

Relatório de Análise de Série Temporal: Nascimentos Femininos

Ubiratan da Silva Tavares

December 12, 2025

1 Introdução

Este relatório apresenta a análise da série temporal de nascimentos femininos diários. O objetivo é compreender a dinâmica da série, identificar padrões como sazonalidade e tendência, diagnosticar a estacionariedade, estimar um modelo de previsão (Suavização Exponencial Simples - SES) e avaliar a presença de *outliers*.

Para garantir a transparência e a replicabilidade dos resultados apresentados neste relatório, todo o código fonte encontra-se disponível publicamente no repositório do GitHub: <https://github.com/ubiratanavares/time-series-forecasting-SES-analysis>.

2 Metodologia

A análise foi conduzida utilizando a linguagem Python e um conjunto de bibliotecas especializadas para ciência de dados e estatística. A seguir, detalha-se a utilização de cada pacote no projeto:

- **pandas**: Utilizado para a manipulação e estruturação dos dados em formato tabular (DataFrames). Fundamental para o tratamento da série temporal, permitindo indexação temporal, tratamento de dados faltantes e operações de fatiamento para divisão entre treino e teste.
- **numpy**: Empregado para operações numéricas de alto desempenho. Serve como base para cálculos matemáticos e vetoriais necessários durante a manipulação dos dados e cálculo de métricas.
- **statsmodels**: Biblioteca central para a modelagem econométrica e estatística. Foi utilizada para:
 - Calcular as funções de autocorrelação (ACF) e autocorrelação parcial (PACF) na Questão 1.
 - Realizar os testes de raiz unitária e estacionariedade (ADF e KPSS) na Questão 2.
 - Estimar e ajustar o modelo de Suavização Exponencial Simples (SES) na Questão 3.

- **matplotlib**: Utilizada para a geração de todas as visualizações gráficas do relatório, incluindo gráficos de linha da série temporal, correlogramas e gráficos de resíduos. Essencial para a inspeção visual dos resultados.
- **seaborn**: Utilizada para aprimorar a estética e o estilo das visualizações gráficas, garantindo gráficos mais informativos e visualmente agradáveis.
- **scikit-learn**: Empregada para o cálculo rigoroso das métricas de avaliação de acurácia do modelo. Forneceu as funções para cálculo do Erro Quadrático Médio (MSE), Erro Absoluto Médio (MAE) e Erro Percentual Absoluto Médio (MAPE).
- **jinja2**: Motor de templates utilizado para a geração automatizada deste relatório. Permite a inserção dinâmica dos resultados estatísticos, tabelas e textos interpretativos diretamente na estrutura do documento LaTeX.

A metodologia analítica seguiu as seguintes etapas:

1. Análise de Autocorrelação (ACF/PACF) para identificação de sazonalidade e dependência temporal.
2. Testes de Estacionariedade: *Augmented Dickey-Fuller* (ADF) e KPSS.
3. Ajuste de Modelo de Suavização Exponencial Simples (SES) e previsão fora da amostra.
4. Diagnóstico de *Outliers* utilizando o método de 3 Desvios Padrão (3-Sigma) nos resíduos.

3 Resultados e Discussões

3.1 Questão 1: Análise de Autocorrelação

A Figura 1 apresenta os correlogramas da série.

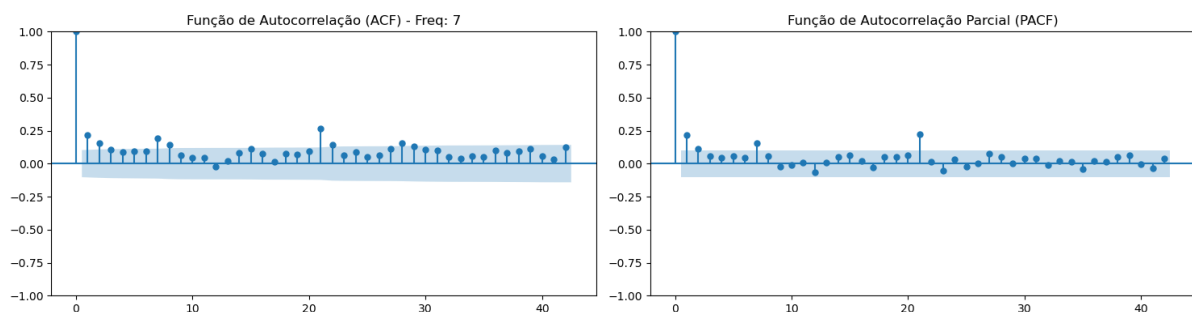


Figure 1: Função de Autocorrelação (ACF) e Autocorrelação Parcial (PACF)

A análise dos correlogramas revela informações importantes sobre a estrutura da série temporal.

Observam-se picos significativos nas defasagens sazonais ([7, 21, 28]), o que sugere fortemente a presença de um componente sazonal na série. Este padrão repetitivo indica que a série segue um ciclo regular (Frequência: 7).

Além disso, a análise da dependência temporal mostra que a autocorrelação permanece significativa para as primeiras 2 defasagens.

O decaimento rápido da função de autocorrelação sugere que a série possui uma dependência de curto prazo e tende à estacionariedade.

3.2 Questão 2: Testes de Estacionariedade

Para confirmar as impressões visuais, foram realizados os testes formais ADF e KPSS.

Table 1: Resultados dos Testes de Estacionariedade

Teste	p-valor	Conclusão (a 5%)
ADF	0.0001	Estacionária
KPSS	0.0100	Não Estacionária

Os resultados indicam que:

Há um conflito entre os testes. O teste ADF indica estacionariedade, enquanto o teste KPSS sugere não estacionariedade. Isso pode indicar processos como estacionariedade por diferença ou tendência determinística.

3.3 Questão 3: Previsão com Suavização Exponencial Simples (SES)

O modelo SES foi ajustado aos dados. A Figura 2 ilustra o ajuste e a previsão.

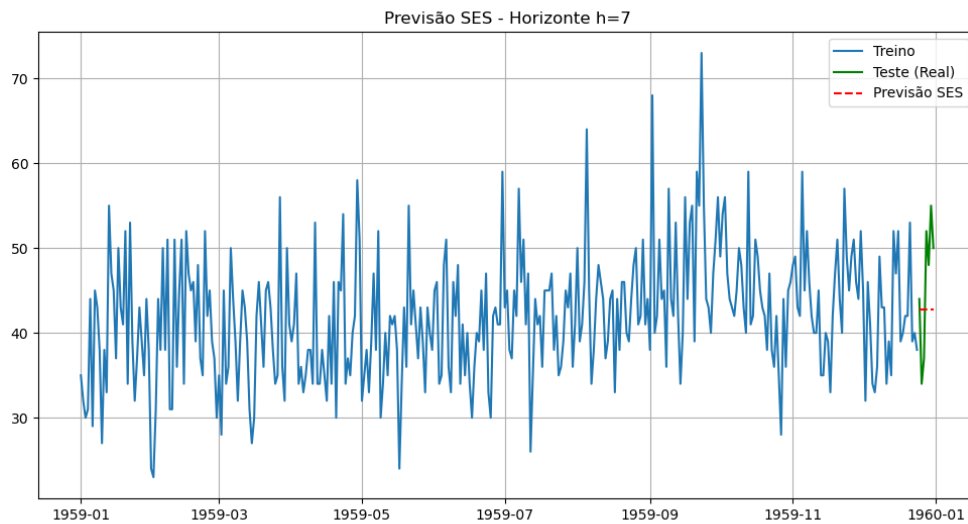


Figure 2: Previsão SES vs Dados Reais

O parâmetro de suavização (α) estimado foi de 0.0495.

Este valor baixo indica que o modelo considera um longo histórico passado, resultando em uma previsão suave e pouco reativa a flutuações recentes.

Quanto à acurácia, o modelo obteve um MAPE de 15.65% e um RMSE de 7.8038.

O MAPE abaixo de 20% sugere que o modelo possui uma boa capacidade preditiva para o horizonte testado.

O método SES, por projetar uma previsão constante, é teoricamente limitado para séries com tendência ou sazonalidade marcantes.

3.4 Questão 4: Diagnóstico de Outliers

A análise de resíduos (Figura 3) permitiu identificar pontos atípicos.

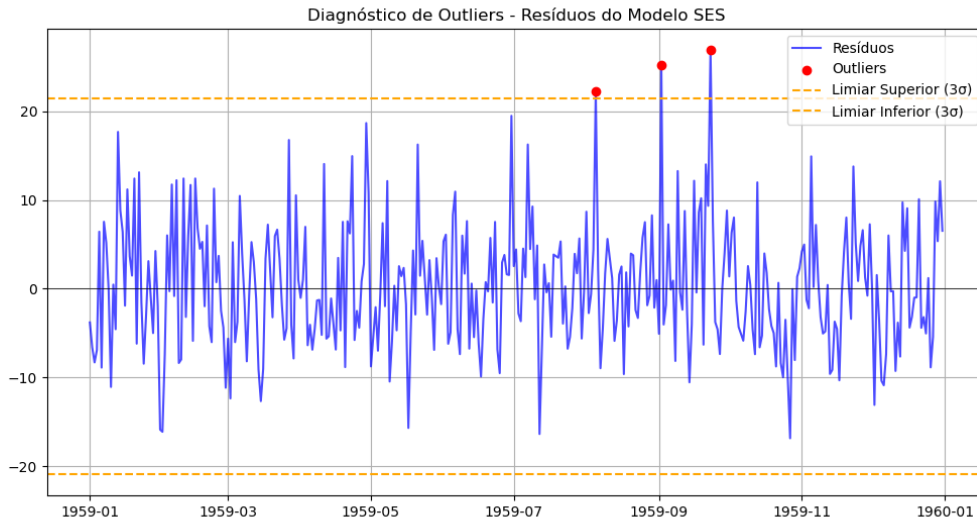


Figure 3: Resíduos do Modelo SES e Outliers Detectados

Utilizando o critério de 3 desvios padrão, foram identificados 3 *outliers*. A presença destes pontos, com um desvio padrão residual de 7.0548, pode impactar a precisão das estimativas de erro (RMSE) e aumentar a incerteza das previsões.

A natureza destes pontos deve ser investigada para determinar se são erros de coleta ou eventos reais atípicos.

4 Conclusões

Com base em todas as análises realizadas, conclui-se que o modelo SES apresenta um desempenho aceitável para previsões de curto prazo. A análise exploratória indicou presença de sazonalidade forte na frequência analisada. Como o SES não modela explicitamente tendência nem sazonalidade, sua aplicação deve ser feita com cautela, especialmente para horizontes de previsão mais longos onde esses componentes estruturais dominariam.