



**UNIVERSIDADE ESTADUAL PAULISTA**  
**“JÚLIO DE MESQUITA FILHO”**  
Câmpus de Presidente Prudente

**THAI CÉU SANTOS**

**ANÁLISE DE CICLOS DA EXPORTAÇÃO DE CAFÉ NO BRASIL VIA MODELOS  
DE DECOMPOSIÇÃO DE SÉRIES TEMPORAIS E ANÁLISE ESPECTRAL  
ECONOMÉTRICA**

Revisado pelo Orientador Prof. Dr.  
Manoel Ivanildo Silvestre Bezerra

\_\_\_\_\_  
Assinatura do Orientador

Data \_\_\_\_/ \_\_\_\_/ 2024

**PRESIDENTE PRUDENTE**

**2024**

**THAIL CÉU SANTOS**

**ANÁLISE DE CICLOS DA EXPORTAÇÃO DE CAFÉ NO BRASIL VIA MODELOS  
DE DECOMPOSIÇÃO DE SÉRIES TEMPORAIS E ANÁLISE ESPECTRAL  
ECONOMÉTRICA**

Relatório Parcial para Trabalho de Conclusão  
de Curso apresentado ao Curso de  
Graduação em Estatística da FCT/Unesp para  
aproveitamento na disciplina TCC 1.

Orientador(a): Prof. Dr. Manoel Ivanildo  
Silvestre Bezerra

**PRESIDENTE PRUDENTE**

**2024**

## RESUMO

O café, como commodity de grande relevância econômica global, é cultivado em mais de 60 países em desenvolvimento e exportado principalmente para mercados desenvolvidos, desempenhando papel crucial nas trocas comerciais internacionais. A competitividade no mercado de café intensificou-se desde os anos 80, demandando maior gestão do conhecimento em sua cadeia produtiva. Dados recentes do CECAFÉ mostram variações nas exportações brasileiras, destacando sua importância econômica. No contexto da análise de séries temporais, o estudo utiliza ferramentas e as séries harmônicas e decomposição temporal, para decompor e compreender padrões nas exportações dos quatro tipos de café brasileiros (Conilon, Arábica, Torrado e Solúvel). Realizamos uma análise descritiva das variáveis Volume, Receita e Preço Médio dos quatro tipos de café (Arábica, Conillon, Solúvel e Torrado), observando métricas estatísticas, gráficos de boxplot, séries temporais e sazonalidade. Aplicamos o teste de estacionaridade de Dickey-Fuller e ajustamos os modelos de Decomposição Temporal e Harmônico para a série de Volume do Arábica, avaliando suas métricas de desempenho (RMSE, MAE,  $R^2$ ) e analisando os resíduos. Identificamos padrões anuais e mensais, variações específicas por tipo de café e comportamentos sazonais. Por fim, comparamos os modelos quanto à precisão e capacidade explicativa para entender as tendências e variações nos dados, e concluímos que os modelos de Decomposição Temporal e Harmônico têm vantagens distintas, com de Decomposição Temporal sendo mais preciso e o Harmônico melhor explicativo. Ambas as abordagens são úteis dependendo do objetivo da análise, seja previsão ou compreensão de variabilidade e tendências.

**Palavras-chave:** Café; Commodity; Exportações; Cadeia produtiva; Conilon; Arábica; Torrado; Solúvel; Séries temporais; Modelos de Decomposição Temporal e Harmônico; Sazonalidade; Tendências.

## ABSTRACT

The coffee, as a commodity of significant global economic relevance, is cultivated in over 60 developing countries and primarily exported to developed markets, playing a crucial role in international trade. Competitiveness in the coffee market has intensified since the 1980s, demanding greater knowledge management in its production chain. Recent data from CECAFÉ shows variations in Brazilian coffee exports, highlighting their economic importance. In the context of time series analysis, this study employs tools such as harmonic models and temporal decomposition to break down and understand patterns in the exports of the four main types of Brazilian coffee (Conilon, Arabica, Roasted, and Soluble). We conducted a descriptive analysis of the variables Volume, Revenue, and Average Price for the four types of coffee (Arabica, Conilon, Soluble, and Roasted), examining statistical metrics, boxplot graphs, time series, and seasonality. The Dickey-Fuller stationarity test was applied, followed by fitting Temporal Decomposition and Harmonic models to the Volume series of Arabica coffee. Model performance metrics (RMSE, MAE,  $R^2$ ) were evaluated, and residuals were analyzed. Annual and monthly patterns, specific variations by coffee type, and seasonal behaviors were identified. Finally, we compared the models in terms of accuracy and explanatory capacity to understand trends and variations in the data. The study concluded that Temporal Decomposition and Harmonic models have distinct advantages, with Temporal Decomposition being more accurate and the Harmonic model offering better explanatory insights. Both approaches are valuable depending on the analysis goal, whether for forecasting or understanding variability and trends.

**Keywords:** Coffee; Commodity; Exports; Supply Chain; Conilon; Arabica; Roasted; Soluble; Time Series; Temporal Decomposition and Harmonic Models; Seasonality; Trends.

## Sumário

1 INTRODUÇÃO .....	1
2 METODOLOGIA.....	3
2.1 Modelo de Regressão Linear Múltipla .....	4
2.3 Decomposição de Modelos de Séries Temporais .....	8
2.3.1 Teste de Dickey-Fuller.....	10
2.3.2 Modelos de Decomposição com Harmônicas .....	12
2.4 Análise de Resíduos .....	14
2.4.1 Teste de Shapiro-Wilk .....	15
2.4.2 Teste de Durbin-Watson.....	16
3 RESULTADOS PARCIAIS .....	17
3.1 Análise Descritiva das variáveis a ser estudada .....	17
3.2 Modelos para a variável Volume (Arábica) .....	34
3.3 Análise de resíduos e adequabilidade dos modelos .....	37
4 CONSIDERAÇÕES PARCIAIS .....	42
REFERÊNCIAS .....	44

## 1 INTRODUÇÃO

A definição de "commodity" refere-se a produtos primários, frequentemente exportados, com baixo grau de industrialização e produzidos em grandes volumes. Esses produtos são reconhecidos por sua dinâmica nas atividades comerciais e desempenham um papel fundamental nas principais trocas comerciais entre os países (Topik; Marichal; Zephyr, 2006).

O mercado de café tem passado por uma crescente competitividade desde a década de 80, e aprimorar a gestão do conhecimento em toda a sua cadeia produtiva se tornou uma prioridade crucial tanto para a indústria quanto para o mercado. O café é um dos principais setores da indústria global de alimentos e bebidas, destacando-se como um dos commodities mais amplamente negociados no cenário internacional (Trauer et al., 2017).

Como um dos principais commodities no comércio global, o café é cultivado em mais de 60 países em desenvolvimento e tem seu principal mercado nos países desenvolvidos, onde pelo menos 14 desses países dependem em mais de 10% de suas receitas provenientes das exportações de café. A Organização Mundial do Consumidor enfatiza que aproximadamente 25 milhões de pessoas dependem diretamente da cafeicultura para sustento, enquanto um número expressivo de cerca de 100 milhões de pessoas estão envolvidas em vários estágios da cadeia produtiva do café, enfatizando assim a importância do desenvolvimento sustentável desse mercado (Trauer et al., 2017).

De acordo com a Sociedade Nacional da Agricultura (2021), em fevereiro deste ano, o Brasil exportou 3,3 milhões de sacas de café, abrangendo todos os tipos de café, como o torrado e moído, café verde e solúvel. Isso representou um aumento de 9% em relação ao mesmo mês do ano de 2020, gerando uma receita cambial de US\$ 423,7 milhões, um acréscimo de 4,70% em relação ao mesmo mês no ano de 2020. Na conversão para a moeda local, esse valor atingiu R\$ 2,3 bilhões, apresentando um aumento de 30,60%. Entretanto, em setembro de 2023, o Brasil exportou 3,294 milhões de sacas de 60 kg de todos os tipos de café, indicando uma queda de 5,3% em relação à 3,480 milhões registradas no mesmo mês de 2022. Em termos de receita, houve uma redução de 23,2% no mesmo período comparativo, resultando em

um valor de US\$ 638 milhões. Esses dados foram retirados do relatório elaborado pelo Conselho dos Exportadores de Café do Brasil (CECAFÉ).

A econometria de séries temporais está profundamente interligada a uma ampla gama de desafios nos domínios econômicos e financeiros. Aqueles que não têm acesso a esse conjunto de ferramentas enfrentam dificuldades consideráveis ao tentar realizar análises econométricas em diversas áreas da economia (Bueno, 2008). Segundo Hoffman (2016), uma série temporal é definida como um conjunto de valores de uma variável dispostos de forma sequencial no tempo. Um exemplo prático disso pode ser encontrado na série dos preços médios mensais de milho em um mercado específico. Isso nos ajuda a perceber como está nossos dados em um determinado espaço de tempo. Somente observar isso não nos ajuda em muita coisa, sendo assim, existem outras ferramentas que nos ajudam a entender melhor nossos dados distribuídos no tempo.

A análise espectral é uma técnica de análise de séries temporais que tem como objetivo identificar as frequências dominantes em uma série e encontrar uma explicação para o sistema a partir do qual as medições foram derivadas (Shumway et al., 2011). A análise espectral envolve a decomposição de uma série temporal em suas componentes de frequência, permitindo a identificação de padrões e tendências periódicas. Para isso, temos algumas ferramentas que nos ajudam a fazer essa decomposição, sendo uma delas a Série Harmônica (Série de Fourier). Aguirre (1995) explica que a série harmônica é uma soma de senoides (senos e cossenos), sendo eles harmonicamente relacionados, e a outra o modelo de Decomposição Temporal. Para se fazer isso, a série precisa ser estacionária (manter uma variância e uma média constante).

Dessa forma, o presente estudo se propõe a compreender, com o suporte dos métodos estatísticos, onde podemos modelar e decompor em componentes independentes e interpretáveis, o comportamento dos quatro tipos de café (Conilon, Arábica, Torrado e Solúvel) exportados pelo Brasil. Através da aplicação da técnica de análise espectral, que será feita no próximo TCC II, iremos detectar a periodicidade da série temporal, o objetivo é observar e explicar as variações ao longo do tempo com base em fatores históricos.

## 2 METODOLOGIA

Com o intuito de entender melhor as exportações de café que o Brasil fez, de janeiro de 2015 até dezembro de 2022, iremos usar os dados fornecidos pelo relatório completo mensal, feito pelo CECAFÉ (Conselho dos Exportadores de Café do Brasil), onde contém os 4 tipos de café exportado (Conilon, Arábica, Torrado e Solúvel), sendo que, para cada um deles, contém o Preço Médio em cada mês e ano (US\$), o Volume exportado em cada mês e ano e o Preço Cambial (US\$) , que seria o valor total de dinheiro que o Brasil obteve nas exportações em cada mês e ano.

Para isso, utilizaremos o Software R, onde serão feitas as análises estatísticas para entender o que se passou, concluindo se teve alguma consequência negativa ou positiva nesse período.

Para cumprirmos nosso objetivo, faremos uma modelagem utilizando regressão de séries temporais para cada tipo de café, do preço médio, volume exportado e preço cambial, e com a ajuda da análise espectral, iremos avaliar o espectro de cada série feita, medindo a importância relativa de cada faixa de frequência. Para utilizar a análise espectral, devemos tentar desvendar se as séries são estacionárias, verificando se necessita de alguma transformação nas séries. Usaremos as séries harmônicas e as de decomposição temporal para chegar no objetivo.



## 2.1 Modelo de Regressão Linear Múltipla

Uma série temporal consiste em um conjunto de observações organizadas cronologicamente. Isso permite que possamos ajustar um modelo de regressão para analisar os dados. Quando lidamos com duas ou mais variáveis correlacionadas e desejamos entender como uma afeta a outra, podemos utilizar a regressão linear para estimar os valores da variável de interesse (chamada de variável resposta) a partir das outras variáveis (conhecidas como variáveis explicativas). Esse método nos ajuda a compreender de que forma as variáveis explicativas influenciam a variável resposta.

No modelo de regressão múltipla, supondo uma variável resposta  $Y$ , que é afetada por duas ou mais variáveis explicativas. Assim, o modelo para  $k$  variáveis explicativas podem ser representadas por:

$$y_n = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \cdots + \beta_k X_{kj} + \varepsilon_j. \quad (2.1)$$

Sendo  $j = 1, 2, \dots, n$

Onde:

- $\beta_0$  é o intercepto.
- $\beta_1, \beta_2, \dots, \beta_k$  são os coeficientes das variáveis explicativas  $X_{1j}, X_{2j}, \dots, X_{kj}$ , respectivamente.
- $\varepsilon_j$  é o erro aleatório, que deve atender a alguns pressupostos importantes: ele precisa ser independente e identicamente distribuído, com  $E(\varepsilon) = 0$  e variância constante desconhecida  $\sigma^2$ .

Neste modelo, cada  $\beta_i$  (para  $i = 1, 2, \dots, k$ ) representa o efeito da variável explicativa  $X_i$  sobre a variável resposta  $Y$ , enquanto o erro aleatório  $\varepsilon$  captura a variação não explicada pelo modelo. Em forma matricial, ficaria:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \quad (2.2)$$

Onde:

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{k1} & x_{n2} & \cdots & x_{kj} \end{bmatrix}; \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}; \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_j \end{bmatrix}.$$

Para estimar  $\hat{\boldsymbol{\beta}}$ , podemos usar o método dos mínimos quadrados, que minimiza a soma dos quadrados das diferenças entre os valores observados  $\mathbf{Y}$  e os valores previstos pelo modelo. As estimativas de  $\hat{\boldsymbol{\beta}}$  são obtidas nos pontos que maximizam a função de mínimos quadrados:

$$Q = \sum_{j=1}^n \varepsilon_j^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \quad (2.3)$$

que, desenvolvendo, encontraremos a diferencial de  $Q$  quando identicamente nula:

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{Y}. \quad (2.4)$$

Para resolver as equações normais, multiplicamos ambos os lados pela inversa de  $\mathbf{X}'\mathbf{X}$ . Assim, o estimador de mínimos quadrados para  $\boldsymbol{\beta}$  é dado por:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}. \quad (2.5)$$

Teremos, então, uma matriz coluna contendo as estimativas. Estimando os parâmetros, podemos calcular os resíduos do nosso modelo, que seria o quanto meu modelo está errando em cada ponto:

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \hat{\mathbf{Y}}. \quad (2.6)$$

A partir dessas estimativas e das matrizes construídas, montaremos a tabela de análise de variância para avaliar a importância dos parâmetros no modelo. As medidas serão calculadas da seguinte forma:

- **Soma de Quadrados Total (SQTotal):** Representa a variação total dos valores de  $Y$  em relação à média  $\bar{y}$ . Calculada por:

$$SQTotal = Y'Y - \frac{(\sum_{j=1}^n y_j)^2}{n}. \quad (2.7)$$

- **Soma de Quadrados da Regressão (SQReg):** Representa a variação explicada pelo modelo de regressão, ou seja, a distância quadrática entre os valores ajustados  $\hat{Y}$  e a média  $\bar{y}$ . Calculada por:

$$SQReg = \hat{\beta}'X'Y - \frac{(\sum_{j=1}^n y_j)^2}{n}. \quad (2.8)$$

- **Soma de Quadrados dos Resíduos (SQRes):** Representa a variação não explicada pelo modelo, ou seja, a distância quadrática entre os valores observados  $Y$  e os valores ajustados  $\hat{Y}$ . Calculada por:

$$SQRes = \epsilon'\epsilon = Y'Y - \hat{\beta}'X'Y. \quad (2.9)$$

Logo:

Tabela 2 - Análise de variância (ANOVA)

Fonte de variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio	$F_0$	P-valor
Regressão	$k = p - 1$	SQReg	$QMReg = \frac{SQReg}{p - 1}$	$\frac{QMReg}{QMRes}$	$P(F \geq F_0)$
Resíduo	$n - p$	SQRes	$QMRes = \frac{SQRes}{n - p}$		
Total	$n - 1$	SQTotal			

Sabemos que adicionar uma variável ao modelo provoca um aumento na soma de quadrados da regressão e uma diminuição na soma de quadrados dos resíduos. No entanto, é importante ter cuidado com essa prática, pois pode diminuir a utilidade

do modelo. Para testar a significância dos parâmetros de forma individual, utilizamos o teste:

- **Hipóteses:**

$$H_0: \beta_j = 0, \text{ para } j = 1, 2, \dots, p$$

$$H_1: \text{Pelo menos um é diferente de } 0$$

Se rejeitarmos a hipótese nula, então o parâmetro é considerado significativo. O teste para essa hipótese é dado por:

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}},$$

onde  $C_{jj}$  é o elemento da diagonal de  $(\mathbf{X}'\mathbf{X})^{-1}$  correspondente a  $\hat{\beta}_j$ . Rejeitamos a hipótese nula  $H_0$  se  $|t_0| > t_{(\frac{\alpha}{2}, n-p)}$ , onde  $t_{(\frac{\alpha}{2}, n-p)}$  é o valor crítico da distribuição  $t$  de Student com  $n - p$  graus de liberdade.

Além disso, podemos construir intervalos de confiança para os parâmetros. Assumimos que cada estatística:

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 C_{jj}}} \quad \text{para } j = 1, 2, \dots, p,$$

segue uma distribuição  $t$  de Student com  $n - p$  graus de liberdade. Com base nisso, podemos construir um intervalo de confiança de  $100(1 - \alpha)\%$  para os coeficientes da regressão  $\beta_j$  como:

$$P \left( \hat{\beta}_j - t_{(\frac{\alpha}{2}, n-p)} \sqrt{\hat{\sigma}^2 C_{jj}} \leq \beta_j \leq \hat{\beta}_j + t_{(\frac{\alpha}{2}, n-p)} \sqrt{\hat{\sigma}^2 C_{jj}} \right) = 1 - \alpha. \quad (2.9)$$

### 2.3 Decomposição de Modelos de Séries Temporais

Neste caso, assumimos que nossa variável resposta é uma série temporal  $Y_t$ , para  $t = 1, 2, \dots, n$ , que varia ao longo do tempo. Podemos descrever essa relação utilizando um modelo de regressão linear simples:

$$Y_t = \beta_0 + \beta_1 t + \varepsilon_t. \quad (2.10)$$

onde  $\varepsilon_t$  é um componente aleatório, que tem média igual a zero, variância constante, e não apresenta covariância entre valores de diferentes períodos, ou seja:

$$\begin{aligned} E(\varepsilon_t) &= 0; \\ \text{Var}(\varepsilon_t) &= \sigma_a^2; \\ \gamma_k = \text{cov}(\varepsilon_t, \varepsilon_{t+k}) &= 0, \text{ para } k \neq 0. \end{aligned}$$

O modelo de decomposição de uma série temporal é dividido em três componentes: tendência ( $D_t$ ), variação cíclica sazonal ( $S_t$ ), e parte aleatória ( $\varepsilon_t$ ):

$$Y_t = D_t + S_t + \varepsilon_t \quad (2.11)$$

Para ajustar um modelo a uma série temporal  $Z_t$ , podemos usar uma regressão linear. Para modelar a tendência  $D_t$ , uma das formas é utilizar um polinômio:

$$D_t = \beta_0 + \beta_1 t + \dots + \beta_m t^m. \quad (2.12)$$

Onde o grau  $m$  do polinômio é significativamente menor que o número de observações  $N$ . Para estimar os parâmetros  $\beta_j$ , utilizamos o método dos mínimos quadrados, com o objetivo de minimizar a função:

$$f(\beta_0, \dots, \beta_m) = \sum_{t=1}^N (Y_t - \beta_0 - \beta_1 t - \dots - \beta_m t^m)^2. \quad (2.13)$$

Assim, obtemos os estimadores de mínimos quadrados usuais  $\widehat{\beta}_0, \widehat{\beta}_1, \dots, \widehat{\beta}_m$ .

Para a sazonalidade  $S_t$ , utilizamos a forma:

$$S_t = \sum_{j=1}^{12} \alpha_j d_{j,t} \quad (2.14)$$

Onde  $d_{j,t}$  são variáveis “dummy”. Neste caso, podemos utilizar senos, cossenos juntamente com a “dummy”.

Em forma matricial, podemos obter os estimadores na seguinte forma:

$$Y = \beta T + D\alpha + \varepsilon, \quad (2.15)$$

onde:

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_t \end{bmatrix}; \mathbf{T} = \begin{bmatrix} 1 & 1 & \dots & \dots & 1^m \\ 1 & 2 & \dots & \dots & 2^m \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & t & \dots & \dots & t^m \end{bmatrix}; \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix};$$

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_j \end{bmatrix}; \mathbf{D} = \begin{bmatrix} 1 & 0 & 0 & \dots & d_{11,1} \\ 0 & 1 & 0 & \dots & d_{11,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{1,t} & d_{2,t} & \dots & \dots & d_{11,t} \end{bmatrix}; \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_t \end{bmatrix}.$$

Sendo que a equação (2.10) pode ser escrita como:

$$Y = \beta T + \varepsilon, \quad (2.16)$$

onde:

$$\mathbf{X} = [\mathbf{T} \quad \mathbf{D}] \text{ e } \hat{\boldsymbol{\gamma}} = \begin{bmatrix} \beta \\ \alpha \end{bmatrix}. \quad (2.17)$$

Sendo:

$$\hat{\boldsymbol{\gamma}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}. \quad (2.18)$$

### 2.3.1 Teste de Dickey-Fuller

Uma série temporal pode ser estacionária ou não, o que significa que ela pode variar em torno de uma média constante ao longo do tempo. Verificar essa condição é crucial para a construção de modelos. O teste de Dickey-Fuller (Morettin e Tolo, 2004) é utilizado para determinar se uma série é estacionária ou possui uma raiz unitária.

Considerando uma série do tipo:

$$y_t = \beta_0 + \beta_1 t + \rho y_{t-1} + \varepsilon_t. \quad (2.19)$$

Ao subtrairmos  $y_{t-1}$  de ambos os lados da equação, obtemos:

$$\Delta y_t = \beta_0 + \beta_1 t + \lambda y_{t-1} + \varepsilon_t, \quad (2.20)$$

onde  $\lambda = \rho - 1$  e  $\varepsilon_t$  é o ruído branco. As hipóteses para o teste são:

- **Hipótese nula ( $H_0$ ):**  $\lambda = 0$
- **Hipótese alternativa ( $H_1$ ):**  $-2 < \lambda < 0$

Para o teste de Dickey-Fuller, temos três opções de aplicação, mas para o nosso objetivo, consideraremos apenas uma:

- **Hipótese nula:** O modelo tem uma tendência determinística linear, uma tendência estocástica e uma raiz unitária.

$$\Delta y_t = \beta_0 + \beta_1 t + \varepsilon_t. \quad (2.21)$$

- **Hipótese alternativa:** O processo não possui raiz unitária.

$$\Delta y_t = \beta_0 + \beta_1 t + \lambda y_{t-1} + \varepsilon_t. \quad (2.22)$$

A estatística do teste é dada por:

$$\tau = \frac{\hat{\lambda}}{s_{\hat{\lambda}}}. \quad (2.23)$$

Onde  $s_{\hat{\lambda}}$  é o erro padrão de  $\hat{\lambda}$ . Os valores críticos do teste estão na tabela 3:

Tabela 3 - Valores críticos do teste de Dickey – Fuller

Modelo	Nível de significância		
	10%	5%	1%
$\Delta y_t = \lambda y_{t-1} + a_t$	-1,62	-1,94	-2,56
$\Delta y_t = \alpha + \beta_1 t + \lambda y_{t-1} + a_t$	-3,13	-3,41	-3,96

Autor: James G. MacKinnon



### 2.3.2 Modelos de Decomposição com Harmônicas

Para modelar a sazonalidade de uma série temporal, podemos usar funções periódicas, como senos e cossenos, que ajudam a ajustar a curva à periodicidade da série.

Considerando o modelo de regressão matricial, com  $y$  representando o vetor de observações e uma matriz de delineamento  $X = [T \mid H]$ , onde:

$$T = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & t \\ \vdots & \vdots \\ 1 & N \end{bmatrix},$$

$$H = \begin{bmatrix} \cos(\omega_1) & \cos(\omega_2) & \dots & \cos(\omega_6) & \sin(\omega_1) & \dots & \sin(\omega_6) \\ \cos(\omega_1 2) & \cos(\omega_2 2) & \dots & \cos(\omega_6 2) & \sin(\omega_1 2) & \dots & \sin(\omega_6 2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \cos(\omega_1 t) & \cos(\omega_2 t) & \dots & \cos(\omega_6 t) & \sin(\omega_1 t) & \dots & \sin(\omega_6 t) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \cos(\omega_1 N) & \cos(\omega_2 N) & \dots & \cos(\omega_6 N) & \sin(\omega_1 N) & \dots & \sin(\omega_6 N) \end{bmatrix},$$

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}, \text{ e } \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \alpha_1 \\ \vdots \\ \alpha_6 \\ \gamma_1 \\ \gamma_6 \end{bmatrix} = \begin{array}{l} \text{Intercepto} \\ \text{Tendência Linear} \\ \cos(\omega_1 t) \\ \vdots \\ \cos(\omega_6 t) \\ \sin(\omega_1 t) \\ \sin(\omega_6 t) \end{array}$$

Onde  $\omega_k = \frac{2\pi tk}{12}$  para  $k = 1, \dots, 6$  e  $t = 1, \dots, N$ .

O modelo de regressão com harmônicas é:

$$Y_t = \beta_0 + \beta_1 t + \sum_{k=1}^6 \alpha_k \cos(\omega_k t) + \sum_{k=1}^6 \gamma_k \sin(\omega_k t) + \varepsilon_t. \quad (2.25)$$

As estimativas dos parâmetros são obtidas pelo método dos mínimos quadrados (2.3) e as estimativas das observações são:

$$\hat{Y} = X\hat{\beta}. \quad (2.26)$$

Para verificar a significância estatística do modelo, utilizamos os testes de hipóteses:

- **Hipóteses:**

- **Hipótese nula ( $H_0$ ):**  $\beta_1 = \alpha_1 = \dots = \alpha_6 = \gamma_1 = \dots = \gamma_6 = 0$
- **Hipótese alternativa ( $H_1$ ):** Pelo menos um parâmetro é diferente de zero

A estatística  $F$  segue uma distribuição de Fisher-Snedecor:

$$F \sim F(p - 1, n - p)$$

Rejeita-se  $H_0$  ao nível de significância  $\alpha$  se o valor observado de  $F$  for igual ou superior ao valor crítico  $F$  tabelado, ou se o valor-p for menor ou igual ao nível de significância  $\alpha = 0,05$ .

Para avaliar a significância dos parâmetros individualmente, utilizamos a estatística  $t$  de Student, conforme descrito na Seção 2.2.

## 2.4 Análise de Resíduos

Após construir nosso modelo, é essencial verificar os resíduos, que são as diferenças entre os valores reais e os valores estimados pelo modelo. Formalmente, os resíduos são definidos como  $\varepsilon_i = y_i - \hat{y}_i$ , para  $i = 1, 2, \dots, n$ , representando a parte da variação que o modelo não conseguiu explicar.

Os resíduos devem satisfazer alguns pressupostos fundamentais, conforme descrito por Montgomery, Peck e Vining (2006):

- **Linearidade:** A relação entre a variável resposta  $y$  e as variáveis explicativas deve ser aproximadamente linear.
- **Média Zero:** Os resíduos  $\hat{\varepsilon}$  devem ter média zero, ou seja,  $E(\hat{\varepsilon}) = 0$ .
- **Variância Constante:** Os resíduos  $\hat{\varepsilon}$  devem ter variância constante, ou seja, a variância dos resíduos deve ser igual a  $\sigma^2$ .
- **Não Correlação:** Os resíduos devem ser não correlacionados entre si.
- **Distribuição Normal:** Os resíduos devem seguir uma distribuição normal,  $\hat{\varepsilon} \sim N(0, \sigma^2)$ .

Com base nesses pressupostos, podemos avaliar a adequação do nosso modelo por meio de diversos testes e análises.

### 2.4.1 Teste de Shapiro-Wilk

Para verificarmos a normalidade dos resíduos, podemos realizar o teste de Shapiro-Wilk, cujas hipóteses são:

$H_0$ : Os resíduos seguem distribuição normal

$H_1$ : Os resíduos não seguem distribuição normal

A estatística do teste é dada por:

$$W_{\text{obs}} = \frac{[\sum_{i=1}^k a_{n-i+1}(y_{n-i+1} - y_i)]^2}{[\sum_{i=1}^n (y_i - \bar{y})^2]^2}. \quad (2.27)$$

Para a interpretação dos resultados, é necessário o uso de tabelas apresentadas por Shapiro e Wilk (1965).

Rejeitamos  $H_0$  se  $W_{\text{tab}} < W_{\text{obs}}$ , ao nível de significância  $\alpha$ .

### 2.4.2 Teste de Durbin-Watson

Outra suposição que fazemos é a de independência entre os resíduos. Para verificar isso, aplicaremos o teste de Durbin-Watson, conforme apresentado em Draper e Smith (1968).

Vamos supor um modelo dado por:

$$y_u = \beta_0 + \sum_{i=1}^k \beta_i X_{iu} + \varepsilon_u, \quad u = 1, 2, \dots, n. \quad (2.28)$$

Testaremos as seguintes hipóteses:

$$H_0: \rho_s = 0$$

$$H_1: \rho_s \neq 0$$

Onde  $\rho_s$  é a correlação serial dos resíduos. A estatística do teste é dada por:

$$d = \frac{\sum_{u=2}^n (\varepsilon_u - \varepsilon_{u-1})^2}{\sum_{u=1}^n \varepsilon_u^2}. \quad (2.29)$$

Entretanto, para rejeitar ou não a hipótese nula, é necessário observar dois valores críticos,  $d_L$  e  $d_U$ , conforme apresentados em Draper e Smith (1968). A partir disso, utilizaremos a seguinte regra de decisão para um teste bicaudal com nível de significância  $\alpha$ :

- Se  $d < d_L$  ou  $4 - d < d_L$ , rejeitamos a hipótese nula  $H_0$  ao nível de significância  $2\alpha$ .
- Se  $d > d_U$  e  $4 - d > d_U$ , não rejeitamos a hipótese nula  $H_0$  ao nível de significância  $2\alpha$ .
- Caso contrário, o teste é inconclusivo.

### 3 RESULTADOS PARCIAIS

Fizemos uma análise descritiva, avaliando o Volume, Receita e Preço Médio de quatro tipos de café (Arábica, Conillon, Solúvel e Torrado), destacando padrões temporais e métricas de variabilidade. Testes de estacionaridade e modelos temporais foram aplicados ao Volume do Arábica, com destaque para os modelos Dummie e Harmônico. Comparamos os modelos também a partir do RMSE, MAE e  $R^2$ . Também fizemos gráficos de séries e boxplots, para identificar padrões sazonais e variações específicas.

#### 3.1 Análise Descritiva das variáveis a ser estudada

Para a análise descritiva, iremos utilizar algumas medidas que já conhecemos para nos aprofundar ainda mais nos dados dos 4 tipos de café. Primeiramente, iremos analisar o Volume.

Para esta variável, pela Tabela 4, o café que foi exportado mais em questão de volume, foi o Arábica, tendo o valor de máximo maior do que os outros. Observamos também, pelas medidas, que a média mais explicativa é a do Solúvel, tendo seu C.V. total de, aproximadamente, 13%. Porém o do Arábica também consegue ser explicativo, tendo seu C.V. total de, aproximadamente, 18%. Graficamente, pela Figura 1 e na Figura 2, podemos sentir essas informações com mais clareza, pois os dados do Arábica e do Solúvel estão bem mais próximos do que os outros.

Tabela 4 - Estatísticas descritivas da variável Volume (em sacas de 60 kg)

Descritiva	Conillon	Arábica	Torrado	Solúvel
<b>Min.</b>	9.862	1.440.854	573	185.288
<b>1st Qu.</b>	61.876	2.265.582	1.825	297.254
<b>Median</b>	215.454	2.570.784	2.373	318.565
<b>Mean</b>	226.254	2.612.259	2.696	318.158
<b>3rd Qu.</b>	351.268	2.954.878	3.245	346.566
<b>Max.</b>	676.325	4.080.675	6.447	456.506
<b>S.D.</b>	170.842,6	487.109,9	1.230,675	43.593,09
<b>C.V.</b>	0,7550918	0,1864708	0,4610619	0,1370172

Figura 1 - Histograma do Volume (em sacas de 60 kg)

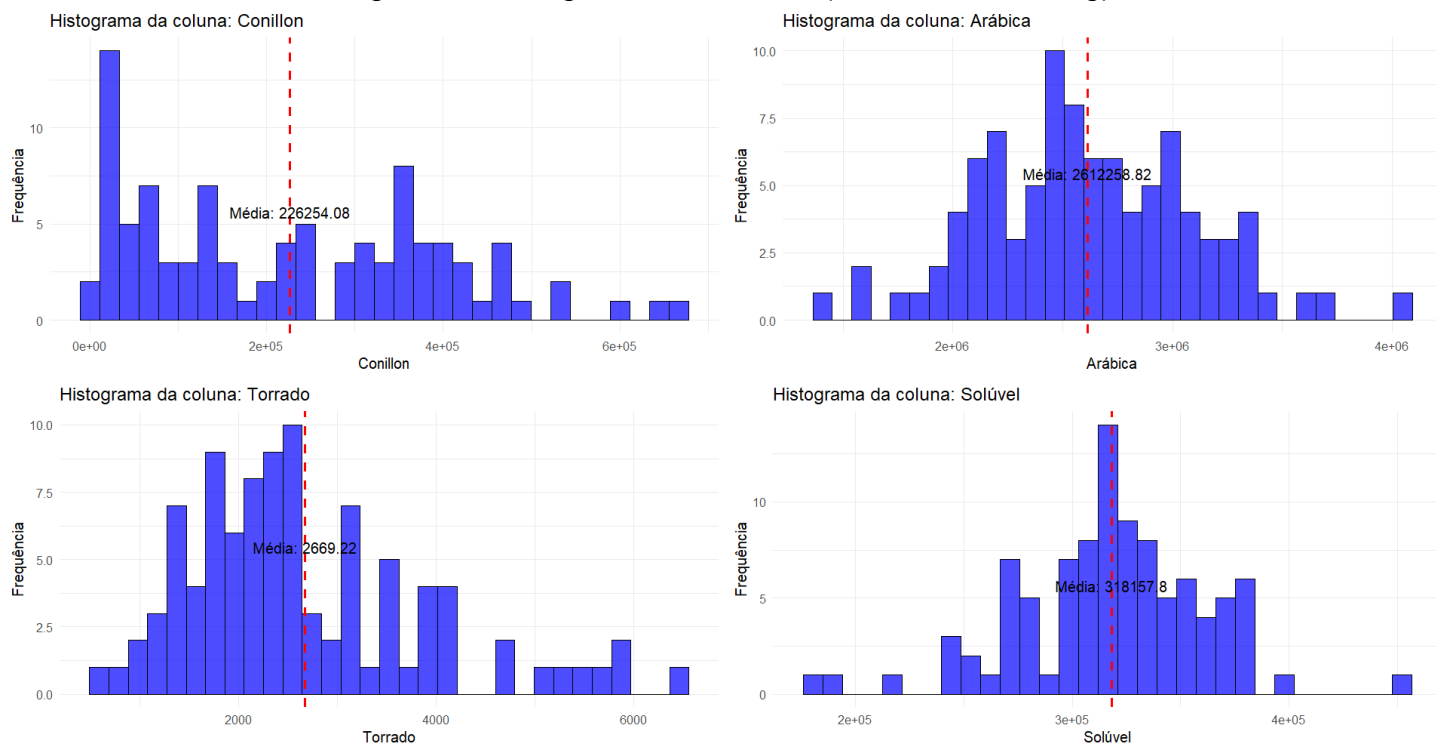
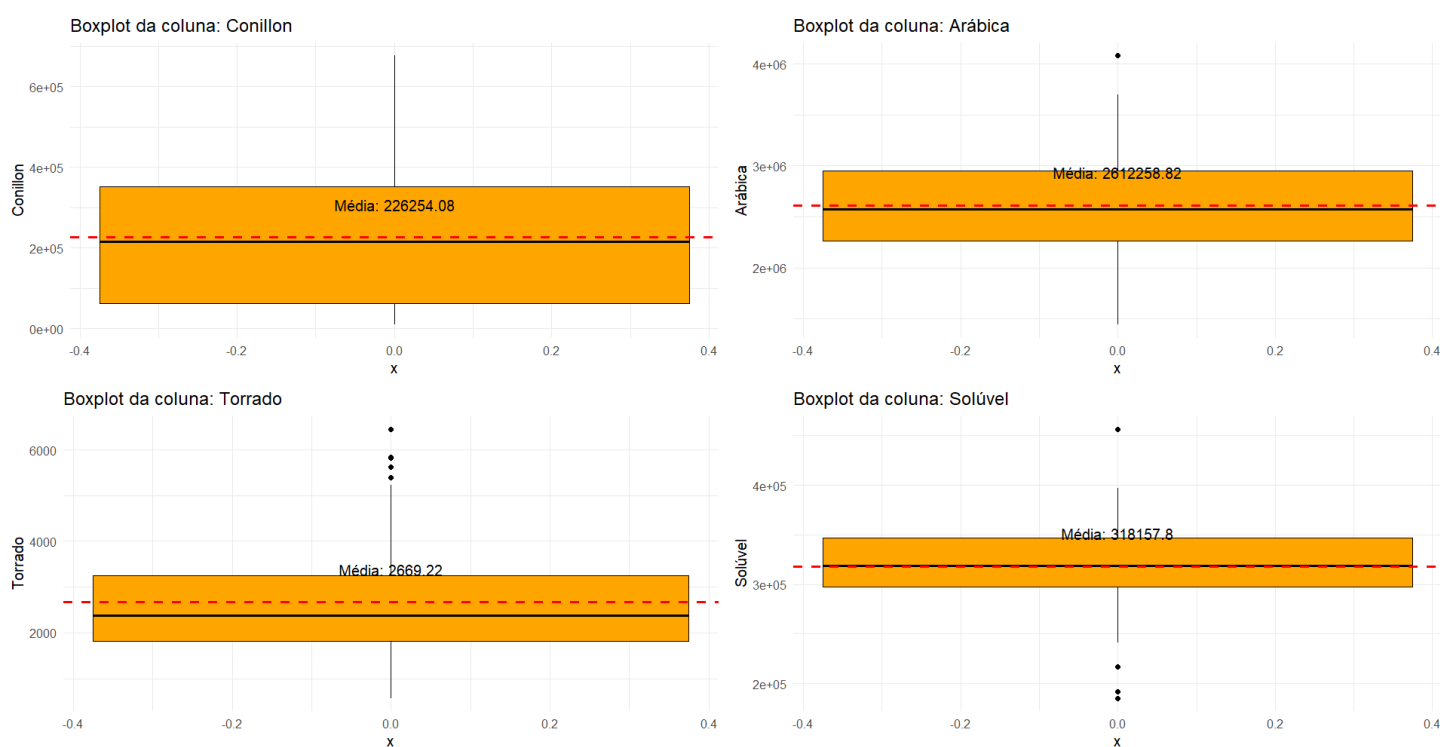


Figura 2 - Box-plot do Volume (em sacas de 60 kg)



Para a variável Receita, a maior venda, pela Tabela 5, que tivemos foi também do café Arábica, tendo o valor máximo como comparação. Observando a variação dos

dados, o C.V. mais explicativo é a do Solúvel e a do Arábica, sendo aproximadamente 17% e 30%, respectivamente. A Figura 3 e Figura 4 consegue expressar muito bem como está a distribuição dos dados em torno da média.

Tabela 5 - Estatísticas descritivas da variável Receita (em US\$ 1000)

Descritiva	Conillon	Arábica	Torrado	Solúvel
<b>Min.</b>	1.538	219.900	209,4	33.142
<b>1st Qu.</b>	7.300	346.591	646,3	45.200
<b>Median</b>	21.778	407.524	943,3	49.395
<b>Mean</b>	22.237	429.613	1.120,2	50.240
<b>3rd Qu.</b>	34.018	461.869	1.501,3	54.453
<b>Max.</b>	58.662	829.654	2.860,8	74.701
<b>S.D.</b>	14.805,34	129.130,2	605,6191	8.577,977
<b>C.V.</b>	0,6658058	0,300573	0,5406358	0,1707406

Figura 3 - Histograma da Receita (em US\$ 1000)

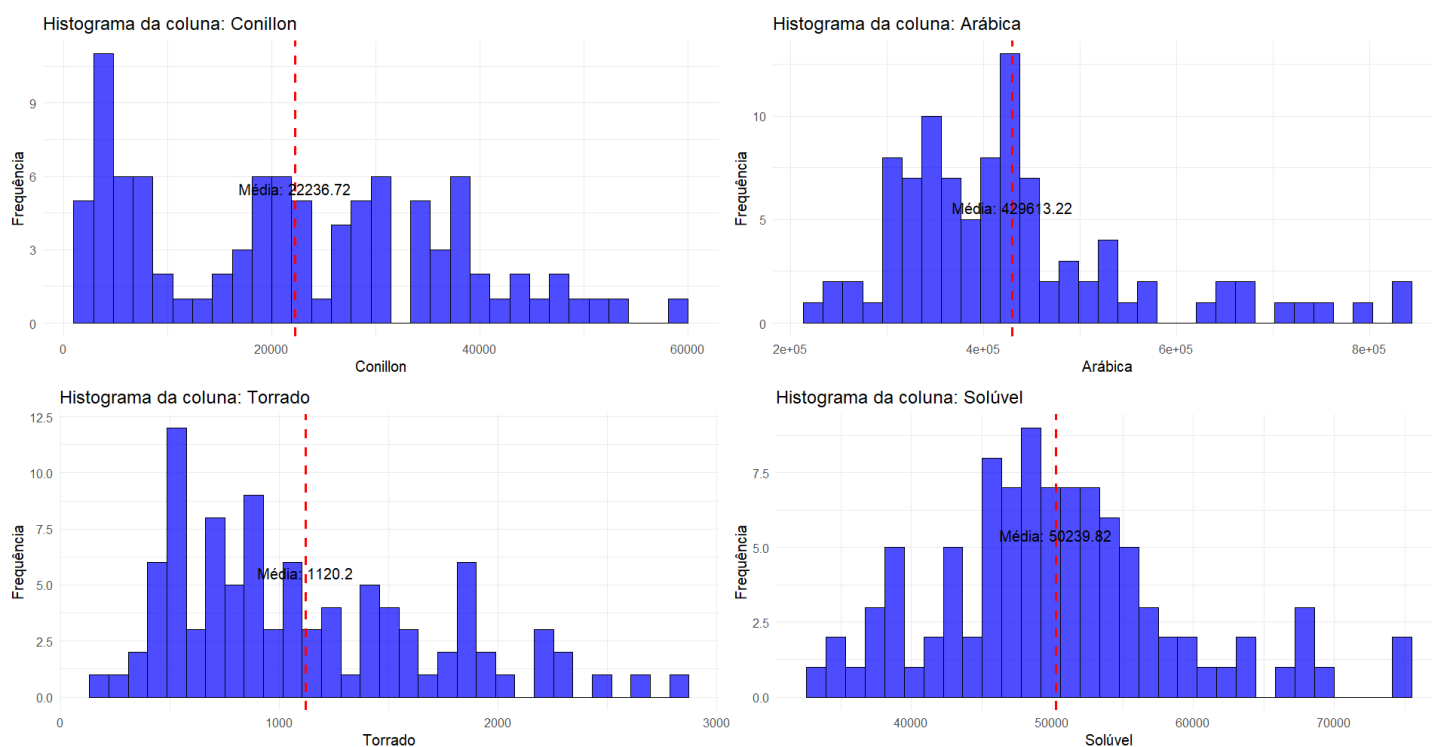
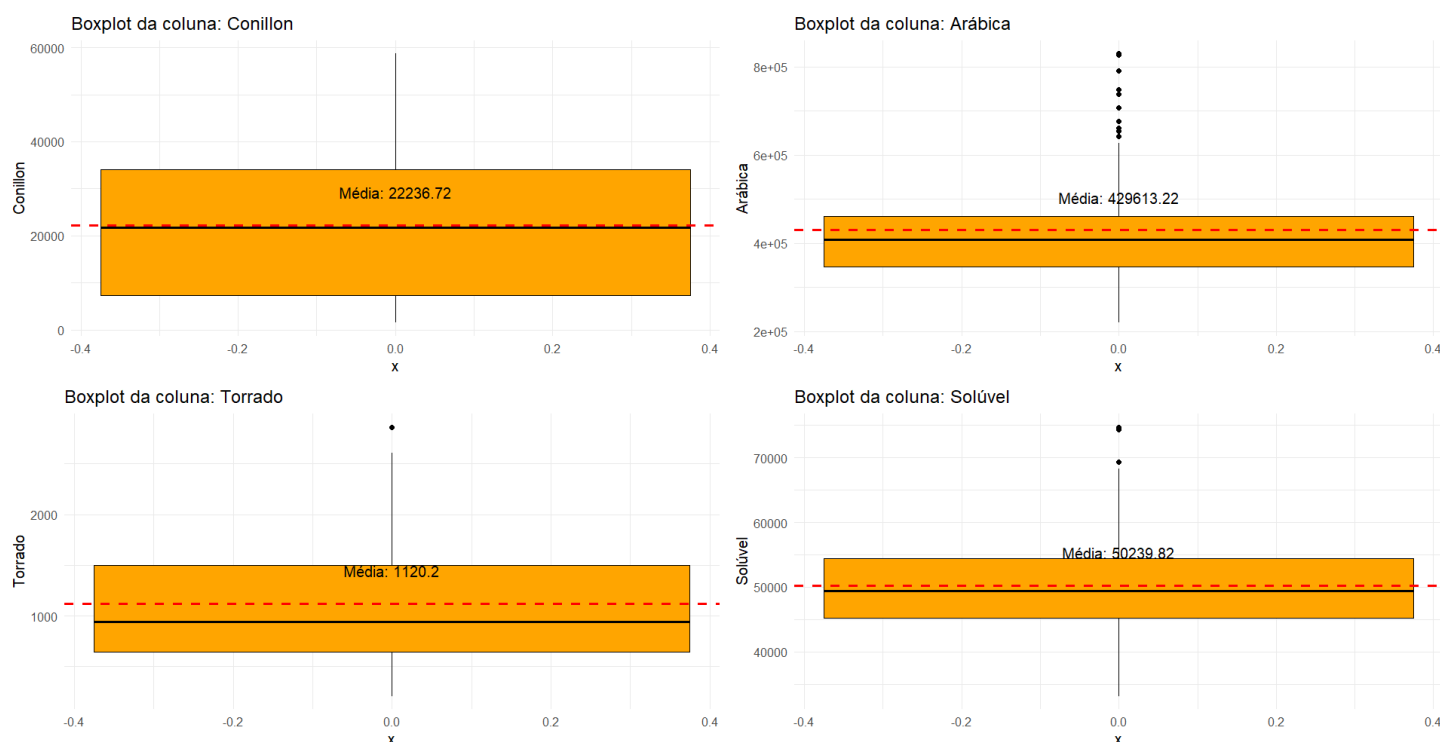




Figura 4 - Box-plot da Receita (em US\$ 1000)



Para a variável Preço Médio, o que teve seu máximo foi o Torrado, e os mais explicativos, novamente, foi o Arábica e o Solúvel, com 22% e 14%, respectivamente e aproximadamente. Pela Figura 5 e Figura 6, também sentimos a distribuição dos dados em torno da média, por mais que exista valores discrepantes, a média ainda consegue explicar bastante dos dados.

Tabela 6 - Estatísticas descritivas da variável Preço Médio (US\$ / Saca de 60kg)

Descritiva	Conillon	Arábica	Torrado	Solúvel
<b>Min.</b>	74,19	119,1	186,5	119,1
<b>1st Qu.</b>	87,09	138,0	315,9	143,9
<b>Median</b>	109,44	155,3	385,2	155,8
<b>Mean</b>	114,18	164,4	427,7	158,8
<b>3rd Qu.</b>	141,17	175,5	499,6	175,0
<b>Max.</b>	176,30	251,0	978,7	231,6
<b>S.D.</b>	29,98622	36,24818	167,2282	22,6532
<b>C.V.</b>	0,262613	0,2204249	0,3909963	0,1426503

Figura 5 - Histograma do Preço Médio (US\$ / Saca de 60kg)

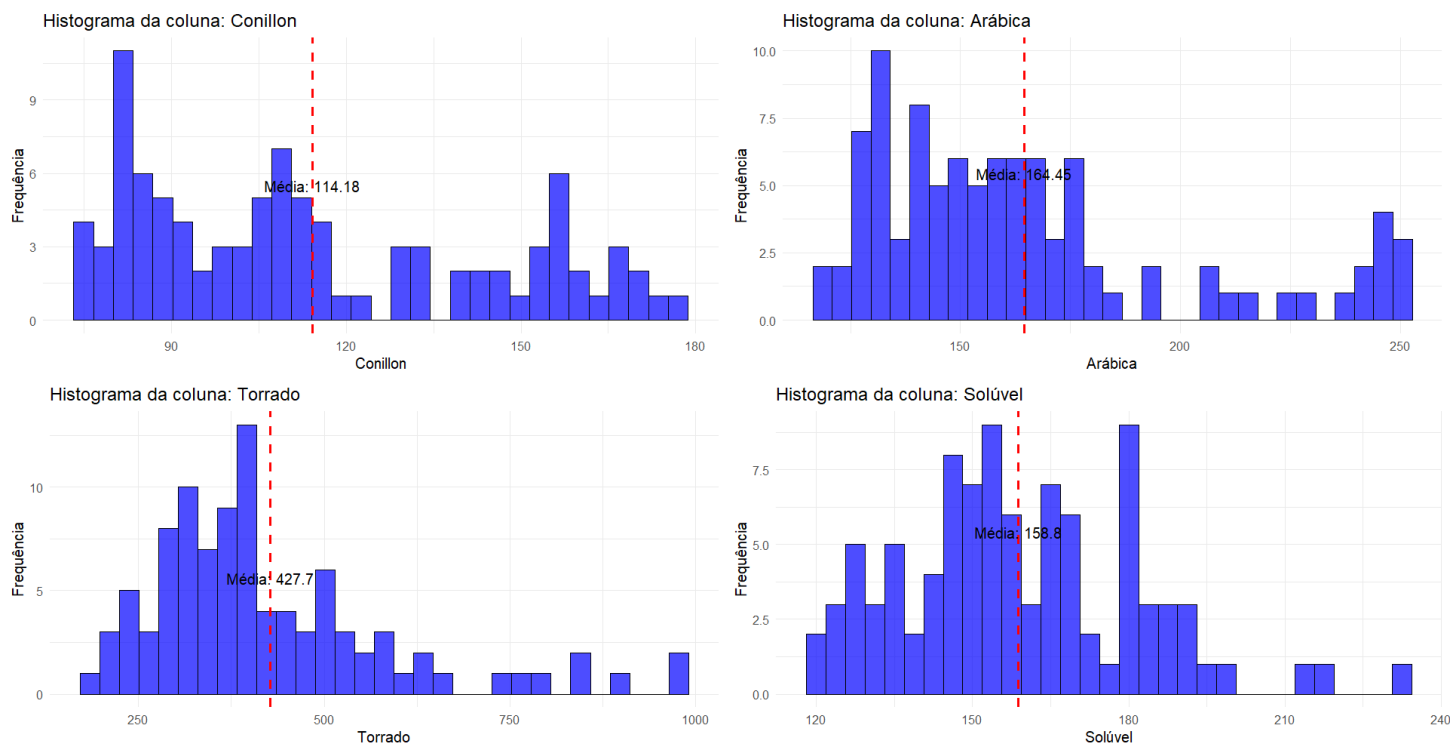
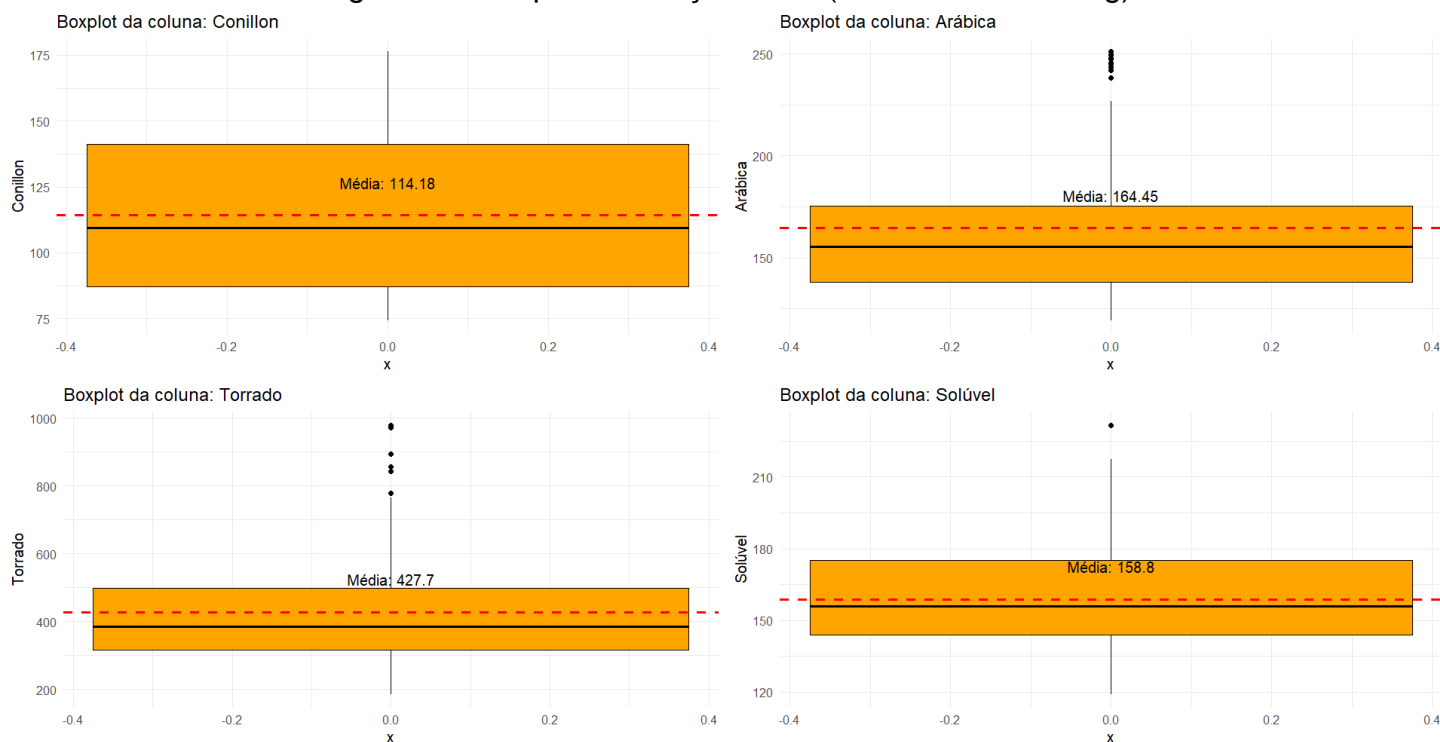


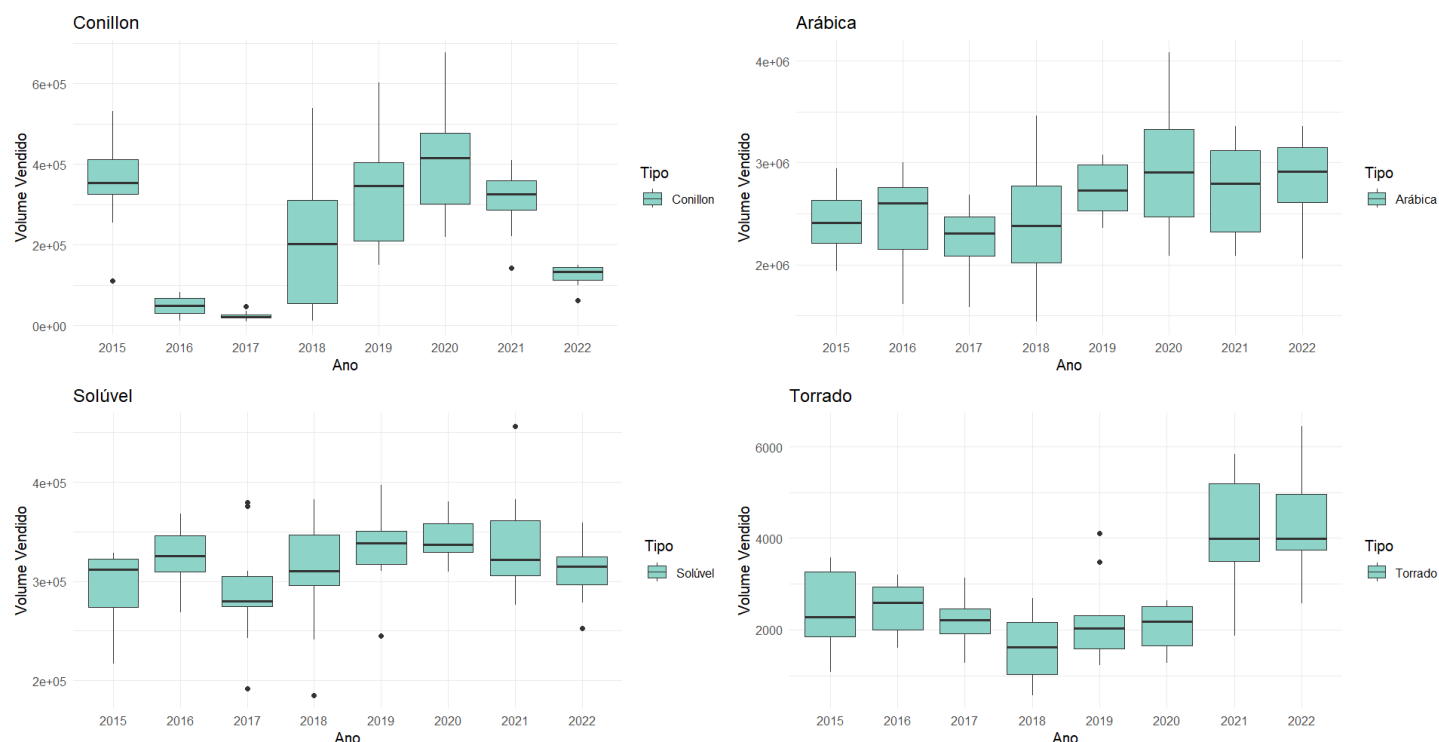
Figura 6 - Box-plot do Preço Médio (US\$ / Saca de 60kg)



Agora, iremos analisar a distribuição anualmente de cada variável e de cada café.

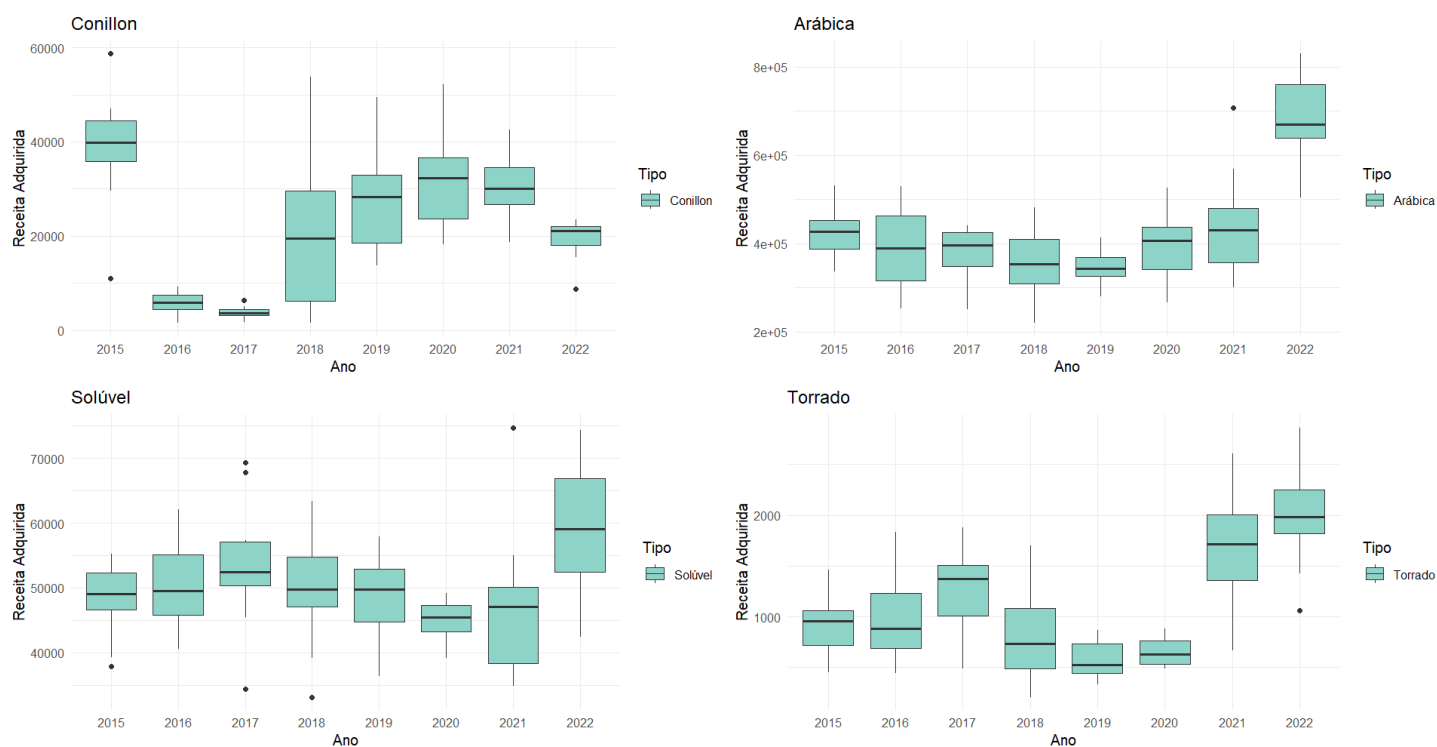
Na Figura 7, temos o Box-plot do Volume, onde cada caixa evidencia a distribuição dos dados em cada ano. Vemos que a distribuição do Arábica e Solúvel anualmente aparenta ser mais constante do que a dos outros, sendo que no Conillon e Torrado, temos alguns anos que é mais exportado e menos exportado, evidenciado com a caixa sendo puxado muito para cima ou para baixo. Existe alguns anos no Conillon que a exportação do café não é muito dispersa, se concentrando muito perto da mediana (evidenciado em 2016, 2017 e 2022).

Figura 7 - Box-plot do Volume (em sacas de 60kg) por ano



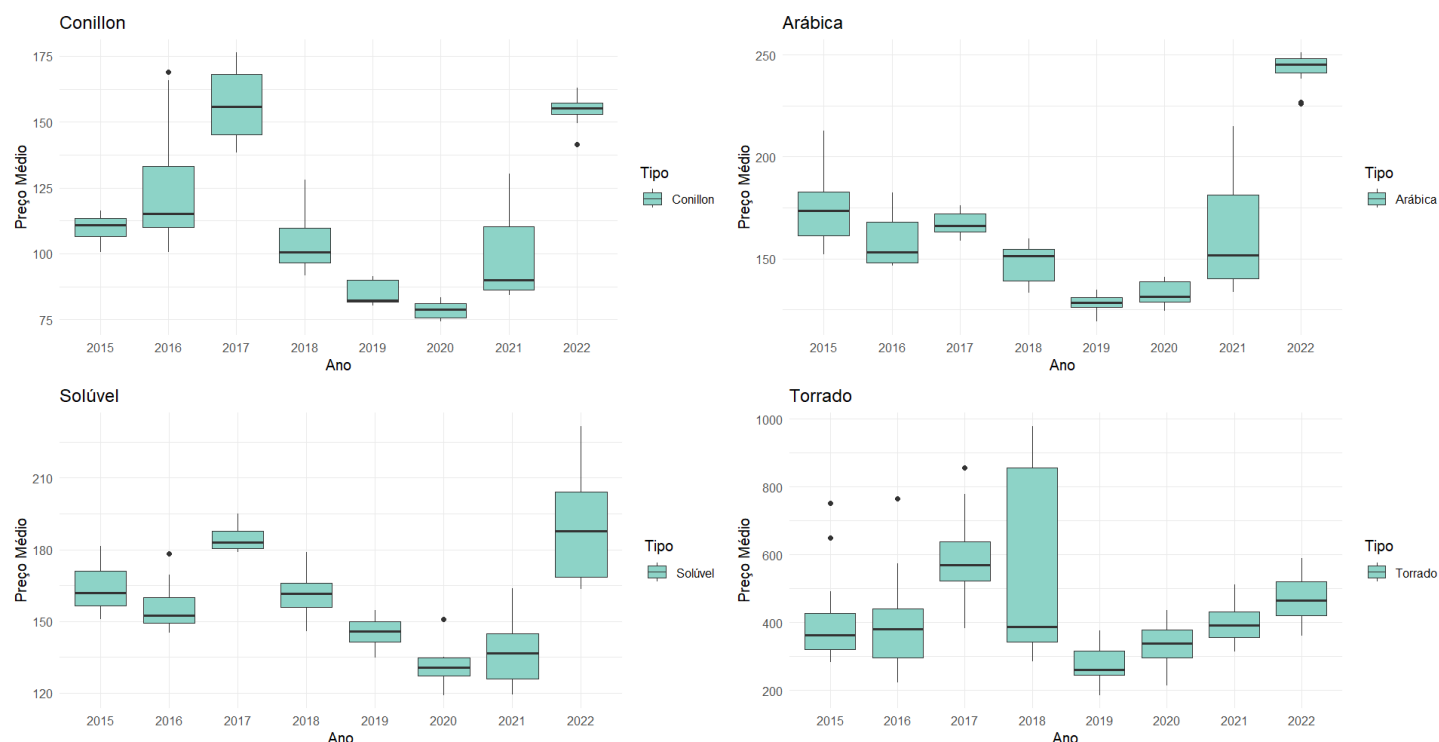
Na Figura 8, temos os Box-plot da Receita. O Conillon, em comparação ao gráfico do Box-plot do Volume, eles são praticamente idênticos, em questão de dispersão dos dados (as caixas não mudam tanto), obtendo as mesmas conclusões (2016, 2017 e 2022 com menos dispersão e 2018 com uma dispersão enorme). Para o Arábica e o Solúvel, a dispersão dos dados muda em 2022, sendo que neste ano, o café foi vendido bem mais do que nos outros, fazendo a caixa ficar um pouco mais acima das outras. A do Torrado, a única mudança foi no tamanho das caixas, ou seja, houve bastante dispersão nas vendas do que nas exportações (se pegarmos os primeiros anos, as caixas da Receita são um pouco maiores do que a do Volume)

Figura 8 - Box-plot da Receita (em US\$ 1000) por ano



Na Figura 9, temos os Box-plots do Preço Médio. Em cada um dos cafés, percebemos que há menos dispersão dos dados nos anos, sendo que a dispersão do Torrado em 2018, foi bem maior comparado com a das outras caixas do mesmo café. Os cafés não parecem ter uma constância na dispersão (tem mais anos que a dispersão era menor do que anos que a dispersão era maior) e a posição de cada caixa nos mostra que houve uma variabilidade alta.

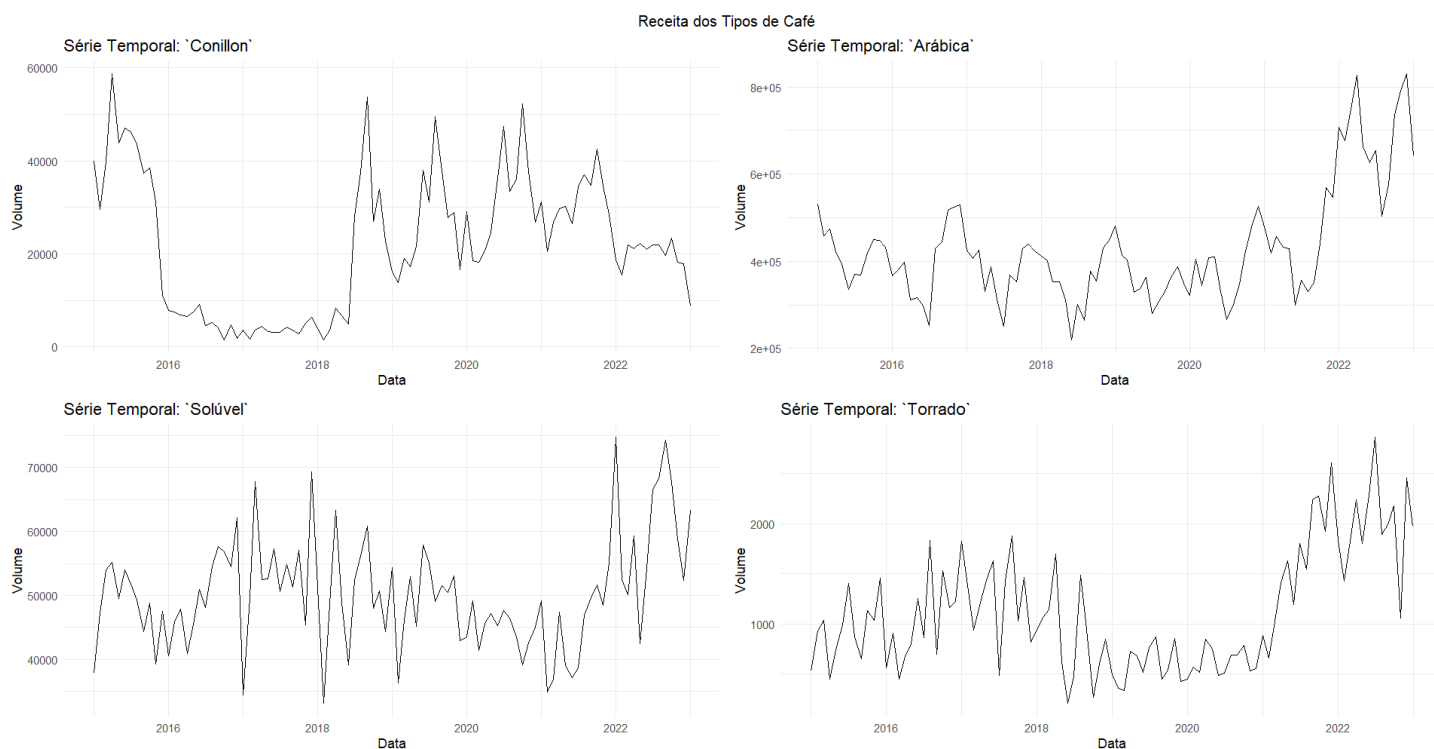
Figura 9 - Box-plot do Preço Médio (US\$ / Saca de 60kg) por ano



Partiremos agora para os gráficos das séries (linhas), onde as conclusões que obtemos dos Box-plots serão complementadas com eles.

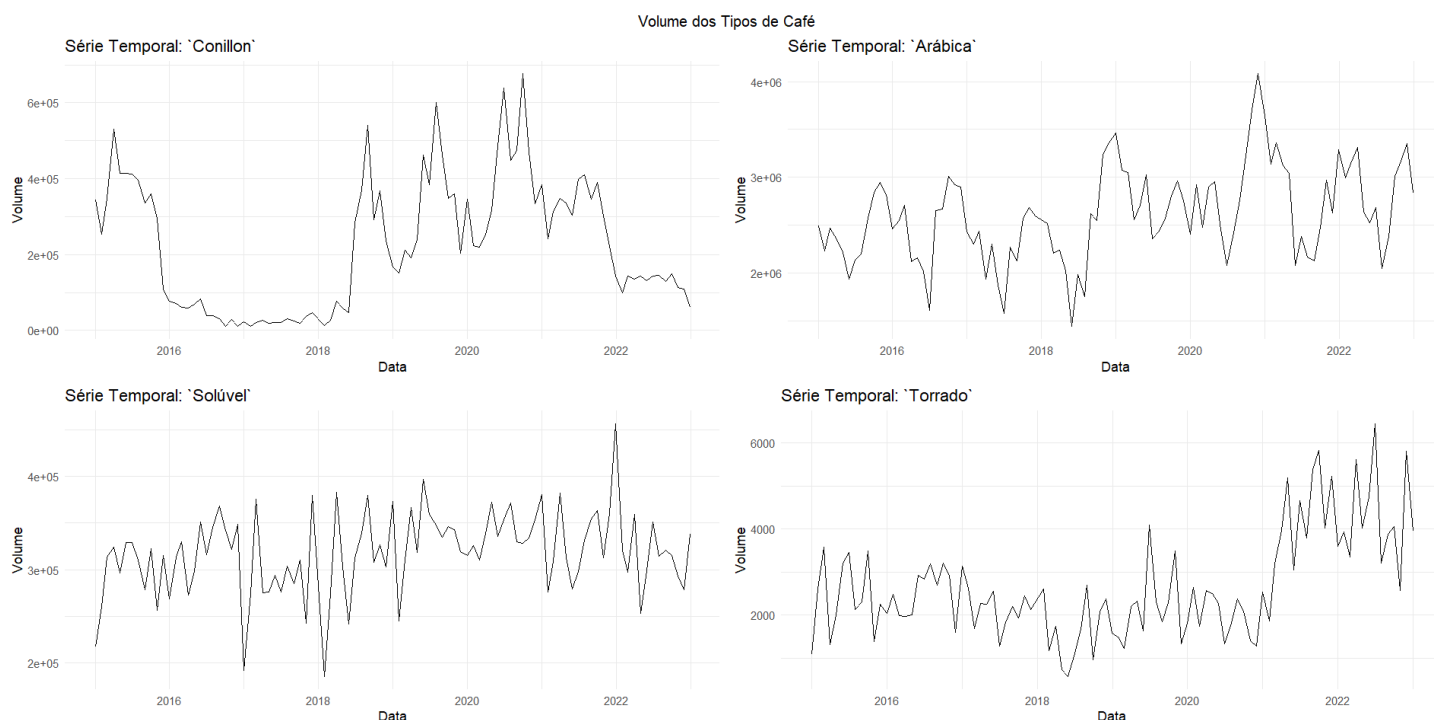
Na Figura 10, temos o gráfico de séries da Receita. Conillon tem uma queda muito preocupante a partir de 2016, evidenciando que neste período, não se vendeu tanto deste café. Para os outros, nos anos de 2021 e 2022, houve um aumento de vendas, onde os picos se mantiveram alto nestes anos. Para o Torrado, também há uma época que ele vende muito menos, sendo seu início em 2019.

Figura 10 - Série da Receita (em US\$ 1000)



Na Figura 11, temos as séries do Volume. Nela, para o Conillon, temos as mesmas conclusões que obtemos para o gráfico da série da Receita (2016 até 2018, houve uma baixa exportação deste café). Para o Solúvel, conseguimos ver alguns picos baixos e altos (baixas exportação em 2017 e 2018, alta exportação em 2022). Para o Torrado, obtemos a mesma conclusão que fizemos a Receita (2021 houve uma exportação mais alta, evidenciando mais vendas) e o Solúvel, temos um gráfico praticamente constante, aumentando suas exportações feitas a partir de 2021.

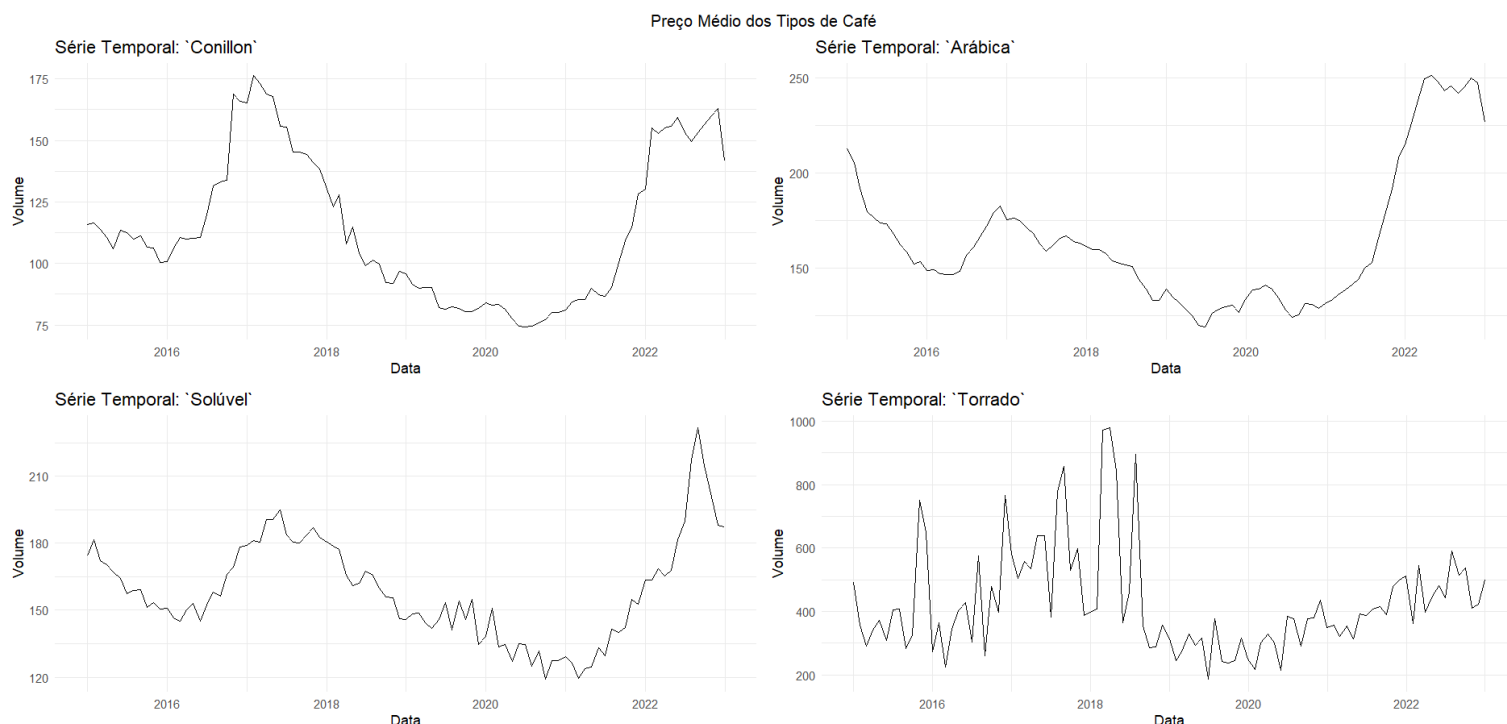
Figura 11 - Série do Volume (em sacas de 60kg)



Para a Figura 12, temos os gráficos da série do Preço Médio. Para todos eles, em 2020 houve uma diminuição da média dos preços, voltando a aumentar a partir de 2022. Não aparentam ter constância as séries, tendo muitos períodos de altos e baixos. Para os cafés Conillon, Arábica e Solúvel, seus picos altos de preço médio estão nos anos de 2017 e 2022, enquanto para o Torrado, o pico mais alto acontece no ano de 2018. Por mais que a escala de cada gráfico esteja diferente, se notarmos o movimento da série, todas aparentam está fazendo o mesmo movimento (começa com baixos preços médios, aumenta nos anos de 2017 a 2018, e diminui a partir de 2019).



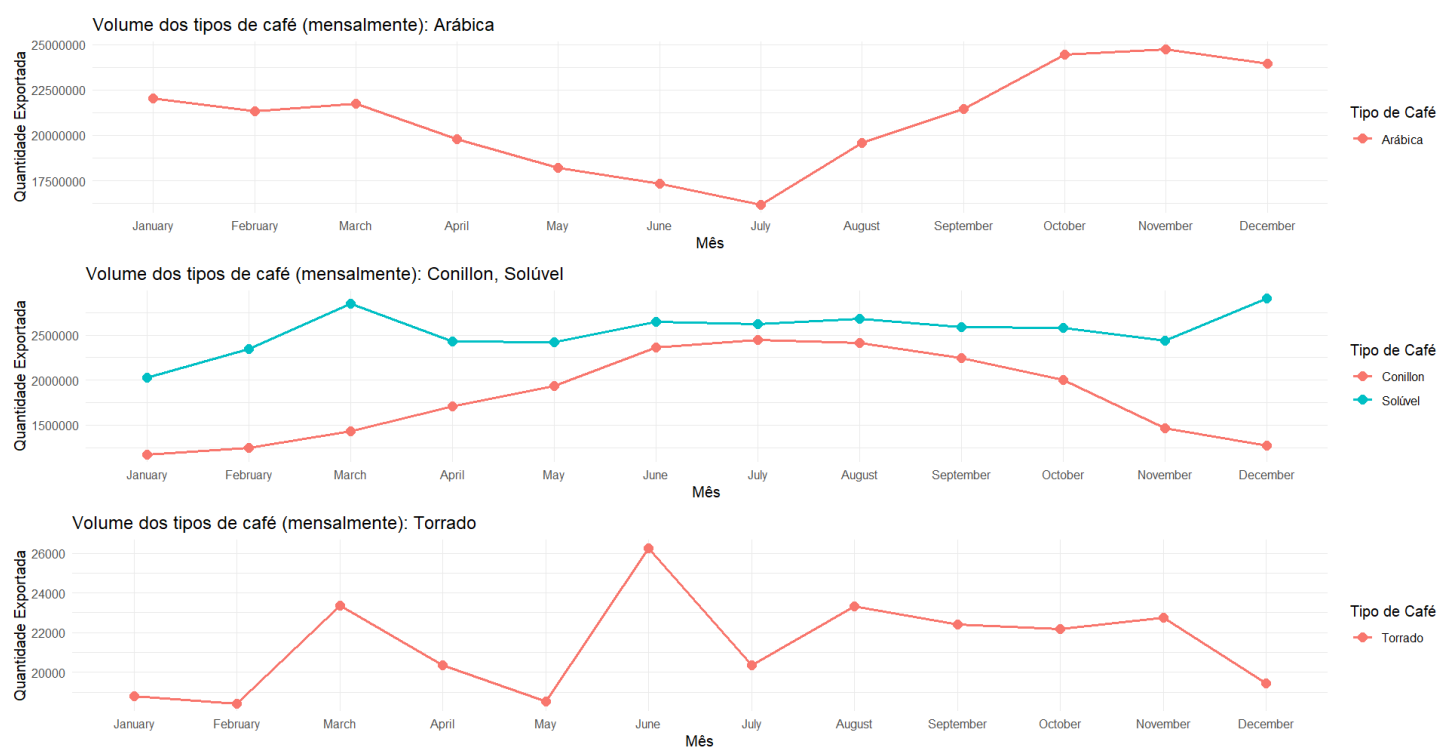
Figura 12 - Série do Preço Médio (US\$ / Saca de 60kg)



Agora, iremos analisar mensalmente cada gráfico. Este gráfico foi feito somando todos os meses iguais, criando somente um ponto, por exemplo, somamos todos os meses de janeiro e criamos um dado representando todos os meses de janeiro.

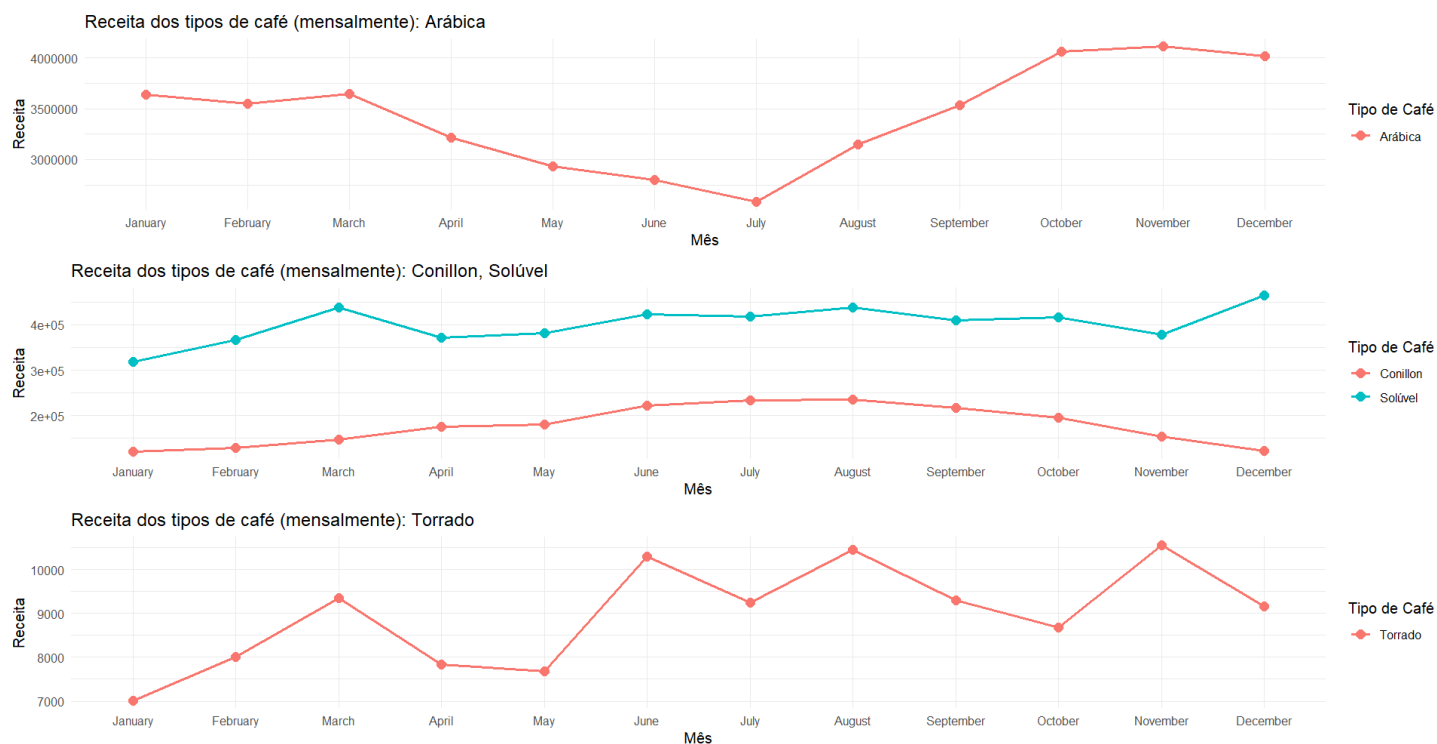
Sendo assim, começando com a variável Volume, percebemos que o café Arábica é o que mais exporta mensalmente, sendo seus pontos altos nos meses de outubro, novembro e dezembro, e seus baixos o mês de julho. Para o Solúvel e o Conillon, o que mais exporta entre eles é o Solúvel, com seus pontos máximos no mês de março e dezembro e seus baixos em abril e novembro. Para o Conillon, temos seu pico em julho e seus baixos em janeiro, fevereiro e dezembro. Note que, para os dois, vemos que a exportação deles quase se iguala nos meses de junho, julho e agosto. Para o Torrado, seus altos estão nos meses de junho e seus baixos, janeiro, fevereiro e dezembro.

Figura 13 - Série do Volume mensalmente (em sacas de 60kg)



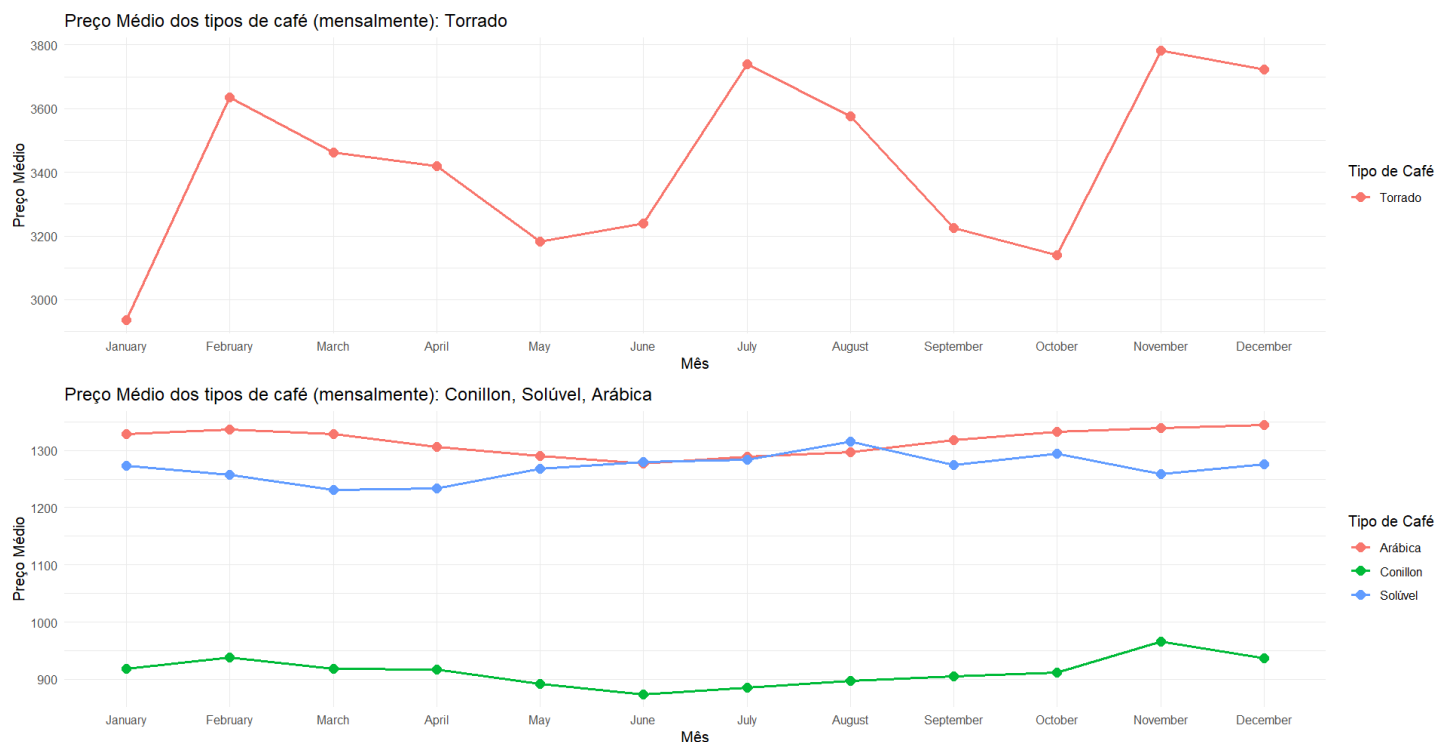
Para a Receita, sobre quem vende mais, segue o mesmo ranking que a da variável Volume. Temos para o Arábica os mesmos altos que o do Volume, o que é lógico pois foram os meses que mais exportaram deste café, e consequentemente, os mais que venderam. Para o Solúvel, temos que os altos são em março e dezembro e o baixo, janeiro. O Conillon, os meses com mais vendas foram em junho, julho e agosto, e os baixos, dezembro e janeiro. Para o Torrado, temos os altos nos meses de junho, agosto e novembro, e os baixos, temos o mês de janeiro.

Figura 14 - Série da Receita mensalmente (em US\$ 1000)



Para o Preço Médio, temos algo diferente. A média do torrado é maior que a dos outros, sendo a menor, a do Conillon. Para o Torrado, o mês de novembro foi o preço médio mais alto, e o janeiro, o menor. O Torrado foi o único que teve variação na média dos meses, sendo que os outros estão um pouco mais contantes. Para o Arábica e o Solúvel, temos determinados meses que as médias se cruzam, sendo eles junho e julho.

Figura 15 - Série do Preço Médio mensalmente (US\$ / Saca de 60kg)



Agora, temos os gráficos de boxplot, que mostra a distribuição mensal. Arábica apresenta a maior quantidade exportada e pouca variação ao longo do ano. Contudo, há alguns outliers, como em maio e outubro, que indicam valores baixos ou altos.

O café Conillon, apresenta maior variabilidade durante o ano, especialmente em meses como maio e agosto, acompanhada por outliers. O café solúvel tem uma distribuição mais consistente, com menor variação, embora também apresente outliers em meses como abril e agosto.

Por fim, o café torrado registra as menores quantidades exportadas. Há meses com maior variação, como julho e setembro, e alguns outliers em meses como julho e outubro.

Figura 16 - Boxplot do Volume mensalmente (em sacas de 60kg)



Sobre a Receita, o café Arábica apresenta receitas significativamente maiores em comparação aos demais. Há poucos outliers em meses como agosto e outubro, indicando um comportamento consistente nas receitas.

No caso do café Conillon, as receitas apresentam maior variabilidade ao longo dos meses, especialmente em julho e setembro. O café Solúvel contém outliers dispersos em meses como junho e agosto. O café Torrado gera as menores receitas. Apesar disso, observa-se uma maior amplitude em meses como agosto e setembro, indicando maior variação nas receitas nesse período. Os outliers, presentes em diversos meses, destacam receitas anormalmente altas ou baixas.

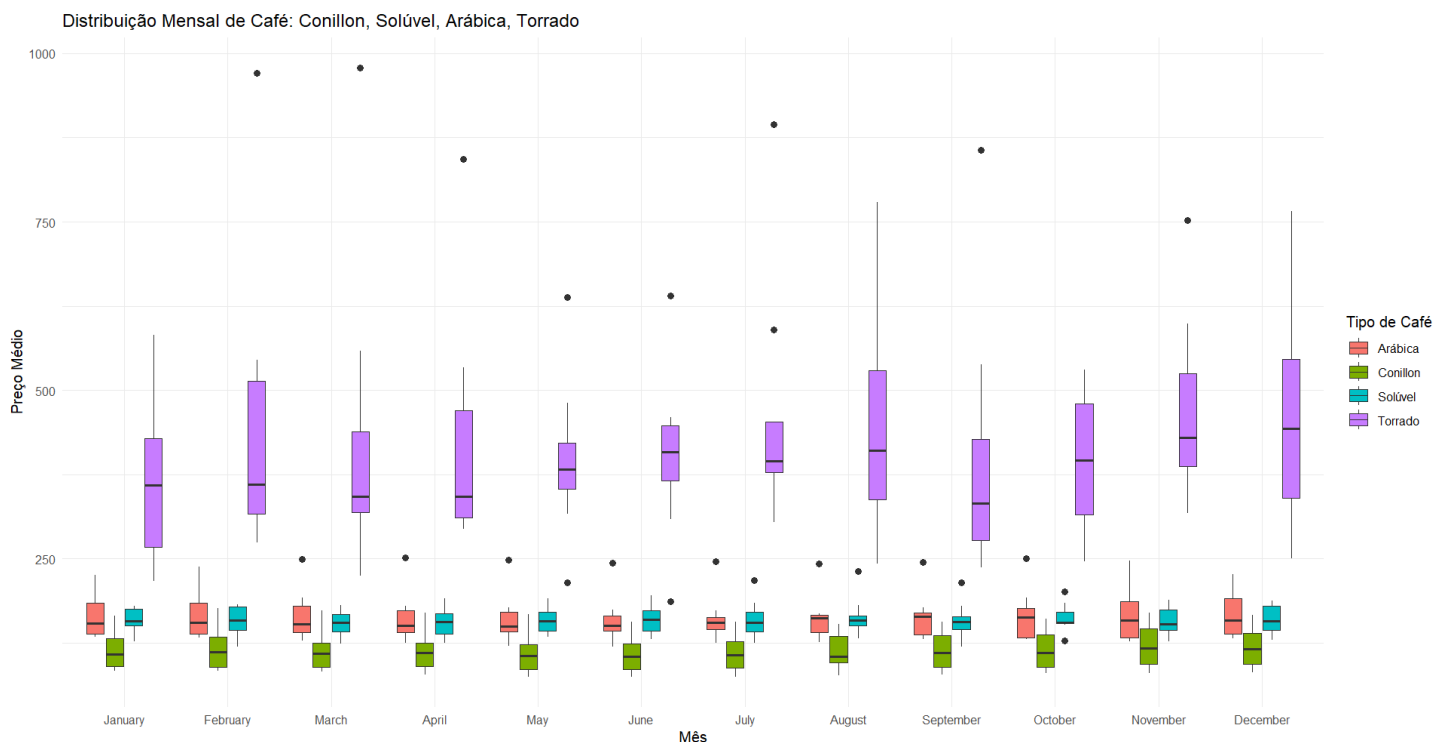
Figura 17 - Boxplot da Receita mensalmente (em US\$ 1000)



Do Preço Médio, o café Torrado apresenta os preços médios mais elevados e grande variabilidade, evidenciada por amplas amplitudes e a presença de outliers em diversos meses, como março, agosto e dezembro. Do café Arábica temos menor variabilidade, sugerindo um comportamento mais estável nos preços.

Do Conillon e Solúvel, apresentam padrões semelhantes, com preços médios próximos entre si. No entanto, o café solúvel apresenta menor variabilidade, com amplitudes mais estreitas, enquanto o café Conillon exibe maior dispersão, evidenciada por outliers em meses como abril e setembro.

Figura 18 - Boxplot do Preço Médio mensalmente (US\$ / Saca de 60kg)



### 3.2 Modelos para a variável Volume (Arábica)

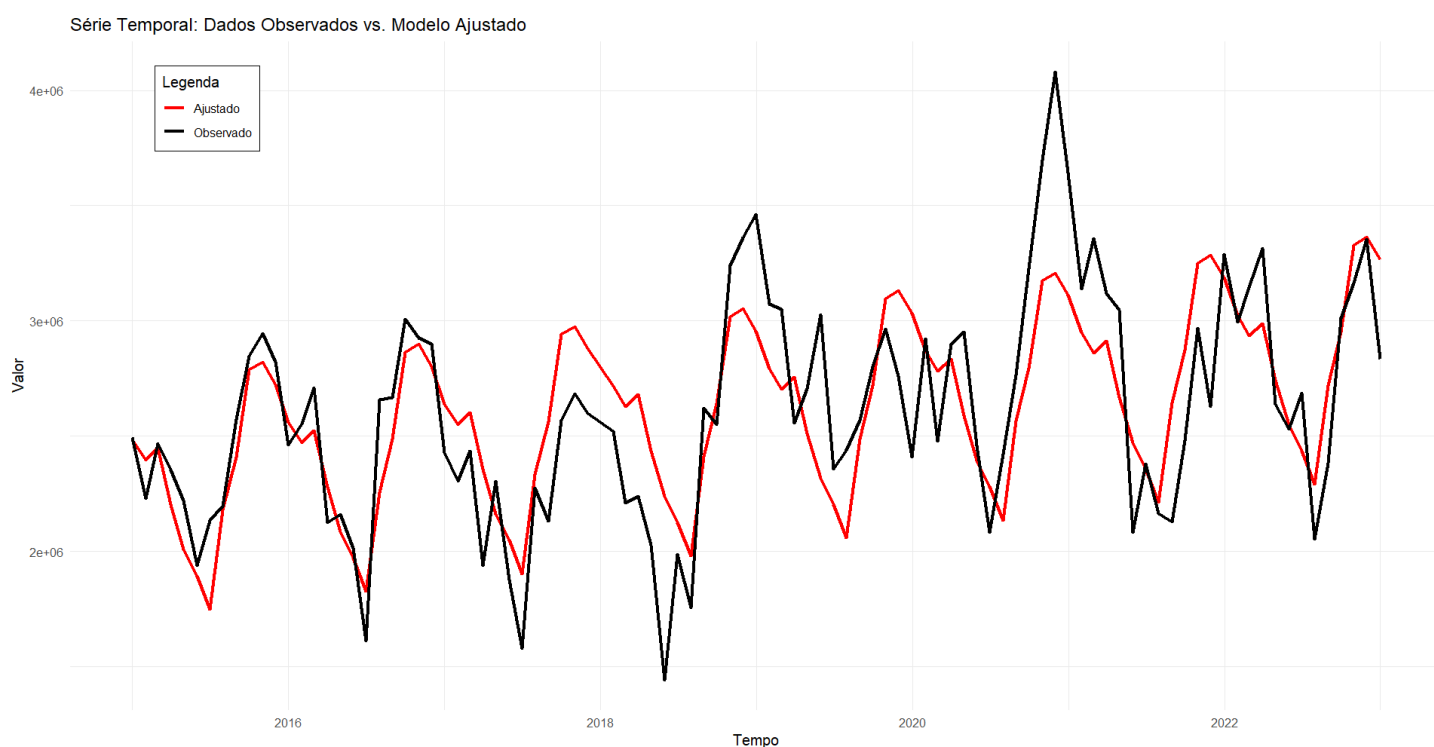
Para esta parte, ao analisarmos detalhadamente cada uma das séries e de cada variável, decidimos começar com a variável Volume do café Arábica. Primeiramente, começamos com o teste de Dickey-Fuller, onde devemos rejeitar  $H_0$  em favor da  $H_a$  para verificarmos se é estacionária. Ao fazermos, ele nos dá um p-valor de 0.01, indicando que a série é estacionária.

Ao verificarmos, fizemos os modelos descritos anteriormente nas seções 2.3 e 2.3.2, sendo assim, o primeiro modelo que iremos apresentar é o de Decomposição Temporal, logo verificamos alguns parâmetros significativos, sendo eles apresentados na Tabela 6, e em seguida, temos na Figura 19 o resultado graficamente do modelo na série.

Tabela 7 - Parâmetros do modelo de Decomposição Temporal

Coeficientes	Parâmetro	Desvio Padrão	P-valor
$\beta_0$	2520030	76129	2e-16
$\beta_1$	6790	1251	4.91e-07
<i>Dummy_abril</i>	-359582	128374	0.00625
<i>Dummy_maio</i>	-561455	128345	3.30e-05
<i>Dummy_junho</i>	-833212	128328	8.01e-07
<i>Dummy_julho</i>	-833212	128324	4.66e-09
<i>Dummy_agosto</i>	-409490	128332	0.00196

Figura 19 - Série temporal do modelo de Decomposição Temporal versus Dados Observados



E o modelo ficou desta forma:

$$\begin{aligned} \hat{y}_t = & 2520030 + 6790t - 359582\text{Dumme}_{\text{abril}} - 561455\text{Dumme}_{\text{maio}} \\ & - 833212\text{Dumme}_{\text{junho}} - 833212\text{Dumme}_{\text{julho}} \\ & - 409490\text{Dumme}_{\text{agosto}} \end{aligned}$$

A Figura 19 apresenta uma análise visual, mostrando que o modelo captura bem a tendência geral e o comportamento sazonal dos dados ao longo do tempo, mas



apresenta algumas discrepâncias, especialmente em períodos com variações abruptas ou picos acentuados.

Entre 2016 e 2017, o modelo ajusta-se bem aos dados, mantendo proximidade entre as linhas. No entanto, o modelo pode não capturar adequadamente extremos ou mudanças rápidas. Apesar disso, a linha ajustada acompanha o padrão sazonal, confirmando que as dummies mensais mostradas na Tabela 7 têm impacto significativo.

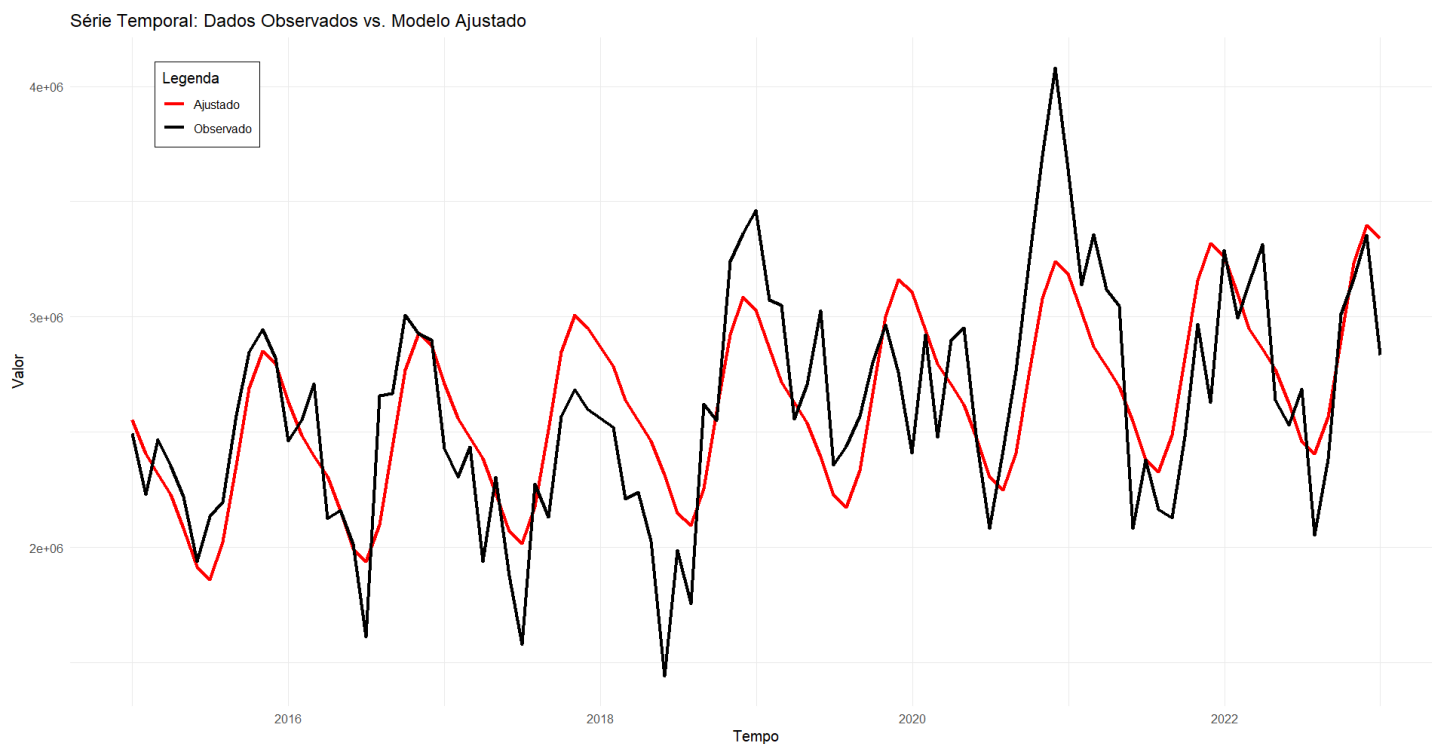
A tendência geral crescente também é capturada pelo modelo, reforçando a contribuição do coeficiente  $\beta_1$  da Tabela 7.

A seguir, iremos ver o modelo harmônico:

Tabela 8 - Parâmetros do modelo Harmônico

Coeficientes	Parâmetro	Desvio Padrão	P-valor
$\beta_0$	2298111	66563	2e-16
$\beta_1$	6477	1192	4.49e-07
$\alpha_1$	422224	46670	2.31e-14
$\gamma_2$	-134749	46700	0.00487

Figura 20 - Série temporal do modelo Harmônico versus Dados Observados



E o modelo fica:

$$\hat{y}_t = 2298111 + 6477t + 422224 \cos\left(\frac{2\pi t}{12}\right) - 134749 \sin\left(\frac{4\pi t}{12}\right)$$

Na Figura 20, observa-se que o modelo captura bem a tendência geral e a sazonalidade da série, acompanhando padrões semelhantes ao longo do tempo. No entanto, há discrepâncias em alguns momentos.

No geral, o modelo é razoavelmente bom para descrever a dinâmica da série.

### 3.3 Análise de resíduos e adequabilidade dos modelos

As Figuras 21 e 22 apresentam os resíduos do modelo Dummie. O gráfico superior da Figura 21, mostra os resíduos ao longo do tempo, observando variações, indicando possíveis padrões ou tendências, o que sugere que os resíduos podem não ser totalmente aleatórios, violando um dos pressupostos básicos do modelo. O gráfico ACF (autocorrelação), localizado no canto inferior esquerdo, revela autocorrelações significativas em defasagens iniciais, o que indica que os resíduos podem estar correlacionados. O histograma dos resíduos, no canto inferior direito, apresenta uma distribuição aproximadamente simétrica. Na Figura 22, observa-se que os resíduos se alinham razoavelmente à linha de normalidade.

Figura 21 - Análise dos Resíduos (Decomposição Temporal)

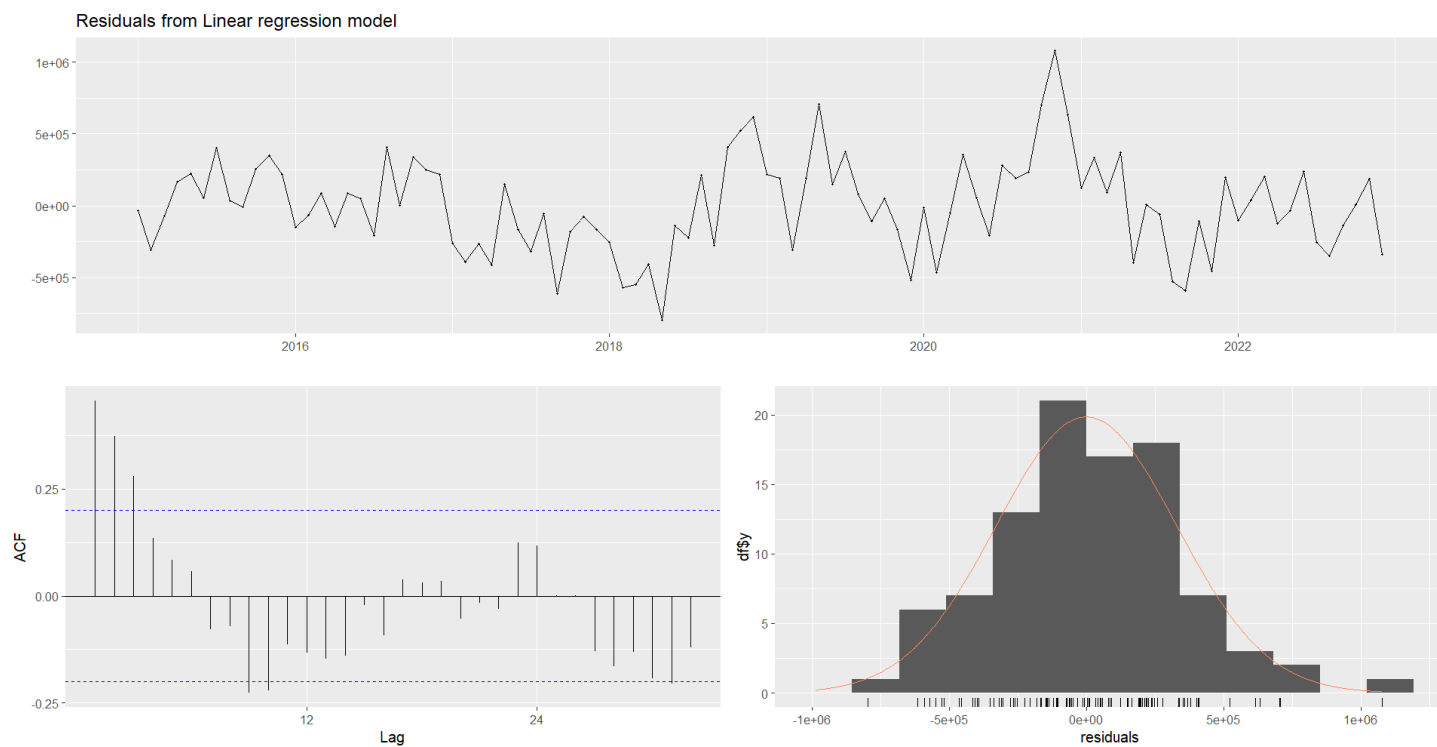
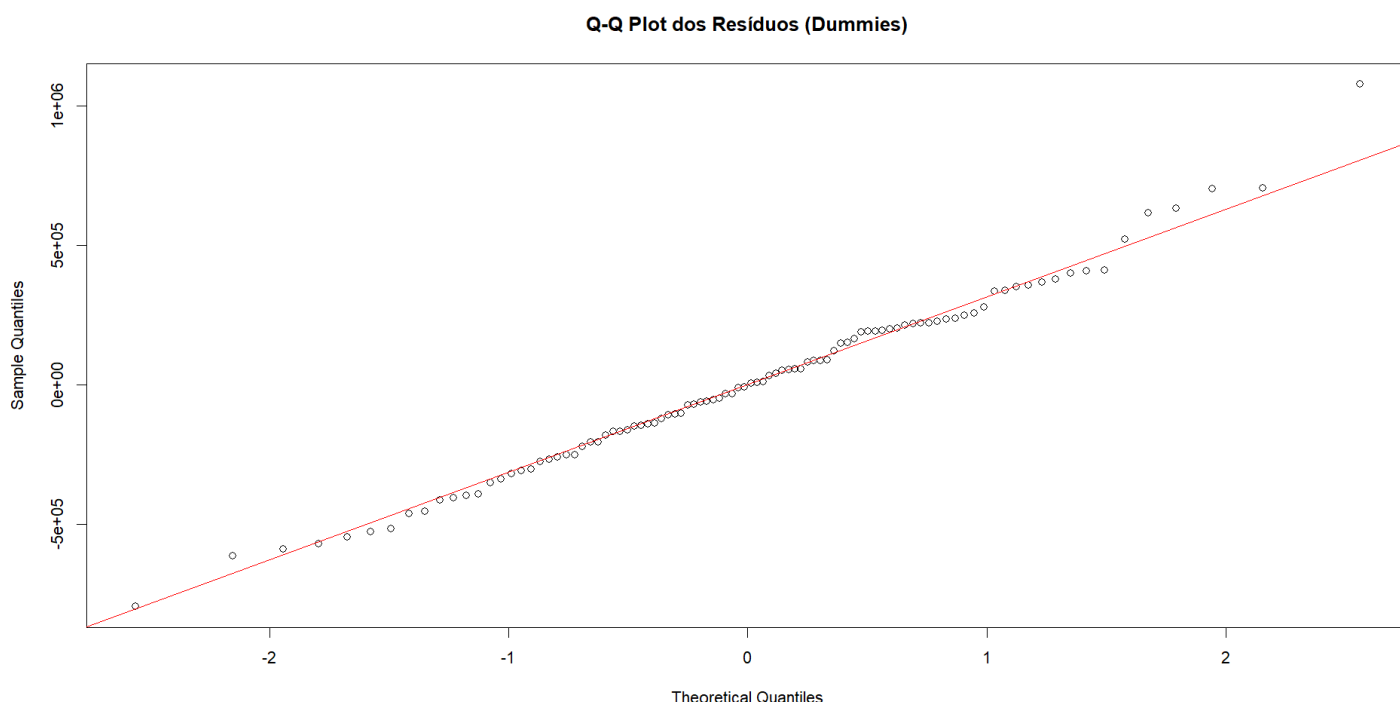


Figura 22 - Q-Q Plot dos Resíduos (Decomposição Temporal)



As Figuras 23 e 24 conseguimos ver os resíduos do modelo harmônico. O gráfico da Figura 23, mostra os resíduos ao longo do tempo, com variações significativas e possíveis padrões ou tendências, sugerindo que os resíduos podem não ser totalmente aleatórios.

O gráfico de autocorrelação (ACF), no canto inferior esquerdo, revela a presença de autocorrelações significativas nos primeiros lags, ultrapassando os limites de significância (linhas pontilhadas). Isso pode violar os pressupostos de independência dos resíduos.

O histograma dos resíduos (inferior direito) apresenta uma distribuição aproximadamente simétrica.

Figura 23 - Análise dos Resíduos (Harmônica)

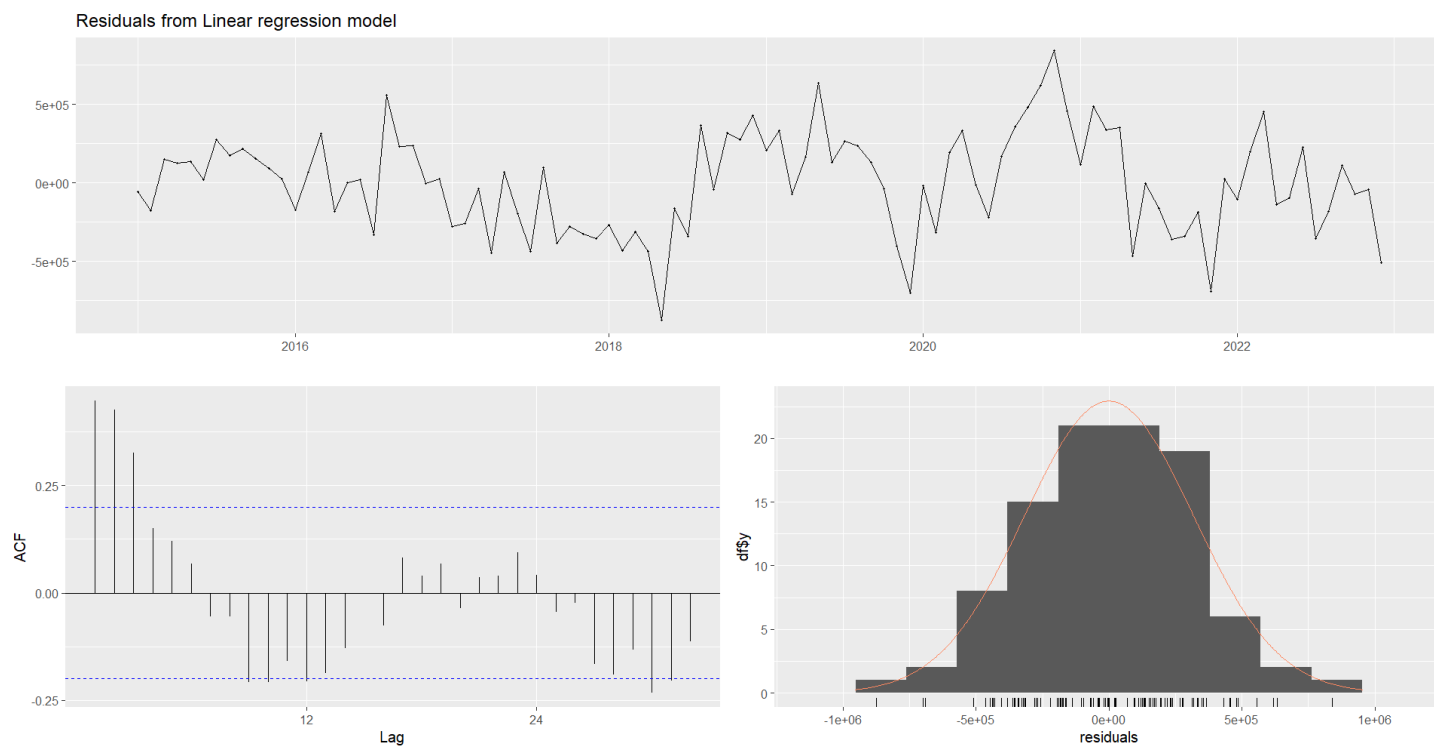
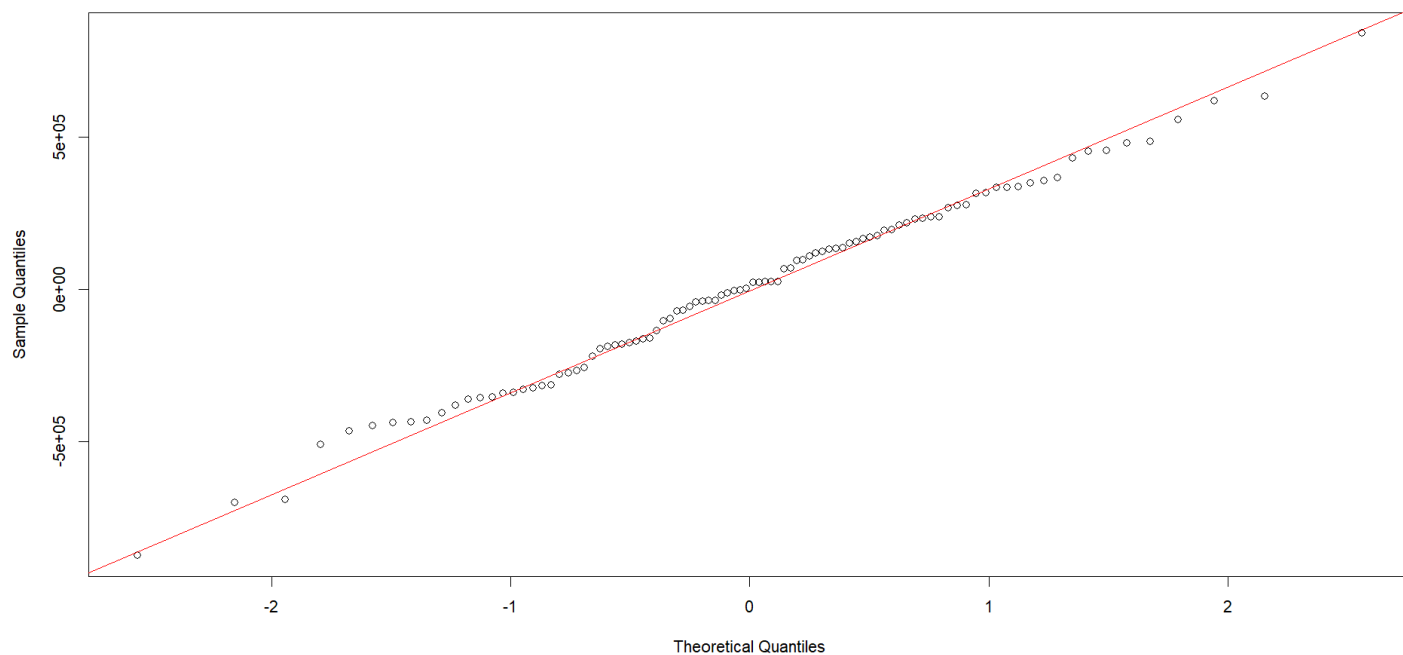


Figura 24 - Q-Q Plot dos Resíduos (Harmônica)

Q-Q Plot dos Resíduos (Harmonica)



Fazendo o teste de Durbin – Watson para os dois modelos, para a harmônica, obtivemos um p-valor 2.978e-07, e para o modelo dummie, 8.445e-07, indicando que os resíduos dos dois modelos não são independentes e existe uma autocorrelação positiva para os dois. Para solucionarmos este problema de autocorrelação, se aplicarmos uma transformação Box-Cox ou a de Cochrane-Orcutt, talvez os resíduos fiquem sem autocorrelação.

Agora, vamos observar as métricas de ajuste. Na Tabela 9, os dois modelos, Decomposição Temporal e Harmônica, foram avaliados com base em métricas de desempenho como RMSE, MAE e  $R^2$ . O modelo Dummie apresentou um RMSE de 305.118,1, enquanto o Harmônica teve um RMSE de 316.424,8, indicando que o Dummie possui erros ligeiramente menores. No MAE, o Dummie também se destaca, com 245.421,9 frente a 255.357,2 do Harmônica, reforçando sua menor discrepância média nas previsões. Por outro lado, o modelo Harmônica apresentou um  $R^2$  de 0,5597, superior ao 0,5142 do Dummie, o que indica que ele explica melhor a variabilidade dos dados. Embora o Dummie tenha métricas de erro ligeiramente melhores, o Harmônica possui um maior poder explicativo.

Tabela 9 - Medidas de diagnóstico dos modelos

Medidas	Dummie	Harmônica
<i>RMSE</i>	<b>305118.1</b>	316424.8
<i>MAE</i>	<b>245421.9</b>	255357.2
$R^2$	0.5142	<b>0.5597</b>

## 4 CONSIDERAÇÕES PARCIAIS

A análise detalhada dos dados dos quatro tipos de café (Arábica, Conillon, Solúvel e Torrado) nos permitiu compreender melhor as características de cada variável avaliada (Volume, Receita e Preço Médio) e os padrões temporais associados a elas.

O café Arábica lidera em exportação, com volumes elevados e maior constância ao longo dos anos, seguido pelo Solúvel. Ambos mostram boa explicação pela média devido aos coeficientes de variação relativamente baixos (CV de 18% para Arábica e 13% para o Solúvel). Já o Conillon e Torrado apresentam maior variabilidade, evidenciada pela dispersão nos gráficos de boxplot e outliers em meses específicos.

O Arábica também domina em termos de Receita, com comportamento consistente ao longo dos anos. O Solúvel mantém uma explicação robusta pela média, com um CV de 17%. O Conillon apresenta maior variabilidade e quedas preocupantes em anos como 2016, enquanto o Torrado demonstra maior dispersão nos primeiros anos e picos isolados.

O Torrado possui o maior preço médio e maior variabilidade, enquanto o Arábica e o Solúvel mantêm padrões mais estáveis. O Conillon apresenta preços médios mais baixos, com variação significativa em alguns períodos. A sazonalidade no Preço Médio é aparente em todos os cafés, com variações de aumento e redução em determinados meses e anos, reforçando a influência de fatores sazonais e econômicos.

Os modelos ajustados a variável Volume do Arábica (Decomposição Temporal e Harmônico) capturam adequadamente a tendência e sazonalidade, embora ambos apresentem limitações em relação à autocorrelação e aleatoriedade dos resíduos, como evidenciado pelos testes de Durbin-Watson e análise de resíduos. O modelo de Decomposição Temporal se destacou em métricas de erro (RMSE e MAE), indicando maior precisão nas previsões. Já modelo Harmônico, embora com maior erro, apresentou melhor capacidade explicativa da variabilidade dos dados, evidenciada pelo  $R^2$  mais elevado. Os modelos de Decomposição Temporal e Harmônico têm vantagens distintas, com o de Decomposição sendo mais preciso e o Harmônico melhor explicativo. Ambas as abordagens são úteis dependendo do objetivo da análise, seja previsão ou compreensão de variabilidade e tendências. Logo, para o

TCC II, iremos continuar com os modelos Harmônicos e de Decomposição Temporal, para as outras séries e também partiremos com a análise espectral.



## REFERÊNCIAS

- AGUIRRE, Antonio. uma introdução à análise espectral de séries temporais econômicas. **Nova Economia**. vol. 5. p. 42-43. Disponível em: <https://revistas.face.ufmg.br/index.php/novaeconomia/article/download/2284/1224/7444>. Acesso em 15 out. 2023.
- BUENO, R. D'LOSSO S. **Econometria de Séries Temporais**. CENGAGE Learning, São Paulo, 2008.
- CECAFÉ. **Conselho dos Exportadores de Café do Brasil 2023**. Disponível em: <https://www.cecafe.com.br/publicacoes/relatorio-de-exportacoes/>. Acesso em: 15 out. 2023.
- DRAPER, Norman Richard; SMITH, Harry. **Applied regression analysis**. 3rd ed. New York: John Wiley & Sons. 1998. 706 p.
- HOFFMANN, Rodolfo. **Análise de regressão**: uma introdução à econometria. Universidade de São Paulo. Escola Superior de Agricultura Luiz de Queiroz, 2016. DOI: <https://doi.org/10.11606/9788592105709> Disponível em: [www.livrosabertos.sibi.usp.br/portaldelivrosUSP/catalog/book/73](http://www.livrosabertos.sibi.usp.br/portaldelivrosUSP/catalog/book/73) . Acesso em 15 outubro. 2023.
- MACKINNON, J. G. **Critical Values for Cointegration Tests**. In R. F. Engle & C. W. J. Granger (Eds.), *Long-Run Economic Relationships: Readings in Cointegration* (pp. 267-276). Oxford University Press. 1991.
- MONTGOMERY, Douglas C.; PECK, Elizabeth A.; VINING, G. Geoffrey. **Introduction to linear regression analysis**. 4th ed. New Jersey: John Wiley & Sons, 2006. 612 p.
- MORETTIN, P.A.; TOLOI, C.M. **Análise de Séries Temporais**. São Paulo, Edgard Blücher, 2004.
- SHAPIRO, S. S., & WILK, M. B. **An analysis of variance test for normality (complete samples)**. *Biometrika*, 52(3-4), 591–611. 1965
- SHUMWAY, R. H; STOFFER, D. S. **Time Series Analysis and Its Applications – With R Examples**. 2a. edition. Springer. New York .2006.
- TOPIK, Steven; MARICHAL, Carlos; ZEPHYR, Frank (Org.). **From silver to cocaine**: latin american commodity chains and the building of the world economy, 1500-2000. Durham e Londres: Duke University Press, 2006.

TRAUER, Eduardo et al. **O Conhecimento e a Cadeia Produtiva do Café.** Congresso Internacional de Conhecimento e Inovação. 12 set. 2017. Disponível em: <https://proceeding.ciki.ufsc.br/index.php/ciki/article/view/235/94>. Acesso em: 17 out. 2023.

SOCIEDADE NACIONAL DA AGRICULTURA (SNA). **Exportações de café do Brasil totalizam 3,3 milhões de sacas em fevereiro.** Disponível em: <https://sna.agr.br/exportacoes-de-cafe-do-brasil-totalizam-3-3-milhoes-de-sacas-em-fevereiro>. 2021.