



Universidade da Integração Internacional da Lusofonia
Afro-Brasileira
Instituto de Engenharias e Desenvolvimento Sustentável

Análise de Séries Temporais para Casos de Dengue no município de Fortaleza

Alexandre Lima de Oliveria
Ithalo Luz Mota
Ivina Lorena Oliveira Moura
João Pedro Rodrigues Dutra
Leticia da Costa Silva

Acarape – CE
Julho de 2025

Sumário

| | |
|----------------------------------------------------------|-----------|
| Sumário | 2 |
| 0.1 Introdução | 3 |
| 0.1.1 Tipos de séries temporais | 3 |
| 0.1.1.1 Série temporal univariada e multivariada | 3 |
| 0.1.1.2 Série temporal regular e irregular | 3 |
| 0.1.2 Estacionariedade | 3 |
| 0.1.3 Componentes de uma série temporal | 4 |
| 0.1.4 Defasagem(Lag) | 4 |
| 0.1.5 Autocorrelação | 5 |
| 0.1.6 Modelos Autoregressivos | 5 |
| 0.1.7 Covariância | 5 |
| 0.2 Objetivos | 5 |
| 0.2.1 Objetivos específicos | 6 |
| 0.3 Análise de séries temporais | 6 |
| 0.3.0.1 Autocorrelação em Python | 7 |
| 0.3.0.2 Autocorrelação e Autocorrelação Parcial em Excel | 9 |
| 0.3.1 Filtragem | 9 |
| 0.3.2 Escolha da ordem do modelo AR | 12 |
| 0.3.3 Previsão AR(13) | 16 |
| 0.4 Conclusão | 18 |
| Bibliografia | 19 |

0.1 Introdução

Séries temporais consistem em sequências de dados obtidos em momentos sucessivos e igualmente espaçados no tempo. A análise dessas séries tem como objetivo identificar comportamentos recorrentes, como tendências, sazonalidades e outros padrões, com o intuito de apoiar a estimativa de valores futuros [4]. Uma série temporal pode ser decomposta em três componentes não observáveis: tendência (T_t), sazonalidade (S_t) e ruído branco (a_t), que representa a variação aleatória. Uma abordagem inicial comum para modelar esses elementos é considerar uma relação aditiva entre eles [6, 11], expressa como:

$$Z_t = T_t + S_t + a_t \quad (1)$$

Antes da Inteligência Artificial se tornar popular o principal método para a modelagem de séries temporais eram as análises estatísticas.

0.1.1 Tipos de séries temporais

0.1.1.1 Série temporal univariada e multivariada

A principal diferença entre uma série temporal univariada e uma multivariada está na quantidade de variáveis observadas em cada instante de tempo. Enquanto a univariada registra apenas um valor por timestamp, a multivariada contempla múltiplas variáveis simultaneamente.

0.1.1.2 Série temporal regular e irregular

1. Regular: A diferença principal entre uma série temporal regular e irregular é a forma como esses dados são coletados. Como o próprio nome sugere, uma série temporal regular é coletada de forma consistente e regular.
2. Irregular: Já uma série temporal irregular pode ter intervalos de tempo diferentes entre registros. Exemplo: dados simulados de um sensor que gera registros de alerta cada vez que a maré passa de um determinado nível.

0.1.2 Estacionariedade

Uma série temporal estacionária é aquela em que características estatísticas, como a média e a variância, não mudam significativamente com o passar do tempo. Por outro lado, numa série temporal não estacionária as características estatísticas mudam ao longo do tempo. Para tornar a série estacionária uma das técnicas é a diferenciação: onde,

$$D(Y_t) = Y_t - Y_{t-1} \quad (2)$$

Y_t é o valor da série original no momento t .

Y_{t-1} é o valor da série original no momento $t-1$.

$D(Yt)$ é o valor da série diferenciada no momento t .

Tabela 1 – Objetivos e aplicações da diferenciação em séries temporais

| Objetivo da Diferenciação | Quando Usar |
|----------------------------|------------------------------------------|
| Remover tendência linear | Série com crescimento constante |
| Estacionarizar série | Antes de aplicar ARIMA ou ACF/PACF |
| Reduzir sazonalidade leve | Em séries com padrão repetitivo |
| Destacar flutuações locais | Quando os níveis brutos mascaram padrões |

0.1.3 Componentes de uma série temporal

Séries temporais podem ser decompostas em 4 componentes principais: tendência, sazonalidade, ciclo e ruído [10].

1. Tendência em séries temporais:

A tendência é a direção geral dos dados ao longo do tempo. O componente de tendência é frequentemente associado ao ciclo de tendência, sendo que os ciclos são caracterizados como padrões de longa duração e comportamento irregular [5].

2. Sazonalidade em séries temporais:

A sazonalidade é uma variação dos dados que sempre ocorre (se repete) em um período específico, é uma variação dos dados que sempre ocorre em um período conhecido.

3. Ciclos em séries temporais:

Ciclo também é uma variação dos dados que ocorre em um período específico e se repete, mas não tem uma frequência fixa como nos exemplos da sazonalidade.

4. Ruído em séries temporais:

O ruído é a variação aleatória dos dados que não pode ser explicada pelos outros componentes.

0.1.4 Defasagem(Lag)

Defasagem ou *Lag* é o deslocamento de um valor anterior de uma série para um período posterior. O *Lag* muda apenas deslocamentos, não a série temporal de origem. Portanto, um atraso de -2 elimina os primeiros dois elementos [8].

0.1.5 Autocorrelação

A autocorrelação, também conhecida como correlação serial, é uma técnica utilizada na análise de séries temporais para verificar a existência de dependência entre os valores atuais e seus valores passados ao longo do tempo [9]. A compreensão sobre a autocorrelação à compreensão da estrutura temporal dos dados. É necessário calcular para identificar a dependência temporal, o que ajuda a entender padrões repetitivos ou estruturas cíclicas nos dados.

$$\rho(h) = \frac{\text{Cov}(X_{t+h}, X_t)}{\sqrt{\text{Var}(X_{t+h})\text{Var}(X_t)}} \quad (3)$$

Para avaliar a estacionariedade, a autocorrelação tende a cair rapidamente para zero conforme o "lag" (defasagem) aumenta. Em séries não estacionárias, a autocorrelação persiste ao longo do tempo, indicando que transformações como diferenciação podem ser necessárias.

Para escolher um modelo apropriado é necessário analisar os gráfico de autocorrelação (ACF) e autocorrelação parcial (PACF).

0.1.6 Modelos Autoregressivos

- Definição

Um modelo autorregressivo de ordem p (AR(p)) usa uma combinação dos valores passados da própria série para prever o valor atual.

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_p X_{t-p} + E_t \quad (4)$$

O termo de perturbação E_t vem de um processo de ruído branco, sendo também estocasticamente independente de $X_{t-1}, X_{t-2}[\dots]$ [3].

0.1.7 Covariância

Ou variância conjunta $\text{Cov}(X_t, X_{t+1})$ é uma medida do grau de interdependência LINEAR entre duas variáveis aleatórias. Em uma série temporal, medidas adjacentes tendem a ser correlacionadas.

$$\text{Cov}(X, Y) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \quad (5)$$

0.2 Objetivos

Desenvolver uma análise de série temporal dos casos de dengue notificados em Fortaleza no período de 2014 a 2024.

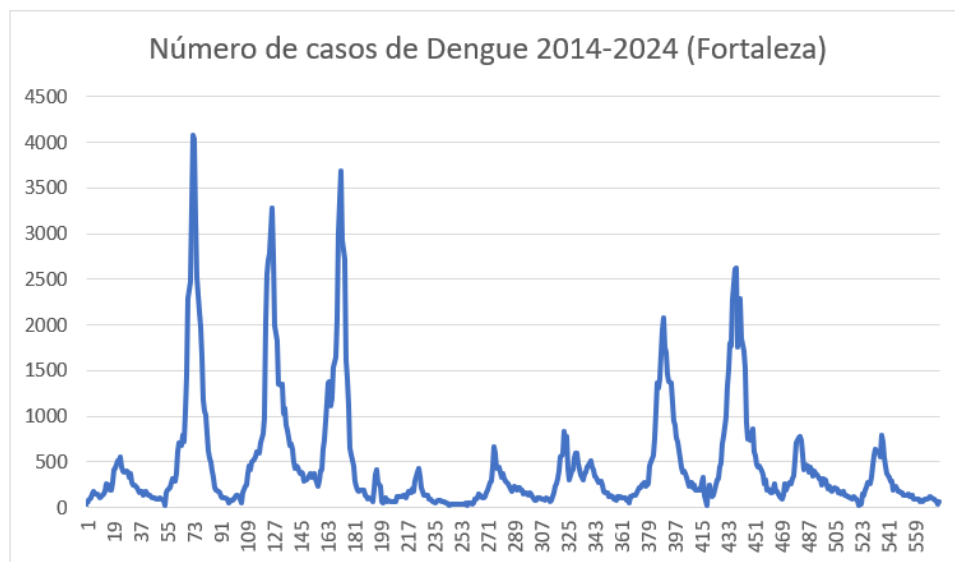
0.2.1 Objetivos específicos

- Identificar tendências
- Detectar padrões sazonais

0.3 Análise de séries temporais

A série temporal analisada refere-se aos casos de dengue registrados em Fortaleza, no período de 2014 a 2024. Os dados foram obtidos a partir da plataforma [InfoDengue](#). A frequência dos dados é semanal.

Figura 1 – Número de casos de Dengue nos períodos de 2014 a 2024



Os dados analisados apresentam assimetria positiva, o que justifica a aplicação do logaritmo na base 10. Essa transformação contribui para aproximar a distribuição dos dados de uma distribuição normal, atendendo a uma premissa fundamental de diversos métodos estatísticos.

Figura 2 – Dados após aplicação do logarítmico na base 10

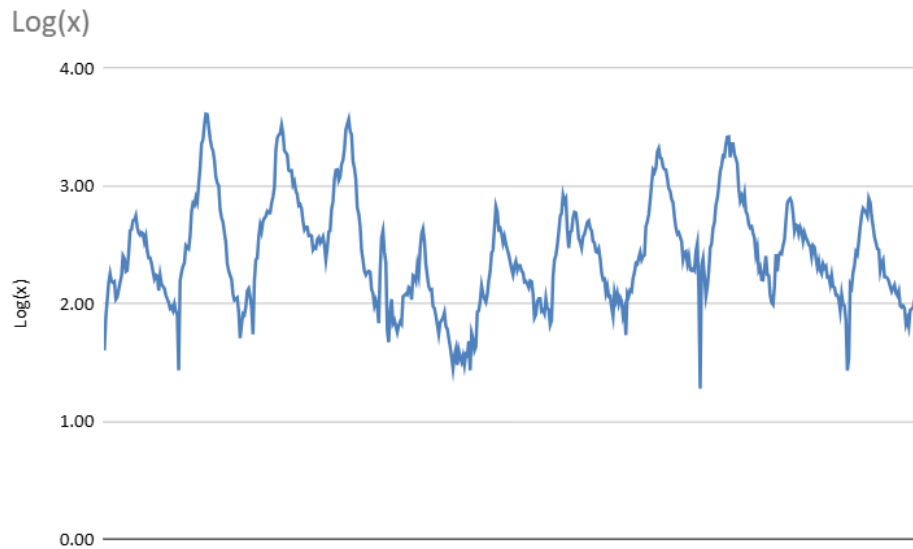
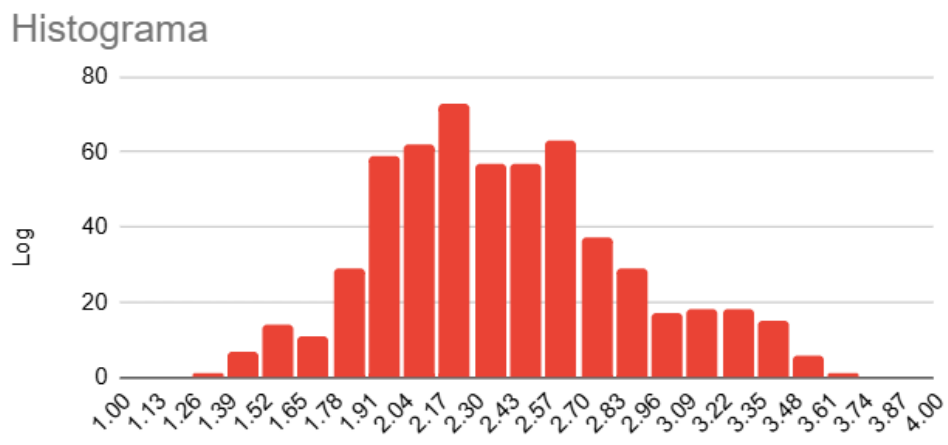


Figura 3 – Histograma após aplicação do logarítmico na base 10



0.3.0.1 Autocorrelação em Python

Para a plotagem do ACF foi usado a linguagem de Python:

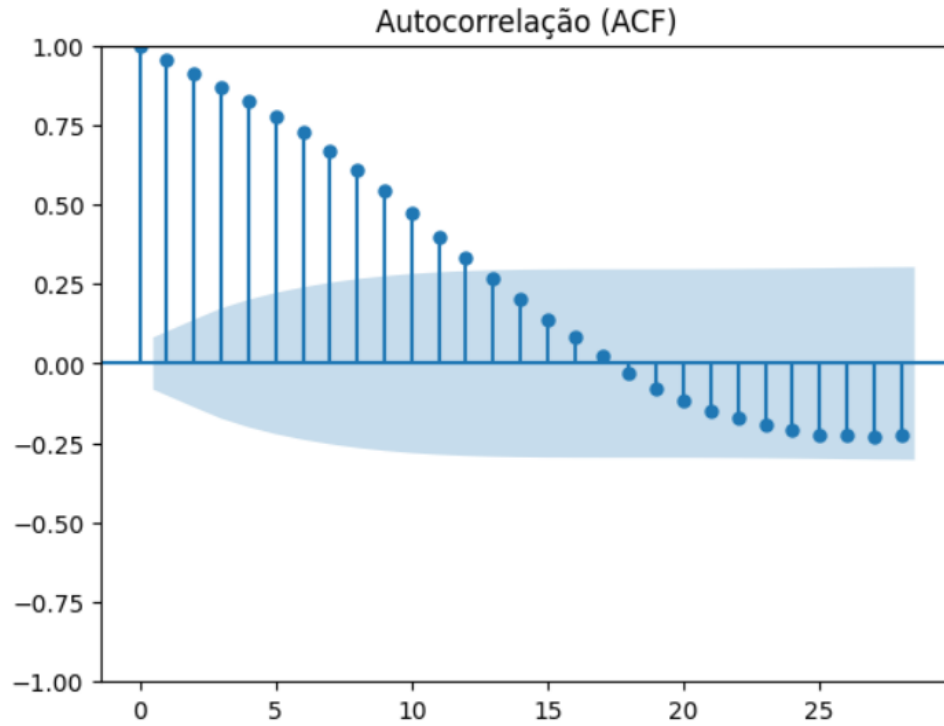
```
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.graphics.tsaplots import plot_acf
```

```
df = pd.read_excel("logacf.xlsx")
```

```
plt.figure(figsize=(10, 4))
plot_acf(df["Log"])
plt.title("Autocorrelacao_(ACF)")
```

```
plt.show()
```

Figura 4 – Autocorrelação (ACF)

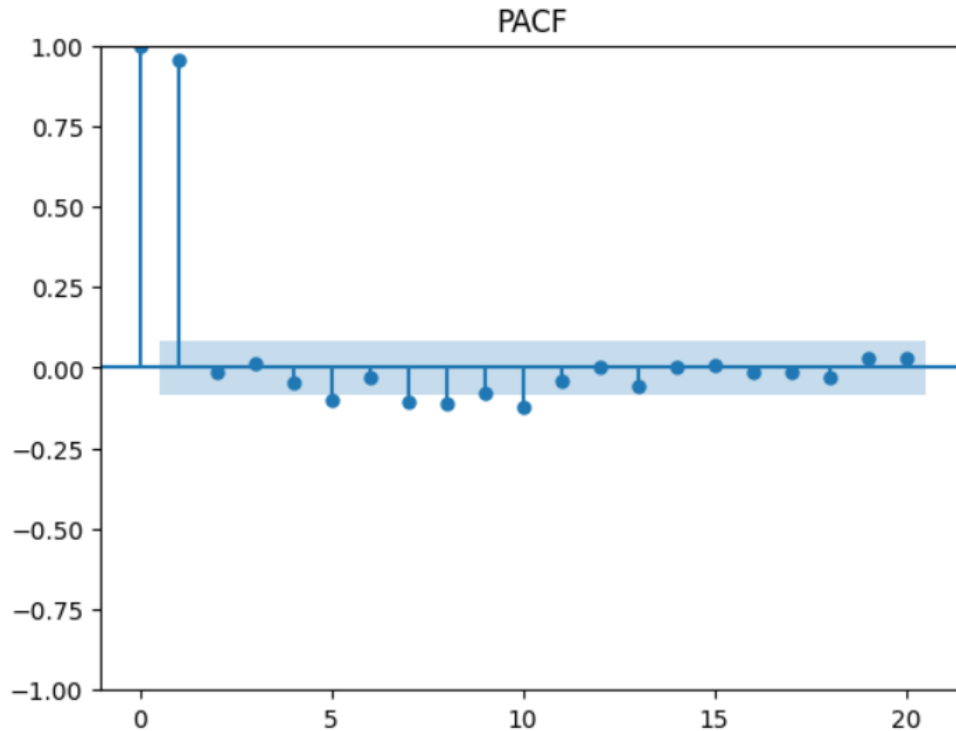


A função de autocorrelação parcial (PACF) é semelhante à autocorrelação simples (ACF), mas com uma diferença importante: ela mede exclusivamente a correlação entre duas observações, eliminando a influência das defasagens intermediárias. Enquanto o gráfico da ACF revela a correlação entre y_t e y_{t-k} para diferentes valores de k , ele não distingue se essa relação é direta ou mediada por valores anteriores. Por exemplo, se y_t está correlacionado com y_{t-1} , e este, por sua vez, com y_{t-2} , pode haver uma correlação aparente entre y_t e y_{t-2} que ocorre apenas por meio de y_{t-1} . O PACF, nesse caso, permite identificar se existe uma correlação direta entre y_t e y_{t-2} , isolando o efeito das defasagens mais curtas [9].

Para a plotagem do PACF:

```
from statsmodels.graphics.tsaplots
import plot_pacf
plot_pacf(df["Log"], lags=20, method='ywm')
plt.title("PACF")
plt.show()
```


Figura 5 – Autocorrelação parcial (PACF)



Pode se observar que o ACF decai lentamente e um PACF possui um corte brusco, o que pode indicar um modelo autorregressivo (AR).

0.3.0.2 Autocorrelação e Autocorrelação Parcial em Excel

Em relação ao cálculo da autocorrelação no Excel, utilizou-se o método *Plug-in* para estimar a função de autocorrelação, uma vez que o método baseado no gráfico de dispersão apresenta algumas limitações ([Aula 17: Por que o Plug-in é o melhor método para a Função de Autocorrelação?](#)).

Na estimativa *Plug-in*:

$$\rho(h) = \frac{\text{Cov}(X_{t+h}, X_t)}{\text{Cov}(X_t, X_t)} \quad (6)$$

A planilha utilizada foi disponibilizada pelo professor Alexandre Cunha [2], no seu blog de séries temporais [1].

0.3.1 Filtragem

Como explicado anteriormente, uma série temporal, ao ser decomposta há três componentes, onde:

- T_t : Tendência-ciclo (variações de longo prazo)
- S_t : Sazonalidade (padrões repetitivos)

- R_t : Resíduo (ruído e variações não explicadas)

(Referência: Hyndman & Athanasopoulos, 2021)

Em métodos avançados como o **STL** (*Seasonal-Trend Decomposition using Loess*), a decomposição pode ser:

- **Aditiva:**

$$Y_t = T_t + S_t + R_t \quad (7)$$

- **Multiplicativa:**

$$Y_t = T_t \times S_t \times R_t \quad (8)$$

A escolha depende da natureza da série temporal, sendo a forma multiplicativa mais adequada quando a amplitude das flutuações sazonais varia com o nível da série [7]. Após a aplicação do logaritmo na base 10 na série (Figura 2), procedeu-se à filtragem dos dados. Primeiramente, foi realizada a remoção da tendência (Figura 6), seguida pela eliminação da sazonalidade (Figura 7). Por fim, obteve-se a série final, resultante da subtração da série transformada em log10 com os componentes de tendência e sazonalidade.

Figura 6 – Tendência da Série

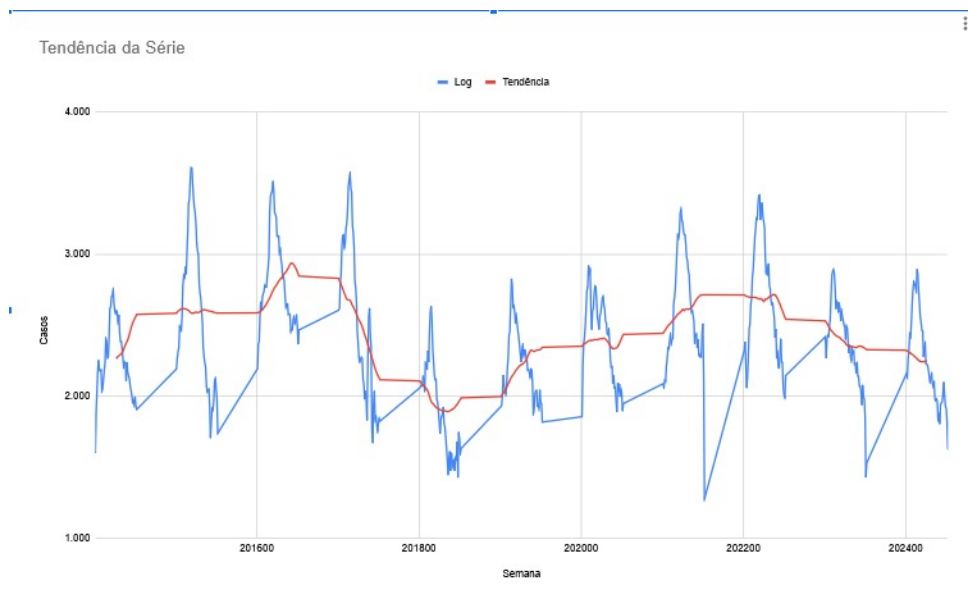
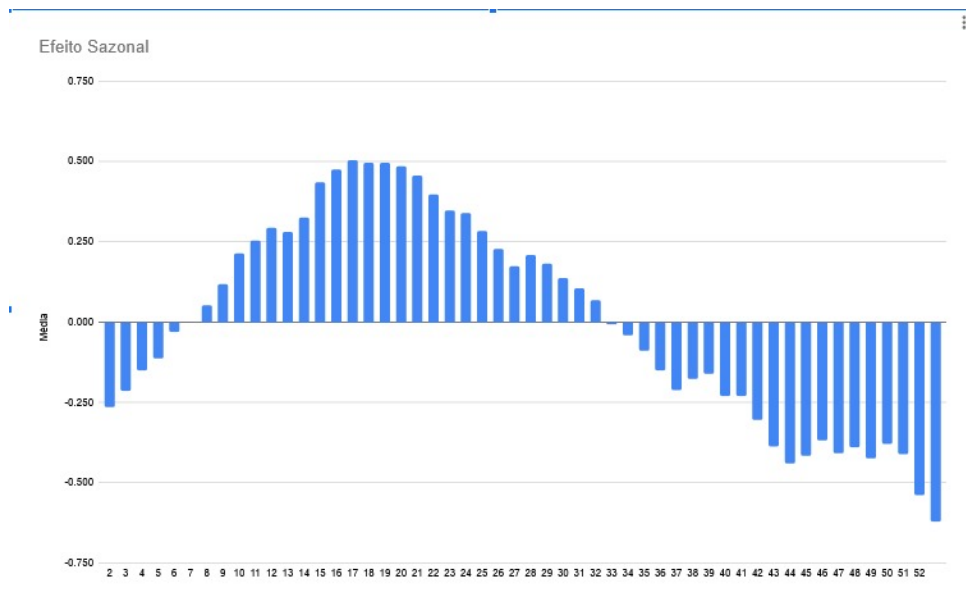
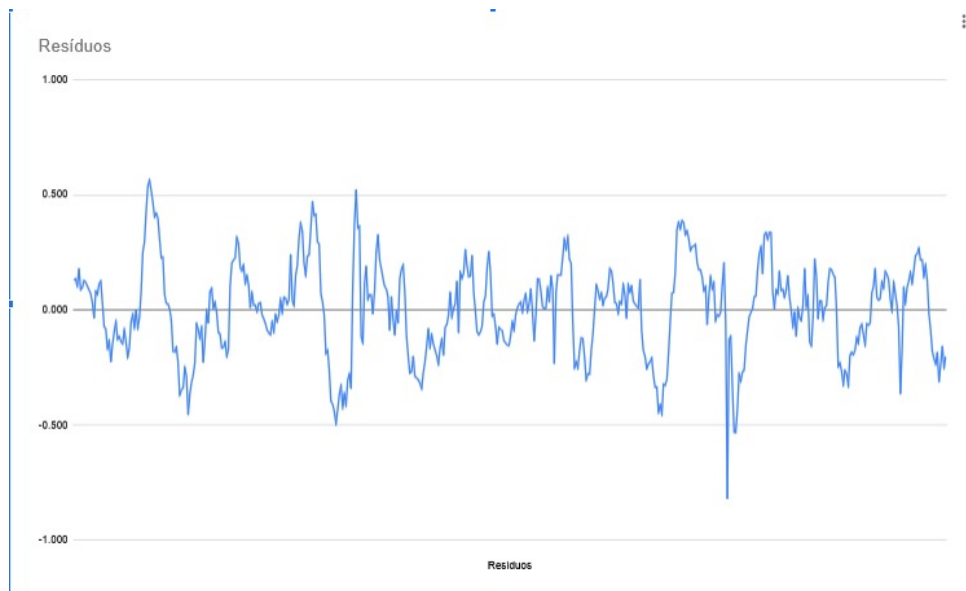


Figura 7 – Sazonalidade da Série



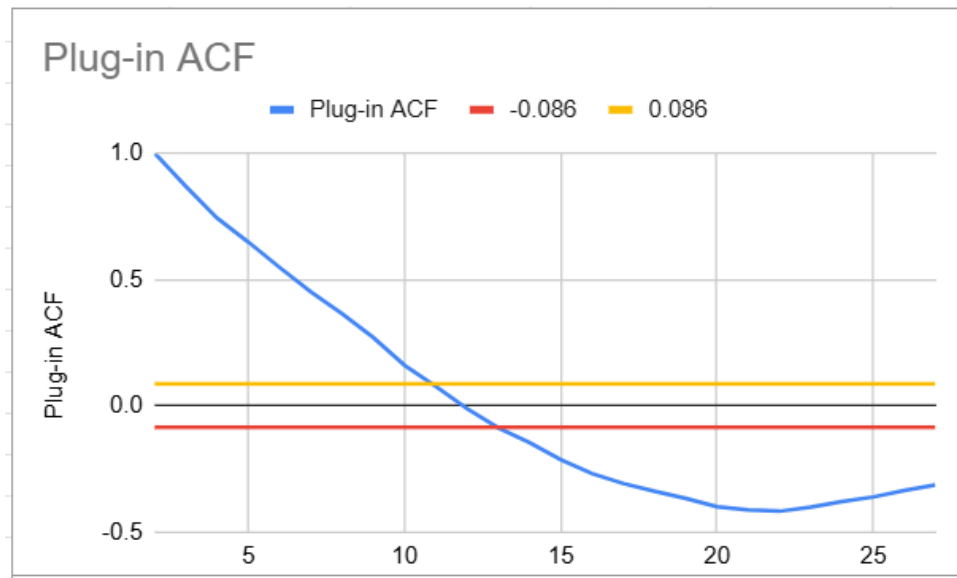
Ao transformar a série com log de base 10 estabiliza-se a variância, ao remover a tendência elimina-se o comportamento de longo prazo e ao remover a sazonalidade eliminam-se padrões repetitivos ao longo do tempo. O que sobra é a parte irregular da série, chamada de componente aleatório, ruído branco ou resíduos de decomposição (Figura 8). A remoção permite tornar a série estacionária, facilitando a modelagem e previsão [7].

Figura 8 – Resíduos da Série



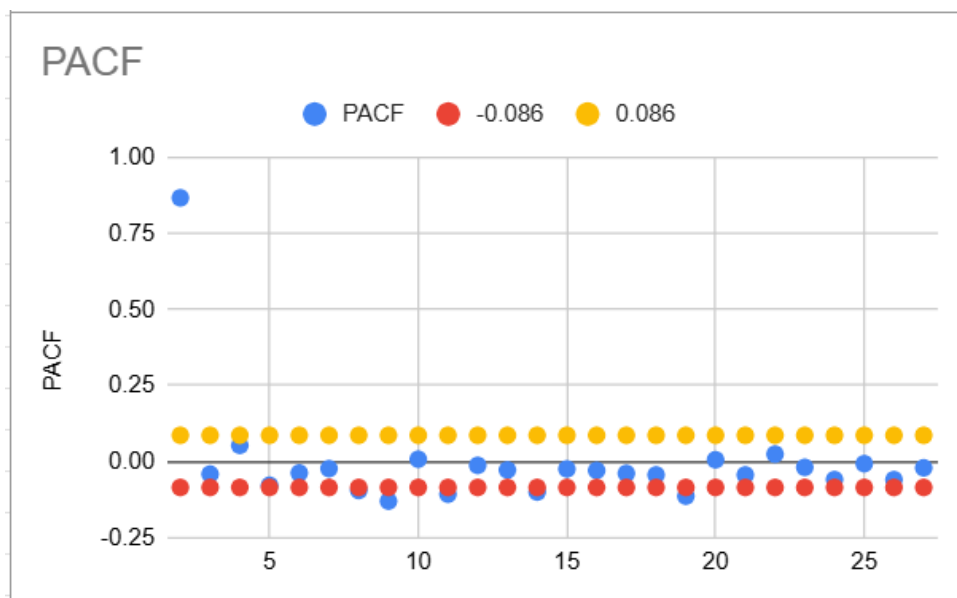
Após a aplicação das funções de autocorrelação (ACF) e autocorrelação parcial (PACF), observa-se na ACF (Figura 9) uma queda exponencial dos coeficientes, indicando a presença de dependência temporal típica de séries não estacionárias.

Figura 9 – Função de autocorrelação (ACF) após a filtragem da série



Já no PACF (Figura 10) observa-se um corte brusco, o que sugere um modelo autor-regressivo (AR).

Figura 10 – Função de autocorrelação parcial (PACF) após a filtragem da série



0.3.2 Escolha da ordem do modelo AR

A escolha da ordem do modelo $AR(p)$ foi baseada na análise conjunta das funções de autocorrelação (ACF) e autocorrelação parcial (PACF), com o objetivo de identificar o número de defasagens significativas na série temporal. As ordens analisadas como mais relevantes foram 8, 13 e 18.

Critérios de Seleção: As ordens $p \in \{8, 13, 18\}$ foram selecionadas considerando:

- Significância estatística ($|\text{PACF}| > 2/\sqrt{n}$)

A autocorrelação do AR(8) é apresentada na Figura 11, e a autocorrelação parcial na Figura 12, após a análise visual da autocorrelação parcial podemos apontar picos nos lags 9 e 18, o que pode sugerir uma presença de correlação restante na série.

Figura 11 – Autocorrelação AR(8)

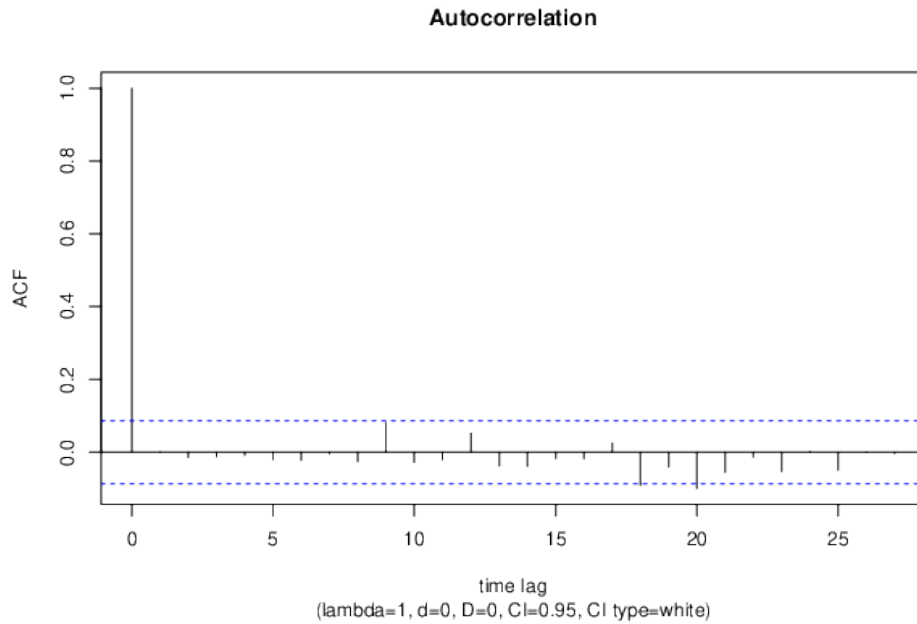
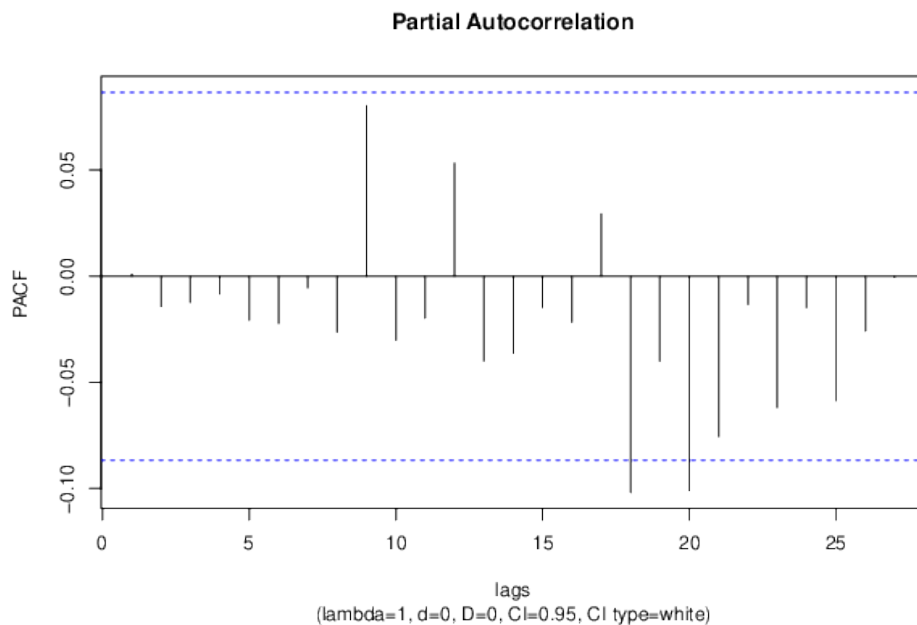


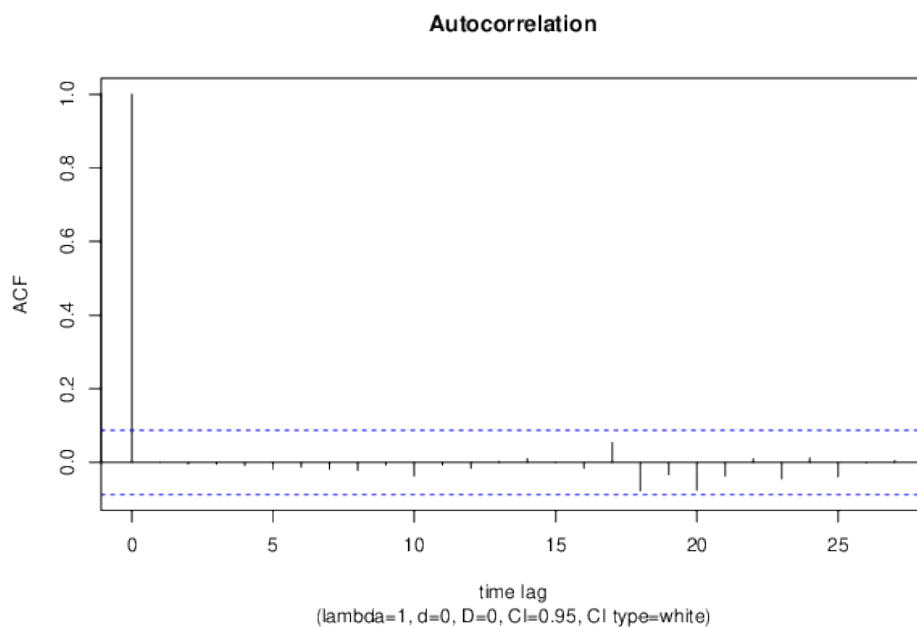
Figura 12 – Autocorrelação Parcial AR(8)



Já na análise do ACF-AR(13), Figura 13, observa-se que todos os coeficientes estão dentro da zona de confiança, logo, interpreta-se que o ACF não apresenta autocorrelações

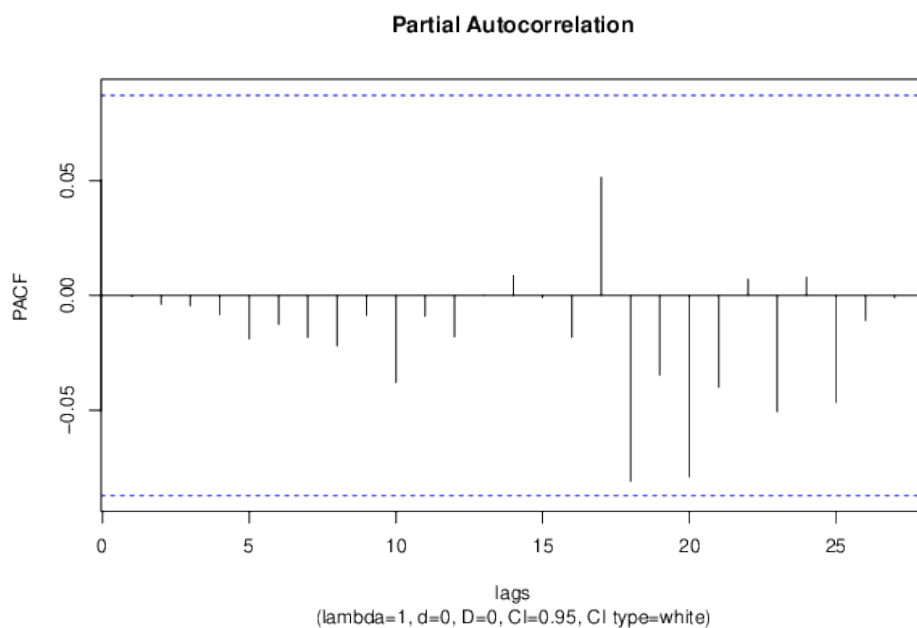
significativas, indicando que os resíduos se comportam como um ruído branco. Isso também sugere que o modelo AR(13) ajustou bem a dependência serial da série original.

Figura 13 – Autocorrelação AR(13)



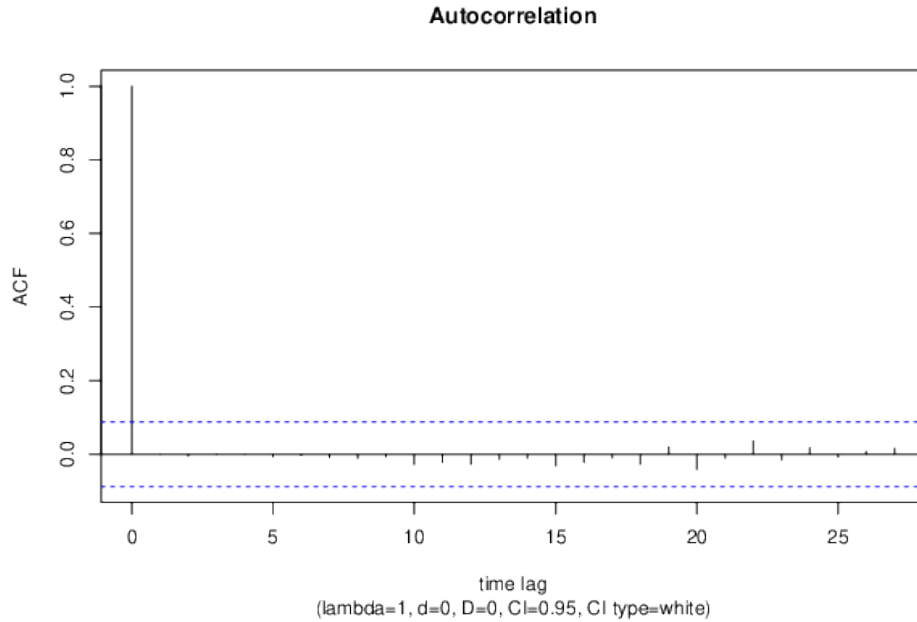
O gráfico de autocorrelação parcial (PACF) do AR(13), Figura 14, não exibe cortes claros fora da ordem definida, o que indica que não há estrutura de dependência direta adicional que o modelo não tenha capturado.

Figura 14 – Autocorrelação Parcial AR(13)



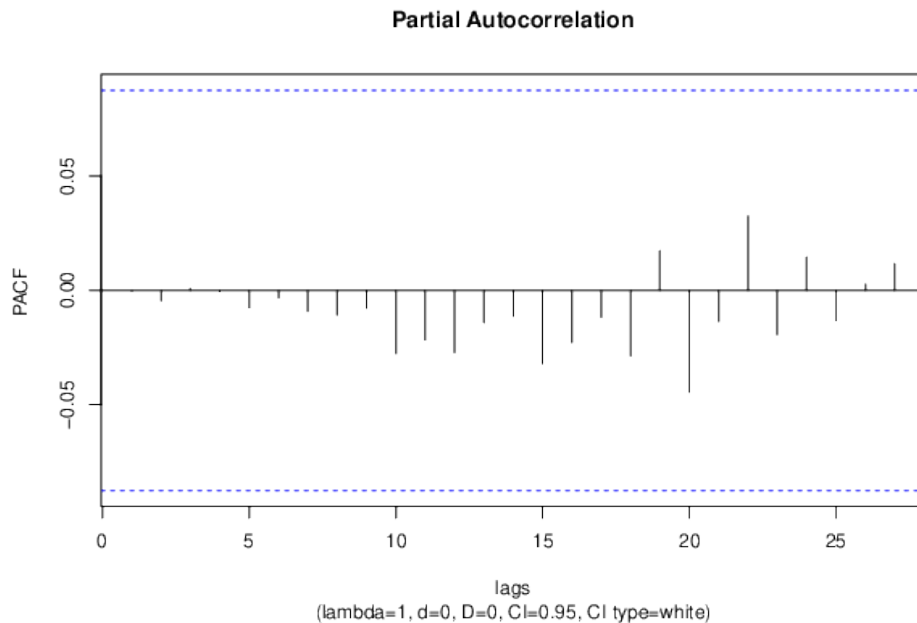
A Figura 15 apresenta o gráfico da função de autocorrelação (ACF) dos resíduos do modelo AR(18), indicando ausência de autocorrelações significativas, pois todos os valores permanecem dentro dos limites de confiança.

Figura 15 – Autocorrelação AR(18)



Da mesma forma, a função de autocorrelação parcial (PACF) (Figura 16) não mostra lags significativos, reforçando que o modelo ajustado explica satisfatoriamente a estrutura temporal da série.

Figura 16 – Autocorrelação Parcial AR(18)



Como observado, as ordens {13 e 18} apresentaram resíduos com comportamento semelhante ao ruído branco, porém para a escolha final da ordem do modelo autorregressivo optou-se pelo AR(13) por ser mais simples e, ainda assim, apresentar bom desempenho estatístico. Como o AIC/BIC foram próximos, isso reforça a escolha pela ordem menor, porque mostra que o ganho de ajuste não compensa o aumento de complexidade. Vale destacar que o coeficiente de determinação (R^2) do modelo AR(13) foi de aproximadamente 77%, indicando que o modelo é capaz de explicar uma parcela significativa da variabilidade da série temporal.

0.3.3 Previsão AR(13)

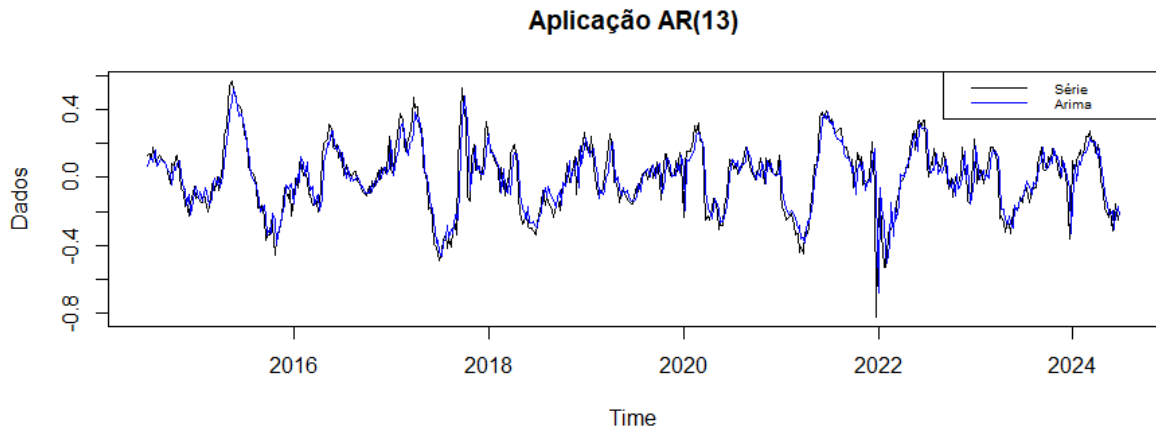
Após a análise exploratória da série temporal e a identificação da estrutura autorregressiva adequada, optou-se pelo modelo ARIMA(13,0,0), ou seja, um modelo puramente autorregressivo de ordem 13. Esta escolha foi fundamentada na análise do PACF, bem como nos critérios de informação (AIC e BIC), que indicaram melhor desempenho estatístico para este modelo em comparação com outras especificações. Com o modelo ajustado, prosseguimos com a etapa de previsão (ou precisão) para avaliar sua capacidade de capturar a dinâmica da série e produzir estimativas futuras confiáveis. Esta etapa é essencial para validar a utilidade prática do modelo e sua adequação aos dados observados.

Para o início da previsão, foi utilizado o modelo ARIMA (13,0,0) onde:

- AR(13): A componente autorregressiva de ordem 13 captura dependências lineares das 13 defasagens anteriores.
- I(0): Indica que a série não necessita de diferenciação (já é estacionária).
- MA(0): Não inclui termos de média móvel.

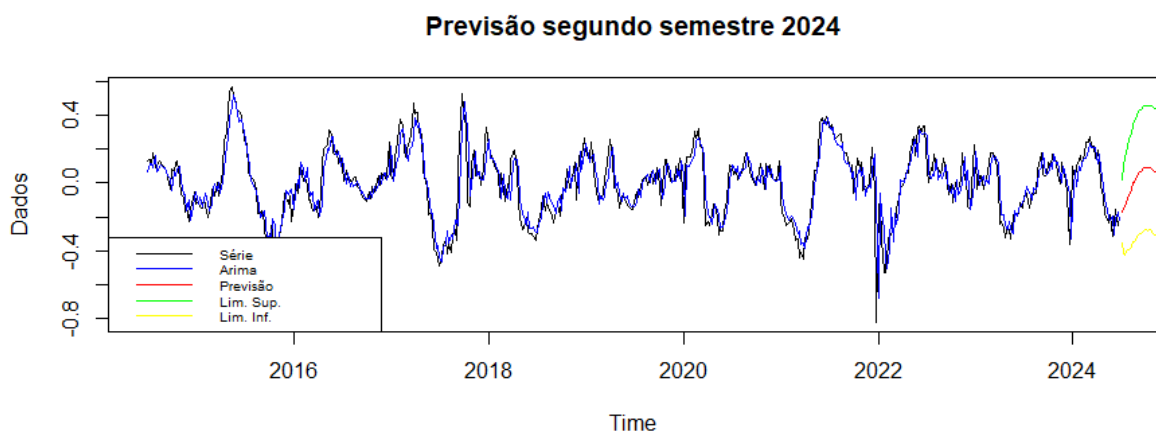
A Figura 17 apresenta a aplicação do modelo ARIMA(13,0,0), evidenciando o comportamento ajustado da série com base no componente autorregressivo de ordem 13.

Figura 17 – Aplicação AR(13)



A Figura 18 apresenta a previsão da série temporal para o segundo semestre de 2024 com base no modelo ARIMA(13,0,0). A linha preta representa os dados observados da série original, enquanto a linha azul mostra os valores ajustados pelo modelo ARIMA ao longo do período histórico. A projeção para o segundo semestre de 2024 está destacada em vermelho, representando os valores previstos. A linha em verde (limite superior) e amarelo (limite inferior) indicam os intervalos de confiança da previsão, com 95% de confiança. Essa faixa reflete a incerteza associada às estimativas futuras. É possível observar que o modelo mantém a tendência e o padrão sazonal da série nos períodos anteriores, e as previsões se mantêm dentro de um intervalo razoável de variação, indicando que o modelo conseguiu capturar adequadamente a estrutura da série até o momento da previsão.

Figura 18 – Previsão da série temporal para o segundo semestre de 2024



0.4 Conclusão

Este trabalho teve como objetivo analisar a série temporal dos casos notificados de dengue em Fortaleza entre 2014 e 2024. Através do modelo ARIMA(13,0,0), foi possível identificar tendências de longo prazo e padrões sazonais recorrentes. O modelo selecionado apresentou bom desempenho na previsão dos casos para o segundo semestre de 2024, mantendo consistência com o comportamento histórico da série. Esses resultados reforçam a importância da análise temporal como ferramenta de apoio ao planejamento e à prevenção em saúde pública.

A versão completa do trabalho, incluindo gráficos e dados em Excel, está disponível no repositório do [GitHub](#).

Bibliografia

- [1] Alexandre Cunha Costa. *Análise de séries temporais*. Acessado em: 29 de jun. 2025. 2020. URL: <https://www.analisedeseriestemporais.com>.
- [2] Alexandre Cunha Costa. *Aula 18: Automatização do Método Plug-in para a Autocorrelação Usando Google Planilha*. Acessado em: 29 jun. 2025. 2020. URL: <https://www.analisedeseriestemporais.com/post/aula-18-automatiza%C3%A7%C3%A3o-do-m%C3%A9todo-plug-in-para-a-autocorrela%C3%A7%C3%A3o-usando-google-planilha>.
- [3] Alexandre Cunha Costa. *Aula 24: Saiba o que é o Modelo Autoregressivo, o Modelo mais Popular de Séries Temporais*. Acessado em: 29 de jun. 2025. 2020. URL: <https://www.youtube.com/watch?v=uDYrYyxrNhg&list=PLSDVadsS1XTCVcg95hQsEOVRnVwgaPTRC&index=24>.
- [4] Data Science Academy. *Análise e Modelagem de Séries Temporais: Técnicas Estatísticas, IA e Suas Aplicações*. Acessado em: 08 jun. 2025. Mar. de 2024. URL: <https://blog.dsacademy.com.br/analise-e-modelagem-de-series-temporais-tecnicas-estatisticas-ia-e-suas-aplicacoes/>.
- [5] DataCamp. *Decomposição de séries temporais: Entenda as tendências, a sazonalidade e o ruído*. <https://www.datacamp.com/pt/tutorial/time-series-decomposition>. Acesso em: 25 jun. 2025. n.d.
- [6] B. S. Everitt. *The Cambridge Dictionary of Statistics in the Medical Sciences*. Cambridge: Cambridge University Press, 1995.
- [7] R. J. Hyndman e G. Athanasopoulos. *Forecasting: Principles and Practice*. 3ª ed. Acesso em: 01 jul. 2025. Melbourne, Austrália: OTexts, 2021. URL: <https://otexts.com/fpp3/>.
- [8] IBM. *Função LAG — IBM Informix 12.10*. <https://www.ibm.com/docs/pt-br/informix-servers/12.10.0?topic=routines-lag-function>. Acesso em: 25 jun. 2025. n.d.
- [9] IBM. *O que é autocorrelação e como ela funciona?* Português. <https://www.ibm.com/br-pt/think/topics/autocorrelation>. Acesso em: 10 out. 2023. IBM, 2023. URL: <https://www.ibm.com/br-pt/think/topics/autocorrelation>.
- [10] I.R. de Nardi et al. “Análise de séries temporais na operação de sistema de tratamento de águas residuárias de abatedouro de frango”. Em: *Engenharia Sanitária e Ambiental* 10.4 (out. de 2005). Available from: <https://doi.org/10.1590/S1413-41522005000400010>, pp. 339–346. DOI: 10.1590/S1413-41522005000400010. URL: <https://doi.org/10.1590/S1413-41522005000400010>.

- [11] Maria do Rosário Dias de Oliveira Latorre e Maria Regina Alves Cardoso. *Análise de séries temporais em epidemiologia: uma introdução sobre os aspectos metodológicos / Time series analysis in epidemiology: an introduction to methodological aspects*. Portuguese. <mailto:mdrddola@usp.br>. Apoio: CNPq (bolsa pesquisador 300328/97-9). Av. Dr. Arnaldo, 715, 01246-904 São Paulo - SP, n.d.