

DOUGLAS VINÍCIUS GONÇALVES ARAÚJO

**Análise de Séries Temporais: Aplicação do  
modelo ARIMA para previsão do Preço da Soja**

JI-PARANÁ

2022

DOUGLAS VINÍCIUS GONÇALVES ARAÚJO

**Análise de Séries Temporais: Aplicação do modelo  
ARIMA para previsão do Preço da Soja**

Relatório de Pesquisa na disciplina de Séries Temporais, do curso de Bacharel em Estatística da Universidade Federal de Rondônia, *campus* Ji-Paraná, como requisito de avaliação.

UNIVERSIDADE FEDERAL DE RONDÔNIA – UNIR  
DEPARTAMENTO DE MATEMÁTICA E ESTATÍSTICA  
RELATÓRIO DE PESQUISA

JÍ-PARANÁ

2022

*"Os livros servem para nos lembrar quanto somos estúpidos e tolos. São o guarda pretoriano de César, cochichando enquanto o desfile ruge pela avenida: Lembre-se, César, tu és mortal. A maioria de nós não pode sair correndo por aí, falar com todo mundo, conhecer todas as cidades do mundo, não temos tempo, dinheiro ou tantos amigos assim. As coisas que você está procurando, Montag, estão no mundo, mas a única possibilidade que o sujeito comum terá de ver noventa e nove por cento delas está num livro".*

*- Fahrenheit 451 de Ray Douglas Bradbury*

# Resumo

O objetivo deste trabalho tem como propor a análise do comportamento dos preços médios mensais da da *commodities* soja por meio da metodologia Box & Jenkins. Com isso foram utilizados .

**Palavras-chaves:** Modelo ARIMA, Preço da Soja, Séries Temporais.

# Lista de ilustrações

Figura 1 – Produção de Soja por continentes em 2020 . . . . .	9
Figura 2 – Produção de Soja por ano no Brasil de 1994 a 2020 . . . . .	10
Figura 3 – Família de trajetórias de processos estocásticos. . . . .	12
Figura 4 – Um processo estocástico interpretado como uma família de variáveis. .	12
Figura 5 – Indicador de preço da soja no período de outubro de 1997 até outubro de 2022 . . . . .	15

# Lista de tabelas

Tabela 1 – Estística Descritiva . . . . .	15
---	----

# Lista de abreviaturas e siglas

EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
CEPEA	Centro de Estudos Avançados em Economia Aplicada
CONAB	Companhia Nacional de Abastecimento
v.a.	Variável aleatória
fdp	Função densidade de probabilidade

# Lista de símbolos

$\Gamma$	Letra grega Gama
$\Lambda$	Lambda
$\zeta$	Letra grega minúscula zeta
$\in$	Pertence



# Sumário

1	INTRODUÇÃO . . . . .	9
2	SÉRIES TEMPORAIS . . . . .	11
2.1	Processos Estocásticos . . . . .	11
2.2	Processo Estocástico Estacionário . . . . .	13
2.3	Decomposição de um Série Temporal . . . . .	13
2.4	Modelos ARIMA . . . . .	13
3	METODOLOGIA . . . . .	14
4	RESULTADOS E DISCUSSÕES . . . . .	15
5	CONSIDERAÇÕES FINAIS . . . . .	17
	REFERÊNCIAS . . . . .	18
	APÊNDICES . . . . .	19
	APÊNDICE A – SCRIPT R . . . . .	20

# 1 Introdução

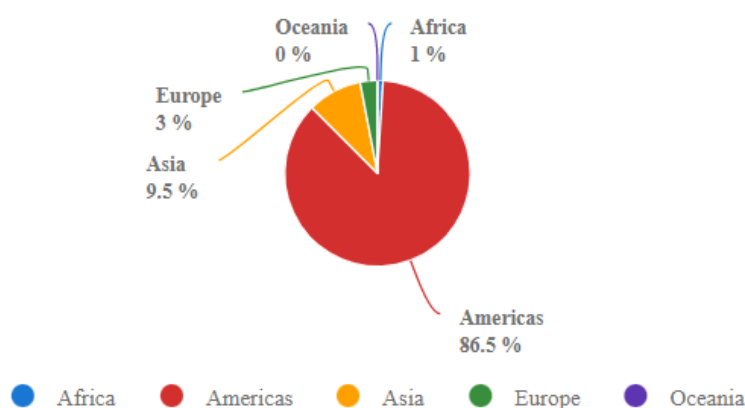
A soja derivou na costa leste da Ásia, nas proximidades do rio Yangtse na China. Antigamente a soja eram plantas rasteiras, bem distinta de hoje que cultivamos que foram modificadas e melhoradas pelo chineses. Apesar de ser conhecida e consumida pela civilização oriental por milhares de anos, ocorreu a sua introdução na Europa no final do século XV.

No Brasil, na década de 60, dois fatores internos que fizeram a produção de soja expandir, fatores estes que influenciaram o cenário mundial da produção de grãos: *i*) com o esforço da produção de suínos e aves assim gerava demanda pelo farelo de soja, já era uma necessidade estratégica; *ii*) na década de 70 houve a explosão do preço da soja no mercado mundial e os agricultores e o governo beneficiavam de uma vantagem competitiva em relação de outros países, pois o escoamento da safra brasileira ocorre na entressafra americana, quando os preços atingem as maiores cotações, em conformidade com (EMBRAPA, 2022).

A produção de soja avançou significativamente no mundo nestes últimos 20 anos, com taxas de crescimento de 3,7% ao ano. Os três maiores produtores de soja mundial, juntos, respondem por mais de 80% da produção, que são Brasil, Estados Unidos (EUA) e Argentina. E destes, o país que mais apresentou taxa de crescimento da produção nessas últimas décadas é o Brasil, com aumento de 5,9% a. a., superior entre os seus concorrentes (EUA avançou 2,7% a. a. e a Argentina 1,6% a. a.), segundo (CEPEA, 2022).

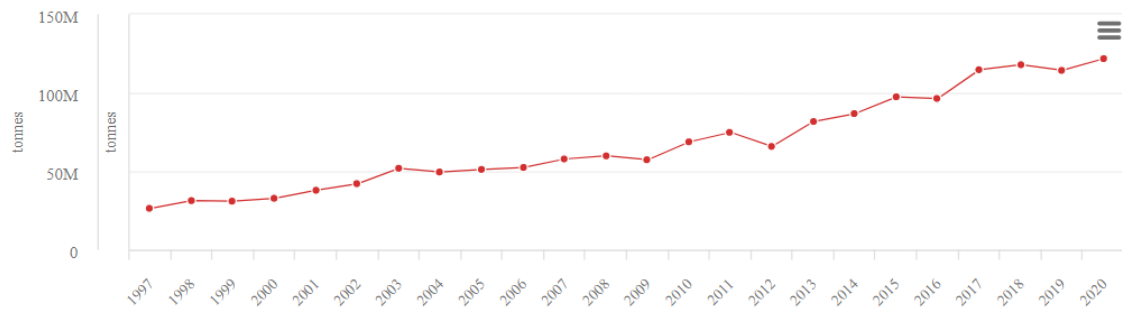
Neste mesmo período o Brasil deixou o posto de segundo lugar para ser o maior

Figura 1 – Produção de Soja por continentes em 2020



Fonte: (FAOSTAT, 2022)

Figura 2 – Produção de Soja por ano no Brasil de 1994 a 2020



Fonte: ([FAOSTAT](#), 2022)

produtor de soja - na safra de 2001/2002, a produção brasileira era aproximadamente 52 milhões de toneladas, nesta mesma safra o EUA era aproximadamente 75 milhões de toneladas, agora na safra 2021/2022 o Brasil alcançou perto de 122 milhões toneladas, superior ao EUA que entorno de 120,7 milhões de toneladas, conforme dados da ([CONAB](#), 2022).

Diante do exposto deste trabalho, pretende à aplicação de modelos **ARIMA** (Autoregressivos Integrados de Médias Móveis) nos índices de preços da *commodities* soja com a finalidade de compreender o comportamento da comercialização deste produto.

## 2 Séries Temporais

Qualquer conjunto de observações ordenadas no tempo, é uma série temporal. Por exemplos:

- i)* temperaturas médias diárias de uma cidade;
- ii)* vendas mensais de uma empresa;
- iii)* valores de fechamento diários da IBOVESPA;
- iv)* preços diários de *commodities*;
- v)* valores mensais do IPCA (Índice Nacional de Preços ao Consumidor Amplo).

As séries temporais podem ser tanto discreta quanto contínua, ou seja, discreta é o intervalo entre as observações pertencentes a um conjunto discreto, e contínua é o intervalo entre as observações pertencentes a um conjunto contínuo. Observa-se que quando dizemos que uma série é discreta, estamos fazendo referência ao tempo entre as observações e não a escala da variável.

Além disso, temos dois enfoques usados na análise de séries temporais. O objetivo de ambos é construir modelos para as séries. No primeiro enfoque é feita análise no *domínio temporal* e os modelos sugeridos são *modelos paramétricos* (números de parâmetros finitos), o segundo é conduzido no *domínio de frequências* e os propostos são *modelos não-paramétricos* (MORETTIN; TOLOI, 2006). O modelo utilizado neste estudo, **ARIMA** é um modelo paramétrico.

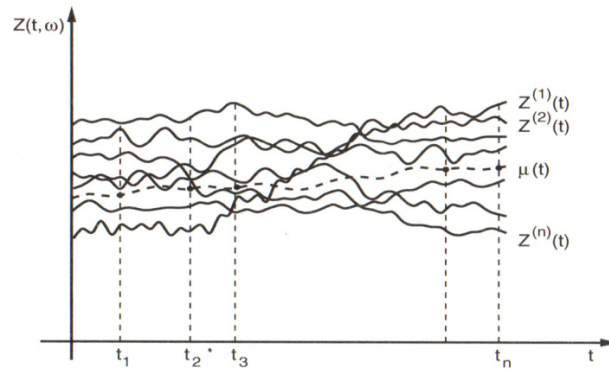
### 2.1 Processos Estocásticos

Podemos definir um processo estocástico como um conjunto cronológico de observações de um determinado fenômeno de forma que o seu comportamento pode ser descrito por uma ou mais distribuições de probabilidade. Conforme (MORETTIN; TOLOI, 2006),

**Definição 2.1.1.** Seja  $T$  um conjunto arbitrário, um processo estocástico é uma família de variáveis aleatórias  $\{Z(t), t \in T\}$ , tal que,  $\forall t \in T$ ,  $Z(t)$  é uma variável aleatória.

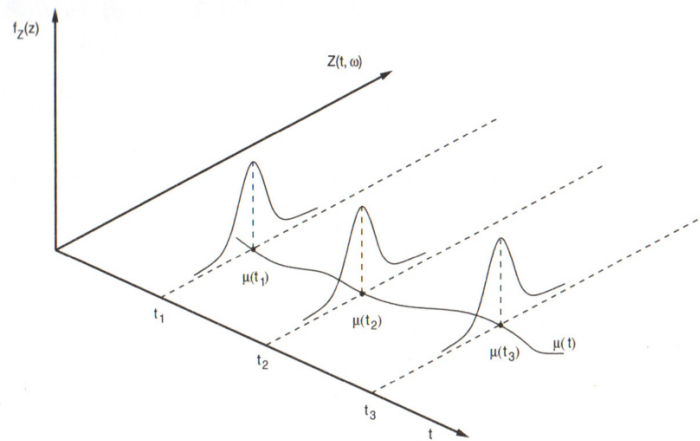
Neste caso, um processo estocástico é uma família de variáveis aleatórias, definidas num mesmo espaço de probabilidade  $(\Omega, \mathcal{A}, \mathcal{P})$ , tendo o conjunto  $T$ , normalmente tomado como conjunto dos inteiros  $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$  ou conjunto dos reais  $\mathbb{R}$ .

Figura 3 – Família de trajetórias de processos estocásticos.



Fonte: (MORETTIN; TOLOI, 2006)

Figura 4 – Um processo estocástico interpretado como uma família de variáveis.



Fonte: (MORETTIN; TOLOI, 2006)

A Figura 4 ilustra uma interpretação de um processo estocástico, ou seja, para  $t \in T$ ,  $Z(t)$  é uma variável aleatória definida em  $\Omega$ , e  $Z(t)$  é uma função de dois argumentos  $Z(t, \omega)$ ,  $t \in T$ ,  $\omega \in \Omega$ . Na figura, cada  $t \in T$ , temos uma v.a.  $Z(t, \omega)$ , com uma distribuição de probabilidade, é possível que a função densidade de probabilidade (fdp) no instante  $t_1$  seja distinta da fdp no instante  $t_2$ , entre os instante  $t_1$  e  $t_2$  quaisquer, mas usualmente é aquela em que a fdp de  $Z(t, \omega)$  é a mesma,  $\forall t \in T$ .

## 2.2 Processo Estocástico Estacionário

## 2.3 Decomposição de um Série Temporal

## 2.4 Modelos ARIMA

## 3 Metodologia

Por

Para obtenção das análises estatísticas dos dados, fez-se o uso do *software* **R** versão 4.2.1

A

O índice de Preços da *commodities* Soja

## 4 Resultados e Discussões

A Tabela 1 apresenta as principais medidas descritivas obtidas a partir da série mensal condizente ao indicador de preços em reais da soja (reais por saca de 60 kg) compreendido no período de outubro de 1997 a outubro de 2022, proporcionando uma visão panorâmica do comportamento da série em estudo.

Tabela 1 – Estatística Descritiva

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
13.87	31.34	46.8	58.63	72.91	195.8

A Figura 5 apresenta a série original do indicador de preço em reais da soja no intervalo de outubro de 1997 a outubro de 2022

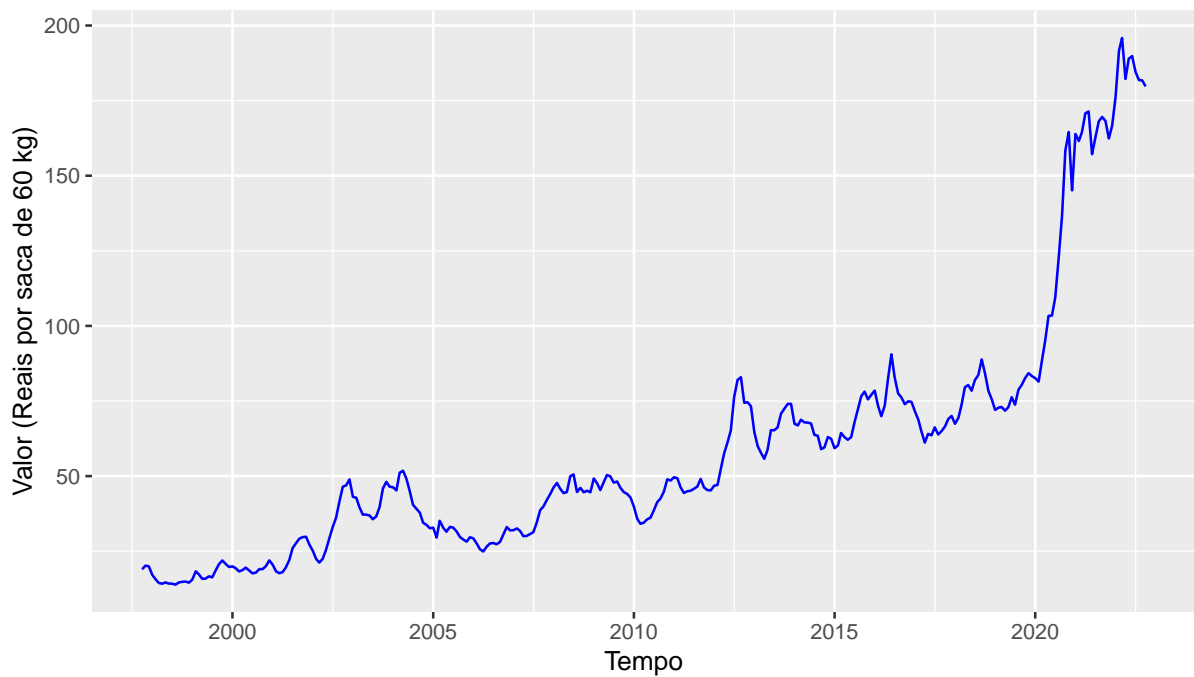
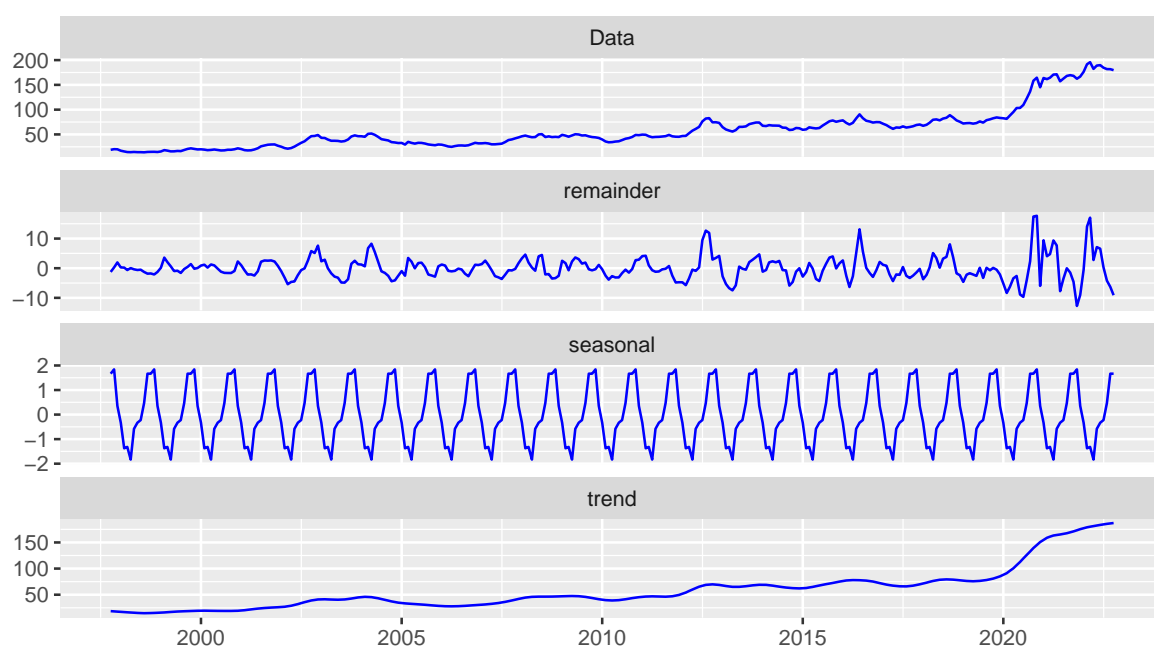


Figura 5 – Indicador de preço da soja no período de outubro de 1997 até outubro de 2022





## 5 Considerações Finais

O presente trabalho teve como objetivo estudar o comportamento do preço da Soja no Brasil, e testar a metodologia Box & Jenkins para previsão.

## Referências

CEPEA. 2022. Acesso em 26 de novembro de 2022. Disponível em: <<https://www.cepea.esalq.usp.br/br>>. Citado na página 9.

CONAB. 2022. Acesso em 26 de novembro de 2022. Disponível em: <<https://www.conab.gov.br/busca?searchword=soja&searchphrase=all>>. Citado na página 10.

EMBRAPA. 2022. Acesso em 26 de novembro de 2022. Disponível em: <<https://www.embrapa.br/soja>>. Citado na página 9.

FAOSTAT. 2022. Acesso em 26 de novembro de 2022. Disponível em: <[www.fao.org/faostat/en/#data/QCL/visualize](http://www.fao.org/faostat/en/#data/QCL/visualize)>. Citado 2 vezes nas páginas 9 e 10.

MORETTIN, P.; TOLOI, C. Análise de séries temporais 2ª edição revista e ampliada. *ABE-Projeto Fisher*, Editora Edgar Blücher, 2006. Citado 2 vezes nas páginas 11 e 12.

## Apêndices

# APÊNDICE A – SCRIPT R

```
library(forecast) #Ferramentas para exibir e analisar previsões de séries temporais univariadas.
library(timeSeries) #Ferramentas para séries temporais financeiras.
library(urca) #Testes de raiz unitária e de cointegração.
library(randtests) #Testes de aleatoriedade não paramétricos para sequências numéricas.
library(tseries) #Análise de séries temporais e finanças computacionais.
library(stats) #Funções para cálculos estatísticos e geração de números aleatórios.
library(readxl) #Importe arquivos Excel para R.
library(tidyverse) #Conjunto de pacotes que funcionam em harmonia para data analysis.
library(ggfortify) #Ferramentas de plotagem unificadas.
library(magrittr) #Fornece o encadeamento de comandos, pipeline (%>%).
library(kableExtra)

soja <- read_xls("../dataset/soja.xls") #Importar o conjunto de dados.

soja <- ts(soja$Preço, start = c(1997,10), frequency = 12) #Criar o objeto de série temporal.

class(soja) #Retornar os valores do atributo de classe de um objeto.
as.ts(soja) #Testar se o objeto é um série temporal.

###Informações descritiva da série (resumo).
summary(soja)

###Plotagem da série.
autoplot(soja, xlab = "Tempo", ylab = "Valor (Reais por saca de 60 kg)", ts.colour = 'blue')

###Decomposição da série e plotagem.
soja %>%
  stl(, s.window = "periodic") %>%
  autoplot(, ts.colour = "blue")
```