



**Universidade de Brasília
Departamento de Estatística**

**Análise de Séries Temporais das Infrações de Trânsito Relacionados ao
Consumo de Substâncias Alcoólicas Ocorridas no Brasil de 2007 a 2022**

Júlia Birbeire Machado

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2023**

Júlia Birbeire Machado

Análise de Séries Temporais das Infrações de Trânsito Relacionados ao Consumo de Substâncias Alcoólicas Ocorridas no Brasil de 2007 a 2022

Orientador(a): Prof. Dr. José Augusto Fiorucci

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2023**

Agradecimentos

Primeiramente, gostaria de agradecer a minha família, por estar sempre presente e me apoiando. Agradeço também aos meus amigos, em especial ao Daniel, Luana, Marcelo, Sabrina e Stephany, que estiveram ao meu lado ao longo do curso, pelo companheirismo e por compartilharem comigo tantos momentos que me permitiram crescer não só como formando, mas também como pessoa. Também, quero agradecer aos professores, especialmente ao meu orientador, o Prof. Dr. José Augusto Fiorucci, por todo o apoio e confiança durante a elaboração deste trabalho.

Resumo

O estudo relacionado as infrações de trânsito é importante para maior segurança nas rodovias do país. Este trabalho consiste na análise das multas relacionadas ao descumprimento do artigo 165 do Código de Trânsito Brasileiro, isto é, infrações relacionadas ao consumo de álcool. Os dados utilizados foram disponibilizados pela Polícia Rodoviária Federal e conta com observações de 2007 até 2022. Nesse estudo, foi trabalhado com a série semanal, sendo consideradas 835 semanas. Definiu-se três modelos, um SARIMA, Regressão Dinâmica Harmônica e MSTL + ARIMA e a partir do estudo de desempenho preditivo por janela deslizando, verificou-se que o harmônico obteve o melhor resultado.

Palavras-chaves: Infração de trânsito, séries temporais, modelo SARIMA, modelo harmônico, decomposição MSTL, janela deslizando.

Lista de Tabelas

| | | |
|----|---|----|
| 1 | Infrações por alcoolemia em feriados | 23 |
| 2 | Teste KPSS para verificar estacionariedade | 26 |
| 3 | Coefficientes do modelo SARIMA | 28 |
| 4 | Testes estatísticos dos resíduos do modelo SARIMA | 29 |
| 5 | Coefficientes do modelo harmônico | 30 |
| 6 | Testes estatísticos dos resíduos do modelo harmônico | 32 |
| 7 | Coefficientes do modelo MSTL + ARIMA | 32 |
| 8 | Testes estatísticos dos resíduos do modelo MSTL + ARIMA | 34 |
| 9 | Critérios de informação dos modelos | 34 |
| 10 | Erros absolutos por horizonte de previsão | 35 |

Lista de Figuras

| | | |
|----|---|----|
| 1 | Exemplo de correlograma de uma série não estacionária | 11 |
| 2 | Exemplo do gráfico pacf de uma série temporal | 13 |
| 3 | Esquema Janela Deslizante | 17 |
| 4 | Gráfico das infrações de alcoolemia por Unidade da Federação | 20 |
| 5 | Gráfico das infrações de alcoolemia por ano | 21 |
| 6 | Gráfico das infrações de alcoolemia por mês | 22 |
| 7 | Gráfico das infrações de alcoolemia por dia da semana | 22 |
| 8 | Gráfico das infrações de alcoolemia em feriados | 23 |
| 9 | Série temporal semanal das infrações de alcoolemia de 2007 a 2022 | 24 |
| 10 | Decomposição via MSTL da série das infrações de alcoolemia de 2007 a 2022 | 24 |
| 11 | Série temporal semanal das infrações de alcoolemia de 2007 a 2022, aplicada a transformação de Box-Cox | 25 |
| 12 | Decomposição via MSTL da série das infrações de alcoolemia de 2007 a 2022 com transformação de Box-Cox | 25 |
| 13 | Gráfico dos resíduos do modelo SARIMA(1,1,4)(2,0,0) | 28 |
| 14 | Análise de resíduos do modelo SARIMA(1,1,4)(2,0,0) | 29 |
| 15 | Gráfico de resíduos do modelo harmônico | 31 |
| 16 | Análise de resíduos do modelo harmônico | 31 |
| 17 | Análise de resíduos do modelo MSTL + ARIMA | 33 |
| 18 | Análise de resíduos do modelo MSTL + ARIMA | 33 |
| 19 | Gráfico dos horizontes e seus erros de previsão | 35 |
| 20 | Previsão do Modelo de Regressão Dinâmico Harmônico | 36 |

Sumário

| | |
|--|----|
| 1 Introdução | 8 |
| 2 Revisão Bibliográfica | 10 |
| 2.1 Séries Temporais | 10 |
| 2.1.1 Componentes de Séries Temporais | 10 |
| 2.1.2 Estacionariedade | 11 |
| 2.1.3 Transformação de Box-Cox | 12 |
| 2.2 Modelos Autorregressivos: $AR(p)$ | 12 |
| 2.3 Modelos de Média Móvel: $MA(q)$ | 13 |
| 2.4 Modelos Autorregressivo Integrado de Médias Móveis: ARIMA. | 13 |
| 2.5 Modelo SARIMA. | 14 |
| 2.5.1 MSTL + ARIMA | 14 |
| 2.6 Modelos de Regressão Dinâmica | 14 |
| 2.7 Modelos de Regressão Harmônica. | 15 |
| 2.8 Adequação do Modelo | 15 |
| 2.9 Estudo de Desempenho Preditivo. | 16 |
| 2.9.1 Janela Deslizante | 16 |
| 2.9.2 Erro Absoluto Médio | 17 |
| 3 Metodologia | 18 |
| 3.1 Banco de Dados | 18 |
| 3.2 Técnicas Utilizadas. | 18 |
| 4 Resultados | 20 |
| 4.1 Análise Exploratória | 20 |
| 4.2 Análise da Série | 24 |
| 4.3 Definição do Modelo | 26 |
| 4.3.1 Variáveis Explicativas | 26 |
| 4.3.2 Modelo SARIMA | 27 |

| | | |
|----------|--|-----------|
| 4.3.3 | Modelo Harmônico | 30 |
| 4.3.4 | Modelo MSTL + ARIMA | 32 |
| 4.4 | Comparação dos Modelos | 34 |
| 4.4.1 | Critérios de Informação | 34 |
| 4.4.2 | Estudo de Desempenho Preditivo | 35 |
| 5 | Conclusão | 37 |
| | Referências. | 38 |

1 Introdução

Práticas de desrespeito as leis de trânsito são muito frequentes na vida dos brasileiros e essas atitudes provocam milhares de mortes, feridos e danos materiais (TORCATE et al., 2021). O Instituto de Pesquisa Econômica Aplicada (IPEA) estima que, no Brasil, ocorrem cerca de 45 mil mortes anualmente e são gastos R\$ 50 bilhões com despesas materiais provocadas por acidentes de trânsito (TV SENADO, 2022).

No Brasil, as pessoas são muito dependentes do sistema rodoviário para sua mobilidade diária, sendo assim, com objetivo de garantir a segurança no trânsito por meio de normas de comportamento, a Lei nº 9.503 criou o Código de Trânsito Brasileiro (CTB) (CZERWONKA, 2016). No Art. 161 do CTB, infração de trânsito está definido como o descumprimento de qualquer norma estabelecida no Código.

Neto, Iglesias e Günther (2013) mostram que, diante do crescente problema dos acidentes de trânsito, é preciso desenvolver medidas de precaução que procurem reduzir sua incidência. As infrações de trânsito desempenham um papel muito importante na garantia da segurança das vias públicas, uma vez que ajudam a acabar com comportamentos de risco intencionais dos motoristas e com isso, na redução de mortes e ferimentos no trânsito (BRASIL, 2022).

Segundo Carvalho (2020), o consumo de álcool antes de dirigir, diminui a capacidade psicomotora do condutor e muitas vezes, além de levar o motorista a cometer outras infrações, torna-se um grande fator nas causas de acidentes. Por isso, no artigo 165 do Código de Trânsito Brasileiro, dirigir sob influência de álcool é estabelecido como uma infração de trânsito.

Nesse contexto, a análise de dados relacionados a violação das leis de trânsito, especialmente as associadas ao consumo de álcool, é de extrema importância para que se consiga melhorar a segurança nas rodovias (SIQUEIRA, 2015). A polícia rodoviária federal (PRF) afirma que esses dados são essenciais para tomadas de decisões em relação a políticas públicas (PRF, 2022).

Devido a essa necessidade de se analisar os dados a respeito das infrações cometidas, esse projeto torna-se necessário para que os órgãos responsáveis pela fiscalização das rodovias possuam mais fundamentos em suas tomadas de decisões e assim consigam deixar o trânsito mais seguro.

Neste trabalho serão aplicadas conhecimentos na área de séries temporais para analisar os padrões das infrações de trânsito relacionadas ao consumo de álcool ocorridas

no Brasil, buscando conhecer o mecanismo da série para que se consiga fazer previsões. Além disso, procura-se verificar a existência de padrões ao longo das semanas e dos meses do ano, e também, busca-se analisar se o comportamento da série muda em semanas com feriados ou eventos regionais.

A partir dos dados compartilhados pela Polícia Rodoviária Federal, serão definidos três modelos a serem estudados. A série trabalhada conta com dados semanais, sendo esperado que exista sazonalidade e por isso, o primeiro modelo ajustado será o SARIMA.

O segundo modelo estudado será um modelo de regressão dinâmico harmônico, pois como os dados são semanais, a série conta com ciclo complexo. Outra maneira escolhida para abordar a sazonalidade complexa, é a retirada da sazonalidade via MSTL (Multiple Seasonal Decomposition of Time Series) e então será ajustado o modelo ARIMA. A comparação do desempenho preditivo dos modelos estudados será realizada via janela deslizante.

Dado os objetivos deste trabalho, na próxima seção, será apresentada a revisão bibliográfica desse trabalho, indicando a teoria necessária para compreensão do estudo e na seção três, além da explicação do banco de dados, serão descritas as técnicas utilizadas. A seção quatro apresenta os resultados encontrados, ela conta com uma análise descritiva dos dados, a análise da série e o estudo dos modelos. Já na seção cinco será feita a conclusão dos resultados observados.

2 Revisão Bibliográfica

Nessa seção serão apresentados os principais conceitos e técnicas necessários para a melhor compreensão do estudo realizado.

2.1 Séries Temporais

Fischer (1982), define uma série temporal como um conjunto de observações dispostas sequencialmente no tempo, e as classifica como discreta, caso esse conjunto seja finito ou infinito enumerável e como contínua, quando é infinito não enumerável. Já Morrettin e Tolo (2018) afirmam que, uma série temporal é uma das várias trajetórias que poderiam ter sido observadas em um processo aleatório.

2.1.1 Componentes de Séries Temporais

Uma série temporal, pode ser vista como uma combinação de três componentes principais: tendência, sazonalidade e resíduos (FISCHER, 1982). A tendência é o comportamento principal da série, indicando o sentido e a velocidade de crescimento ou decrescimento das observações no tempo (SOUSA et al., 2021). Já a sazonalidade é a variação periódica da série e os resíduos são os comportamentos aleatórios, que não estão relacionados as observações anteriores, tendência e nem sazonalidade (FISCHER, 1982).

Hyndman e Athanasopoulos (2018), mostram que, considerando uma série (y_t) e sendo S_t a componente de sazonalidade, T_t a de tendência e R_t os resíduos, as séries podem ser decompostas pelos métodos:

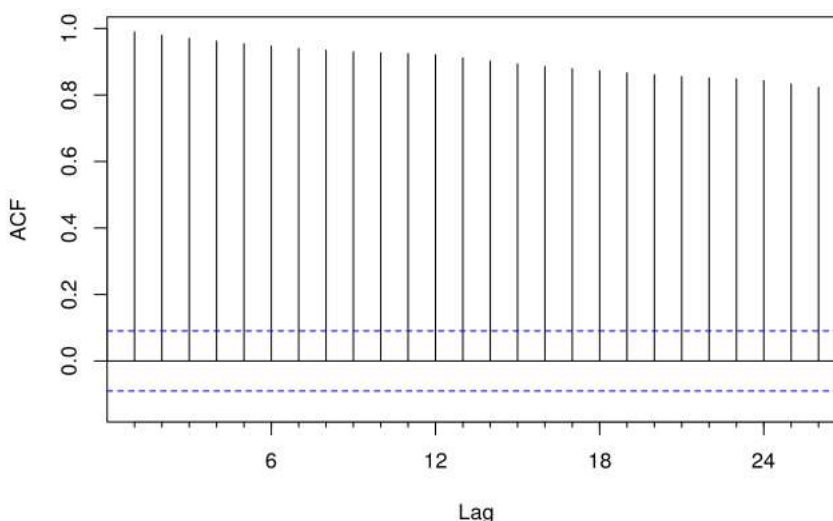
- Aditivo: $y_t = S_t + T_t + R_t$, mais indicado quando a magnitude da sazonalidade ou a tendência não varia ao longo do tempo.
- Multiplicativo: $y_t = S_t \times T_t \times R_t$, usado nos casos em que a variação do padrão sazonal ou da tendência aparenta ser proporcional ao longo do tempo na série.
- Seasonal Trending using Loess (STL): A componente sazonal pode mudar ao longo do tempo. É a mais indicada para lidar com outliers.

2.1.2 Estacionariedade

Uma série é dita estacionária quando se desenvolve aleatoriamente ao redor de uma média constante ao longo do tempo (SOUSA et al., 2021). Existem maneiras de verificar se a série é ou não estacionária, sendo as principais: a análise do gráfico da função de autocorrelação (fac) e o teste KPSS (JOSE, 2022).

A fac é definida por: $\rho_\tau = \frac{COV(Y_t, Y_{t-\tau})}{Var(Y_t)}$ e mede a relação linear entre dois pontos da série, variando entre -1 e +1 (JOSE, 2022). Fischer (1982) afirma que, ao analisar o correlograma, caso se verifique que o coeficiente de autocorrelação não diminui rapidamente a medida que a defasagem aumenta, indica que a série não é estacionária. A seguir um exemplo de um correlograma de uma série não estacionária, apresentado por Yakubu e Saputra (2022).

Figura 1: Exemplo de correlograma de uma série não estacionária



Fonte: Yakubu e Saputra, 2022

Em relação ao teste KPSS (Kwiatkowski Philips Schmidt e Shin), Silveira, Matos e Konrath (2016) apresentam a estatística de teste como: $\sum_{t=1}^N \frac{S_t^2}{N^2 \hat{\sigma}_s^2}$, e definem as seguintes hipóteses:

$$\begin{cases} H_0 : \text{A série é estacionária} \\ H_1 : \text{A série não é estacionária} \end{cases}$$

Verificar se as séries são estacionárias é muito importante, uma vez que grande parte das análises de séries temporais supõem estacionariedade. Por isso, existem técnicas para transformar séries em estacionárias, sendo a mais comum tomar diferenças sucessivas da série original, até que esta seja estacionária (MORETTIN; TOLOI, 2018).

2.1.3 Transformação de Box-Cox

Em alguns casos, os dados apresentam diferentes padrões de variação ao longo do tempo e então, pode ser útil aplicar alguma transformação (ATKINSON; RIANI; CORBELLINI, 2021). A transformação de Box-Cox, definida por Box e Cox (1964), consiste em aplicar a série a transformação:

$$y^{(t)} = \begin{cases} \frac{y^{\lambda}-1}{\lambda}, (\lambda \neq 0), \\ \log y, (\lambda = 0), \end{cases}$$

sendo y a série original, $y^{(t)}$ a série transformada e λ um número real. Para definir o melhor valor de λ , busca-se o valor em que λ maximiza a equação $L_{max(\lambda)} = -\frac{n}{2} \log \frac{SQE(\lambda)}{n}$, sendo $L_{max(\lambda)}$ a função de máxima verossimilhança de λ e $SQE(\lambda)$ a soma de quadrados residual (AZEVEDO; JUNIOR; FERNANDES2, 2016).

2.2 Modelos Autorregressivos: AR(p)

Morettin (2014) define um processo autorregressivo de ordem p como um processo estacionário X_t que satisfaz:

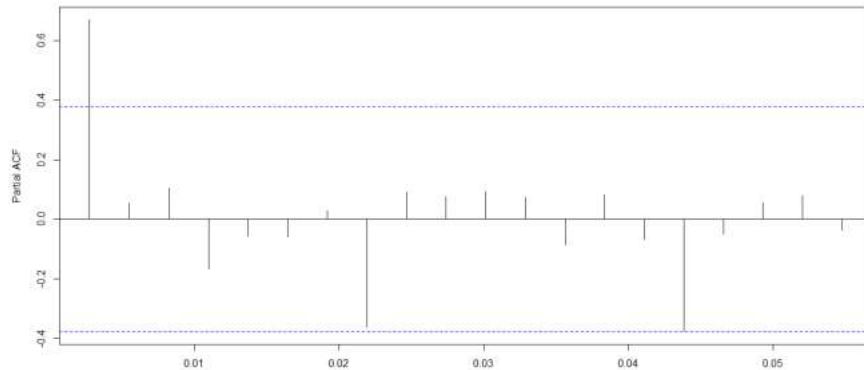
$$x_t = \mu + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \epsilon_t$$

onde $\mu, \phi_1 \dots \phi_p$ são parâmetros e ϵ_t é um ruído branco com média zero e variância $\hat{\sigma}^2$. E ainda, Morettin e Tolo (2018) apresentam a equação como: $\phi(B)x_t = \epsilon_t$, onde $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ é o polinômio autorregressivo de ordem p e sendo B o operador de retardo ($B^p x_t = x_{t-p}$).

Reimbold et al. (2017) afirmam que um processo AR(p) é estacionário se o módulo das raízes do polinômio autorregressivo são maiores do que um, ou seja, não possui raiz unitária.

A partir da análise do gráfico da função de autocorrelação parcial (facp), que se obtém ao plotar os valores de ϕ_p obtidos nos modelos AR(p) para os possíveis valores de p , é possível determinar a ordem "p" do modelo verificando quais lags estão fora dos limites (YAKUBU; SAPUTRA, 2022). A seguir um exemplo de um gráfico facp, apresentado por Yakubu e Saputra (2022).

Figura 2: Exemplo do gráfico pacf de uma série temporal



Fonte: Yakubu e Saputra, 2022

Para identificar um modelo $AR(p)$: a série deve ser estacionária, a fac decai de forma amortizada para zero e a facp é zero a partir do lag p .

2.3 Modelos de Média Móvel: $MA(q)$

Segundo Morettin e Tolo (2018), um processo linear $X = \{X_t : t = 1, 2, \dots\}$ é dito um processo de médias móveis de ordem "q" se, sendo ϵ_t um ruído branco com média zero e variância $\hat{\sigma}^2$ e sendo $\theta_1, \dots, \theta_q$ constantes, pode ser escrito como:

$$x_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}$$

Podendo também ser escrito como: $X_t = \Theta(B)\epsilon_t$, onde $\Theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$ é o polinômio de médias móveis de ordem q e B é o operador de retardo (SOUSA et al., 2021).

Para identificar um modelo $MA(q)$: a série deve ser estacionária, a fac converge para zero com corte a partir de q e a facp converge para zero de forma amortizada.

2.4 Modelos Autorregressivo Integrado de Médias Móveis: ARIMA

O modelo ARIMA é composto pela combinação do modelo de médias móveis com o autorregressivo, aplicado a uma transformação da série temporal (YAKUBU; SAPUTRA, 2022). Morettin (2014) define um processo X_t como $ARIMA(p, d, q)$ se:

$$\phi(B)\Delta^d X_t = \theta(B)\epsilon_t$$

onde $\phi(B)$ é o polinômio autorregressivo, $\theta(B)$ o polinômio de médias móveis e Δ^d o operador de diferença de ordem d .

Em relação a ordem do modelo, Hyndman e Athanasopoulos (2018) definem "d" como o número de diferenciações necessárias até que a série se torne estacionária, "p" é a ordem da parte autorregressiva e "q" a da parcela de médias móveis.

2.5 Modelo SARIMA

O modelo SARIMA é o modelo ARIMA quando se considera as relações sazonais, sendo definido X_t um SARIMA quando:

$$\phi(B)\Phi(B)\Delta^d\Delta^D X_t = \theta(B)\Theta(B)\epsilon_t$$

sendo $\phi(B)$ o polinômio autorregressivo não sazonal, $\Phi(B)$ o autorregressivo sazonal, $\theta(B)$ o polinômio de médias móveis não sazonal, $\Theta(B)$ o de médias móveis sazonal, Δ^d o operador de diferença não sazonal e Δ^D o de diferença de sazonal (BARROS et al., 2018).

Representado por $ARIMA(p, d, q)x(P, D, Q)_S$, onde (p, q, d) representa a parte sem sazonalidade e $(P, D, Q)_S$ a parcela sazonal do modelo (DUBEY et al., 2021).

2.5.1 MSTL + ARIMA

Em alguns casos, quando a série possui múltipla sazonalidade ou até ciclos complexos, isto é, não inteiros, faz-se a previsão via MSTL + ARIMA (CLEVELAND et al., 1990). Nesse caso, primeiro retira-se a sazonalidade via MSTL e utilizando o modelo ARIMA, faz-se a previsão com a sazonalidade ajustada, e por fim, replica-se a sazonalidade no ultimo ciclo da previsão (CLEVELAND et al., 1990).

2.6 Modelos de Regressão Dinâmica

Modelos de regressão dinâmica são uma combinação da série com a influencia de variáveis explicativas, sendo utilizados quando verifica-se que a hipótese de independência dos erros não é aceita e, também, quando a variável de interesse é dependente de variáveis causais (HYNDMAN; ATHANASOPOULOS, 2018).

Os modelos são definidos como:

$$\begin{cases} y_t = \beta_0 + \beta_1 x_{1,t} + \dots + \beta_k x_{k,t} + \eta_t \\ \eta_t = \phi_1 \eta_{t-1} + \dots + \phi_p \eta_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \end{cases}$$

sendo y_t a variável resposta, $x_t = (x_{1,t}, \dots, x_{k,t})$ o vetor de variáveis explicativas, $\{\eta_t\}$ um processo ARMA(p, q), $\{\varepsilon_t\}$ um processo independente e identicamente distribuído com média zero e variância constante, $\beta = (\beta_0, \beta_1, \dots, \beta_k)'$ o vetor dos parâmetros regressores $(\phi_1, \dots, \phi_p)'$ o vetor dos parâmetros autoregressores e $(\theta_1, \dots, \theta_q)'$ o vetor dos parâmetros de médias móveis (ALMEIDA, 2014).

2.7 Modelos de Regressão Harmônica

O modelo de regressão harmônica representa séries sazonais com tendência linear (CAMELO, 2017). Nestes modelos, a sazonalidade é modelada utilizando séries de Fourier e por isso são indicados para casos de múltipla sazonalidade ou sazonalidade complexa (HYNDMAN; ATHANASOPOULOS, 2018).

Definido pela formula:

$$Y_t = \beta_0 + \beta_1 t + T_t + S_t + C_t + \epsilon_t$$

em que β_0 e β_1 são os coeficientes de regressão linear, T_t a tendência, S_t a sazonalidade, C_t o ciclo e ϵ_t um processo independente e identicamente distribuído com média zero e variância constante (SILVESTRE; BEZERRA, 2015).

Nesse modelo, a componente sazonal e os ciclos complexos são modelados utilizando séries de Fourier, sendo a sazonalidade $S_t = \sum (a_{it} \cos(\omega_i t) + b_{ij} \sin(\omega_i t))$, onde a_{it} e b_{it} são parâmetros estocásticos de variação do tempo, ω_t a frequência associada a sazonalidade da série e sendo o ciclo $C_i = \sum (\alpha_{it} \cos(f_i t) + \beta_{it} \sin(f_i t))$, onde α_{it} e β_{it} são parâmetros estocásticos de variação do tempo e f_i as frequências associadas a componente cíclica (YOUNG, 1999).

2.8 Adequação do Modelo

Uma técnica utilizada para verificar se o modelo escolhido é adequado é por meio da análise dos resíduos, isto é, conferir se os pressupostos sobre ϵ_t foram atendidos (MORETTIN; TOLOI, 2018). Além da análise gráfica, certos testes estatísticos podem

ser utilizados para verificar os pressupostos.

O pressuposto da estacionariedade pode ser analisado pelo teste KPSS, apresentado anteriormente (SOUSA et al., 2021). Já a normalidade pode ser verificada pela aplicação do teste de normalidade de Shapiro-Wilk, que a partir da estatística $W = \frac{1}{\sum (X_i - \bar{X})^2} [\sum a_i (X_{(n-i+1)} - X_{(i)})]$, busca verificar as hipóteses:

$$\begin{cases} H_0 : \text{A variável segue uma distribuição Normal} \\ H_1 : \text{A variável segue outro modelo} \end{cases}$$

Em relação a autocorrelação, Morettin (2014) indica o teste de Ljung-Box como o mais adequado. Esse teste verifica as hipóteses:

$$\begin{cases} H_0 : \rho_1^2 = \rho_2^2 = \dots = \rho_k^2 = 0 \\ H_1 : \rho_k^2 \neq 0 \end{cases}$$

a partir da estatística: $Q = n(n+2) \sum \frac{\rho_k^2}{n-k}$ que segue uma distribuição χ^2 com $K - p - q$ graus de liberdade.

2.9 Estudo de Desempenho Preditivo

2.9.1 Janela Deslizante

O método de validação cruzada, é utilizado para avaliar o desempenho preditivo de um modelo. A técnica consiste em separar os dados em duas partes, sendo uma usada para definir o modelo e a outra para validá-lo, a partir de métricas que avaliam a qualidade do ajuste (BROWNE, 2000).

A janela deslizante, é um método de validação cruzada, em que cada horizonte de previsão é avaliado mais de uma vez, por meio de janelas de treinamento e de teste (CHU, 1995). Nesse método, a cada novo período, modelo é estimado novamente e repete-se o procedimento, como é ilustrado na Figura 3.

Figura 3: Esquema Janela Deslizante



Fonte: Notas de aula do Prof. Dr. José Augusto Fiorucci

2.9.2 Erro Absoluto Médio

O Erro Absoluto Médio (EMA) mede a média da diferença absoluta entre os valores previstos pelo modelo e os valores observados, sendo uma métrica utilizada para avaliar o desempenho do modelo e definida como:

$$EMA = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

onde y_i é o valor real observado e \hat{y}_i o valor previsto pelo modelo e n o número de previsões realizadas (WILLMOTT, 2005).

3 Metodologia

Uma vez que esse trabalho foca na avaliação de modelos de séries temporais para prever o número de incidentes de multas de trânsito por consumo de álcool, essa seção diz respeito ao banco de dados e as técnicas utilizadas durante o estudo.

3.1 Banco de Dados

Os dados utilizados na realização deste trabalho são disponibilizados pela Polícia Rodoviária Federal, podendo ser encontrados no site da PRF.

No site, está disponível um banco de dados para cada mês de cada ano. Para a realização desse estudo, esses arquivos foram combinados, por linhas, no software R, ficando então um banco com todas as observações de 2007 a 2022.

Depois, foram selecionadas apenas as observações relativas ao artigo 165 do Código de Trânsito Brasileiro, que define como uma infração gravíssima: "Dirigir sob a influência de álcool ou de qualquer outra substância psicoativa que determine dependência" e também: "Recusar-se a ser submetido a teste, exame clínico, perícia ou outro procedimento que permita certificar influência de álcool ou outra substância psicoativa" (Brasil, 2008).

Sendo assim, o banco de dados utilizado para a realização das análises conta com dados de 2007 a 2022, contendo 556093 observações, sendo todas ocorrências relacionadas ao artigo 165 do CTB.

3.2 Técnicas Utilizadas

Em relação as técnicas utilizadas, vale destacar que, assim como as manipulações realizadas no banco de dados, todas as análises foram feitas utilizando o software R.

Nesse trabalho, a partir do banco definido na Seção 3.1, após a análise exploratória, os dados serão agrupados por semanas, sendo uma semana considerada de domingo a sábado, e será estabelecida uma série temporal semanal, ou seja, com frequência complexa igual a $365/7$. A série será trabalhada utilizando o pacote `forecast` do R.

Depois, serão considerados três modelos, sendo um SARIMA, um modelo de regressão dinâmica harmônica e MSTL + ARIMA, e a partir da função `auto.arima`, serão definidas a ordem de cada um.

Após a definição dos modelos, será feito um estudo de desempenho preditivo por janela deslizante, utilizando a função tsCV. Escolhido o modelo com o melhor comportamento, será feita a previsão de infrações para o ano de 2023.

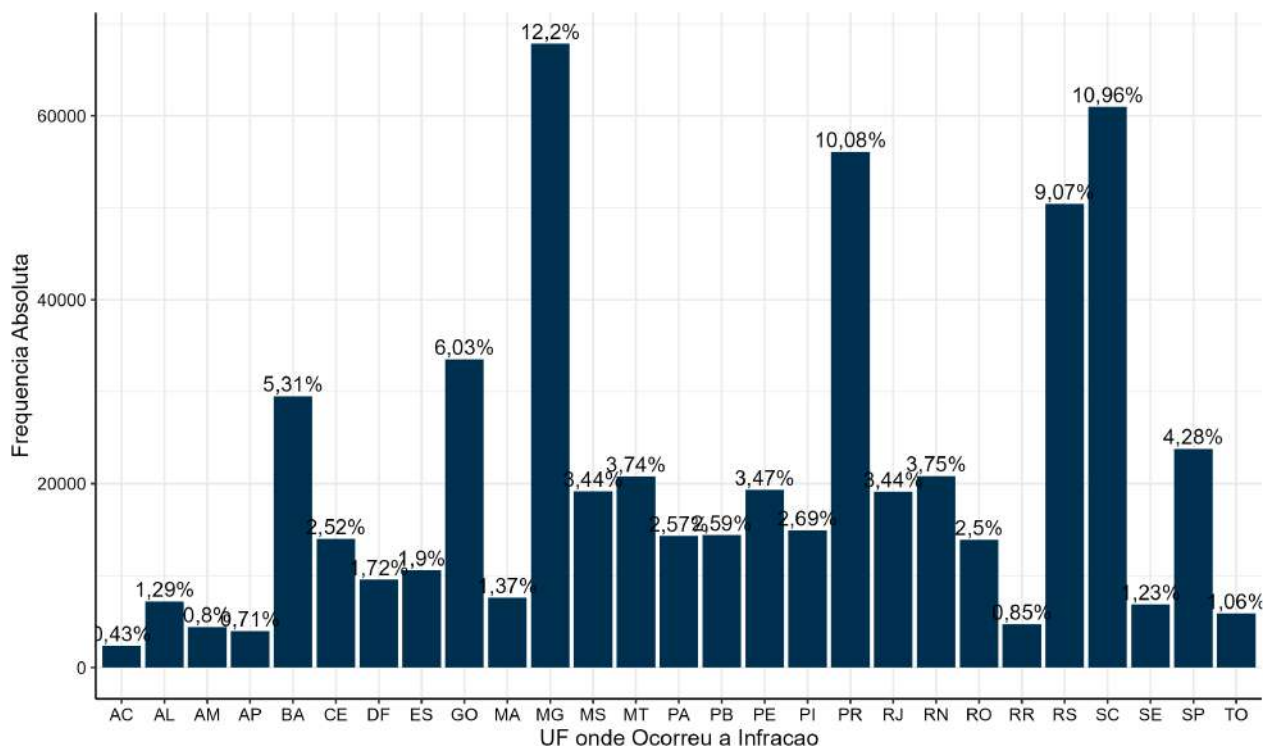
4 Resultados

Nessa seção, será apresentada a análise descritiva dos dados, bem como o ajuste dos modelos e um estudo do desempenho preditivo, afim de definir o modelo que melhor se ajusta a série observada.

4.1 Análise Exploratória

Na Figura 4 é apresentado o percentual de ocorrência de infrações por Unidade Federativa.

Figura 4: Gráfico das infrações de alcoolemia por Unidade da Federação



A partir da análise da Figura 4, percebe-se que Minas Gerais foi o estado com maior número de infrações relacionadas a ingestão de álcool de 2007 a 2022 (12,2%), seguido do estado de Santa Catarina (10,96%) e Paraná (10,0%). Apenas 1,72% das multas ocorreram no Distrito Federal e os estados com o menor percentual foram o Acre (0,43%) e Roraima (0,85%).

A Figura 5 apresentada a seguir, mostra o número de infrações por ano.

Figura 5: Gráfico das infrações de alcoolemia por ano

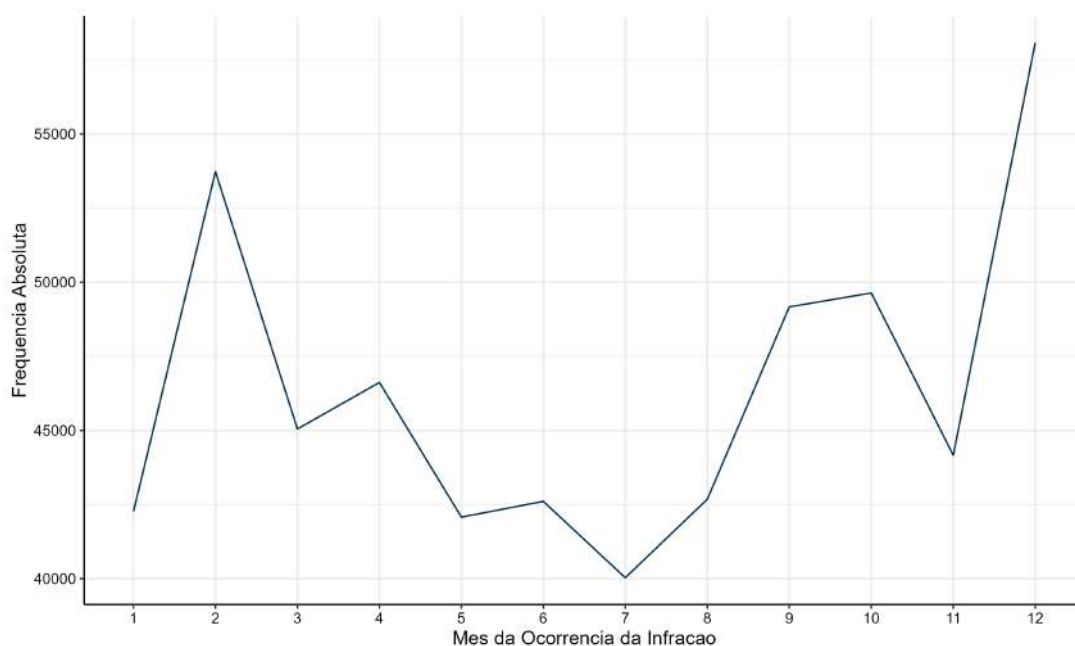


Ao analisar a Figura 5, percebe-se que a quantidade de infrações tende a aumentar ao longo dos anos. Vale notar o efeito de algumas mudanças do artigo 165, como por exemplo, em junho de 2008, a norma passou a considerar como infração dirigir com qualquer concentração de álcool (BUSCH et al., 2023).

Em 2020, o Brasil foi afetado pela pandemia do corona vírus, gerando um grande impacto na vida das pessoas, fazendo com que muitos ficassem em casa (BARCELLOS; XAVIER, 2022), o que provocou uma queda nas infrações, como pode ser visto no gráfico. Percebe-se que em 2022, o número de multas volta a crescer.

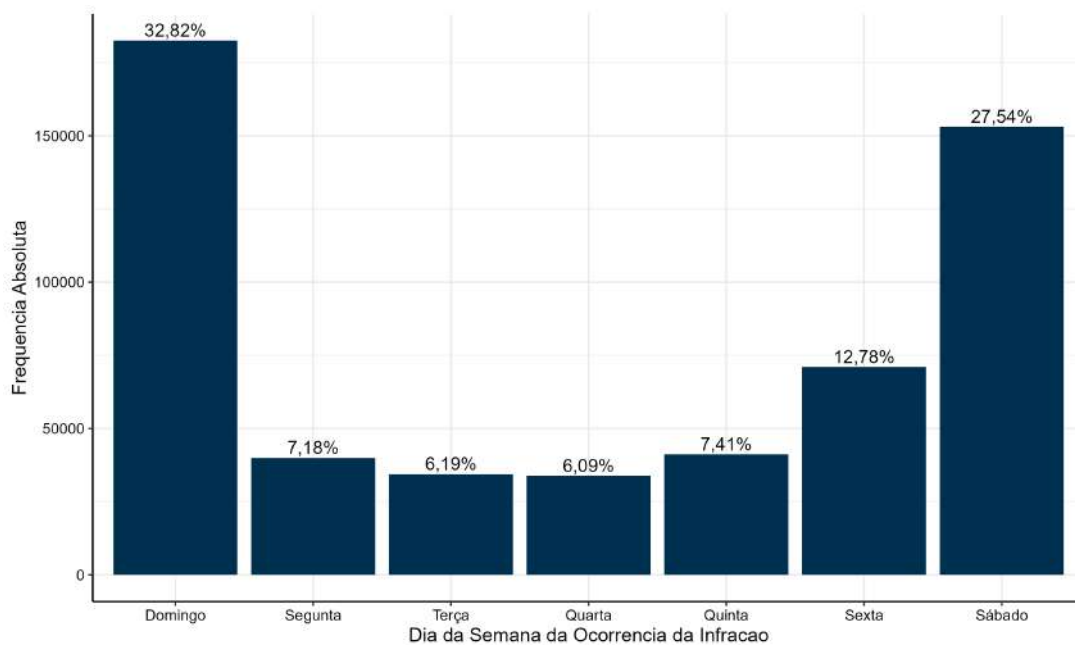
Ainda em relação a data em que ocorreram as infrações, a Figura 6 mostra que das infrações cometidas de 2007 a 2022, 58080 (10,45%) foram observadas em dezembro e 53726 (9,66%) em fevereiro, sendo esses os meses com maior quantidade de multas. Já julho, apresentou 40030 (7,20%), sendo o mês com a menor frequência.

Figura 6: Gráfico das infrações de alcoolemia por mês



Já em relação ao dia da semana, a Figura 7 apresenta essa informação.

Figura 7: Gráfico das infrações de alcoolemia por dia da semana



Ao analisar a Figura 7, nota-se que cerca de 60% das infrações foram cometidas em finais de semana, sendo 32,82% em domingos e 27,54% em sábados. Sexta-feira é o terceiro dia com mais observações, representando 12,78%, já a quarta-feira é o dia da semana com a menor frequência (6,09%).

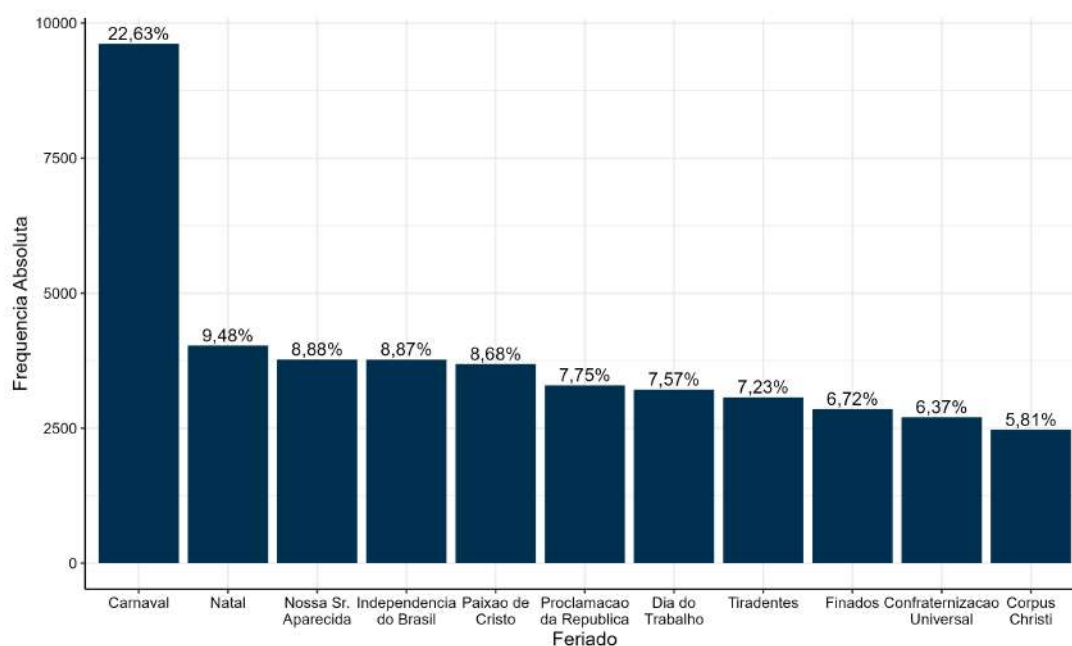
Além da notória diferença do número de infrações aos finais de semana para os outros dias da semana, verifica-se na Tabela 1, apresentada a seguir, que a média de infrações em feriados (221) é 2,5 vezes maior do que em um dia comum (90).

Tabela 1: Infrações por alcoolemia em feriados

| Estatística | Feriado | Não Feriado |
|---------------|---------|-------------|
| Média | 221,22 | 90,87 |
| Desvio Padrão | 128,78 | 99,89 |
| Máximo | 781 | 1135 |
| Total | 42475 | 513618 |

Sendo assim, a Figura 8 mostra a frequência das violações de trânsito por feriado.

Figura 8: Gráfico das infrações de alcoolemia em feriados

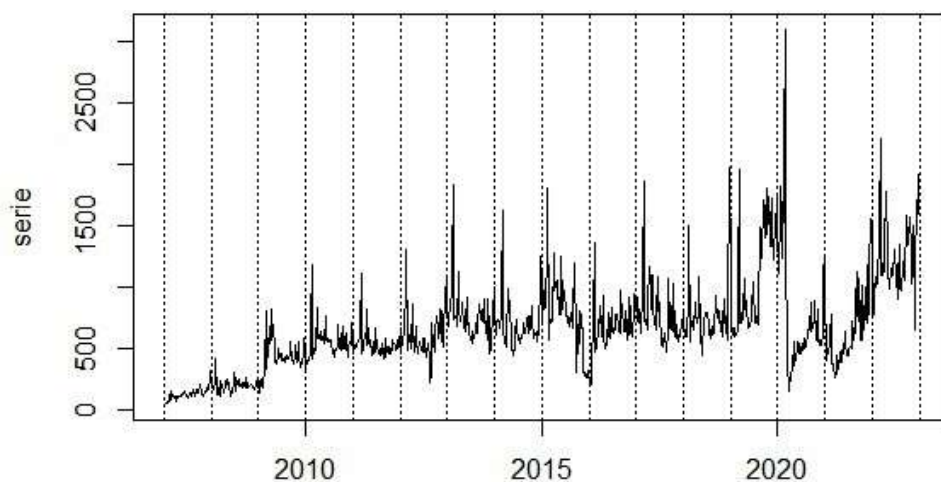


Ao observar a Figura 8, verifica-se que o carnaval apresenta a maior frequência, representado 22,63% das infrações cometidas em feriados, seguido do natal com 9,48%. Já em relação aos demais feriados, Corpus Christi apresentou a menor frequência com apenas 5,81% das observações.

4.2 Análise da Série

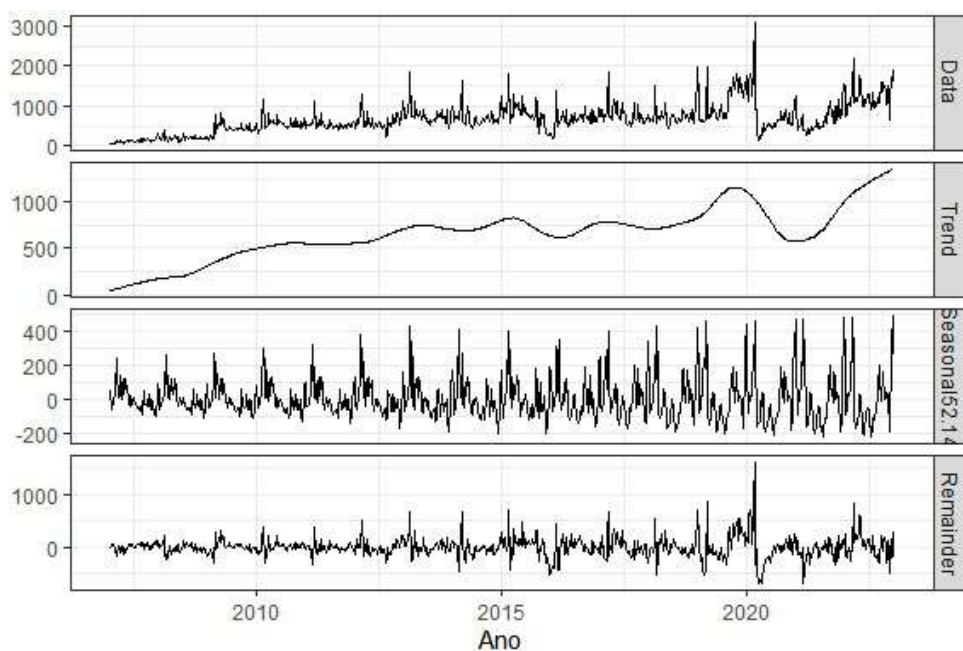
A Figura 9 apresenta a série temporal das infrações de trânsito por conduzir o veículo sob influência de álcool ou outras substâncias entorpecente, cometidas de 2007 a 2022.

Figura 9: Série temporal semanal das infrações de alcoolemia de 2007 a 2022



A série é formada por dados semanais, contando com 835 semanas. Para melhor visualização, foi feita a decomposição via MSTL, apresentada na Figura 10.

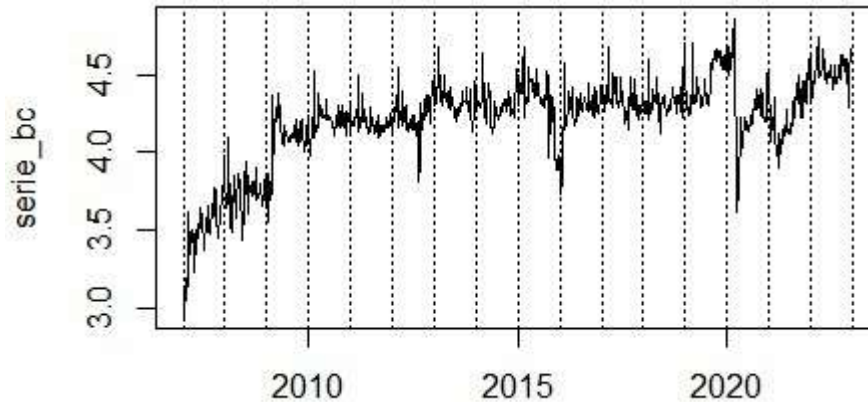
Figura 10: Decomposição via MSTL da série das infrações de alcoolemia de 2007 a 2022



A partir da decomposição, percebe-se que a série apresenta tendência de cres-

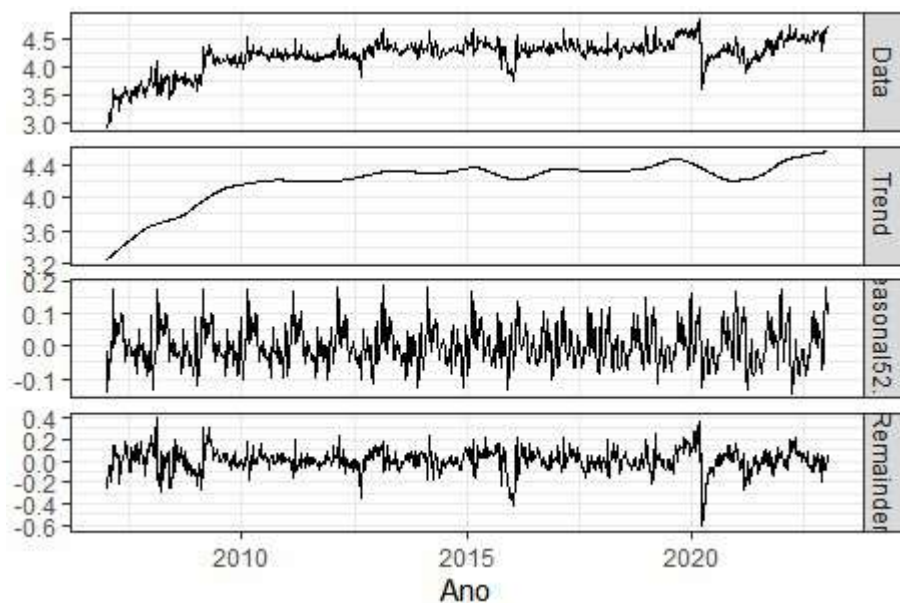
cimento linear e ciclo sazonal semanal. Também, é possível perceber que a série possui diferentes padrões de variação ao longo do tempo, então para estabilizá-la, é aplicada uma transformação de Box-Cox, considerando o λ de -0,13, escolhido automaticamente pelo software R. Após a transformação, tem-se a série a seguir.

Figura 11: Série temporal semanal das infrações de alcoolemia de 2007 a 2022, aplicada a transformação de Box-Cox



Para melhor visualização, foi feita a decomposição via MSTL (Multiple Seasonal Decomposition of Time Series), apresentada na Figura 12.

Figura 12: Decomposição via MSTL da série das infrações de alcoolemia de 2007 a 2022 com transformação de Box-Cox



Ao observar a decomposição da série após a transformação, nota-se que os dados não aparentam oscilar sobre uma média constante ao longo do tempo, indicando que a série não é estacionária. Para verificar essa hipótese, foi feito o teste KPSS, cujo resultado

está apresentado na tabela a seguir.

Tabela 2: Teste KPSS para verificar estacionariedade

| | Teste | Estatística | P-Valor |
|--|---------------------------|-------------|---------|
| | KPSS da série original | 5,85 | 0,01 |
| | KPSS após a diferenciação | 0,02 | 0,10 |

Verifica-se na Tabela 2 que o teste apresentou um p-valor menor do que 0,05, dessa forma, à um nível de significância de 5%, rejeita-se a hipótese de estacionariedade e por isso, a série precisa ser transformada. Foi feita uma diferenciação simples na série e observou-se um p-valor de 0,1, ou seja, não rejeita-se a hipótese de estacionariedade.

4.3 Definição do Modelo

4.3.1 Variáveis Explicativas

Como foi exposto anteriormente e ao observar a Figura 11, é possível notar uma grande mudança no comportamento da série a partir do ano de 2020, o que ocorre devido a pandemia do corona vírus. Por isso, acredita-se que o período da pandemia deve ser considerado como uma variável explicativa do modelo.

Diante disso, foram consideradas as quatro fases da pandemia, sendo a primeira fase de março a agosto de 2020, a segunda fase de setembro a janeiro de 2021, a terceira de fevereiro a junho de 2021 e a quarta de julho a novembro de 2021 (BARCELLOS; XAVIER, 2022), e se estabeleceu as seguintes variáveis:

$$Primeira = \begin{cases} 1, & \text{se a semana pertence a primeira fase da pandemia} \\ 0, & \text{se a semana não pertence a primeira fase da pandemia} \end{cases}$$

$$Segunda = \begin{cases} 1, & \text{se a semana pertence a segunda fase da pandemia} \\ 0, & \text{se a semana não pertence a segunda fase da pandemia} \end{cases}$$

$$Terceira = \begin{cases} 1, & \text{se a semana pertence a terceira fase da pandemia} \\ 0, & \text{se a semana não pertence a terceira fase da pandemia} \end{cases}$$

$$Quarta = \begin{cases} 1, & \text{se a semana pertence a quarta fase da pandemia} \\ 0, & \text{se a semana não pertence a quarta fase da pandemia} \end{cases}$$

Além disso, foi visto anteriormente que existe uma grande diferença no compor-

tamento das multas em semanas com feriados e por isso, foi criada a variável a seguir:

$$Feriado = \begin{cases} 1, & \text{se a semana conta com algum feriado} \\ 0, & \text{se a semana não conta com feriados} \end{cases}$$

Também, percebe-se que o carnaval se destaca em relação aos outros feriados e dias comuns, então foi definida uma variável indicando a semana do carnaval.

$$Carnaval = \begin{cases} 1, & \text{se é semana de carnaval} \\ 0, & \text{se não é semana de carnaval} \end{cases}$$

Sendo assim, define-se as variáveis explicativas: primeira, segunda, terceira e quarta fase da pandemia, feriado na semana e semana de carnaval.

4.3.2 Modelo SARIMA

A série trabalhada é sazonal e não estacionária, então, o primeiro modelo a ser considerado é o SARIMA. Dada a transformação de Box-Cox com $\lambda = -0,13$ e as variáveis explicativas definidas anteriormente, a partir da função `auto.arima` do R, encontrou-se uma diferença simples ($d = 1$), $p = 1$, $q = 4$, $P = 2$ e $Q = 0$, diferenças sazonais não foram necessárias, ou seja, $D = 0$.

Define-se então, um $SARIMA(1, 1, 4)(2, 0, 0)_{[52]}$ com os coeficientes apresentados na Tabela 3.

Tabela 3: Coeficientes do modelo SARIMA

| | Coeficientes | Desvio-Padrão |
|----------|--------------|---------------|
| AR(1) | -0.90 | 0,12 |
| MA(1) | 0.69 | 0,13 |
| MA(2) | -0.48 | 0,05 |
| MA(3) | -0.32 | 0,06 |
| MA(4) | -0.07 | 0,03 |
| sAR(1) | 0.04 | 0,03 |
| sAR(2) | 0.02 | 0,03 |
| Fase 1 | -0.28 | 0,06 |
| Fase 2 | -0.18 | 0,08 |
| Fase 3 | -0.19 | 0,08 |
| Fase 4 | -0.05 | 0,06 |
| Feriado | 0.08 | < 0,01 |
| Carnaval | 0.19 | 0,01 |

Ao observar os coeficientes de regressão apresentados na Tabela 3, nota-se que as variáveis relacionadas a pandemia apresentaram valor negativo, ou seja, durante esse período, a quantidade de infrações semanais diminui, sendo a primeira fase a que representa a maior variação.

Além disso, verifica-se que o fato de possuir feriado na semana aumenta em média 0,08 o número de infrações cometidas. Em particular, na semana do carnaval, observa-se o aumento de em média 0,19 em relação as outras semanas.

Para verificar a adequação do modelo, é feita a análise de resíduos a seguir.

Figura 13: Gráfico dos resíduos do modelo SARIMA(1,1,4)(2,0,0)

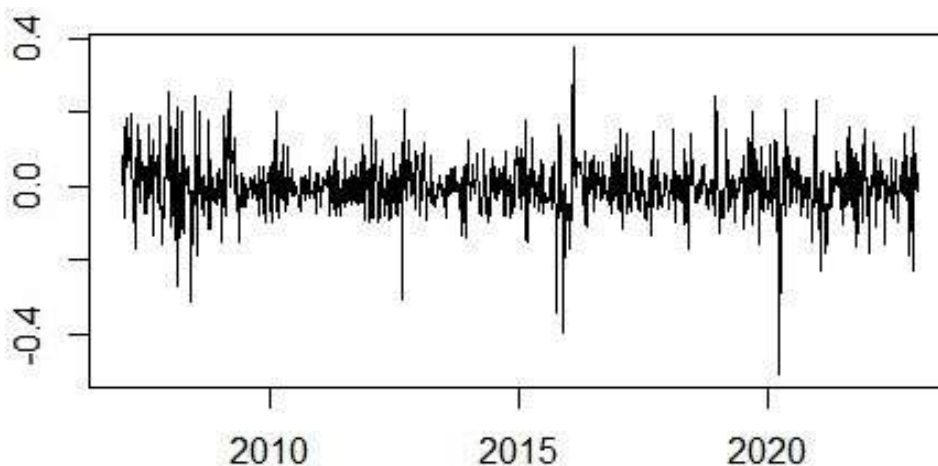
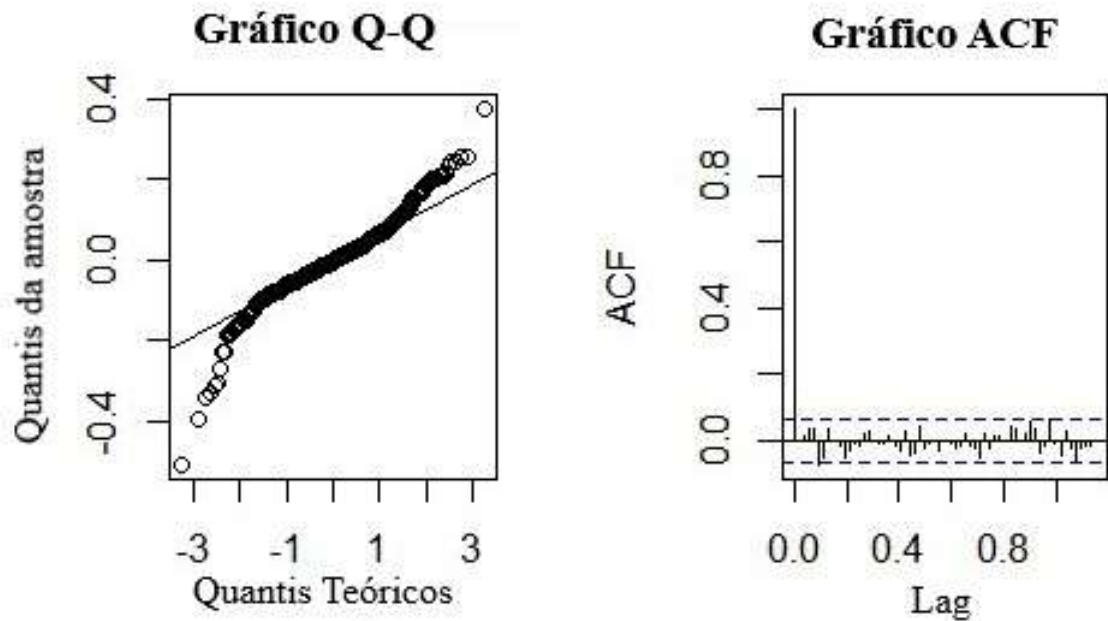


Figura 14: Análise de resíduos do modelo SARIMA(1,1,4)(2,0,0)



Ao observar a Figura 13, os resíduos aparentam oscilar em torno da média igual a zero e com variância constante. Porém, é interessante que notar na Figura 14 os resíduos não aparentam seguir normalidade. Em relação ao gráfico ACF verifica-se que não sobrou autocorrelação nos resíduos. Para confirmar a análise visual, foram realizados testes cujos resultados estão exibidos na tabela a seguir.

Tabela 4: Testes estatísticos dos resíduos do modelo SARIMA

| Teste | Estatística | P-Valor |
|--------------|-------------|---------|
| KPSS | 0,27 | 0,10 |
| Shapiro-Wilk | 0,94 | < 0,01 |
| Ljung-Box | 14,48 | 0,48 |

A Tabela 4 mostra que o teste KPSS apresentou um p-valor de 0,10, ou seja, não rejeita-se a hipótese de estacionariedade dos resíduos. Em relação a independência, o resultado do teste Ljung-Box indica que não há motivos para rejeitar a hipótese de independência.

Já o teste de Shapiro-Wilk, indica que deve-se rejeitar a hipótese de normalidade, porém, isso não é um problema, uma vez que previsões intervalares podem ser obtidas de modo alternativo via o procedimento de Bootstrap não paramétrico.

4.3.3 Modelo Harmônico

O segundo modelo testado é o modelo de regressão dinâmica harmônico, sendo escolhido devido ao fato da série contar com ciclo sazonal complexo, isto é, não inteiro, o que não é tratado por modelos SARIMA tradicionais. Considerando a transformação de Box-Cox com $\lambda = -0,13$ e as variáveis explicativas definidas, o modelo apresentou os coeficientes indicados na Tabela 5.

Tabela 5: Coeficientes do modelo harmônico

| | Coeficientes | Desvio-Padrão |
|----------|--------------|---------------|
| AR(1) | 0,62 | 0,07 |
| AR(2) | -0,13 | 0,04 |
| AR(3) | 0,11 | 0,04 |
| MA(1) | -0,85 | -0,27 |
| Fase 1 | -0,27 | 0,06 |
| Fase 2 | -0,19 | 0,07 |
| Fase 3 | -0,22 | 0,08 |
| Fase 4 | -0,05 | 0,06 |
| Feriado | 0,08 | < 0,01 |
| Carnaval | 0,19 | 0,01 |
| S1 - 52 | 0,14 | 0,14 |
| C1 - 52 | -0,004 | 0,01 |
| S2 - 52 | 0,01 | < 0,01 |
| C2 - 52 | -0,01 | < 0,01 |
| S3 - 52 | 0,01 | < 0,01 |
| C3 - 52 | -0,004 | < 0,01 |
| S4 - 52 | -0,006 | < 0,01 |
| C4 - 52 | -0,01 | < 0,01 |
| S5 - 52 | -0,01 | < 0,01 |
| C5 - 52 | 0,002 | < 0,01 |

A Tabela 5 mostra que o coeficiente relacionado a variável fase 1 é igual a -0,27, ou seja, nas semanas da primeira fase da pandemia, as infrações diminuíram, em média, 0,27. Em relação as outras variáveis relacionadas a pandemia, a variável fase 2 observou-se um coeficiente de -0,19, a fase 3 de -0,22 e a fase quatro de -0,05.

Percebe-se que nas semanas em que a variável feriado é igual a 1, isto é, nas

semanas que possuem feriados, ocorre um aumento médio de 0,08 no número de infrações observadas. Já em relação a variável carnaval, esse aumento médio é de 0,19.

Após o ajuste do modelo, encontrou-se os resíduos apresentados na Figura 15. Para verificar a adequação do modelo, é feita a análise de resíduos a seguir.

Figura 15: Gráfico de resíduos do modelo harmônico

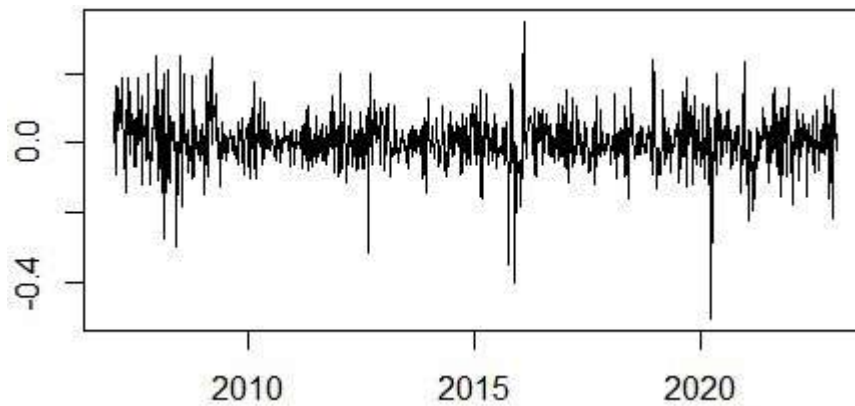
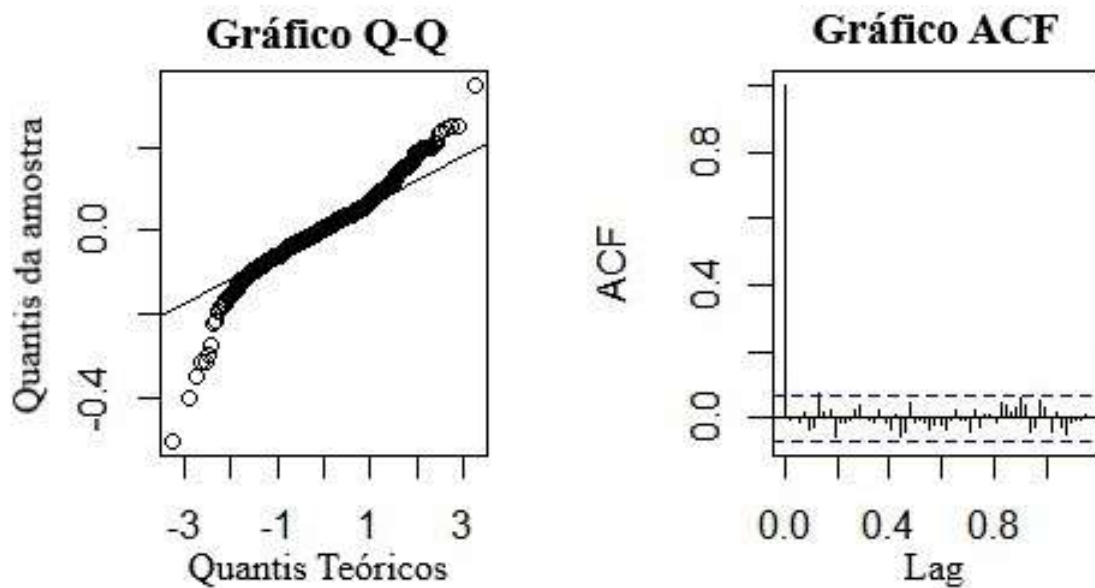


Figura 16: Análise de resíduos do modelo harmônico



A Figura 15 mostra que os resíduos do modelo parecem variar de maneira constante em torno da média zero, porém, nota-se na Figura 16 que eles não aparentam seguir normalidade. Além disso, percebe-se que não sobrou autocorrelação no gráfico ACF. Para complementar a análise visual, fez-se os testes estatísticos apresentados na tabela 6.

Tabela 6: Testes estatísticos dos resíduos do modelo harmônico

| Teste | Estatística | P-Valor |
|--------------|-------------|---------|
| KPSS | 0,42 | 0,06 |
| Shapiro-Wilk | 0,95 | < 0,01 |
| Ljung-Box | 12,03 | 0,67 |

O resultado do teste de KPSS, indica que não deve-se rejeitar a hipótese de estacionariedade dos resíduos, uma vez que apresentou o p-valor de 0,10. Apesar de não ser um problema, é interessante notar que, no que se refere a normalidade, o teste de Shapiro-Wilk aponta que deve-se rejeitar essa hipótese. Quanto ao teste de Ljung-Box, o p-valor de 0,67 sugere que não há evidências estatísticas para rejeitar a hipótese de independência dos resíduos.

4.3.4 Modelo MSTL + ARIMA

Uma vez que a série estudada é semanal, ou seja, conta com um ciclo complexo, o terceiro modelo testado é MSTL + ARIMA. Dada a transformação de Box-Cox com $\lambda = -0,13$ e as variáveis explicativas definidas anteriormente, a partir da função `auto.arima` do R, encontrou-se MSTL + ARIMA(3,1,1). A Tabela 7 apresenta os coeficientes encontrados para o modelo.

Tabela 7: Coeficientes do modelo MSTL + ARIMA

| | Coeficientes | Desvio-Padrão |
|----------|--------------|---------------|
| AR(1) | 0,62 | 0,05 |
| AR(2) | -0,08 | 0,04 |
| AR(3) | 0,16 | 0,04 |
| MA(1) | -0,90 | 0,04 |
| Fase 1 | -0,22 | 0,06 |
| Fase 2 | -0,21 | 0,07 |
| Fase 3 | -0,18 | 0,07 |
| Fase 4 | -0,05 | 0,05 |
| Feriado | 0,05 | < 0,01 |
| Carnaval | 0,18 | 0,01 |

Ao analisar a Tabela 7, nota-se que as variáveis relacionadas a fase 1, fase 2 e

fase 3 da pandemia apresentam valores próximos, indicando uma variação média similar no número de infrações nesses três períodos.

Além disso, verifica-se que a variável feriado, que indica se ocorreu algum feriado na semana observada, eleva, em média, a quantidade de infrações em 0,05. Já a variável carnaval, aumenta, em média, 0,18 a quantidade de infrações observadas na semana.

Para verificar a adequação do modelo, é feita a análise de resíduos a seguir.

Figura 17: Análise de resíduos do modelo MSTL + ARIMA

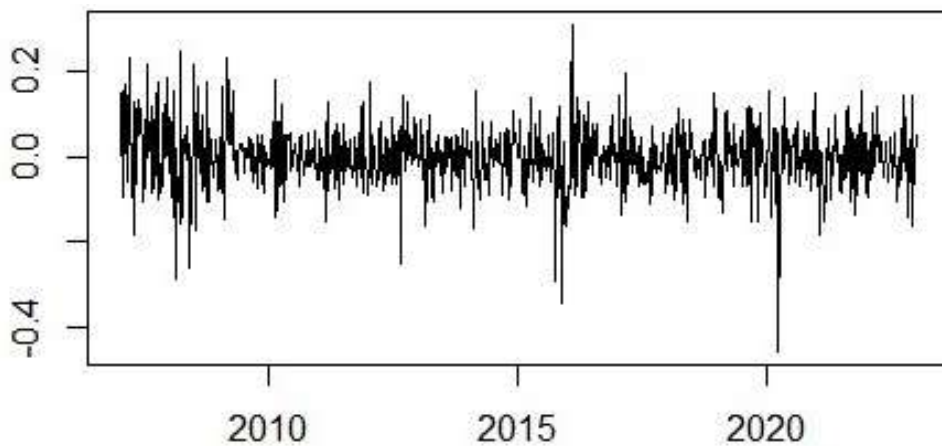
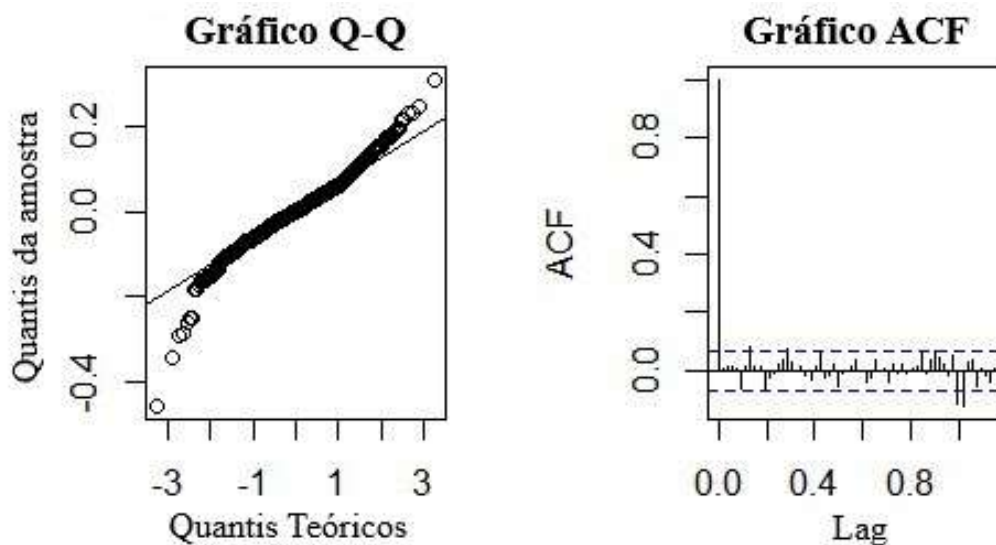


Figura 18: Análise de resíduos do modelo MSTL + ARIMA



Ao analisar a Figura 17, pode-se perceber que os resíduos do modelo aparentam oscilar sobre a média zero com variância constante. Além disso, verifica-se na Figura 18 que não sobrou autocorrelação no gráfico ACF mas que os resíduos não aparentam seguir normalidade. Para complementar a análise dos resíduos, fez-se os testes apresentados na

Tabela 8.

Tabela 8: Testes estatísticos dos resíduos do modelo MSTL + ARIMA

| Teste | Estatística | P-Valor |
|--------------|-------------|---------|
| KPSS | 0,60 | 0,01 |
| Shapiro-Wilk | 0,96 | < 0,01 |
| Ljung-Box | 20,96 | 0,16 |

Em relação ao resultado dos testes, a Tabela 8 mostra que tanto o KPSS quanto o Shapiro-Wilk apresentaram p-valor menor do que 0,05, indicando que a um nível de significância de 5%, deve-se rejeitar as hipóteses de estacionariedade e normalidade. Já o teste de Ljung-Box apresentou o p-valor de 0,16, ou seja, não rejeita-se a hipótese de independência dos resíduos.

4.4 Comparação dos Modelos

4.4.1 Critérios de Informação

Após a definição dos modelos, verificou-se os critérios de informação de Akaike e Bayesiano, os resultados se encontram na tabela a seguir.

Tabela 9: Critérios de informação dos modelos

| Modelo | AICc | BIC | σ^2 |
|--------------|----------|----------|------------|
| SARIMA | -1806,32 | -1736,01 | 0,006 |
| Harmônico | -1810,92 | -1712,81 | 0,006 |
| MSTL + ARIMA | -1949,87 | -1893,54 | 0,005 |

Ao analisar a Tabela 9, nota-se que os três modelos apresentaram σ^2 muito similar. O SARIMA apresentou AICc de -1806,32, o modelo de regressão dinâmica harmônica de -1810,92 e o MSTL + ARIMA de -1949,87.

Ainda que o modelo MSTL + ARIMA obteve os menores critérios de parcimônia, vale destacar que a comparação entre os modelos via esses critérios não é um procedimento correto, uma vez que esses modelos não são casados (um não é exatamente caso particular do outro).

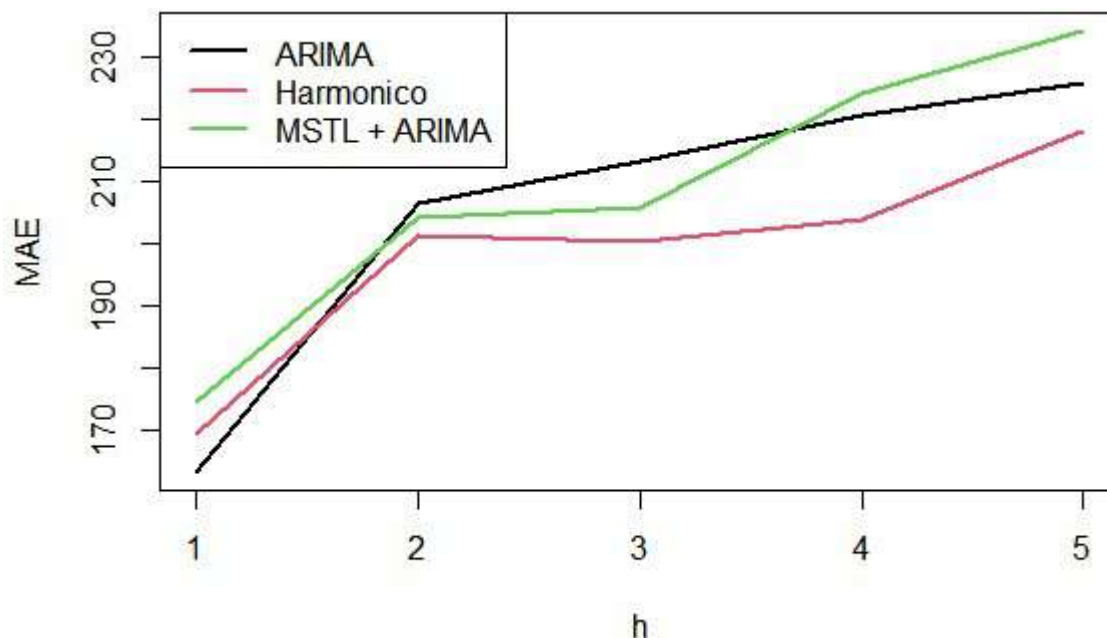
4.4.2 Estudo de Desempenho Preditivo

Para elaboração do estudo de desempenho preditivo por janela deslizante, considera-se uma janela de treinamento inicial composta por 783 semanas e um horizonte de previsão de 5 semanas. Considerando os três modelos definidos anteriormente para construir as funções de previsão, a Tabela 10 e a Figura 19 apresentam os resultados dos erros absolutos por horizonte de previsão.

Tabela 10: Erros absolutos por horizonte de previsão

| Horizonte | ARIMA | Harmônico | MSTL + ARIMA |
|-----------|--------|-----------|--------------|
| h = 1 | 163,47 | 169,35 | 174,52 |
| h = 2 | 206,56 | 201,32 | 204,13 |
| h = 3 | 213,07 | 200,32 | 205,82 |
| h = 4 | 220,62 | 204,01 | 224,17 |
| h = 5 | 225,58 | 218,04 | 234,09 |

Figura 19: Gráfico dos horizontes e seus erros de previsão



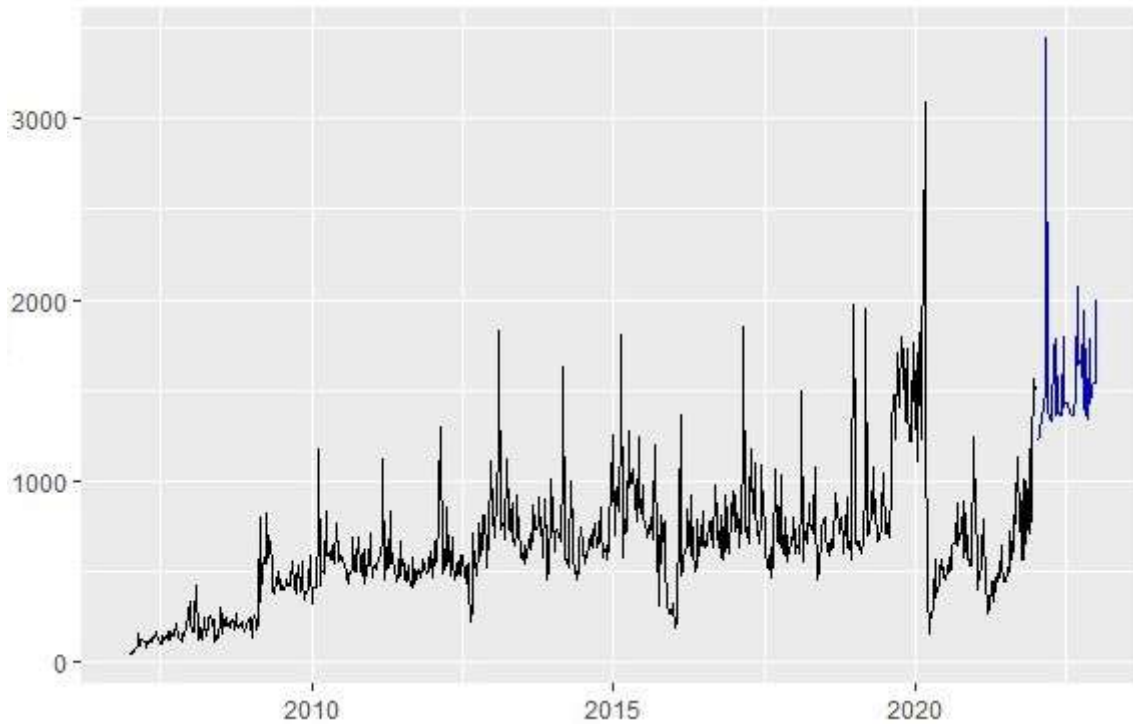
A Tabela 10 e a Figura 19, mostram que ao observar os dois primeiros horizontes de previsão, os três modelos apresentaram MAE similar, mas, a partir de $h=3$, nota-se uma diferença maior do modelo de harmônico para os demais.

Além disso, percebe-se que o modelo harmônico obteve os menores erros médios

para previsão em quase todos os horizontes, sendo então, o modelo com o melhor comportamento. Já o ARIMA e o MSTL + ARIMA ficaram praticamente empatados em todos os cinco horizontes de previsão.

Uma vez que o modelo harmônico tende a ter previsões melhores para a série, a Figura 20 apresenta a sua previsão para as próximas 52 semanas, isto é, até 2023.

Figura 20: Previsão do Modelo de Regressão Dinâmico Harmônico



5 Conclusão

Nesse trabalho, foi estudado o comportamento da série temporal das infrações de trânsito semanais, relativas ao artigo 165 do Código de Trânsito Brasileiro. Verificou-se que a série conta com diferentes padrões da variação ao longo do tempo, então, foi aplicada a transformação de Box-Cox. Além disso, observou-se que a série possui ciclo complexo, isto é, não inteiro, e tendência crescente.

Ao observar o comportamento da série ao longo dos anos, em especial em semanas com feriados, notou-se a necessidade de criar variáveis indicando a presença de feriados, especialmente o carnaval. Além disso, a série apresentou um comportamento muito peculiar durante o período de pandemia, e por isso, definiu-se quatro variáveis para abordá-la no modelo.

O primeiro modelo trabalhado foi o SARIMA, que a partir da função `auto.arima()` do R, foi definido como SARIMA(1,1,2)(2,0,0). Na análise desse modelo, não rejeitou-se as hipóteses de estacionariedade e de independência dos resíduos. O segundo modelo abordado foi o harmônico, que também não rejeitou as hipóteses de independência e estacionariedade dos resíduos.

Por último, foi utilizada a técnica MSTL + ARIMA, em que retira-se a sazonalidade via MSTL e se ajusta um modelo ARIMA, depois faz-se a previsão com a sazonalidade ajustada e replica-se a sazonalidade no último ciclo de previsão. Nesse caso, não rejeitou-se a hipótese de independência dos resíduos mas rejeitou-se a de estacionariedade.

Em relação ao três modelos estudados, o estudo por janela deslizante indicou que o melhor modelo definido foi o harmônico, uma vez que apresentou os menores erros médios para previsão.

Referências

- ALMEIDA, T. R. C. de. Regressão dinâmica: Modelo estratégico para acompanhamento e previsão do índice de volume do pib de minas gerais. Convibra, 2014. Disponível em: https://convibra.org/congresso/res/uploads/pdf/2014.32_9091.pdf.
- ATKINSON, A. C.; RIANI, M.; CORBELLINI, A. The box-cox transformation: Review and extensions. *Statist. Sci.*, Institute of Mathematical Statistics, v. 36, p. 239–255, 2021.
- AZEVEDO, A. M.; JUNIOR, V. C. A.; FERNANDES2, J. S. Transformação box-cox na homocedasticidade e normalidade uni e multivariada em experimentos de batata-doce. *Horticultura Brasileira*, Associação Brasileira de Horticultura, v. 34, p. 93–101, 2016.
- BARCELLOS, C.; XAVIER, D. R. As diferentes fases, os seus impactos e os desafios da pandemia de covid-19 no brasil. *Reciis – Revista Eletrônica de Comunicação, Informação Inovação em Saúde*, v. 16, p. 221–226, 2022.
- BARROS, A. C. et al. *Análise de Séries Temporais em R: Curso introdutório*. [S.l.]: ELSEVIER, 2018.
- BOX, G. E. P.; COX, D. R. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, Royal Statistical Society, v. 26, p. 211–252, 1964.
- BROWNE, M. W. Cross-validation methods. *Mathematical Psychology*, The Ohio State University, v. 44, p. 108–132, 2000.
- BUSCH, R. et al. Revisão narrativa sobre causas e consequências da embriaguez nos crimes de trânsito. *Journal of Education Science and Health*, v. 3, p. 1–12, 2023. Disponível em: <https://bio10publicacao.com.br/jesh/article/view/190>.
- CAMELO, H. do N. Utilização de regressão dinâmica harmônica através do software livre r para previsão de séries temporais de velocidade do vento em regiões do nordeste brasileiro. *Revista de Engenharia e Tecnologia*, uepg, v. 9, p. 55–63, 2017.
- CARVALHO, L. E. S. Alcoolismo e trânsito. *Escola Anna Nery*, v. 4, p. 185–206, 2020.
- CHU, C.-S. J. Time series segmentation: A sliding window approach. *Information Sciences*, NORTH- HOLLAND, v. 85, n. 1, p. 147–173, 1995.
- CLEVELAND, R. B. et al. A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, v. 6, p. 3–73, 1990. Disponível em: <https://www.wessa.net/download/stl.pdf>.
- CZERWONKA, M. *CTB completa 18 anos marcados por conquistas e desafios*. 2016. Disponível em: <https://www.portaldotransito.com.br/noticias/ctb-completa-18-anos-marcados-por-conquistas-e-desafios-2/>.
- DUBEY, A. K. et al. Study and analysis of sarima and lstm in forecasting time series data. *ELSEVIER, Sustainable Energy Technologies and Assessments*, v. 47, p. 101–474, 2021.

- FISCHER, S. *Séries Univariantes de Tempo - metodologia de Box & Jenkins*. [S.l.]: Porto Alegre: Fundação de Economia Estatística, 1982.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting Principles and Practice*. [S.l.]: Texts, 2018.
- JOSE, J. Introduction to time series analysis and its applications. *ResearchGate*, v. 1, p. 1–12, 2022.
- MORETTIN, P. A. *Econometria financeira: Um curso em séries temporais financeiras*. [S.l.]: Blucher, 2014.
- MORETTIN, P. A.; TOLOI, C. M. C. *Análise de Séries Temporais*. [S.l.]: Blucher, 2018.
- NETO, I. L.; IGLESIAS, F.; GÜNTHER, H. Uma medida de justificativas de motoristas para infrações de trânsito. *PSICO*, PUCRS, v. 43, p. 7–13, 2013. Disponível em: <https://revistaseletronicas.pucrs.br/ojs/index.php/revistapsico/article/view/11088/7613>.
- PRF, P. R. F. *Observatório de Dados da PRF*. 2022. Disponível em: <https://www.gov.br/prf/pt-br/acesso-a-informacao/dados-abertos/observatorio-de-dados-da-prf>.
- REIMBOLD, M. M. P. et al. Aplicação de teste de raiz unitária às variáveis de propulsores eletromecânicos. *Revista Eletrônica de Extensão da URI, Vivências*, v. 13, p. 46–54, 2017.
- SILVEIRA, A. G.; MATTOS, V. L. D. de; KONRATH, A. C. Avaliação da estacionariedade e teste de cointegração em séries temporais o caso da demanda de energia elétrica residencial no Brasil. *RETEC - Revista de Tecnologias*, Fatec, v. 9, p. 75–79, 2016.
- SILVESTRE, M. R.; BEZERRA, M. I. S. Modelos decomposição temporal e de regressão harmônica: uma comparação para a série mensal da temperatura mínima e máxima média de presidente prudente (sp). *Revista Brasileira de Meteorologia*, scielo, v. 30, p. 457–466, 2015.
- SOUSA, A. R. dos S. et al. *Análise de Séries Temporais*. [S.l.]: sagah, 2021.
- TORCATE, A. S. et al. Mineração de dados para análise e predição das infrações de trânsito na cidade do Recife. *Revista de Engenharia e Pesquisa Aplicada*, v. 6, p. 1–11, 2021.
- WILLMOTT, K. M. C. J. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *CLIMATE RESEARCH*, Inter-Research Science, v. 30, p. 79–82, 2005.
- YAKUBU, U. A.; SAPUTRA, M. P. A. Time series model analysis using autocorrelation function (acf) and partial autocorrelation function (pacf) for e-wallet transactions during a pandemic. *International Journal of Global Operations Research*, IJGOR, v. 3, p. 80–85, 2022.
- YOUNG, P. C. Nonstationary time series analysis and forecasting. *Progress in Environmental Science*, v. 1, p. 3–48, 1999.