima(dados,order=c(0,1,2))
AIC(M02log)
M02log
mm1<-0.0957/ 0.1726
mm1
mm2<-0.1128/ 0.1922
mm2
dt(mm1, length(difdados))
dt(mm2, length(difdados))

```

```

M10log=stats::arima(dados,order=c(1,1,0))
AIC(M10log)
M10log
ar1<--0.1557/ 0.1486
ar1
dt(ar1, length(difdados))

```

```

M11log=stats::arima(dados,order=c(1,1,1))
AIC(M11log)
M11log
ar1<--0.5521/ 0.3402
ar1
mm1<-0.3852/ 0.3518
mm1
dt(ar1, length(difdados))
dt(mm1, length(difdados))

```

```

M12log=stats::arima(dados,order=c(1,1,2))
AIC(M12log)
M12log
ar1<--0.4574/ 0.3337
ar1
mm1<-0.3528/ 0.3157
mm1
mm2<-0.1834/ 0.1957
mm2
dt(ar1, length(difdados))
dt(mm1, length(difdados))

```

```

dt(mm2, length(difdados))

M20log=stats::arima(dados,order=c(2,1,0))
AIC(M20log)
M20log
ar1<--0.1453/ 0.1149
ar1
ar2<-0.1149/ 0.1481
ar2
dt(ar1, length(difdados))
dt(ar2, length(difdados))

M21log=stats::arima(dados,order=c(2,1,1))
AIC(M21log)
M21log
ar1<- -0.3730/ 0.4578
ar1
ar2<-0.0948/ 0.1692
ar2
mm1<-0.2329/ 0.4389
mm1
dt(ar1, length(difdados))
dt(ar2, length(difdados))
dt(mm1, length(difdados))

M22log=stats::arima(dados,order=c(2,1,2))
AIC(M22log)
M22log
ar1<--0.8900/ 0.1633
ar1
ar2<--0.8693 / 0.0950
ar2
mm1<-0.8615/ 0.1431
mm1
mm2<-1/ 0.2165
mm2
dt(ar1, length(difdados))
dt(ar2, length(difdados))
dt(mm1, length(difdados))
dt(mm2, length(difdados))

M14log=stats::arima(dados,order=c(4,1,1))
M14log
ar1<--0.8355 / 0.1979
ar1

```

```

ar2<-0.0374 / 0.1844
ar2
ar3<--0.0925 / 0.1805
ar3
ar4<--0.3077 / 0.1401
ar4
mm1<-0.7279/ 0.1577
mm1
dt(ar1, length(difdados))
dt(ar2, length(difdados))
dt(ar3, length(difdados))
dt(ar4, length(difdados))
dt(mm1, length(difdados))

```

#PRIMEIRA DIFERENÇA

```

M01log=stats::arima(milho,order=c(0,1,1))
AIC(M01log)
M01log
mm1<--0.2483/ 0.1285
mm1
dt(mm1, length(dif))

```

```

M02log=stats::arima(milho,order=c(0,1,2))
AIC(M02log)
M02log
mm1<--0.2537/ 0.1364
mm1
mm2<-0.0281/ 0.1182
mm2
dt(mm1, length(dif))
dt(mm2, length(dif))

```

```

M10log=stats::arima(milho,order=c(1,1,0))
AIC(M10log)
M10log
ar1<--0.2542/ 0.1327
ar1
dt(ar1, length(dif))

```

```

M11log=stats::arima(milho,order=c(1,1,1))
AIC(M11log)
M11log
ar1<--0.2812/ 0.6566
ar1

```

```
mm1<-0.0289/ 0.6927
mm1
dt(ar1, length(dif))
dt(mm1, length(dif))
```

```
M12log=stats::arima(milho,order=c(1,1,2))
AIC(M12log)
M12log
ar1<-0.9998/ 0.0008
ar1
mm1<--1.6751/ 0.1269
mm1
mm2<-0.6817/ 0.1256
mm2
dt(ar1, length(dif))
dt(mm1, length(dif))
dt(mm2, length(dif))
```

```
M20log=stats::arima(milho,order=c(2,1,0))
AIC(M20log)
M20log
ar1<--0.2531/ 0.1373
ar1
ar2<-0.0045/ 0.1368
ar2
dt(ar1, length(dif))
dt(ar2, length(dif))
```

```
M21log=stats::arima(milho,order=c(2,1,1))
AIC(M21log)
M21log
ar1<--1.0785/ 0.2507
ar1
ar2<--0.2573/ 0.1371
ar2
mm1<-0.8160/ 0.2216
mm1
dt(ar1, length(dif))
dt(ar2, length(dif))
dt(mm1, length(dif))
```

```
M22log=stats::arima(milho,order=c(2,1,2))
AIC(M22log)
M22log
```

```

ar1<--1.6390/ 0.1592
ar1
ar2<--0.8714/ 0.1915
ar2
mm1<-1.4682/ 0.2426
mm1
mm2<-0.6940/ 0.2993
mm2
dt(ar1, length(dif))
dt(ar2, length(dif))
dt(mm1, length(dif))
dt(mm2, length(dif))

```

```

M14log=stats::arima(milho,order=c(4,1,1))
M14log
ar1<--0.8805 / 0.2693
ar1
ar2<--0.1694 / 0.1887
ar2
ar3<--0.1055 / 0.1772
ar3
ar4<--0.2062 / 0.1368
ar4
mm1<-0.6435/ 0.2475
mm1
dt(ar1, length(dif))
dt(ar2, length(dif))
dt(ar3, length(dif))
dt(ar4, length(dif))
dt(mm1, length(dif))

```

```

M12log=stats::arima(milho,order=c(1,1,2))
AIC(M12log)
M12log
ar1<-0.9998/ 0.0008
ar1
mm1<--1.6751/ 0.1269
mm1
mm2<-0.6817/ 0.1256
mm2
dt(ar1, length(dif))
dt(mm1, length(dif))
dt(mm2, length(dif))

```

#Diferença:(0,1,1); (1,1,2) e (2,1,2); Log (1,1,2) e (2,1,2).

```

aM01log=stats::arima(dados,order=c(0,1,1))
aM12log=stats::arima(dados,order=c(1,1,2))
aM22log=stats::arima(dados,order=c(2,1,2))
bM12log=stats::arima(milho,order=c(1,1,2))
bM22log=stats::arima(milho,order=c(2,1,2))

#Seleção do Modelo
AIC(aM01log)
AIC(aM12log)
AIC(aM22log)
AIC(bM12log)
AIC(bM22log)

#eqmP e MAPE
M01log=stats::arima(dados,order=c(0,1,1))
TF01<-dados[1:46,1]
TF01
M01TF=stats::arima(TF01,order=c(0,1,1))
forecast(M01TF, 10, level=c(95))
a01<-
c(40.02923,40.02923,40.02923,40.02923,40.02923,40.02923,40.02923,40.0
2923,40.02923)
a01
eM01<-dados[47:56,1]
Erro01<-eM01-a01
EQM01<-(1/length(Erro01))*sum(Erro01^2)
EQM01

MAPE01<-(1/length(eM01))*sum(abs(Erro01)/eM01)
MAPE01

M12log=stats::arima(dados,order=c(1,1,2))
TF12<-dados[1:46,1]
TF12
M12TF=stats::arima(TF12,order=c(1,1,2))
forecast(M12TF, 10, level=c(95))
a12<-
c(38.60802,38.60802,39.37643,39.68033,39.54008,39.60480,39.57493,39.58872,39.5
8236,39.58529)
a12
eM12<-dados[47:56,1]
Erro12<-eM12-a12
EQM12<-(1/length(Erro12))*sum(Erro12^2)
EQM12

```

```
MAPE12<-(1/length(eM12))*sum(abs(Erro12)/eM12)
MAPE12
```

```
M22log=stats::arima(dados,order=c(2,1,2))
TF22<-dados[1:46,1]
TF22
M22TF=stats::arima(TF02,order=c(2,1,2))
forecast(M22TF, 10, level=c(95))
a22<-
c(39.42135,39.76693,39.76693,39.76693,39.76693,39.76693,39.76693,39.76693,39.7
6693,39.76693)
a22
eM22<-dados[47:56,1]
Erro22<-eM22-a22
EQM22<-(1/length(Erro22))*sum(Erro22^2)
EQM22
MAPE22<-(1/length(eM22))*sum(abs(Erro22)/eM22)
MAPE22
```

```
M12log=stats::arima(milho,order=c(1,1,2))
TF12<-milho[1:46,1]
M12TF=stats::arima(TF12,order=c(1,1,2))
M12TF<-forecast(M12TF, 10, level=c(95))
forecast(M12TF, 10, level=c(95))
a12<-
c(3.792144,3.819295,3.846409,3.873486,3.900525,3.927526,3.954490,3.981416,4.00
8306,4.035158)
exp(a12)
eM12<-dados[47:56,1]
eM12
Erro12<-eM12-(exp(a12))
EQM12<-(1/length(Erro12))*sum(Erro12^2)
EQM12
MAPE12<-(1/length(eM12))*sum(abs(Erro12)/eM12)
MAPE12
```

```
M22log=stats::arima(milho,order=c(2,1,2))
TF22<-milho[1:46,1]
M22TF=stats::arima(TF22,order=c(2,1,2))
M22TF<-forecast(M22TF, 10, level=c(95))
forecast(M22TF, 10, level=c(95))
```

```

a22<-
c(3.647470,3.689277,3.700754,3.639341,3.732911,3.634248,3.709482,3.679410,3.65
5654 ,3.726340)
exp(a22)
eM22<-dados[47:56,1]
eM22
Erro22<-eM22-(exp(a22))
EQM22<-(1/length(Erro22))*sum(Erro22^2)
EQM22
MAPE22<-(1/length(eM22))*sum(abs(Erro22)/eM22)
MAPE22

```

#Verificação dos Resíduos (Ruido Branco)

```

par(mfrow=c(1,1))
hist(M12log$residuals) #Resíduos seguem distribuição normal assitótica a dir
ts.plot(M12log$residuals) #variância constante
lillie.test(M12log$residuals) #O teste de Lilliefors é uma adaptação do teste de
Kolmogorof-Smirnoff, usado para verificação de normalidade de um conjunto de
dados. A diferença básica entre os testes é que este testa a normalidade a partir de
média e desvio fornecidos, enquanto aquele usa a média e o desvio-padrão, calculados
no próprio conjunto de dados.O teste de Lilliefors é usado para verificar a aderência
dos dados a uma distribuição normal qualquer
M12log$residuals
acf(M12log$residuals, lag.max = NULL)
Box.test(M12log$residuals, type=c("Box-Pierce")) #Como  $X^2 < P$ -valor concluímos que
o Modelo 2 é adequado. Use a estatística Ljung-Box q para testar se uma série de
observações ao longo do tempo é aleatória e independente. Se as observações não
forem independentes, um aox pode ser correlacionada com outra observação k
unidades de tempo depois, uma relação chamada autocorrelação. A autocorrelação
pode reduzir a precisão de um modelo preditivo baseado no tempo, como um gráfico
de série temporal, e levar a uma interpretação incorreta dos dados.
cpgram(M12log$residuals)#Como M2 está no Intervalo de Confiança concluímos que o
Modelo 2 é adequado.

```

```

M12log=stats::arima(milho,order=c(1,1,2))
modelo12log<-forecast(M12log, 10, level=c(95))
modelo12log
resa<-modelo12log$fitted[1:28,1]
resa
resb<-modelo12log$fitted[29:56,1]
resb
var.test(resa,resb)

```

```

prev<-read.table("prev.txt", header=T)
prev

```

```

Previsao<-exp(prev)
Previsao
Previsao=ts(Previsao,start=2016,frequency=1)
Previsao

modelo12log

#previsão
plot(modelo12log, lwd=2, col="black", main="ARIMA(1,1,2)", xlab="Ano",
ylab="Produção")
ts.plot(dados,b,ylab="Produção",xlab="Tempo",lwd = 2,col=c("blue","red"))
legend(1970,80,c("Dados Originais", "Modelo"),col =c("blue","red"), pch=rep(20,2))
b
M12log

inf<-read.table("lwr.txt", header=T)
sup<-read.table("uwr.txt", header=T)
inf<-c(exp(inf))
sup<-c(exp(sup))
inf
sup

inf[1]=63.78449
inf[2]=65.47380
inf[3]=67.20919
inf[4]=68.99196
inf[5]=70.82273
inf[6]=72.70135
inf[7]=74.62832
inf[8]=76.60559
inf[9]=78.63289
inf[10]=80.71141
sup[1]=102.87970
sup[2]=108.30160
sup[3]=114.00360
sup[4]=120.00220
sup[5]=126.31390
sup[6]=132.95490
sup[7]=139.94510
sup[8]=147.30280
sup[9]=155.04880
sup[10]=163.20380

plot(T12,lwd=2,xlim=c(1961,2025),ylim=c(10,165))
dv<-seq(2017,2026,1)

```

```
tv<-Previsao

lines(dv,tv,col="blue")

polygon(c(dv[1],dv[1],dv[10],dv[10],dv[1]),c(inf[1],sup[1],sup[10],inf[10],inf[1]),col='gray90',border='gray90')
lines(dv,tv,col="blue",lwd=2)

abline(v =2016)
```