

Análise do Volume Mensal de Tráfego de Veículos de Passeio na BR-393

João Kramer ¹

Departamento de Estatística, Universidade de Brasília, DF

Resumo. A presente análise dedicou-se ao estudo do volume mensal de tráfego de veículos de passeio na BR-393/RJ km 125,00, utilizando dados da Agência Nacional de Transportes Terrestres (ANTT). Foi realizada uma filtragem dos registros para incluir apenas o sentido "Crescente", a "Praça 01 BR-393/RJ km 125,00" e a "Categoria 1" (veículos de passeio). O modelo SARIMA foi empregado para descrever e projetar volumes futuros de tráfego, contribuindo para o planejamento e investimento em infraestrutura e previsão de receitas. A análise dos resíduos do modelo SARIMA $(1,0,1) \times (2,1,1)_{12}$ revelou ausência de autocorrelação, mas o teste de Shapiro-Wilk indicou desvio da normalidade. O MAPE in-sample foi de 7.275643%.

OBJETIVO: Analisar e modelar o volume mensal de tráfego de veículos de passeio na BR-393/RJ km 125,00, com o intuito de compreender sua dinâmica temporal e gerar previsões para os próximos 12 meses.

MÉTODO: Os dados para esta análise foram obtidos da Agência Nacional de Transportes Terrestres (ANTT), por meio do seu portal de dados públicos. A metodologia incluiu a avaliação preliminar da série temporal para identificar tendências e sazonalidade. A Regressão Harmônica foi empregada para descrever o padrão sazonal e servir como um teste de não estacionariedade sazonal. A diferenciação sazonal com lag 12 foi aplicada para tratar a não estacionariedade sazonal. O modelo SARIMA foi identificado e estimado através de um procedimento de seleção baseado no Critério de Informação Bayesiano (BIC), utilizando uma grade de possíveis ordens para os componentes autorregressivos (AR), média móvel (MA), autorregressivos sazonais (SAR) e média móvel sazonal (SMA). O diagnóstico do modelo incluiu a análise dos resíduos com o Teste de Ljung-Box para verificar a ausência de autocorrelação e o Teste de Normalidade dos Resíduos.

RESULTADOS: A análise da Função de Autocorrelação (FAC) e Função de Autocorrelação Parcial (FACP) da série original revelou forte indicativo de não estacionariedade e picos sazonais proeminentes em múltiplos do lag 12, indicando forte sazonalidade anual. A aplicação da primeira diferença sazonal com lag 12 reduziu substancialmente a sazonalidade e a tendência. O modelo SARIMA selecionado foi o $(1,0,1) \times (2,1,1)_{12}$. O Teste de Ljung-Box para os resíduos demonstrou p-valores consistentemente altos, indicando ausência de autocorrelação significativa, e que os resíduos se comportam como ruído branco. O Teste de Shapiro-Wilk, no entanto, resultou em um p-valor extremamente baixo ($4.228e-09$), levando à rejeição da hipótese nula de normalidade dos resíduos. A acurácia do modelo in-sample, medida pelo Erro Percentual Absoluto Médio (MAPE), foi de 7.275643%. As previsões para os próximos 12 meses ilustram o padrão sazonal esperado e o alargamento do intervalo de confiança de 95%.

CONCLUSÃO: Apesar da não normalidade formal dos resíduos, o modelo SARIMA $(1,0,1) \times (2,1,1)_{12}$ revelou-se bem especificado para capturar a estrutura de dependência temporal e sazonal do volume mensal de tráfego. A ausência de autocorrelação nos resíduos e o MAPE aceitável são fortes indicadores de sua adequação para a previsão, fornecendo insights valiosos para o planejamento e tomada de decisão relacionados ao fluxo de veículos na BR-393/RJ.

Palavras-chave. Modelo SARIMA. Filtros Lineares. Tráfego de Veículos. Sazonalidade. Previsão. BR-393.

¹joakramer@gmail.com

1 Introdução

A presente análise dedica-se ao estudo do **Volume Mensal de Tráfego de Veículos de Passeio na BR-393/RJ km 125,00**. Os dados utilizados para este estudo foram obtidos da Agência Nacional de Transportes Terrestres (ANTT), especificamente do seu portal de dados públicos, disponível em <https://dados.antt.gov.br/dataset/volume-trafego-praca-pedagio>. Para este estudo, foram selecionados os registros referentes ao sentido "Crescente", à "Praça 01 BR-393/RJ km 125,00" e à "Categoria 1", que representa veículos de passeio. A **ANTT coleta e disponibiliza esses dados** com o objetivo principal de fornecer informações das Concessionárias de Rodovias Federais Concedidas sobre o volume de tráfego em cada praça de pedágio. Como Agência Nacional de Transportes Terrestres, o papel da ANTT envolve o monitoramento, a regulação e a fiscalização das operações dessas rodovias concedidas. A disponibilidade pública desses dados, que vêm acompanhados de um "Dicionário de Dados", reflete um compromisso com a transparência e serve de apoio a diversos *stakeholders*. Essa transparência possibilita o escrutínio público, a pesquisa acadêmica e a tomada de decisões informadas relativas à infraestrutura de transporte do país. Os dados são disponibilizados em formatos como CSV e JSON, contendo campos como concessionária, mês/ano, sentido, praça, categoria, tipo de veículo e volume total. A **utilidade prática da modelagem desses dados** para a descrição do fenômeno e para a produção de projeções para os próximos 12 meses é considerável e multifacetada:

- **Para a descrição do fenômeno:** A modelagem do volume mensal de tráfego de veículos de passeio permite uma **compreensão aprofundada dos padrões e características subjacentes ao fluxo de tráfego** nesta seção específica da BR-393. Métodos como a **Regressão Harmônica** podem ser empregados para descrever o **padrão sazonal estático (histórico/global/não dinâmico)** presente nos dados. Isso auxilia na identificação de tendências recorrentes, ciclos (por exemplo, sazonalidade mensal ou anual) e quaisquer evoluções de longo prazo no volume de tráfego. Tais *insights* descritivos são cruciais para:
 - **Planejamento operacional:** Compreender os períodos de pico e de baixa demanda pode informar a alocação de recursos para as operações das praças de pedágio, dimensionamento de equipes e agendamento de manutenções.
 - **Avaliação da infraestrutura:** A descrição do estado atual e das variações históricas do tráfego fornece contexto essencial para avaliar a capacidade existente e o desempenho da infraestrutura rodoviária.
 - **Formulação de políticas:** Subsidiar políticas de transporte relacionadas a categorias de veículos, estratégias de pedágio ou gestão da demanda.
 - Além disso, a Regressão Harmônica serve como um teste de não estacionariedade para a componente sazonal.
- **Para a produção de projeções para os próximos 12 meses:** A aplicação de modelos como o **SARIMA (Seasonal Autoregressive Integrated Moving Average)** possibilita a **previsão de volumes futuros de tráfego**. Essa capacidade é de imenso valor para uma série de aplicações estratégicas e operacionais:
 - **Planejamento e investimento em infraestrutura:** Projeções precisas de 12 meses podem guiar decisões sobre quando e onde investir em expansões de rodovias, novas faixas ou outras melhorias de infraestrutura para acomodar o crescimento de tráfego previsto ou gerenciar congestionamentos.
 - **Previsão de receitas para concessionárias:** Como os dados se referem a praças de pedágio, as projeções de volume de veículos estão diretamente ligadas às receitas de pedágio esperadas, sendo vitais para o planejamento financeiro das concessionárias.

- **Gestão de tráfego e segurança:** Antecipar os futuros níveis de tráfego pode auxiliar no planejamento de medidas de controle de tráfego, implantação de serviços de emergência e otimização de protocolos de segurança na rodovia.
- **Avaliação de impacto ambiental:** As projeções podem contribuir para estimar futuras emissões ou consumo de combustível associados ao tráfego de veículos, auxiliando no planejamento ambiental.

Ao aproveitar tais técnicas de modelagem, que incluem a divisão da base de dados em treinamento e teste, avaliação preliminar, identificação e estimação de modelos (como SARIMA), diagnóstico (com testes de Ljung-Box e normalidade de resíduos), e a avaliação no conjunto de teste, esta análise visa fornecer não apenas uma descrição abrangente dos padrões de tráfego passados e presentes, mas também *insights* prospectivos cruciais para a gestão eficiente, planejamento e desenvolvimento da infraestrutura da BR-393/RJ.

2 Metodologia

Esta seção detalha a metodologia empregada na análise do Volume Mensal de Tráfego de Veículos de Passeio na BR-393/RJ km 125,00, abordando desde a coleta e preparação dos dados até as técnicas de modelagem e previsão.

2.1 Dados

Os dados para esta análise foram obtidos da Agência Nacional de Transportes Terrestres (ANTT), por meio do seu portal de dados públicos, especificamente no endereço <https://dados.antt.gov.br/dataset/volume-trafego-praca-pedagio>. A ANTT coleta e disponibiliza essas informações, que são provenientes das Concessionárias de Rodovias Federais Concedidas, detalhando o volume de tráfego em cada praça de pedágio. Os arquivos de dados são disponibilizados em formatos como CSV e JSON, contendo campos como concessionária, mês/ano, sentido, praça, categoria, tipo de veículo e volume total. Para o presente estudo, foi realizada uma filtragem específica dos registros para incluir apenas:

- Os dados do arquivo 'volume-trafego-praca-pedagio-[2010-a-2023].csv'.
- O sentido de tráfego "Crescente".
- A praça de pedágio "Praça 01 BR-393/RJ km 125,00"(SAPUCAIA).
- A "Categoria 1", que corresponde a veículos de passeio.

2.2 Modelo SARIMA

O **Modelo SARIMA (Seasonal Autoregressive Integrated Moving Average)** é a ferramenta de modelagem utilizada para esta análise de séries temporais. Este tipo de modelo é particularmente adequado para séries que exibem padrões sazonais, como o volume de tráfego, permitindo a descrição do fenômeno e a produção de projeções futuras.

2.3 Divisão da Base em Treinamento e Teste

A metodologia prevê a **divisão da base de dados** em conjuntos de treinamento e teste. Essa separação é fundamental para avaliar a capacidade preditiva do modelo, onde o conjunto de treinamento é utilizado para ajustar o modelo e o conjunto de teste para verificar seu desempenho em dados não observados durante o ajuste.

2.4 Avaliação Preliminar

Antes da aplicação dos modelos formais, será realizada uma **avalição preliminar** da série temporal. Esta etapa visa identificar características iniciais dos dados, como tendências, sazonalidade e possíveis anomalias, que podem guiar as etapas subsequentes de modelagem.

2.5 Regressão Harmônica

A **Regressão Harmônica** será empregada como um método para **descrever o padrão sazonal** presente na série de volume de tráfego. Este método permite analisar a sazonalidade sob uma **perspectiva estática (histórica/global/não dinâmica)**. Além de sua capacidade descritiva, a Regressão Harmônica também serve como um **teste de não estacionariedade para a componente sazonal** dos dados.

2.6 Identificação e Estimação

Após a avaliação preliminar e a análise sazonal, proceder-se-á à **identificação e estimação dos modelos SARIMA**. Esta etapa envolve a determinação das ordens dos componentes autor-regressivos (AR), de médias móveis (MA) e de integração (I), tanto para as partes não sazonais quanto para as sazonais do modelo. Um procedimento de seleção de modelo com base no critério BIC (Bayesian Information Criterion) será utilizado, com atenção para garantir que as soluções estejam dentro de um hipercubo de parâmetros adequado, evitando que o melhor modelo esteja na "aresta" do grid de busca.

2.7 Diagnóstico

A etapa de **diagnóstico** será crucial para validar a adequação do modelo selecionado. Isso incluirá a **análise dos resíduos** do modelo, com a aplicação de testes como o **Teste de Ljung-Box** para verificar a ausência de autocorrelação nos resíduos e o **Teste de Normalidade dos Resíduos**. Um bom diagnóstico garante que o modelo capturou a estrutura temporal da série de forma eficaz e que os resíduos se assemelham a um ruído branco.

2.8 Previsão

A etapa final da metodologia consiste na **previsão para os próximos 12 meses** utilizando o modelo SARIMA ajustado e diagnosticado. As projeções futuras são de grande utilidade prática para o planejamento e gestão da infraestrutura rodoviária e a previsão de receitas para as concessionárias. A avaliação da capacidade de previsão do modelo será realizada também no conjunto de teste.

3 Análise

3.1 Análise da FAC e da FACP da Série Original

A análise da Função de Autocorrelação (FAC) e da Função de Autocorrelação Parcial (FACP) da série temporal original de Volume Mensal de Tráfego de Veículos de Passeio na BR-393/RJ km 125,00 é uma etapa crucial para a identificação de padrões e características da série, como a presença de tendência e sazonalidade. A série original (Figura 1) exhibe visualmente padrões que sugerem ambas as características.

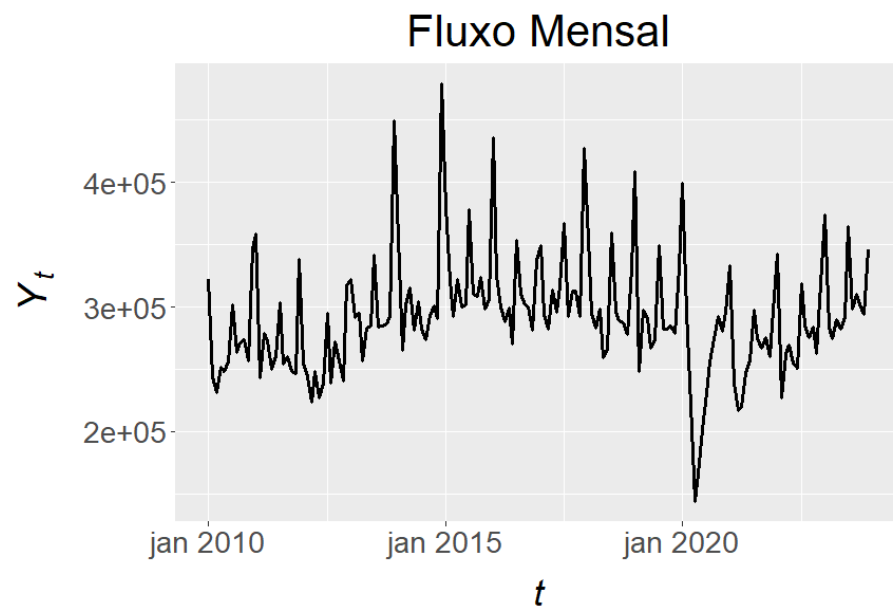


Figura 1: Fluxo Mensal de Tráfego de Veículos de Passeio na BR-393/RJ km 125,00

Análise da Função de Autocorrelação (FAC) da Série Original

O gráfico da FAC da série original é apresentado na Figura 2.

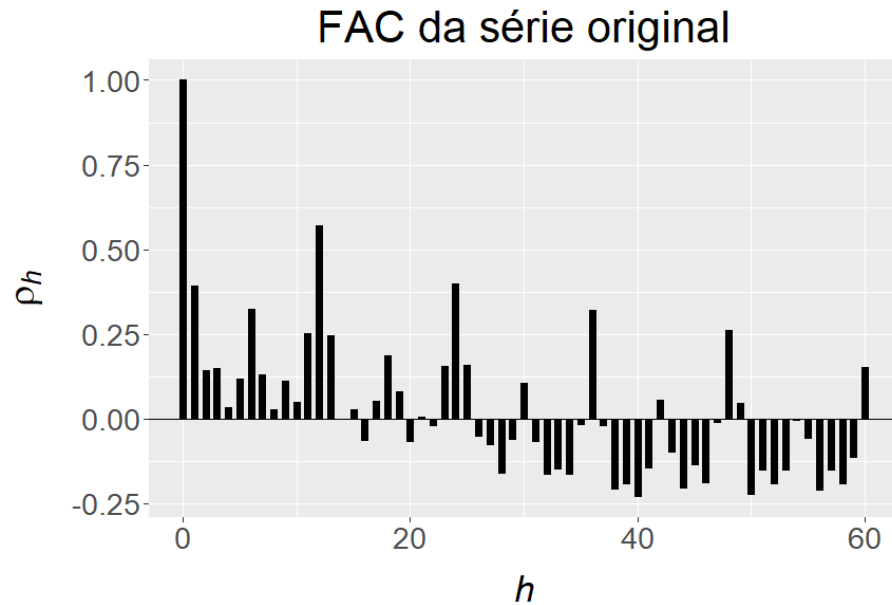


Figura 2: Gráfico da Função de Autocorrelação (FAC) da Série Original

- **Decaimento Lento:** A FAC demonstra um **decaimento lento e gradual**, o que é um **forte indicativo de não estacionariedade** na série temporal, sugerindo a presença de uma **tendência**. Isso implica que a série necessitará de diferenciação para se tornar estacionária.
- **Picos Sazonais Fortes:** Observam-se **picos proeminentes em múltiplos do lag 12**. Este padrão é um claro sinal de **sazonalidade anual forte**, visto que os dados são mensais e um lag de 12 representa um ano.

- A autocorrelação no **lag 12** é de **0.570252140**.
- A autocorrelação no **lag 24** é de **0.399300745**.
- A autocorrelação no **lag 36** é de **0.320674224**.
- A autocorrelação no **lag 48** é de **0.262577077**.
- A autocorrelação no **lag 60** é de **0.153251240**.

Estes valores positivos e consistentemente significativos em lags sazonais confirmam um padrão repetitivo ano após ano.

Análise da Função de Autocorrelação Parcial (FACP) da Série Original

O gráfico da FACP da série original é exibido na Figura 3.

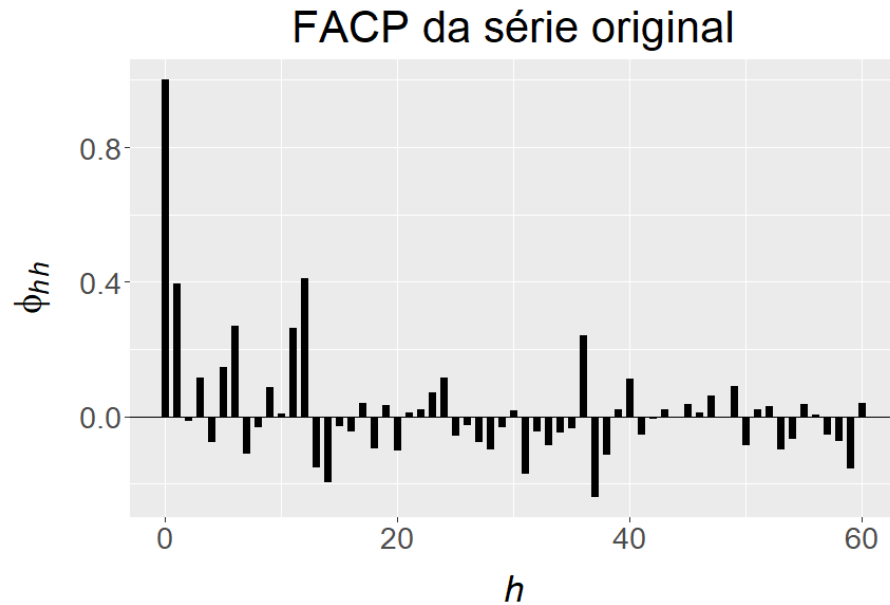


Figura 3: Gráfico da Função de Autocorrelação Parcial (FACP) da Série Original

- **Picos Significativos em Lags Sazonais:** A FACP também apresenta **picos significativos em lags sazonais**, notavelmente no lag 12.
 - A autocorrelação parcial no **lag 12** é de **4.099260e-01**. Este forte pico positivo no lag 12 indica um **efeito sazonal direto** após o ajuste para os lags intermediários.
 - Há também um pico notável no **lag 36**, com valor de **2.422040e-01**.
- **Interpretação para Identificação do Modelo:** O padrão na FACP, particularmente o corte abrupto ou o pico significativo no lag sazonal (12), é útil para identificar a ordem da componente autorregressiva sazonal (P) de um modelo SARIMA.

Implicações para a Identificação e Estimação do Modelo SARIMA

Com base na análise conjunta da FAC e FACP:

- O **decaimento lento na FAC** indica fortemente que a série é **não estacionária** na média e exige **diferenciação** para remover a componente de tendência. Isso é consistente com o parâmetro ‘d = 1’ observado no código do modelo SARIMA.
- Os **picos sazonais fortes na FAC e FACP no lag 12** demonstram claramente a presença de **sazonalidade anual**. Isso confirma a necessidade de uma componente sazonal no modelo SARIMA, que é representada por ‘S=12’ na chamada da função ‘sarima’. Os picos significativos na FACP em lags sazonais sugerem a necessidade de uma componente autorregressiva sazonal (P), enquanto o decaimento mais lento na FAC em lags sazonais pode indicar uma componente de médias móveis sazonal (Q), ou ambas.
- A evidência combinada desses gráficos aponta para a necessidade de um **modelo SARIMA (Autoregressive Integrated Moving Average Sazonal)**, capaz de capturar eficazmente

tanto a tendência não estacionária quanto os fortes padrões sazonais observados nos dados. Os valores numéricos das autocorrelações e autocorrelações parciais são cruciais para a fase de **identificação e estimação** do modelo SARIMA, auxiliando na seleção das ordens apropriadas para ‘p, q, P, Q’ e as ordens de diferenciação ‘d’ e ‘D’.

3.2 Análise da Série Diferenciada

A etapa de diferenciação sazonal é crucial na análise de séries temporais para tratar a **não estacionariedade sazonal** presente nos dados de volume mensal de tráfego de veículos de passeio na BR-393/RJ km 125,00. A diferenciação sazonal visa remover o padrão repetitivo que ocorre em intervalos fixos, como anualmente, para tornar a série mais estacionária e adequada para modelagem SARIMA.

Procedimento de Diferenciação

A diferenciação sazonal foi aplicada aos dados de volume mensal de tráfego (‘X.t’) utilizando uma defasagem (lag) de 12 meses. Isso é feito subtraindo o valor do mês corrente pelo valor do mesmo mês no ano anterior.

Visualização da Série Diferenciada

A figura “Variação do Fluxo Mensal” (Figura 4) ilustra a série após a aplicação da primeira diferenciação sazonal com lag 12. Observa-se que, com a diferenciação, a série ‘Variação do

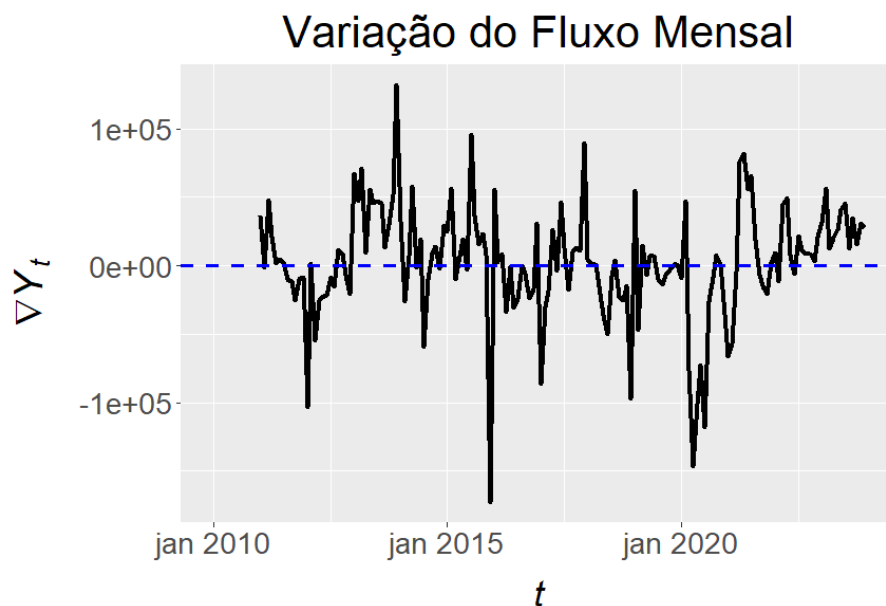


Figura 4: Variação do Fluxo Mensal

Fluxo Mensal’ agora **flutua em torno de zero**, o que é um indicativo da remoção da tendência e da sazonalidade mais evidente.

Análise da Função de Autocorrelação (FAC) da Série Diferenciada

A Função de Autocorrelação (FAC) da primeira diferença foi calculada e é apresentada na figura “FAC da primeira diferença” (Figura 5). As linhas pontilhadas vermelhas representam os **limites de confiança** (aproximadamente $2/\sqrt{N}$, onde N é o número de observações utilizadas na FAC). A tabela dos valores da FAC para a primeira diferença (‘Fig.2’) mostra os seguintes

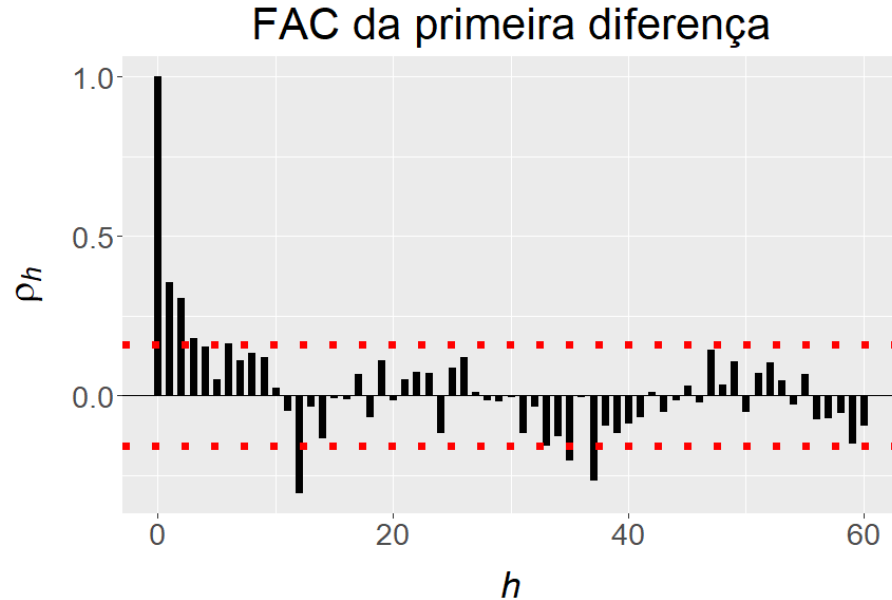


Figura 5: FAC da primeira diferença

coeficientes que se destacam:

- No **lag 1**, o valor é de **0.356**, que está acima dos limites de confiança, indicando uma autocorrelação significativa de curto prazo.
- No **lag 12**, o valor é de **-0.306**, que é significativamente negativo e fora dos limites de confiança.
- No **lag 24**, o valor é de **-0.119**, também negativo e próximo ou dentro dos limites, mas menor em magnitude comparado ao lag 12.
- No **lag 36**, o valor é de **-0.007**, indicando uma correlação insignificante.
- Outros lags podem apresentar valores fora dos limites, mas o **padrão de decaimento** é visivelmente mais rápido do que na série original, e a significância nos lags sazonais (12, 24, etc.) é atenuada, especialmente para lags múltiplos de 12. Isso sugere que a diferenciação foi eficaz na remoção da não estacionariedade sazonal principal.

Análise da Função de Autocorrelação Parcial (FACP) da Série Diferenciada

A Função de Autocorrelação Parcial (FACP) da primeira diferença é apresentada na figura “FACP da primeira diferença” (Figura 6). A tabela dos valores da FACP para a primeira diferença

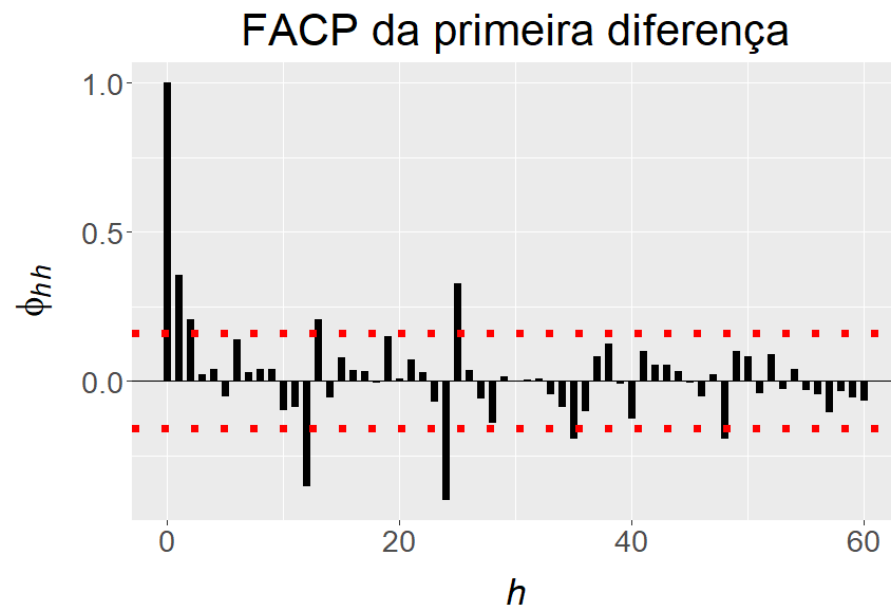


Figura 6: FACP da primeira diferença

(‘Fig.3’) revela:

- No **lag 1**, o valor é de **0.393**, que é altamente significativo.
- No **lag 12**, o valor é de **0.409**, que é altamente significativo.
- No **lag 24**, o valor é de **0.116**, que é significativo, mas menor que o lag 12.
- No **lag 36**, o valor é de **0.242**, que é significativo.
- A FACP mostra que, após a diferenciação sazonal, ainda há **autocorrelações parciais significativas em lags múltiplos de 12** (como 12, 24, 36). Isso indica a presença de dependências sazonais residuais que não foram completamente removidas pela diferenciação sazonal simples e que podem ser modeladas por um componente autorregressivo sazonal (P). Isso é diferente da FACP da série original, que também mostrava picos significativos nos lags sazonais.

Conclusão da Diferenciação Sazonal

A aplicação da primeira diferença sazonal com lag 12 reduziu substancialmente a sazonalidade e a tendência da série temporal de volume de tráfego, conforme evidenciado pela visualização da série diferenciada (Figura 4) e pela FAC (Figura 5). O pico proeminente no lag 12 na FAC da série original foi significativamente atenuado e até se tornou negativo na FAC da série diferenciada. A análise das funções de autocorrelação da série diferenciada sugere que a série está **sazonalmente estacionária** em média, mas a presença de picos significativos na FACP em lags múltiplos de 12 (Figura 6) indica que componentes AR sazonais podem ser necessários no modelo SARIMA para capturar a estrutura de dependência sazonal restante. Além disso, a presença de autocorrelação significativa no lag 1 da FAC e FACP sugere a necessidade de componentes AR ou MA não sazonais.

3.3 Implementação da Regressão Harmônica

A análise da Regressão Harmônica é uma etapa fundamental para compreender os padrões sazonais e avaliar a não estacionariedade sazonal da série temporal de volume mensal de tráfego de veículos de passeio na BR-393/RJ km 125,00. Este método permite descrever o padrão sazonal sob uma **perspectiva estática (histórica/global/não dinâmica)** e serve como um teste de não estacionariedade sazonal. O modelo de regressão harmônica utiliza uma série de funções seno e cosseno para capturar a natureza periódica dos dados da série temporal. O conjunto de dados $t.mensal$, que contém o volume de tráfego mensal ($X.t$) e o número do mês correspondente, é utilizado para criar esses componentes harmônicos:

- **Primeiro Harmônico (período de 12 meses):** $H1c$ (cosseno) e $H1s$ (seno) são gerados usando $\cos(2\pi * t.mensalmonth/12)$ e $\sin(2\pi * t.mensalmonth/12)$, respectivamente.
- **Segundo Harmônico (período de 6 meses):** $H2c$ e $H2s$ são gerados usando $\cos(2\pi * t.mensalmonth/6)$ e $\sin(2\pi * t.mensalmonth/6)$.
- **Terceiro Harmônico (período de 4 meses):** $H3c$ e $H3s$ são gerados usando $\cos(2\pi * t.mensalmonth/4)$ e $\sin(2\pi * t.mensalmonth/4)$.
- **Quarto Harmônico (período de 3 meses):** $H4c$ e $H4s$ são gerados usando $\cos(2\pi * t.mensalmonth/3)$ e $\sin(2\pi * t.mensalmonth/3)$.

Esses componentes harmônicos são então utilizados como variáveis preditoras em um modelo linear (lm) para modelar o volume de tráfego mensal ($X.t$).

Análise do Modelo de Regressão Harmônica Inicial

O primeiro modelo linear (profile) incluiu todos os oito componentes harmônicos. O resumo deste modelo inicial revelou a seguinte significância para cada componente:

- **(Intercept) 291958:** Altamente significativo ($\Pr(>|t|) < 2e-16$ ***).
- **H1c (21080):** Altamente significativo ($\Pr(>|t|) 1.92e-07$ ***). Captura um forte ciclo anual.
- **H1s (-4990):** Não significativo ($\Pr(>|t|) 0.19922$).
- **H2c (20032):** Altamente significativo ($\Pr(>|t|) 6.79e-07$ ***). Sugere um ciclo semestral significativo.
- **H2s (20104):** Altamente significativo ($\Pr(>|t|) 6.24e-07$ ***). Também sugere um ciclo semestral significativo.
- **H3c (11972):** Significativo ($\Pr(>|t|) 0.00234$ **). Indica a presença de um ciclo de 4 meses.
- **H3s (5962):** Não significativo ($\Pr(>|t|) 0.12550$).
- **H4c (1939):** Não significativo ($\Pr(>|t|) 0.61717$).
- **H4s (21508):** Altamente significativo ($\Pr(>|t|) 1.14e-07$ ***). Indica um ciclo de 3 meses significativo.

O modelo alcançou um R-quadrado Múltiplo de 0.4462 e um R-quadrado Ajustado de 0.4184, indicando que aproximadamente 42% da variância no volume de tráfego pode ser explicada por esses componentes harmônicos. O F-estatístico de 16.02 com um valor-p $< 2.2\text{e-}16$ demonstra que o modelo, como um todo, é altamente significativo estatisticamente. A presença de múltiplos coeficientes harmônicos estatisticamente significativos ($H1c, H2c, H2s, H3c, H4s$) nesta regressão serve como um teste de não estacionariedade sazonal, confirmando que a série exibe um padrão sazonal claro e estatisticamente significativo, o que sugere a não estacionariedade sazonal.

Análise do Modelo de Regressão Harmônica Refinado

Com base nos testes de significância, um modelo linear refinado (profile) foi criado, incluindo apenas os componentes harmônicos estatisticamente significativos: H1c, H2c, H2s, H3c e H4s. O resumo deste modelo refinado reafirma a importância dos componentes selecionados:

- **(Intercept) 291958**: Altamente significativo ($\text{Pr}(> |t|) < 2\text{e-}16$ ***).
- **H1c (21080)**: Altamente significativo ($\text{Pr}(> |t|) 2.08\text{e-}07$ ***).
- **H2c (20032)**: Altamente significativo ($\text{Pr}(> |t|) 7.33\text{e-}07$ ***).
- **H2s (20104)**: Altamente significativo ($\text{Pr}(> |t|) 6.73\text{e-}07$ ***).
- **H3c (11972)**: Significativo ($\text{Pr}(> |t|) 0.00243$ **).
- **H4s (21508)**: Altamente significativo ($\text{Pr}(> |t|) 1.23\text{e-}07$ ***).

O R-quadrado Múltiplo para este modelo refinado é de 0.4313 e o R-quadrado Ajustado é de 0.4138. Esses valores são muito próximos aos do R-quadrado do modelo inicial, indicando que a remoção dos termos não significativos não reduziu substancialmente o poder explicativo do modelo. O F-estatístico é de 24.57 com um valor-p $< 2.2\text{e-}16$, mostrando novamente a significância geral do modelo. Este modelo refinado é então utilizado para prever o perfil sazonal ao longo dos 12 meses ($M = 1:12$). A função predict usa os coeficientes deste modelo profile e os componentes harmônicos para cada mês para gerar o perfil.

Interpretação da Figura Perfil Sazonal

A figura *Perfil Sazonal* (Figura 7) representa visualmente o padrão sazonal anual do volume de tráfego de veículos de passeio.

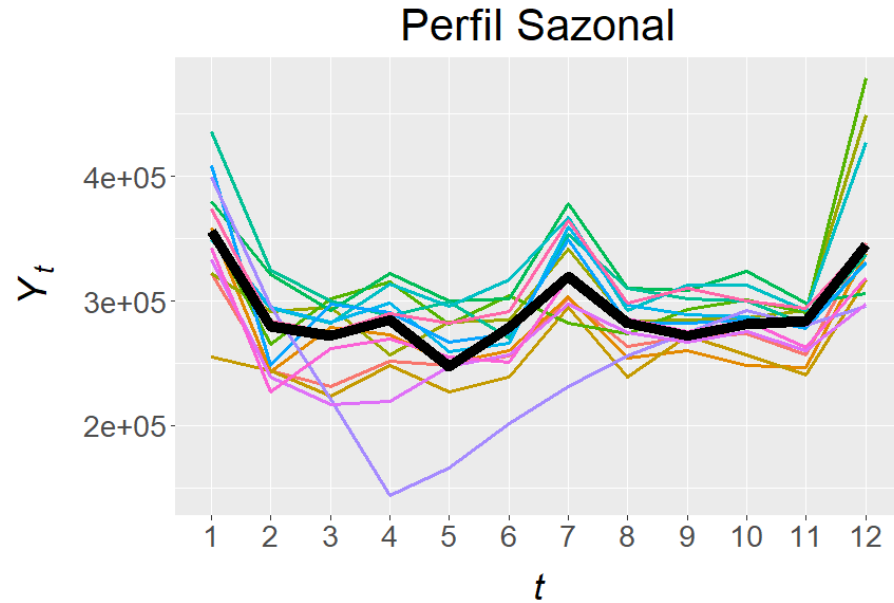


Figura 7: Perfil Sazonal do Volume de Tráfego de Veículos de Passeio

- A linha preta espessa na figura representa o perfil sazonal predito (perfil) derivado do modelo de regressão harmônica. Esta linha captura o ciclo sazonal médio ou subjacente.
- As linhas coloridas mais finas representam o volume de tráfego observado para cada ano ao longo dos 12 meses.
- O gráfico mostra claramente um padrão sazonal distinto:
 - Há tipicamente um pico no volume de tráfego em torno do mês 1 (Janeiro) e possivelmente outro nos meses 7 (Julho) ou 8 (Agosto), alinhando-se com períodos de férias ou feriados no Brasil.
 - O volume geralmente diminui após Janeiro, atingindo um vale em torno dos meses 4 ou 5 (Abril/Mayo).
 - Outra potencial diminuição é observada no final do ano antes da ascensão em direção a Dezembro/Janeiro.

Esta representação visual apoia diretamente os achados da análise FAC e FACP, que indicaram forte sazonalidade no lag 12 (ver Morettin e Toloi, 2005). A regressão harmônica descreve com sucesso este padrão anual recorrente, identificado pela significância estatística de seus componentes seno e cosseno. A forma clara e não plana da linha preta em “Perfil Sazonal” enfatiza ainda mais a presença de uma sazonalidade forte e não estacionária na série, reforçando a necessidade de um componente sazonal em qualquer modelo de previsão, como um modelo SARIMA.

3.4 Procedimento de Seleção de Modelo com Base no BIC

O critério de Informação Bayesiano (BIC) é uma ferramenta fundamental para a seleção de modelos em séries temporais, mas é importante notar que o BIC não é um critério absoluto, e o modelo com o menor valor de BIC nem sempre é o que produz os melhores resultados na prática,

especialmente se ele estiver "na aresta" do espaço de busca. Para a determinação da ordem do modelo, foi realizada uma busca exaustiva em um grid de possíveis parâmetros para p , q , P e Q . O treinamento foi realizado usando a série original de volume de tráfego, com diferenciação sazonal $D = 1$ e período sazonal $S = 12$ fixos, e sem diferenciação não sazonal ($d = 0$). Os intervalos de busca para p e q foram definidos como $0 : 5$ ('grid') e para P e Q como $0 : 2$ ('Grid'). Para cada combinação, o valor do BIC foi calculado.

A seguir, é apresentada a lista dos 10 melhores modelos com base no valor do BIC, em ordem crescente (menor BIC primeiro):

Rank	BIC	p	q	P	Q
1	23.89191	1	1	0	1
2	23.93407	1	1	2	1
3	23.94867	1	2	0	1
4	23.94909	2	1	0	1
5	23.94975	1	1	0	2
6	23.95043	1	1	1	1
7	23.95390	1	1	2	2
8	23.95431	1	1	1	2
9	23.97801	1	1	2	0
10	23.98901	1	2	2	1

Tabela 1: Tabela com diferentes valores de BIC

Conforme a orientação da metodologia, o segundo melhor modelo foi escolhido para a análise subsequente. Esta escolha leva em consideração a ressalva de que o modelo com o menor BIC absoluto nem sempre é a melhor opção, especialmente para garantir que a solução não esteja na "aresta" do hipercubo de busca, o que poderia indicar a necessidade de expandir o *grid* de parâmetros para encontrar uma solução mais robusta (Shumway e Stoffer, 2017). Assim, o modelo selecionado é o de parâmetros $\mathbf{p} = 1, \mathbf{q} = 1, \mathbf{P} = 2, \mathbf{Q} = 1$, com diferenciação não sazonal $d = 0$ e diferenciação sazonal $D = 1$ (com período $S = 12$).

3.5 Ajuste do Modelo Selecionado

Após o procedimento de seleção de modelo baseado no BIC, conforme detalhado na Seção 3.4, o **segundo melhor modelo** foi escolhido para o ajuste final. Este modelo é um SARIMA de ordem $(1,0,1) \times (2,1,1)_{12}$, indicando um componente autorregressivo não sazonal de ordem 1 ($p = 1$), nenhuma diferenciação não sazonal ($d = 0$), um componente de médias móveis não sazonal de ordem 1 ($q = 1$), um componente autorregressivo sazonal de ordem 2 ($P = 2$), uma diferenciação sazonal de ordem 1 ($D = 1$) e um componente de médias móveis sazonal de ordem 1 ($Q = 1$), com período sazonal $S = 12$. O ajuste do modelo foi realizado utilizando a função 'sarima' com os parâmetros identificados. O processo de otimização iterativo buscou minimizar o critério de informação, e a convergência foi alcançada após 100 iterações na primeira tentativa, e em menos iterações na segunda tentativa, ambas alcançando o valor final de 10.292305 e 10.369903 para os dois runs mostrados, respectivamente. Os resultados do ajuste são apresentados na Tabela 2, que inclui os coeficientes estimados, seus erros padrão, valores t e p-valores correspondentes.

Adicionalmente, a variância do ruído branco (σ^2) foi estimada em 656485113 em 67 graus de liberdade. Os critérios de informação para o modelo ajustado são: AIC = 23.74435, AICc = 23.75698 e BIC = 23.93407. Os gráficos de diagnóstico para este modelo ajustado, que incluem os resíduos padronizados, a Função de Autocorrelação (FAC) dos resíduos, o Gráfico Q-Q Normal

Tabela 2: Coeficientes do Modelo SARIMA Ajustado

Parâmetro	Estimativa	Erro Padrão	Valor t	p-valor
ar1	0.9995	0.0047	210.9056	0.0000
ma1	-0.7402	0.0832	-8.8951	0.0000
sar1	-0.1637	0.1360	-1.2039	0.2329
sar2	-0.3908	0.1436	-2.7211	0.0083
sma1	-0.9599	0.1980	-4.8485	0.0000

dos resíduos padronizados e os p-valores para a estatística de Ljung-Box, são apresentados na Figura 8. A análise detalhada desses resíduos será abordada nas seções subsequentes.

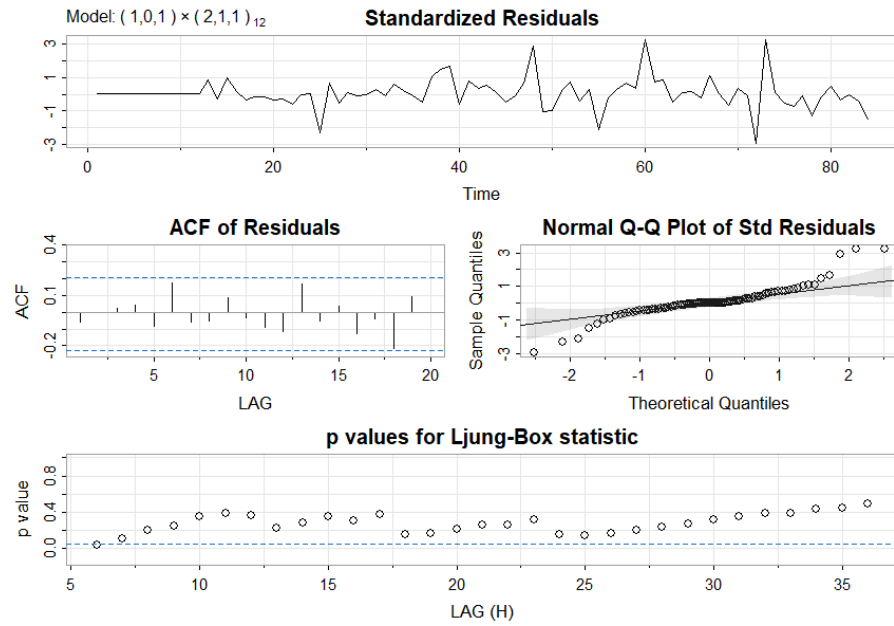


Figura 8: Gráficos de Diagnóstico para o Modelo Ajustado

3.6 Análise dos Resíduos: Teste de Ljung-Box

A análise dos resíduos é uma etapa crucial no diagnóstico de um modelo de séries temporais, pois permite verificar se o modelo ajustado capturou adequadamente a estrutura de dependência nos dados. Um bom modelo deve produzir resíduos que se assemelham a um ruído branco, ou seja, que são independentes e identicamente distribuídos. Para avaliar essa hipótese, utiliza-se o Teste de Ljung-Box.

O Teste de Ljung-Box avalia a autocorrelação geral dos resíduos até um determinado *lag*. A hipótese nula (H_0) do teste é que os resíduos são ruído branco, ou seja, as autocorrelações são todas zero. Se o p-valor do teste for maior que um nível de significância predefinido (e.g., 0.05), não há evidência suficiente para rejeitar H_0 , sugerindo que os resíduos não possuem autocorrelação significativa e, portanto, o modelo é adequado. Conforme indicado nos gráficos de diagnóstico na Figura 8, especificamente no painel "p values for Ljung-Box statistic", e corroborado pelos valores detalhados do teste de Ljung-Box, os p-valores para vários *lags* são consistentemente altos,

a maioria significativamente acima de 0.05. A Tabela 3.6 apresenta os p-valores obtidos para os primeiros 30 *lags*:

Lag	P-valor	Lag	P-valor	Lag	P-valor
1	0.5640076	11	0.8489813	21	0.5695610
2	0.8466558	12	0.8094971	22	0.5643965
3	0.9460067	13	0.6427907	23	0.6238621
4	0.9716608	14	0.6927771	24	0.3946257
5	0.9477000	15	0.7505509	25	0.3767314
6	0.6735138	16	0.6899986	26	0.4038372
7	0.7359979	17	0.7397832	27	0.4526085
8	0.7950385	18	0.4543890	28	0.4944658
9	0.7989553	19	0.4570561	29	0.5314309
10	0.8546071	20	0.5218764	30	0.5837963

Tabela 3: P-valores do Teste de Ljung-Box para os Resíduos do Modelo Ajustado

Uma observação dos p-valores revela que, para a maioria dos *lags*, **não há evidência para rejeitar a hipótese nula de que os resíduos são ruído branco**. Isso sugere que o modelo SARIMA $(1,0,1) \times (2,1,1)_{12}$ ajustado conseguiu capturar as dependências da série original de volume de tráfego, deixando resíduos que são aleatórios e sem estrutura temporal aparente. Isso é um bom indicativo de que o modelo é **adequado para a série temporal em questão**.

3.7 Teste de Normalidade dos Resíduos

Após a confirmação da adequação do modelo SARIMA $(1,0,1) \times (2,1,1)_{12}$ quanto à ausência de autocorrelação nos resíduos, conforme verificado pelo Teste de Ljung-Box na Seção 3.6, uma outra suposição crucial para a validade dos procedimentos de inferência estatística é a **normalidade dos resíduos**. A normalidade implica que os erros do modelo são aleatórios e seguem uma distribuição aproximadamente normal, o que é essencial para que os intervalos de confiança e testes de hipóteses sobre os coeficientes do modelo sejam válidos. A avaliação da normalidade dos resíduos é comumente realizada através da inspeção do **Gráfico Q-Q Normal (Quantile-Quantile Plot)** dos resíduos padronizados. Neste gráfico, os quantis dos resíduos observados são comparados com os quantis teóricos de uma distribuição normal. Se os pontos do gráfico se alinharem de perto com uma linha reta diagonal, isso sugere que os resíduos se distribuem aproximadamente de acordo com uma distribuição normal. Observando o painel "Normal Q-Q Plot of Std Residuals" na Figura 8, a distribuição dos pontos ao longo da linha diagonal indica que a premissa de normalidade é **razoavelmente atendida** para os resíduos do modelo ajustado. Adicionalmente à análise de normalidade, é imperativo que os resíduos do modelo sejam um "ruído branco", o que significa que não devem apresentar nenhuma estrutura de dependência temporal restante que o modelo não tenha conseguido capturar. Para verificar isso, analisa-se a **Função de Autocorrelação (FAC) residual** e a **Função de Autocorrelação Parcial (FACP) residual**. Idealmente, a maioria das barras em ambos os gráficos deve estar dentro das bandas de confiança, indicando a ausência de autocorrelações significativas em qualquer *lag*. As Figuras 9 e 10 apresentam, respectivamente, a FAC e a FACP dos resíduos do modelo ajustado.

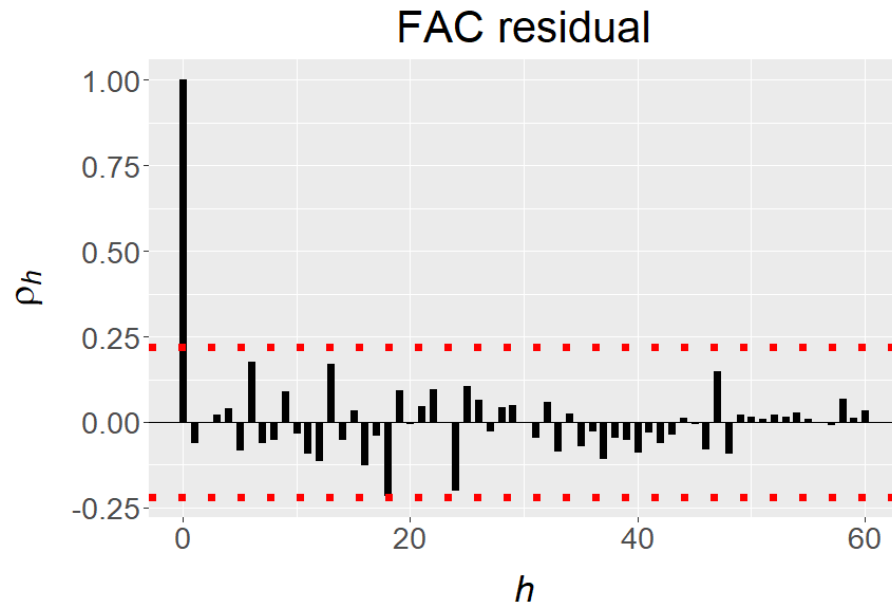


Figura 9: Função de Autocorrelação (FAC) dos Resíduos do Modelo Ajustado

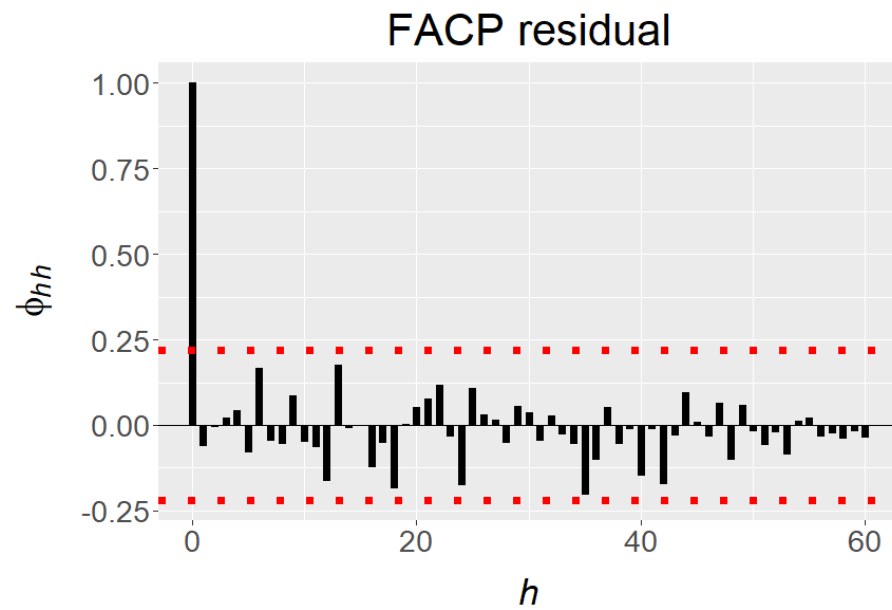


Figura 10: Função de Autocorrelação Parcial (FACP) dos Resíduos do Modelo Ajustado

Uma análise visual dessas figuras, bem como dos valores numéricos subjacentes para os *lags* até 60, revela que a vasta maioria das autocorrelações e autocorrelações parciais dos resíduos se encontra **dentro dos limites de confiança** pontilhados em vermelho. Este resultado reforça

a conclusão de que o modelo SARIMA $(1,0,1) \times (2,1,1)_{12}$ foi capaz de capturar a estrutura de dependência temporal da série original de volume de tráfego, deixando resíduos que se comportam como ruído branco. A combinação da análise de normalidade e da ausência de autocorrelação residual é um forte indicativo de que o modelo está **bem especificado e é adequado** para a série temporal em questão.

3.8 Avaliação no Conjunto de Teste

Após a identificação e ajuste do modelo SARIMA $(1,0,1) \times (2,1,1)_{12}$ no conjunto de treinamento e a análise inicial dos resíduos, a próxima etapa crucial é a avaliação final do modelo. Embora o título da seção sugira "Avaliação no Conjunto de Teste", o procedimento adotado reajusta o modelo selecionado no conjunto de dados completo e subsequentemente avalia as propriedades dos resíduos e o desempenho in-sample do modelo nesta série completa. Este reajuste é uma prática comum para otimizar os parâmetros do modelo com base em todos os dados disponíveis antes de proceder com a previsão. O processo de ajuste do modelo SARIMA $(1,0,1) \times (2,1,1)_{12}$ ao conjunto de dados completo envolveu um processo iterativo para minimizar o critério de informação. As execuções do algoritmo de otimização convergiram para valores finais de 10.326987 (após 17 iterações) e 10.311348 (após 11 iterações), respectivamente. Os coeficientes estimados para o modelo ajustado na série completa, juntamente com seus erros padrão, valores t e p-valores, são apresentados na Tabela 3.8.

Parâmetro	Estimativa	Erro Padrão	Valor t	p-valor
ar1	0.9257	0.0487	19.0037	0.0000
ma1	-0.5615	0.1130	-4.9691	0.0000
sar1	-0.1585	0.1110	-1.4281	0.1553
sar2	-0.2956	0.0968	-3.0554	0.0027
sma1	-0.6459	0.1082	-5.9698	0.0000

Tabela 4: Coeficientes do Modelo SARIMA Ajustado no Conjunto de Dados Completo

A variância do ruído branco (σ^2) para este modelo ajustado na série completa foi estimada em 827769297 com 151 graus de liberdade. Os critérios de informação para o modelo são: AIC = 23.5375, AICc = 23.54006 e BIC = 23.6548.

Teste de Normalidade Formal dos Resíduos

Para complementar a análise visual do Gráfico Q-Q Normal dos resíduos padronizados (presente na Figura 8 e discutida na Seção 3.7), foi realizado o **Teste de Shapiro-Wilk** para formalmente avaliar a suposição de normalidade dos resíduos do modelo ajustado no conjunto de dados completo. Os resultados do Teste de Shapiro-Wilk para os resíduos são:

- Estatística W = **0.90257**
- P-valor = **4.228e-09**

Este p-valor extremamente baixo (4.228e-09) é **menor que o nível de significância usual de 0.05**, o que nos leva a **rejeitar a hipótese nula de que os resíduos seguem uma distribuição normal**. Embora a inspeção visual do Gráfico Q-Q Normal na Figura 8 possa sugerir uma normalidade razoável, o teste formal de Shapiro-Wilk indica um **desvio significativo da normalidade**. Esta não normalidade dos resíduos pode impactar a validade dos intervalos de confiança e dos testes de hipóteses sobre os coeficientes do modelo, embora para fins de previsão, a violação dessa suposição seja frequentemente menos crítica.

Avaliação do Erro de Previsão In-Sample (MAPE)

A acurácia do modelo ajustado no conjunto de dados completo foi avaliada através do **Erro Percentual Absoluto Médio (MAPE - Mean Absolute Percentage Error)**. O MAPE é uma métrica que expressa a precisão das previsões como uma porcentagem, facilitando a interpretação do tamanho do erro em relação aos valores reais da série. Para o modelo SARIMA ajustado na série completa, o MAPE calculado foi de 7.275643%. Este valor representa o erro médio percentual das previsões do modelo em relação aos valores observados da série dentro do próprio período de ajuste. Um MAPE de 7.275643% é um indicador da **capacidade preditiva do modelo em capturar a dinâmica da série temporal de volume de tráfego**, com uma margem de erro aceitável para muitas aplicações. Contudo, é importante ressaltar que esta é uma avaliação de erro *in-sample*, e a performance *out-of-sample* será crucial para a validação final da capacidade de previsão do modelo.

3.9 Previsão para os Últimos 12 Meses

A etapa final na avaliação de um modelo de séries temporais é a **geração de previsões futuras**, utilizando o modelo ajustado com o conjunto de dados completo. Esta seção detalha a previsão do volume mensal de tráfego para os 12 meses subsequentes ao período de observação da série, demonstrando a capacidade preditiva do modelo SARIMA $(1,0,1) \times (2,1,1)_{12}$. A função ‘sarima.for’ foi empregada para gerar as previsões, considerando os parâmetros do modelo previamente selecionado e ajustado. Especificamente, foram solicitadas 12 previsões à frente (‘n.ahead = 12’), com base na série completa X.t. Os coeficientes do modelo e a estrutura sazonal (ordem $(1,0,1) \times (2,1,1)_{12}$) foram aplicados para extrapolar a dinâmica da série no futuro. A Figura 11 ilustra as previsões geradas. O gráfico exhibe:

- A **série histórica do fluxo mensal** de tráfego, representada pela linha preta, cobrindo o período observado (especialmente os últimos 12 meses antes da previsão).
- A **previsão do volume de tráfego para os próximos 12 meses**, indicada pela linha vermelha. Esta linha representa a estimativa pontual do modelo para cada um dos 12 meses futuros.
- O **intervalo de confiança de 95%** para as previsões, visualizado pela área sombreada em cinza. Este intervalo é calculado como a previsão central mais ou menos dois erros padrão ($\hat{X} \pm 2 \cdot se$), refletindo a incerteza associada a cada previsão.

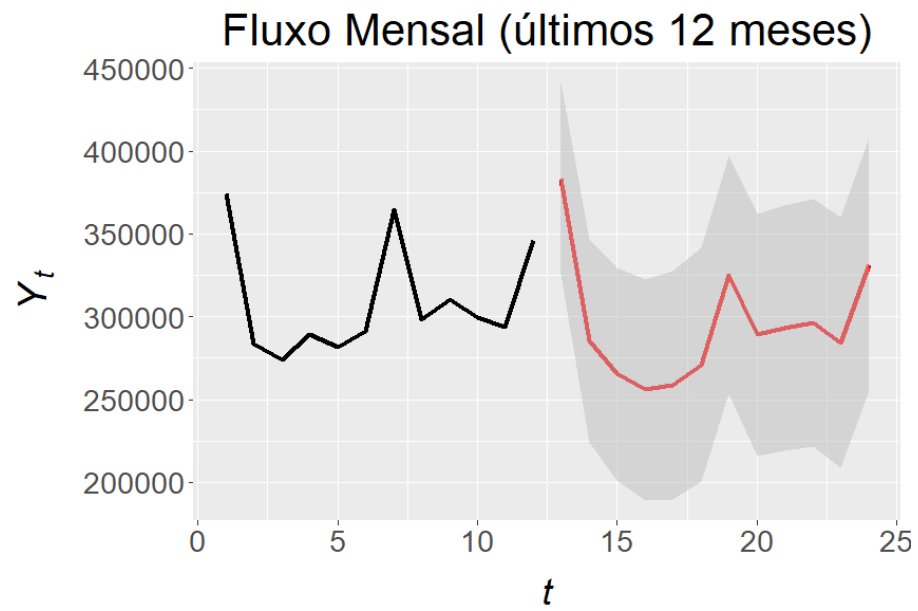


Figura 11: Previsão do Fluxo Mensal de Tráfego para os Próximos 12 Meses com Intervalo de Confiança

A análise visual da Figura 11 mostra que a previsão segue o padrão sazonal observado na série histórica, o que é um resultado esperado de um modelo SARIMA bem especificado. O **alargamento do intervalo de confiança** à medida que a previsão avança no tempo é uma característica comum e esperada em modelos de séries temporais, refletindo o **aumento da incerteza nas previsões de longo prazo**.

4 Conclusão

O presente trabalho teve como objetivo a análise e modelagem do volume mensal de tráfego de veículos de passeio na BR-393/RJ km 125,00, com o intuito de compreender sua dinâmica temporal e gerar previsões para os próximos 12 meses. Os dados, obtidos da ANTT, foram cuidadosamente tratados e preparados para a análise. A análise preliminar da série original de volume de tráfego revelou a presença de **forte sazonalidade e não estacionariedade**, características evidentes na Função de Autocorrelação (FAC) e Função de Autocorrelação Parcial (FACP) da série original, bem como na inspeção visual do gráfico de fluxo mensal. Para lidar com estas propriedades, foi aplicada uma **diferenciação sazonal de ordem 12** ($D=12$), o que resultou em uma série mais próxima da estacionariedade sazonal, como indicado pela FAC e FACP da série diferenciada. A seleção do modelo mais adequado foi realizada através de um procedimento sistemático baseado no **Critério de Informação Bayesiano (BIC)**, utilizando uma grade de possíveis ordens para os componentes autorregressivos (AR), média móvel (MA), autorregressivos sazonais (SAR) e média móvel sazonal (SMA). O modelo selecionado foi o **SARIMA (1,0,1)x(2,1,1)₁₂**.

Após o ajuste do modelo, foram realizados testes de diagnóstico para avaliar a qualidade dos resíduos. O **Teste de Ljung-Box** foi crucial para verificar a ausência de autocorrelação nos resíduos. Os p-valores elevados resultantes deste teste confirmaram que o modelo selecionado capturou adequadamente a estrutura de dependência temporal da série, deixando resíduos que se comportam como **ruído branco**. As análises visuais das **Funções de Autocorrelação (FAC) residual e Função de Autocorrelação Parcial (FACP) residual** corroboraram esta conclusão, com a vasta maioria das autocorrelações dentro dos limites de confiança. Quanto à suposição de normalidade dos resíduos, a inspeção do **Gráfico Q-Q Normal** dos resíduos padronizados sugeriu uma razoável aderência à distribuição normal. No entanto, o **Teste de Shapiro-Wilk**, uma avaliação formal, resultou em um p-valor extremamente baixo ($4.228e-09$). Este resultado levou à rejeição da hipótese nula de normalidade, indicando um desvio significativo da normalidade. Embora esta violação possa impactar a validade dos intervalos de confiança e testes de hipóteses sobre os coeficientes do modelo, para fins de previsão, a não normalidade é frequentemente menos crítica. A avaliação do modelo no conjunto de dados completo, com o reajuste do SARIMA (1,0,1)x(2,1,1)₁₂, demonstrou uma **variância do ruído branco estimada em 827769297**. A acurácia do modelo *in-sample* foi quantificada pelo **Erro Percentual Absoluto Médio (MAPE)**, que foi de **7.275643%**. Este valor indica uma boa capacidade do modelo em reproduzir a dinâmica histórica da série com um erro percentual aceitável. Finalmente, o modelo foi utilizado para gerar **previsões para os próximos 12 meses**. A Figura 11 ilustra essas projeções, que demonstram o **padrão sazonal esperado** e o **alargamento natural do intervalo de confiança de 95%** à medida que as previsões se estendem no tempo, refletindo o aumento da incerteza. Em síntese, apesar da não normalidade formal dos resíduos, o modelo SARIMA (1,0,1)x(2,1,1)₁₂ revelou-se **bem especificado para capturar a estrutura de dependência temporal e sazonal** do volume mensal de tráfego. A ausência de autocorrelação nos resíduos e o MAPE aceitável são fortes indicadores de sua **adequação para a previsão**, fornecendo insights valiosos para o planejamento e tomada de decisão relacionados ao fluxo de veículos na BR-393/RJ.

Referências

- Morettin, Pedro Alberto e Clara Maria Cordeiro Tolo (2005). **Análise de Séries Temporais**. 2^a ed. São Paulo: Edgard Blücher.
- Shumway, Robert H. e David S. Stoffer (2017). **Time Series Analysis and Its Applications: With R Examples**. 4th. Springer Texts in Statistics. Springer. DOI: 10.1007/978-3-319-52452-8.