MODELAGEM E PROJEÇÃO DO NÚMERO DE NASCIDOS VIVOS NO BRASIL (2017-2024): UMA ANÁLISE COMPARATIVA ENTRE SARIMA E XGBOOST PARA APOIO AO PLANEJAMENTO EM SAÚDE

Raphael de Santana Borges[1], Marcello Montillo Provenza[1,2], Igor C. A. Lima[1], Jorge Luiz de Jesus Goulart[1]

[1]Universidade do Estado do Rio de Janeiro, Rio de Janeiro, RJ, Brasil (raphaelborges.uerj@gmail.com)
[2] Faculdade de Formação de Professores - UERJ, São Gonçalo, RJ, Brasil

Resumo: O estudo busca entender qual modelo (SARIMA ou XGBoost) se ajusta melhor a evolução do número de nascidos vivos no Brasil, com dados do SINASC (2017 a 2024). O SARIMA apresentou baixo erro percentual na previsão. Esse declínio pode ser influenciado por fatores como urbanização, escolaridade feminina e acesso a métodos contraceptivos. Os resultados indicam a importância de seu uso para o planejamento de políticas públicas de saúde materno-infantil.

Palavras-chave: Nascidos vivos; Séries temporais; SARIMA; XGBoost; Demografia.

INTRODUÇÃO

O número de nascidos vivos é um dos indicadores demográficos e de saúde mais relevantes para um país, pois reflete tendências sociais, econômicas e sanitárias e orienta o planejamento de políticas públicas, a organização da infraestrutura hospitalar, a alocação de profissionais de saúde e ações voltadas à atenção materno-infantil (Brasil, 2022). No Brasil, essas informações são registradas pelo Sistema de Informações sobre Nascidos Vivos (SINASC), fonte essencial para a vigilância epidemiológica e para a produção de estudos populacionais de longo prazo.

Nas últimas décadas, o país tem passado por uma intensa transição demográfica, caracterizada pela queda consistente da fecundidade, redução dos nascimentos e aceleração do envelhecimento populacional. Esse processo é influenciado por fatores como urbanização, maior escolaridade feminina, inserção das mulheres no mercado de trabalho e ampla disponibilidade de métodos contraceptivos (IBGE, 2021).

Embora diversos estudos já tenham documentado o declínio sustentado dos nascimentos no Brasil, ainda são limitadas as análises que buscam prever essas tendências utilizando metodologias estatísticas consolidadas e técnicas modernas de modelagem (Reis et al., 2016). Nesse contexto, modelos de séries temporais, como o SARIMA e métodos de aprendizado de máquina, como o XGBoost, apresentam-se como abordagens promissoras para analisar a dinâmica dos nascimentos e projetar seu comportamento futuro.

A análise de fenômenos em saúde ao longo do tempo requer abordagens que considerem a dependência temporal dos dados, tornando as séries temporais uma ferramenta essencial para a identificação de padrões, sazonalidades e mudanças estruturais. Essas técnicas têm sido amplamente utilizadas no monitoramento de eventos epidemiológicos e na avaliação de intervenções em saúde pública (Oliveira, 2024; Reis et al., 2016).

Durante a pandemia de COVID-19, por exemplo, modelos de séries temporais desempenharam papel central na análise da evolução dos casos, na identificação de picos de transmissão e no suporte à tomada de decisão por gestores públicos (World Health Organization, 2020). A aplicação dessas metodologias ao estudo dos nascidos vivos permite compreender melhor sua dinâmica temporal e contribuir para o planejamento de políticas de saúde (IBGE, 2021).

Diante desse cenário, o presente estudo tem como objetivo analisar a série histórica do número de nascidos vivos no Brasil, empregando modelos de previsão para projetar tendências futuras. A pesquisa busca fornecer subsídios quantitativos que permitam compreender a dinâmica demográfica recente e antecipar demandas para o planejamento de políticas públicas.

MATERIAL E MÉTODOS

As bases de dados utilizadas neste estudo foram disponibilizadas pelo Sistema Único de Saúde (SUS), por meio do Sistema de Informação sobre Nascidos

Vivos (Sinasc), disponível no OPENDATASUS. O Sinasc consolida, anualmente, informações sobre os nascidos vivos no Brasil, abrangendo um período significativo, estendendo-se de 1996 ao resultado parcial de 2025. Trata-se, portanto, de uma fonte abrangente, capaz de oferecer uma visão de longo prazo sobre a dinâmica dos nascimentos no país.

No âmbito deste projeto, foi definido um recorte temporal de janeiro de 2017 a dezembro de 2024, permitindo contemplar três períodos distintos: pré-pandemia, pandemia e pós-pandemia de COVID-19. Essa abordagem permite analisar como eventos associados ao colapso sanitário, econômico e social podem influenciar os padrões de nascimentos no Brasil, ao mesmo tempo em que mantém a atualidade necessária para uma interpretação consistente e contextualizada dos dados.

Todo o processamento dos dados, desde a integração das bases anuais até as etapas de modelagem, foi realizado na linguagem de programação R, por meio do ambiente RStudio. Essa escolha permitiu a implementação de algoritmos computacionais voltados à análise de séries temporais, à modelagem preditiva e à organização das etapas analíticas propostas (R Core Team, 2024).

Para assegurar a validação dos modelos, as bases de dados foram separadas em duas partes: treino e teste. A base de treinamento abrangeu o período de janeiro de 2017 a abril de 2023, sendo utilizada para o ajuste e a calibração dos modelos estatísticos e computacionais. Já a base de teste reuniu os dados de maio de 2023 a dezembro de 2024, funcionando como conjunto independente destinado a avaliar o desempenho preditivo dos modelos por meio da comparação entre os valores estimados e os valores observados.

A análise de tendência foi empregada para identificar padrões persistentes de aumento ou diminuição dos valores ao longo do tempo, permitindo compreender a evolução gradual da série em horizontes temporais mais longos (Box et al., 2015).

Para investigar a existência de tendência na série, utilizou-se o teste estatístico não paramétrico de Mann-Kendall, cuja hipótese nula considera a inexistência de tendência na série de nascidos vivos (Duarte, 2021). Complementarmente, realizou-se a decomposição da série temporal, com o objetivo de isolar os componentes de tendência, sazonalidade e resíduo, permitindo avaliar como o número de nascidos vivos evoluiu ao longo dos anos (Box et al., 2015).

Antes da análise de tendência e sazonalidade, verificou-se a aderência dos dados à distribuição normal. Para isso, aplicou-se o teste de Shapiro-Wilk, utilizado para avaliar a hipótese nula de que a amostra provém de uma população com distribuição normal. De forma complementar, empregou-se o teste de Kolmogorov-Smirnov, que compara a distribuição empírica dos dados com uma distribuição teórica esperada, neste caso, a distribuição normal (Leotti et al., 2012).

Para investigar diferenças estatisticamente significativas entre os números de nascidos vivos ao longo dos diferentes anos ou estações, empregou-se o teste de Kruskal-Wallis. Trata-se de uma alternativa não paramétrica à ANOVA, adequada para comparar três ou mais grupos independentes quando a premissa de normalidade não é atendida. A hipótese nula do teste considera que as medianas dos grupos são iguais (Conover, 1999).

A sazonalidade foi analisada considerando a presença de padrões ou flutuações que se repetem em intervalos específicos de tempo, como ciclos mensais ou anuais. No presente estudo, utilizou-se a biblioteca stats, da linguagem R, para realizar a decomposição da série temporal, permitindo identificar e isolar componentes sazonais associados a flutuações regulares nos dados (Box et al., 2015; R Core Team, 2024).

Também foram analisadas as Funções de Autocorrelação e Autocorrelação Parcial, respectivamente FAC e FACP, com o objetivo de identificar a estrutura de dependência temporal da série. Essas ferramentas foram utilizadas por serem importantes para a identificação da dinâmica temporal e para auxiliar na escolha de parâmetros de modelos de séries temporais, como os modelos ARIMA e SARIMA (Box et al., 2015).

Foram comparados dois modelos de previsão: o modelo Autorregressivo Integrado de Médias Móveis Sazonal, SARIMA, e o modelo Extreme Gradient Boosting, XGBoost. O modelo SARIMA foi ajustado com auxílio da função auto.arima, proveniente do pacote forecast, a fim de selecionar uma especificação adequada para a série analisada (Hyndman e Khandakar, 2008).

O modelo SARIMA (eq. 1) é uma extensão do modelo ARIMA que incorpora componentes sazonais, sendo apropriado para séries temporais que apresentam padrões sazonais definidos:

$$\phi_p(B)\Phi_P\left(B^S\right)(1-B)^d\left(1-B^S\right)^D y_t = \theta_q(B)\Theta_Q\left(B^S\right)\varepsilon_t \quad (1)$$

Onde s é o período de sazonalidade; $\Phi_P(B^S)$ e $\Theta_Q(B^S)$representam, respectivamente, os componentes autorregressivos e de médias móveis sazonais.

O modelo SARIMA (p,d,q)(P,D,Q)s considera componentes autorregressivos, de diferenciação e de médias móveis, tanto na parte não sazonal quanto na parte sazonal da série. Por se tratar de um modelo

univariado, utiliza apenas a própria dinâmica histórica da variável analisada, sem incorporar variáveis explicativas externas (Rio Grande do Sul, 2022).

O modelo XGBoost foi empregado como abordagem de aprendizado de máquina baseada em gradient boosting, combinando múltiplas árvores de decisão para melhorar a capacidade preditiva. Esse algoritmo constrói modelos de forma sequencial, de modo que cada nova árvore busca reduzir os erros cometidos pelas árvores anteriores. O modelo final é obtido pela soma das contribuições das árvores ajustadas, permitindo capturar relações complexas entre as variáveis (Chen e Guestrin, 2016). O valor previsto pelo meodelo pode ser representado por:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(\mathbf{x_i}) \qquad (2)$$

Onde f_k representa a k-ésima árvore de decisão; K corresponde ao número total de árvores do modelo; e $\mathbf{x_i}$ é o vetor de variáveis explicativas.

O algoritmo busca minimizar uma função objetivo (eq. 3) composta por dois termos: a função de perda associada ao erro de previsão e um termo de regularização que controla a complexidade do modelo (Chen e Guestrin, 2016):

$$L(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \qquad (3)$$

em que $l(y_i, \hat{y}_i)$ representa a função de perda e $\Omega(f_k)$ corresponde ao termo de regularização aplicado às árvores do modelo.

Como o XGBoost não modela diretamente a dependência temporal da série, foi necessário transformar o problema de previsão em uma tarefa de regressão supervisionada. Para isso, foram construídas variáveis defasadas da própria série de nascimentos, representando observações em períodos anteriores. Especificamente, foram utilizadas defasagens de um, dois, três e doze meses, permitindo ao modelo explorar dependências de curto prazo e padrões sazonais anuais (Chen e Guestrin, 2016; Chen et al., 2024). Dessa forma, o modelo pode ser representado por (eq. 4):

$$N_t = f(N_{t-1}, N_{t-2}, N_{t-3}, N_{t-12}) \qquad (4)$$

em que N_t representa o número de nascimentos no período t.

Após a construção das variáveis defasadas, os primeiros registros da base foram removidos, pois não apresentavam valores anteriores suficientes para a composição dos preditores. Em seguida, o modelo XGBoost foi ajustado com a base de treinamento, correspondente ao período de janeiro de 2017 a abril de 2023, e avaliado com a base de teste, correspondente ao período de maio de 2023 a dezembro de 2024.

Para avaliar o desempenho dos modelos, foram utilizadas duas métricas: o Erro Percentual Absoluto Médio, MAPE, e o Erro Absoluto Médio, MAE. O MAPE (eq. 5) expressa o erro médio em termos percentuais, possibilitando a comparação da acurácia do modelo em séries com diferentes escalas.

$$MAPE = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \qquad (5)$$

O MAE (eq. 6) representa a média dos valores absolutos das diferenças entre previsões e valores observados, fornecendo uma medida direta do desvio médio das estimativas em relação aos dados reais (Filho, 2018).

$$MAE = \frac{1}{n} \sum_{t=1}^{n} |y_t - \hat{y}_t| \qquad (6)$$

A análise conjunta do MAPE e do MAE permitiu avaliar o grau de precisão e confiabilidade dos modelos ajustados, contribuindo para a comparação do desempenho preditivo das abordagens estatística clássica e de aprendizado de máquina na previsão da série temporal de nascidos vivos (Filho, 2018).

RESULTADOS E DISCUSSÃO

A série temporal mensal de nascidos vivos no Brasil, correspondente ao período de janeiro de 2017 a dezembro de 2024, indicou tendência de queda ao longo do período analisado (Figura 1). Possíveis explicações para esse comportamento incluem o saldo migratório internacional e transformações sociais associadas à urbanização, à escolaridade, à inserção das mulheres no mercado de trabalho e ao acesso a métodos contraceptivos (IBGE, 2022; Oliveira, 2024).

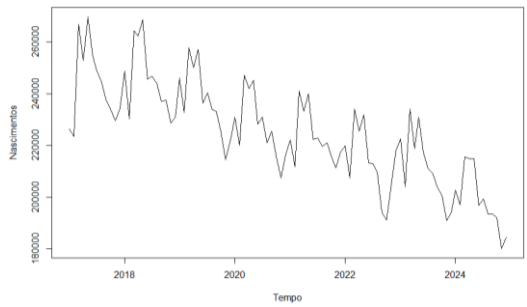


Figura 1. Visualização da série de nascidos vivos (2017-2024).

As estatísticas descritivas da série indicaram média mensal de 225.091,6 nascimentos, mediana de 224.343,5 nascimentos, primeiro quartil de 212.577,5,

terceiro quartil de 237.528,5, valor mínimo de 180.124, valor máximo de 269.777 e desvio padrão de 20.055,6. A proximidade entre média e mediana sugere uma distribuição relativamente simétrica, sem evidências de assimetria acentuada.

A distribuição da série histórica de nascidos vivos apresentou comportamento visual próximo ao de uma distribuição normal (Figura 2). Ao longo do período estudado, a série atingiu valor máximo em maio de 2017, com cerca de 270 mil nascimentos, e valor mínimo no final de 2024. A redução observada foi interpretada no texto original como possível reflexo de transformações demográficas, incluindo queda histórica da fecundidade, aumento da escolaridade feminina, reconfiguração das estruturas familiares e envelhecimento populacional (BRASIL..., 2022; IBGE, 2022; UNFPA, 2023).

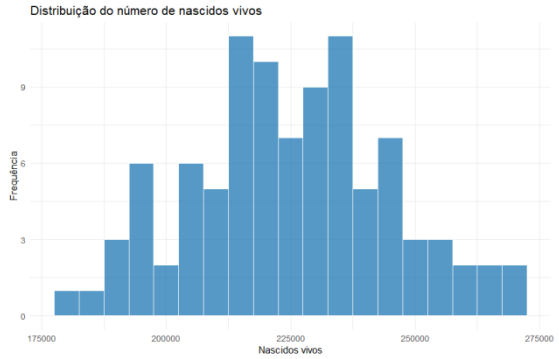


Figura 2. Distribuição da série de nascidos vivos (2017-2024).

Apesar de a série incluir o período da pandemia de COVID-19, visualmente não foi observada uma variação expressiva atribuível exclusivamente a esse evento. Com base nas características observadas na série, foram ajustados modelos de previsão fundamentados em abordagens estatísticas clássicas e em aprendizado de máquina, com o objetivo de comparar seus desempenhos preditivos.

A normalidade dos dados foi avaliada por meio dos testes de Shapiro-Wilk e Kolmogorov-Smirnov. O teste de Shapiro-Wilk apresentou p-valor igual a 0,738, enquanto o teste de Kolmogorov-Smirnov apresentou p-valor igual a 0,995. Considerando nível de significância de 1%, não houve evidências estatísticas suficientes para rejeitar a hipótese nula de normalidade, sugerindo que os dados podem ser tratados como aproximadamente normais no contexto da análise realizada (Leotti et al., 2012).

A presença de tendência foi avaliada inicialmente por inspeção visual e, posteriormente, pelo teste de Mann-Kendall. O teste apresentou p-valor inferior a 0,001, levando à rejeição da hipótese nula de ausência de tendência ao nível de significância de 1%. Esse resultado confirmou estatisticamente a tendência decrescente observada graficamente na série temporal (Duarte, 2021).

O teste de Kruskal-Wallis apresentou p-valor igual a 0,002, levando à rejeição da hipótese nula ao nível de significância de 1%. No texto original, esse resultado foi interpretado como evidência de sazonalidade, em concordância com a inspeção visual da série. Do ponto de vista metodológico, recomenda-se explicitar quais grupos foram comparados — por exemplo, meses, anos ou estações — para que a interpretação do teste fique mais clara e diretamente vinculada ao desenho analítico adotado (Conover, 1999).

A decomposição da série temporal evidenciou a presença de componentes de tendência, sazonalidade e resíduo (Figura 3). O componente de tendência apresentou comportamento decrescente ao longo do período, sugerindo redução gradual no número mensal de nascimentos. O componente sazonal indicou a presença de flutuações recorrentes, evidenciando padrões temporais regulares na série (Box et al., 2015; R Core Team, 2024).

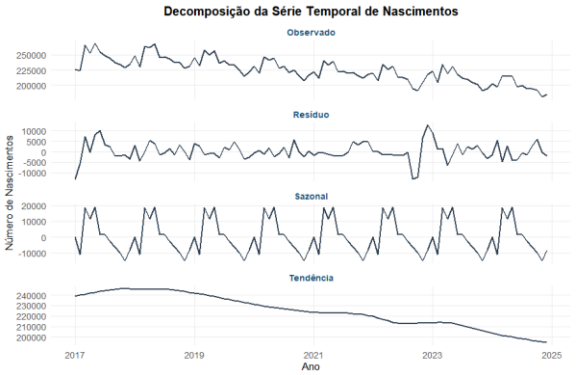


Figura 3. Decomposição da série em resíduo, sazonalidade e tendência.

A presença de sazonalidade na série de nascidos vivos é relevante para a compreensão da dinâmica populacional e para o planejamento de políticas públicas. A identificação desses ciclos pode auxiliar na organização da capacidade hospitalar, no dimensionamento da atenção primária, no planejamento de programas sociais e educacionais e na adequação da gestão de recursos ao longo do tempo.

A análise dos correlogramas foi utilizada como etapa auxiliar na modelagem ARIMA. A FAC apresentou autocorrelações positivas e significativas nos primeiros lags, com decaimento gradual até aproximadamente o lag 12, indicando presença de componente sazonal (Figura 4).

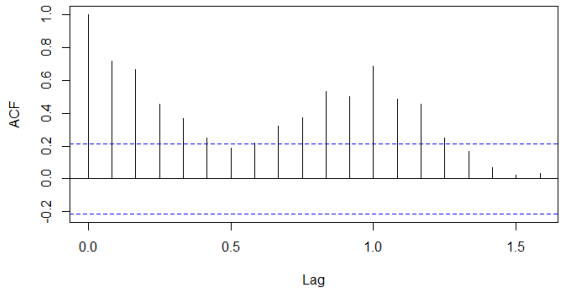


Figura 4. Gráfico de autocorrelação (FAC).

A FACP apresentou pico elevado no primeiro lag, seguido por valores menores e alternância de sinais, padrão compatível com processos que envolvem componentes de médias móveis (Figura 5) (Box et al., 2015).

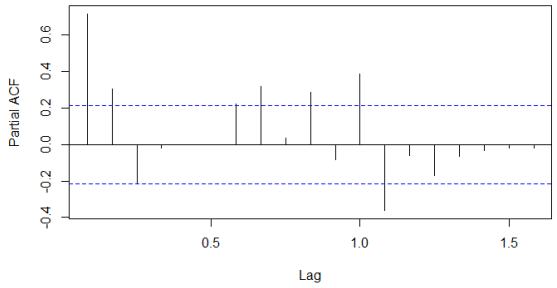


Figura 5. Gráfico de autocorrelação parcial (FACP).

A partir da análise dos gráficos da FAC e da FACP e do ajuste automático realizado pela função auto.arima, foi selecionado o modelo ARIMA 0,0,30,1,112. Esse modelo incorporou uma componente de média móvel não sazonal de ordem 3 e uma componente sazonal de ordem 1, com sazonalidade anual. O ajuste foi realizado por meio do pacote forecast, conforme a abordagem de seleção automática implementada em auto.arima() (Hyndman e Khandakar, 2008; Hyndman et al., 2025).

Para o modelo ARIMA ajustado, as métricas avaliativas finais foram MAPE = 1,35 e MAE = 3.062. Esses valores indicam alta proximidade entre previsões e valores observados. Considerando o volume médio mensal de aproximadamente 225 mil nascimentos, os erros observados foram interpretados como pequenos em termos relativos e absolutos, sugerindo boa capacidade de ajuste e previsão para a série analisada (Figura 6) (Filho, 2018).

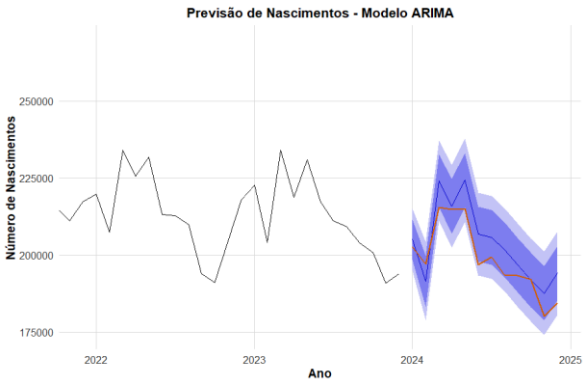


Figura 6. Gráfico de previsão para o ano de 2024.

No caso do XGBoost, a série foi transformada em um problema de regressão supervisionada por meio da criação de variáveis defasadas. Foram utilizadas as defasagens de um, dois, três e doze períodos anteriores, correspondentes a lag1, lag2, lag3 e lag12. Essa estrutura permitiu ao modelo utilizar valores passados da própria série como preditores, buscando capturar dependências de curto prazo e padrões sazonais anuais (Chen e Guestrin, 2016; Chen et al., 2024).

Após a construção das defasagens, os primeiros registros da base foram removidos em razão da ausência de valores anteriores necessários para compor os preditores. Em seguida, os dados foram divididos em treino e teste, utilizando-se os dados de janeiro de 2017 a abril de 2023 para treinamento e os dados de maio de 2023 a dezembro de 2024 para avaliação preditiva.

As métricas avaliativas finais do XGBoost foram MAPE = 9,21 e MAE = 17.862. Esses valores indicam erros médios superiores aos obtidos pelo modelo ARIMA. Em termos absolutos, o erro médio de aproximadamente 18 mil nascimentos representa discrepância considerável em relação à média mensal da série, situada em torno de 225 mil nascimentos (Filho, 2018).

O desempenho inferior do XGBoost pode ser explicado, no contexto do estudo, pela estrutura relativamente estável da série e pela presença de forte componente sazonal, características que tendem a ser bem representadas por modelos estatísticos clássicos de séries temporais. Além disso, o texto original destaca que o modelo XGBoost não foi submetido a um processo extensivo de otimização de hiperparâmetros, o que pode ter limitado seu desempenho preditivo (Chen e Guestrin, 2016; Chen et al., 2024).

A comparação direta entre os modelos ARIMA e XGBoost indicou desempenho preditivo superior do ARIMA para a série de nascidos vivos no Brasil (Figura 7).

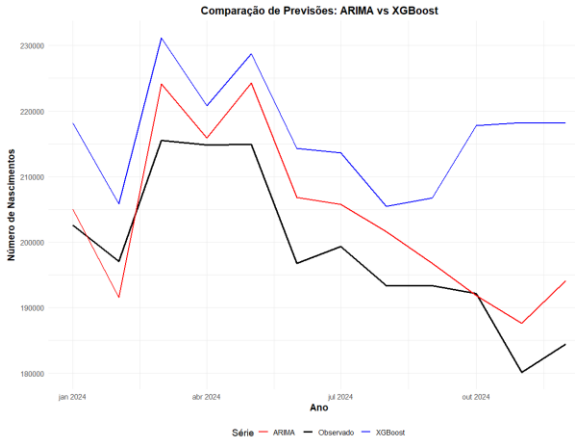


Figura 7. Comparação de previsões: ARIMA vs XGBoost.

De modo geral, os resultados sugerem que, para séries temporais com comportamento relativamente estável e forte componente sazonal, modelos estatísticos clássicos, como o ARIMA, podem apresentar maior precisão preditiva do que abordagens de aprendizado de máquina, especialmente quando estas não são extensivamente ajustadas. Ainda assim, o XGBoost permanece como abordagem complementar, sobretudo em contextos com maior complexidade estrutural, relações não lineares ou disponibilidade de variáveis explicativas adicionais (Chen e Guestrin, 2016; Chen et al., 2024).

CONCLUSÃO

O presente trabalho teve como objetivo o estudo da série temporal de nascidos vivos. No desenvolvimento da análise, o SARIMA foi o modelo que apresentou melhor desempenho no ajuste da série. A análise exploratória da série temporal revelou um comportamento não estacionário, sazonal e com forte tendência de decrescimento confirmado pelos testes de Kruskal-Wallis e Mann-Kendall. Estes achados justificam a escolha da metodologia capaz de capturar simultaneamente tendências, sazonalidade e dependências temporais.

O modelo escolhido apresentou erros baixos e demonstrou boa capacidade de aderência aos dados observados, indicando sua adequação para representar a dinâmica da série. Além disso, as previsões geradas por ele mostraram-se coerentes com as tendências históricas, evidenciando o potencial do modelo para subsidiar análises e planejamento em saúde pública.

A análise descritiva demonstrou possível simetria na distribuição dos nascimentos, com mediana (224.344) e média (225.092) próximas nos dados que posteriormente veio a ser confirmada com os testes Shapiro-Wilk e Kolmogorov-Smirnov. Ademais, essas medidas apontam para uma série relativamente estável, sem flutuações abruptas e com dispersão controlada, o que confere robustez às análises subsequentes realizadas no estudo.

Dessa forma, conclui-se que o uso de modelos SARIMA e XGBoost para séries demográficas são ferramentas eficazes para o monitoramento e para a projeção do número de nascidos vivos. Entretanto, obteve-se métricas melhores no modelo SARIMA. Sendo assim, o SARIMA seria a metodologia que poderia contribuir melhor para o planejamento de políticas públicas voltadas à atenção materno-infantil e à alocação adequada de recursos no sistema de saúde.

Como trabalhos futuros, recomenda-se atualizar o intervalo com dados mais recentes, de modo a aprimorar a capacidade preditiva para os próximos anos. Sugere-se também a exploração de outras modelos de previsão, como redes neurais artificiais, explorar hiper parâmetros do modelo XGBoost e demais técnicas de aprendizado de máquina, ampliando o escopo metodológico e permitindo comparações mais robustas.

REFERÊNCIAS

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. Time series analysis: forecasting and control. 5. ed. Hoboken: Wiley, 2015.

BRASIL. Ministério da Saúde. Sistema de Informações sobre Nascidos Vivos (SINASC): conceitos e aplicações. Brasília: Ministério da Saúde, 2022.

BRASIL presencia queda histórica da fecundidade e reconfiguração dos fluxos migratórios, aponta Censo 2022. Brasília: LAEMI - UnB, 2022. Disponível em: https://laemi.com.br/brasil-presencia-queda-historica-da-fecundidade-e-reconfiguracao-dos-fluxos-migratorios-aponta-censo-2022/. Acesso em: 13 nov. 2025.

CHEN, T.; GUESTRIN, C. XGBoost: a scalable tree boosting system. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 22., 2016, New York. Proceedings [...]. New York: ACM, 2016.

CHEN, T. et al. xgboost: extreme gradient boosting. [S. l.: s. n.], 2024. Disponível em: https://cran.r-project.org/package=xgboost. Acesso em: 13 nov. 2025.

CONOVER, W. J. Practical nonparametric statistics. 3. ed. Nova York: John Wiley & Sons, 1999. 584 p.

DUARTE, A. F. B. Análise de tendências climáticas no Estado do Ceará. 2021. Dissertação (Mestrado em Engenharia Civil - Recursos Hídricos) - Universidade Federal do Ceará, Fortaleza, 2021.

FILHO, M. As métricas mais populares para avaliar modelos de machine learning. [S. l.]: Mario Filho, 2018. Disponível em: https://mariofilho.com/as-metricas-mais-populares-para-avaliar-modelos-de-machine-learning/. Acesso em: 13 nov. 2025.

HYNDMAN, R. J.; KHANDAKAR, Y. Automatic time series forecasting: the forecast package for R. Journal of Statistical Software, [S. l.], v. 27, n. 3, p. 1-22, 2008.

HYNDMAN, R. et al. forecast: forecasting functions for time series and linear models. Versão 8.24.0. [S. l.: s. n.], 2025. Disponível em: https://pkg.robjhyndman.com/forecast/. Acesso em: 13 nov. 2025.

IBGE. Projeção da população do Brasil e das Unidades da Federação. Rio de Janeiro: IBGE, 2021. Disponível em: https://www.ibge.gov.br. Acesso em: 13 nov. 2025.

IBGE. Síntese de indicadores sociais: uma análise das condições de vida da população brasileira. Rio de Janeiro: IBGE, 2022.

LEOTTI, V. B.; COSTER, R.; RIBOLDI, J. Teste de normalidade em estudos de validação de questionários: uma revisão. Jornal Brasileiro de Patologia e Medicina Laboratorial, [S. l.], v. 48, n. 6, p. 405-409, 2012.

OLIVEIRA, L. G. S. de. As surpresas e os alertas que vêm do front demográfico no Brasil. Portal FGV, [S. l.], 27 nov. 2024. Disponível em: https://portal.fgv.br/artigos/surpresas-e-alertas-que-vem-do-front-demografico-brasil. Acesso em: 13 nov. 2025.

R CORE TEAM. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2024. Disponível em: https://www.r-project.org/. Acesso em: 13 nov. 2025.

REIS, C. S. dos; NORONHA, K.; WAJNMAN, S. A transição da fecundidade no Brasil: uma análise por idade e ordem de nascimento das mulheres. Revista Brasileira de Estudos de População, [S. l.], v. 33, n. 3, p. 549-570, 2016.

RIO GRANDE DO SUL. Secretaria da Fazenda. Tesouro do Estado. Modelo SARIMA para projeção do ICMS. Porto Alegre: SEFAZ/RS, 2022. (Texto para Discussão). Disponível em: https://tesouro.fazenda.rs.gov.br/upload/arquivos/202508/29094527-1643376115-artigo-modelo-sarima-texto-de-discussao.pdf. Acesso em: 13 nov. 2025.

UNFPA. State of World Population Report. [S. l.]: UNFPA, 2023.

WORLD HEALTH ORGANIZATION. World health statistics 2020. Geneva: WHO, 2020.