v7.n12.2026

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

MODELAGEM E PROJEÇÃO DE CASOS DE TUBERCULOSE NO BRASIL POR REGRESSÃO LINEAR

MODELING AND PROJECTION OF TUBERCULOSIS CASES IN BRAZIL USING LINEAR REGRESSION

MODELIZACIÓN Y PROYECCIÓN DE CASOS DE TUBERCULOSIS EN BRASIL MEDIANTE REGRESIÓN LINEAL

Francisco José Ribeiro Facchinetti[1], Marcelo Lisboa Rocha[2], Silvanete Maria da Silva[3]

https://doi.org/10.47820/recima21.v7i12.9714

PUBLICADO: 12/2026

RESUMO

Este estudo analisou 25 observações anuais de tuberculose no Brasil (2001–2025) associadas ao SINAN/DATASUS, com as variáveis ano, masculino, feminino e total. Ajustaram-se regressões lineares simples às 22 observações de 2001–2022, reservando-se 2023–2025 à avaliação externa. A série masculina apresentou tendência crescente ($r = 0,6752$; $R^2 = 0,4559$; $p = 0,0006$) e a feminina, decrescente ($r = -0,6846$; $R^2 = 0,4687$; $p = 0,0004$); a do total não foi significativa ($p = 0,2072$). As estimativas subestimaram as contagens de 2023–2025 (MAE nacional = 21.353,67; RMSE = 21.504,72). Uma previsão ingênua baseada em 2022 obteve MAE nacional menor (10.312,33). Embora estatisticamente significativas, as tendências apresentaram poder explicativo moderado e discrepâncias no período pós-ajuste. As extrapolações até 2050 devem ser interpretadas como cenários exploratórios, e não como previsões epidemiológicas validadas.

PALAVRAS-CHAVE: Tuberculose. Epidemiologia. Análise de Regressão. Vigilância em Saúde Pública.

ABSTRACT

This study analyzed 25 annual observations of tuberculosis in Brazil (2001–2025) using data from SINAN/DATASUS, covering the variables year, male, female, and total. Simple linear regressions were fitted to the 22 observations from 2001–2022, with the years 2023–2025 reserved for external evaluation. The male series showed an upward trend ($r = 0.6752$; $R^2 = 0.4559$; $p = 0.0006$), while the female series showed a downward trend ($r = -0.6846$; $R^2 = 0.4687$; $p = 0.0004$); the trend for the total was not statistically significant ($p = 0.2072$). The estimates underestimated the counts for 2023–2025 (national MAE = 21,353.67; RMSE = 21,504.72). A naïve forecast holding 2022 counts constant had a lower MAE (10,312.33). Although statistically significant, the trends exhibited moderate explanatory power and discrepancies during the post-fitting period. Extrapolations up to 2050 should be interpreted as exploratory scenarios rather than validated epidemiological forecasts.

KEYWORDS: *Tuberculosis. Epidemiology. Regression Analysis. Public Health Surveillance.*

RESUMEN

Este estudio analizó 25 observaciones anuales de tuberculosis en Brasil (2001–2025) a partir del SINAN/DATASUS, considerando las variables año, sexo masculino, sexo femenino y total. Se ajustaron regresiones lineales simples a las 22 observaciones correspondientes al periodo 2001–

[1]Doutorando do Programa de Pós-graduação em Governança e Transformação Digital da Universidade Federal do Tocantins (UFT).
[2]Pós-doutor em modelagem computacional pela Universidade do Estado do Rio de Janeiro (UERJ) e Professor titular do Programa de Pós-graduação em Governança e Transformação Digital da Universidade Federal do Tocantins (UFT).
[3]Mestranda em Educação pela Universidade Federal do Tocantins.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

2022, reservando los años 2023–2025 para la evaluación externa. La serie masculina mostró una tendencia creciente (r = 0,6752; R² = 0,4559; p = 0,0006), mientras que la femenina presentó una tendencia decreciente (r = −0,6846; R² = 0,4687; p = 0,0004); la tendencia del total no resultó significativa (p = 0,2072). Las estimaciones subestimaron los recuentos para el periodo 2023–2025 (MAE nacional = 21.353,67; RMSE = 21.504,72). Una previsión ingenua que mantuvo los valores de 2022 obtuvo un MAE menor (10.312,33). Si bien fueron estadísticamente significativas, las tendencias presentaron un poder explicativo moderado y discrepancias en el periodo posterior al ajuste. Las extrapolaciones hasta el año 2050 deben interpretarse como escenarios exploratorios, y no como predicciones epidemiológicas validadas.

PALABRAS CLAVE: Tuberculosis. Epidemiología. Análisis de Regresión. Vigilancia en Salud Pública.

1. INTRODUÇÃO

Este artigo examina a variação das contagens anuais de tuberculose no Brasil por sexo e avalia em que medida regressões lineares ajustadas aos anos de 2001–2022 se aproximam dos registros de 2023–2025.

A literatura recente destaca que a vigilância da tuberculose depende da qualidade e da oportunidade dos registros de notificação, do fortalecimento da atenção primária e de análises que considerem as desigualdades sociais e territoriais (Jesus *et al.*, 2022; Maciel; Sales, 2016; Sales *et al.*, 2018; Silva *et al.*, 2017).

A série do CSV associada ao SINAN/DATASUS reúne 25 observações anuais, de 2001 a 2025, distribuídas por sexo e total nacional. Em 2025, foram contabilizados 116.215 registros (81.633 masculinos e 34.582 femininos), valor tratado como consolidado na versão do CSV utilizada. O estudo combina a descrição das tendências de 2001–2022 com a comparação entre valores estimados e observados em 2023–2025. O horizonte de 2050 foi mantido para ilustrar os efeitos de prolongar as retas históricas, sem atribuir a essas projeções validade preditiva demonstrada.

A questão que orienta o estudo é em que medida a regressão linear simples pode descrever as tendências dos casos de tuberculose no Brasil e estimar cenários até 2050.

A contribuição específica desta investigação consiste em confrontar as tendências lineares ajustadas ao período de 2001 a 2022 com as contagens efetivamente registradas entre 2023 e 2025, que não participaram da estimação dos modelos. Ao combinar análise de tendências, validação temporal retrospectiva, medidas de erro fora da amostra e avaliação dos pressupostos das regressões, o estudo examina não apenas a direção das associações históricas, mas também as limitações de sua utilização para previsões epidemiológicas. Essa

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

abordagem permite distinguir a significância estatística das tendências da precisão das estimativas produzidas para períodos posteriores.

A interpretação de séries nacionais requer considerar as condições econômicas, o acesso aos serviços, a organização do cuidado e as diferenças entre territórios e grupos populacionais, pois esses fatores influenciam as notificações, os desfechos e a transmissão da doença (Costa *et al.*, 2023; Gaspar *et al.*, 2016; Li *et al.*, 2022; Mesquita *et al.*, 2023; Prado *et al.*, 2017; Queiroz *et al.*, 2018).

O objetivo foi descrever as tendências históricas das contagens masculinas, femininas e totais entre 2001 e 2022, comparar suas extrapolações com 2023–2025 e discutir as limitações de cenários lineares estendidos até 2050.

2. REFERENCIAL TEÓRICO

Na análise de séries temporais, a regressão linear simples estima a variação média anual de uma contagem em função do tempo (Paiva Júnior *et al.*, 2021). Os coeficientes descrevem associações observadas no período de ajuste; sua significância estatística, isoladamente, não demonstra precisão para estimar anos posteriores (Gomes; Ferracioli, 2006; Morettin; Bussab, 2017; Santana, 2020).

Estudos de séries temporais e de intervenção realizados no Brasil mostram que os registros do SINAN podem subsidiar a avaliação de tendências, a estimação de cenários e o acompanhamento dos efeitos de mudanças diagnósticas na detecção de casos de tuberculose (Giacomet *et al.*, 2021; Silva; Galvão, 2024; Villalva-Serra *et al.*, 2024).

Em séries anuais, padrões sistemáticos dos resíduos e autocorrelação podem comprometer a interpretação dos erros-padrão e dos testes de significância convencionais (Durbin; Watson, 1950; Morettin; Bussab, 2017). A normalidade residual pode ser examinada pelo teste de Shapiro–Wilk (Shapiro; Wilk, 1965), enquanto a estatística de Durbin–Watson contribui para investigar a dependência serial de primeira ordem.

Os intervalos de confiança (IC) da resposta média expressam a incerteza da média estimada pela regressão; os intervalos de predição também incorporam a variabilidade de uma nova observação. Ao extrapolar uma série, é necessário considerar a continuidade aproximada da relação temporal e a comparabilidade das contagens nos diferentes anos. Essas condições não são asseguradas pela consolidação de um dado observado.

O erro absoluto médio (MAE), a raiz do erro quadrático médio (RMSE) e o erro absoluto médio escalonado (MASE) descrevem a magnitude dos erros das estimativas e permitem

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

comparar alternativas de previsão. Na avaliação de séries temporais, a separação cronológica entre os períodos de ajuste e de avaliação evita o uso de informações futuras no treinamento (Hyndman; Koehler, 2006).

3. METODOLOGIA

O estudo utilizou dados secundários agregados de tuberculose, organizados no arquivo “tuberculose-2001 a 2025.csv”, associado à base “Casos de tuberculose: desde 2001 (SINAN)”, disponibilizada pelo Departamento de Informática do Sistema Único de Saúde (DATASUS), do Ministério da Saúde (Brasil, [s. d.]). O conjunto contém 25 observações anuais, correspondentes ao período de 2001 a 2025, estruturadas nas variáveis Ano, Masculino, Feminino e Total. As 22 observações de 2001 a 2022 foram utilizadas no ajuste das regressões lineares, enquanto os anos de 2023 a 2025 foram reservados à avaliação fora da amostra. A verificação do arquivo confirmou a continuidade da sequência anual, a ausência de valores faltantes nas variáveis analisadas e a correspondência aritmética entre as contagens por sexo e o total informado. Entretanto, por se tratar de uma base agregada, não foi possível recuperar integralmente os parâmetros da consulta original ao sistema, incluindo os critérios de classificação epidemiológica, os procedimentos de identificação de duplicidades e o tratamento dos registros ignorados. Assim, os resultados restringem-se às contagens efetivamente disponibilizadas no arquivo, sem pressupor a equivalência de definições epidemiológicas que não puderam ser verificadas.

O arquivo CSV com as 25 observações anuais de 2001 a 2025 está disponibilizado para conferência e reprodução dos cálculos em: https://drive.google.com/file/d/1LE5Do-PuhJbJKx-S5XVb_tTLpq8q3-Wg/view?usp=drive_link

3.1. Aplicação do método

O CSV foi lido com separador ponto e vírgula e codificação UTF-8. Os cálculos foram executados em Python 3.13, utilizando pandas 2.2.3, NumPy 2.3.5, SciPy 1.17.0 e statsmodels 0.14.6; os gráficos foram produzidos com Matplotlib 3.10.8. Ajustaram-se modelos por mínimos quadrados ordinários com intercepto (OLS), adotando x = ano − 2001 (Lima, 2009; Scavetta, 2022; Silva, 2020).

No período de ajuste, 2001–2022, o CSV apresenta 1.324.466 contagens masculinas e 635.785 femininas, somando 1.960.251. Para a série completa, 2001–2025, o acumulado é de 2.299.312 contagens. As séries por sexo apresentaram comportamentos temporais distintos, motivo pelo qual foram ajustados modelos separados, além de uma regressão para o total

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

nacional (Schmuller, 2019). Foi organizado o conjunto dados_sinan_ano_sexo_total, com as 25 linhas de 2001 a 2025. Para a modelagem foram filtrados os anos até 2022. A Tabela 1 apresenta as correlações de Pearson entre ano, masculino, feminino e total exclusivamente no período de ajuste; o Apêndice A transcreve integralmente as 25 observações do CSV para permitir a conferência dos resultados.

Tabela 1. Matriz de correlação de Pearson entre ano, casos por sexo e total anual de tuberculose no Brasil, 2001–2022

Variável	Ano	Masculino	Feminino	Total
Ano	1,0000	0,6752	−0,6846	0,2798
Masculino	0,6752	1,0000	0,0543	0,8939
Feminino	−0,6846	0,0543	1,0000	0,4961
Total	0,2798	0,8939	0,4961	1,0000

Fonte: cálculos realizados pelos autores com o CSV de 2001–2025, derivado da base SINAN/DATASUS (Brasil, [s. d.]); correlações calculadas em 2001–2022.

As associações envolvendo o total foram descritas com cautela, pois Total = Masculino + Feminino em todos os anos do CSV. As tendências temporais foram avaliadas pelas correlações entre ano e contagem de cada série, e não por correlações aritmeticamente induzidas entre total e componentes.

Para a avaliação da associação linear, calculou-se a correlação de Pearson entre ano e cada contagem anual de 2001 a 2022 (n = 22), com teste t bilateral para H0: r = 0 e nível de 5%. Foram estimadas três regressões OLS com intercepto: masculino, feminino e total como variáveis-resposta, e ano − 2001 como preditor. Relataram-se coeficientes, erros-padrão, estatísticas t e F, valores de p, R^2 e desvio-padrão residual; as correlações entre total e componentes foram tratadas apenas como relações descritivas (Schober *et al.*, 2018).

Os modelos temporais foram ajustados separadamente às três séries anuais (masculino, feminino e total), utilizando o período 2001–2022. Para s pertencente a {masculino, feminino, total}, adotou-se a especificação linear:

$$Y_{s,t} = \beta_{0,s} + \beta_{1,s}\,(t - 2001) + \varepsilon_{s,t}$$

Nessa equação, $Y_{s,t}$ é a contagem anual da série s no ano t; $\beta_{0,s}$ é o intercepto, correspondente à média estimada para 2001; $\beta_{1,s}$ representa a variação média anual, e $\varepsilon_{s,t}$ é o resíduo. A parametrização x = t − 2001 define o ano inicial como x = 0, sem alterar a reta ajustada.

As equações estimadas diretamente a partir das 22 linhas históricas do CSV foram:

$$\hat{Y}_{m,t} = 55.683{,}23 + 430{,}45(t - 2001)$$

$$\hat{Y}_{f,t} = 31.265{,}03 - 225{,}31(t - 2001);\ \hat{Y}_{to,t} = 86.948{,}26 + 205{,}15(t - 2001)$$

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Os coeficientes foram estimados diretamente a partir das 22 contagens anuais de 2001–2022, por OLS com intercepto. Os erros-padrão derivaram da matriz de covariância convencional dos parâmetros; os testes t possuem 20 graus de liberdade residuais e os testes globais correspondem a F(1,20). Para x = ano − 2001, obtiveram-se média temporal de 10,5 e soma de quadrados Sxx = 885,5. Os erros-padrão do intercepto e da inclinação foram, respectivamente, 1.289,90 e 105,14 na série masculina, 658,06 e 53,64 na feminina e 1.930,76 e 157,38 no total nacional. Os desvios-padrão residuais foram 3.128,82, 1.596,22 e 4.683,32 casos, nessa ordem.

Para um ano t, a média extrapolada foi obtida pela substituição de x = t − 2001 nas equações. O IC de 95% da média estimada foi calculado por $\hat{Y}(t) \pm t(0{,}975;20) \times s \times \sqrt{[1/22 + (x - 10{,}5)^2/885{,}5]}$, com $t(0{,}975;20) \approx 2{,}08596$ e $s = \sqrt{[\Sigma(\text{resíduo}^2)/20]}$. Os intervalos convencionais pressupõem linearidade da média, homocedasticidade e independência dos erros; os testes t e F usuais também dependem de condições inferenciais sobre os resíduos. Esses IC não são intervalos de predição de futuras contagens anuais e suas coberturas nominais não ficam asseguradas quando os pressupostos são violados.

Os indicadores de ajuste histórico foram r de Pearson, R^2, F(1,20), estatísticas t, valores de p e erros-padrão dos coeficientes. Para verificar os pressupostos dos três ajustes OLS, os resíduos dos 22 anos de treinamento (2001–2022), ordenados cronologicamente, foram submetidos ao teste de Shapiro–Wilk, cuja hipótese nula corresponde à normalidade dos resíduos (Shapiro; Wilk, 1965). Adotou-se nível de significância de 5% e registraram-se a estatística W e seu valor de p. Para investigar dependência temporal, calculou-se a estatística de Durbin–Watson (DW), definida pela razão entre a soma dos quadrados das diferenças de resíduos consecutivos e a soma dos quadrados dos resíduos (Durbin; Watson, 1950). Valores próximos de 2 são compatíveis com ausência de autocorrelação de primeira ordem, enquanto valores abaixo de 2 sugerem autocorrelação positiva. Complementarmente, calcularam-se as correlações residuais de primeira defasagem e inspecionaram-se gráficos de resíduos em função do ano e dos valores ajustados. Não foi calculado valor de p para DW; a estatística foi interpretada de modo descritivo, sem atribuir a ela uma conclusão inferencial isolada.

Considerando os indícios de autocorrelação temporal identificados nos resíduos, realizou-se uma análise de sensibilidade da inferência estatística mediante o estimador de matriz de covariância robusta à heterocedasticidade e à autocorrelação (HAC), proposto por Newey e West (1987). O procedimento foi aplicado às três regressões lineares ajustadas com as 22 observações de 2001 a 2022, utilizando duas defasagens temporais, ponderação pelo núcleo de Bartlett e correção para amostra pequena. Os testes bilaterais dos coeficientes angulares consideraram a distribuição t de Student com 20 graus de liberdade e nível de significância de 5%. Foram comparados os erros-padrão e os valores de p convencionais dos mínimos

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

quadrados ordinários (OLS) com os respectivos resultados robustos HAC. A análise teve como finalidade verificar a sensibilidade das conclusões sobre as tendências temporais à presença de dependência serial dos resíduos, sem modificar os coeficientes estimados ou as extrapolações dos modelos.

A avaliação fora da amostra utilizou as observações de 2023, 2024 e 2025, excluídas do treinamento. Para cada série, definiu-se erro = estimado − observado e calcularam-se MAE, RMSE e MASE; o denominador do MASE foi a média das diferenças absolutas consecutivas de 2001–2022. Essas métricas descrevem o desempenho nos três anos disponíveis, sem demonstrar capacidade de extrapolação até 2050.

Como referência de desempenho para a avaliação fora da amostra, utilizou-se também uma previsão ingênua de persistência: em cada uma das três séries, a contagem registrada em 2022 foi repetida como estimativa fixa para 2023, 2024 e 2025. Desse modo, o comparador foi definido exclusivamente com dados disponíveis até o fim do treinamento, sem atualização com os observados de 2023 ou 2024. Os erros das duas abordagens foram calculados nos mesmos três anos e resumidos por MAE, RMSE e MASE; para este último, adotou-se em ambos os métodos o mesmo denominador, correspondente à média das diferenças absolutas consecutivas de 2001–2022 (Hyndman; Koehler, 2006). Como avaliação temporal complementar, aplicou-se validação retrospectiva por origem móvel. Para cada ano de 2018 a 2022, o modelo foi ajustado apenas com os anos anteriores e produziu uma previsão de um passo à frente. Os erros dessas cinco previsões foram resumidos por MAE e RMSE.

Os intervalos das Tabelas 2 e 3 foram calculados a partir da matriz de covariância convencional dos modelos OLS. Como os diagnósticos apontaram limitações dos pressupostos residuais, seus níveis nominais de confiança exigem cautela. Além disso, a posição de uma contagem observada em relação ao IC da média não avalia a cobertura de um intervalo de predição. Como o CSV satisfaz Total = Masculino + Feminino em todos os anos, a regressão OLS do total apresenta coeficientes pontuais iguais à soma dos coeficientes das regressões por sexo. Já sua variância residual envolve as duas variâncias por sexo e duas vezes a covariância entre os resíduos; por isso, o IC do total foi calculado diretamente por regressão da coluna Total, e não pela soma dos limites individuais.

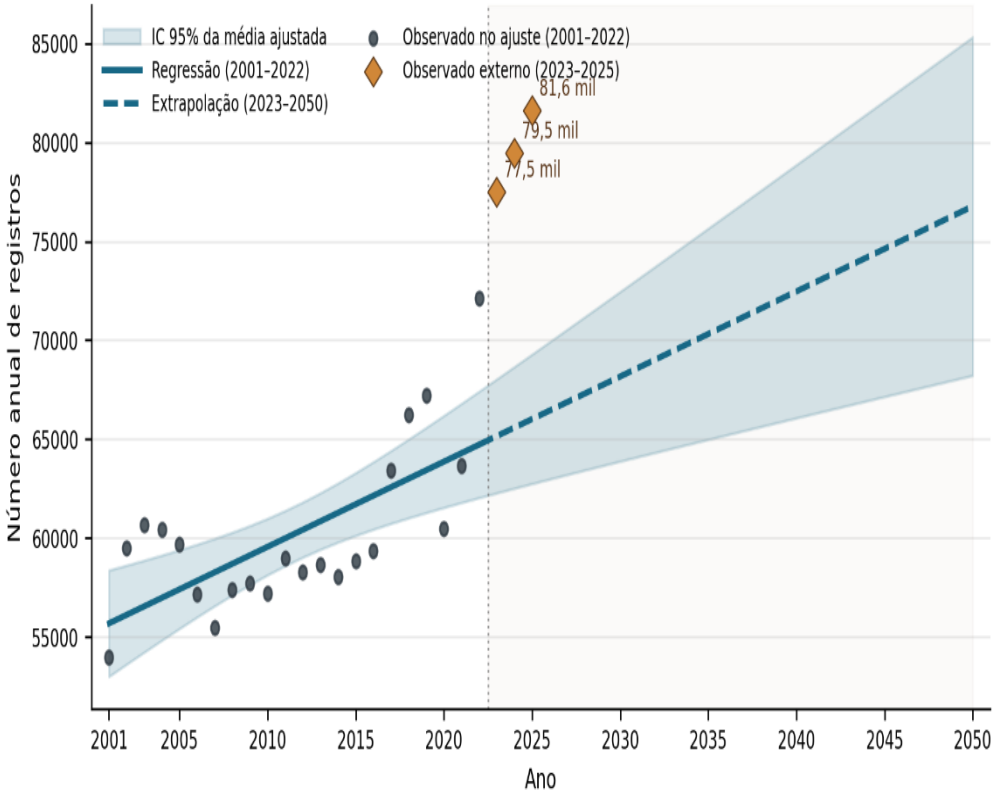
O modelo OLS dos 22 totais anuais apresentou desvio-padrão residual de 4.683,32 casos. As semiamplitudes dos IC de 95% da média para 2023, 2024 e 2025 foram de 4.311,82, 4.602,01 e 4.897,01 casos, respectivamente; os limites calculados sem arredondamento intermediário constam da Tabela 2. Para 2023, a estimativa masculina foi de 65.153,22 casos (IC 95% da média: 62.272,59–68.033,86) e a feminina, de 26.308,30 (24.838,70–27.777,90). Esses resultados utilizaram exclusivamente os coeficientes estimados em 2001–2022.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

A interpretação das estimativas para 2023–2025 considerou a adequação aproximada da relação linear, a estabilidade dos parâmetros do período de ajuste, a comparabilidade dos registros entre os anos e a independência e variância residual exigidas pela inferência OLS convencional. Os testes de Shapiro–Wilk e Durbin–Watson se referem aos resíduos de 2001–2022 e não verificam diretamente a estabilidade da relação temporal após esse período.

Nas Figuras 1 e 2, os pontos escuros representam as contagens efetivamente observadas em 2001–2022; as linhas contínuas indicam as regressões OLS ajustadas e as tracejadas, sua extrapolação até 2050. As bandas correspondem aos IC de 95% da média, calculados diretamente do ajuste, e os losangos destacam as observações externas de 2023–2025. A Figura 3 foi construída com todas as 22 contagens individuais de treinamento, agrupadas em classes; não inclui 2023–2025.

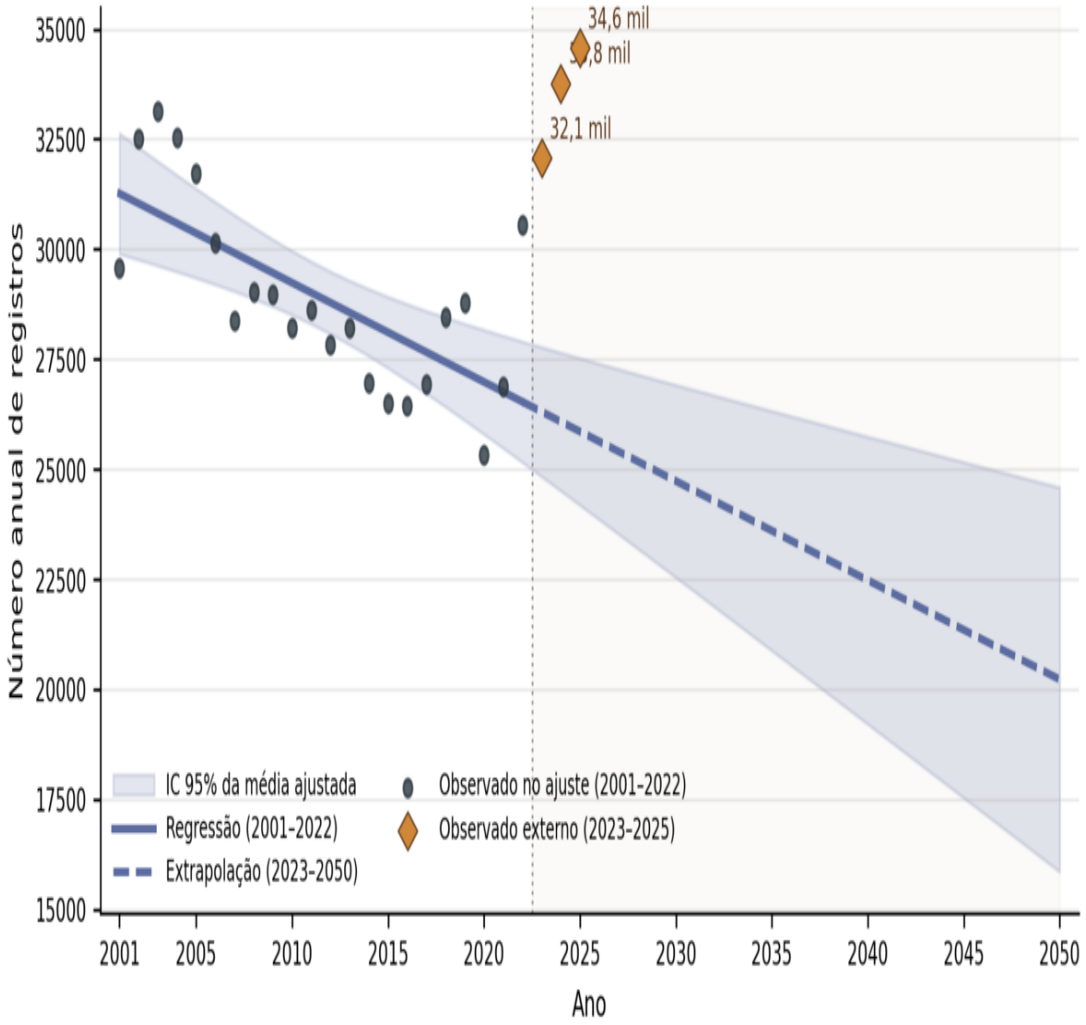
Figura 1. Série masculina: contagens anuais observadas, regressão linear de 2001–2022, IC 95% da média e extrapolação até 2050



Fonte: elaboração dos autores, com base no CSV de 2001–2025 (SINAN/DATASUS).

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Figura 2. Série feminina: contagens anuais observadas, regressão linear de 2001–2022, IC 95% da média e extrapolação até 2050

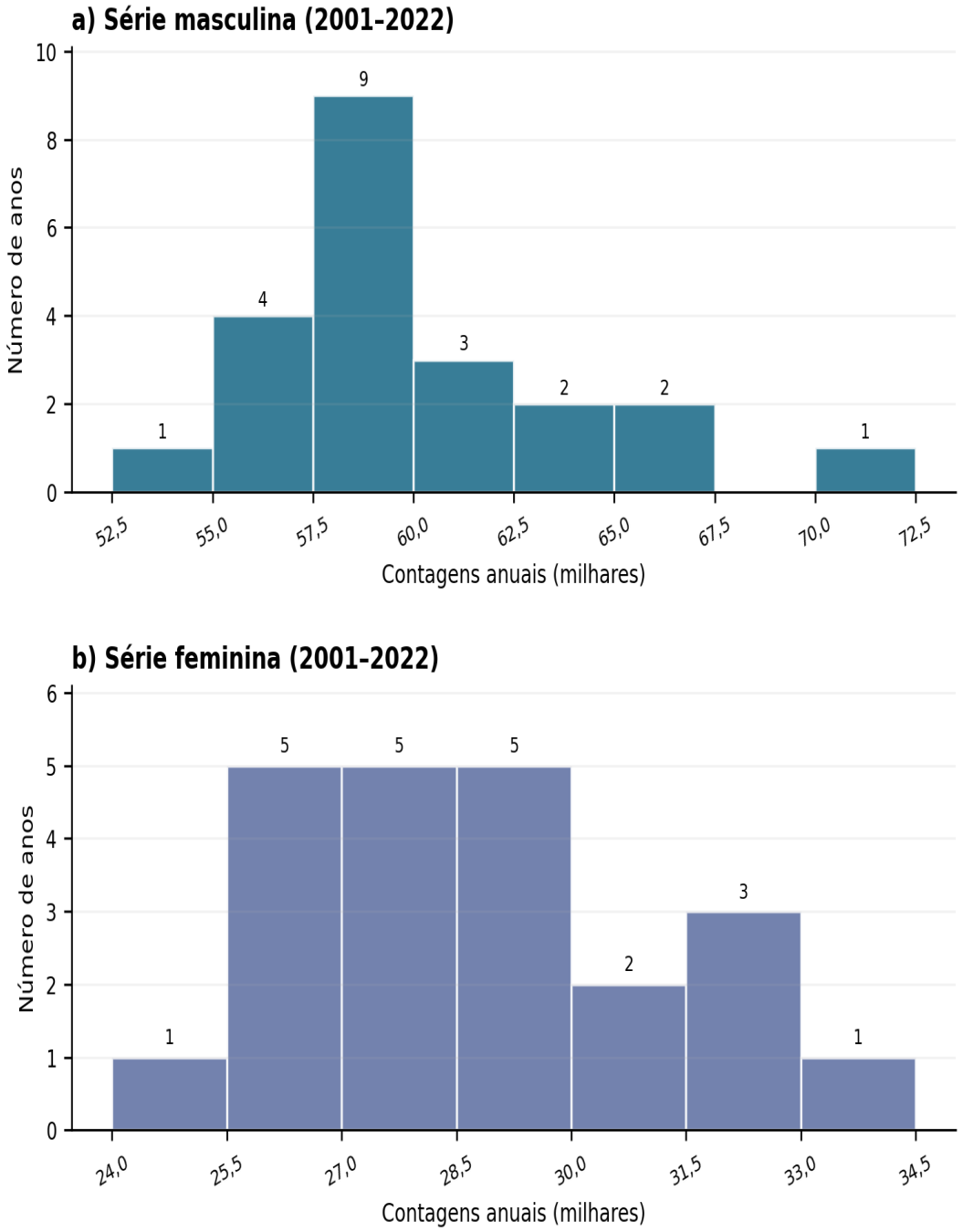


Fonte: elaboração dos autores, com base no CSV de 2001–2025 (SINAN/DATASUS).

A Figura 3 apresenta histogramas construídos a partir das 22 contagens anuais individuais de 2001–2022, com classes regulares por sexo. O eixo horizontal mostra contagens em milhares; o vertical, o número de anos em cada classe. Esses gráficos descrevem a distribuição das contagens durante o treinamento, sem preservar a ordem cronológica; não fornecem teste de normalidade dos resíduos nem demonstram tendência temporal por si sós.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Figura 3. Histogramas das contagens anuais de tuberculose por sexo no Brasil, 2001–2022: (a) masculino; (b) feminino



Fonte: elaboração dos autores a partir dos 22 registros anuais de 2001–2022 presentes no CSV do SINAN/DATASUS. Nota: frequências calculadas diretamente, sem utilização das observações externas de 2023–2025.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

4. RESULTADOS E DISCUSSÃO

Em 2001–2022, as contagens totais anuais variaram de 83.560 a 102.708, com média de 89.102,32 registros. Para homens, a amplitude foi de 53.978 a 72.141 e, para mulheres, de 25.339 a 33.141. As correlações temporais da Tabela 1 foram 0,6752 (masculino), −0,6846 (feminino) e 0,2798 (total).

A Tabela 2 reúne as estimativas pontuais do total para 2023–2025, os respectivos IC de 95% da média e as contagens registradas no CSV. Nenhum desses três anos foi utilizado para estimar os coeficientes.

Tabela 2. Estimativas do total nacional, IC 95% da média e contagens observadas do CSV, Brasil, 2023–2025

Ano	Total estimado	IC 95% da média estimada	Total observado
2023	91.461,52	87.149,70–95.773,34	109.586
2024	91.666,67	87.064,66–96.268,68	113.260
2025	91.871,82	86.974,80–96.768,83	116.215

Fonte: cálculos OLS dos autores a partir das contagens de 2001–2022; observados de 2023–2025 extraídos do CSV associado ao SINAN/DATASUS (Brasil, [s. d.]).

Na Tabela 2, as estimativas para 2023, 2024 e 2025 foram, respectivamente, 91.461,52, 91.666,67 e 91.871,82 registros, diante de 109.586, 113.260 e 116.215 observados no CSV. Os erros (estimado menos observado) foram −18.124,48, −21.593,33 e −24.343,18. O MAE nacional foi de 21.353,67 registros e o RMSE, de 21.504,72. O crescimento dos desvios indica baixa precisão das extrapolações nesses três anos, sem identificar, isoladamente, o motivo do desajuste.

A Tabela 3 apresenta os valores estimados por sexo, com seus IC de 95% da média, e as contagens observadas de 2023–2025. Cada regressão foi ajustada à respectiva série de 2001–2022.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Tabela 3. Estimativas por sexo, IC 95% da média e contagens observadas do CSV, Brasil, 2023–2025

Ano	Masculino estimado	IC 95% masculino	Feminino estimado	IC 95% feminino	Masculino observado	Feminino observado
2023	65.153,22	62.272,59 – 68.033,86	26.308,30	24.838,70 – 27.777,90	77.515	32.071
2024	65.583,67	62.509,17 – 68.658,18	26.082,99	24.514,49 – 27.651,50	79.485	33.775
2025	66.014,13	62.742,54 – 69.285,72	25.857,69	24.188,64 – 27.526,74	81.633	34.582

Fonte: regressões OLS dos autores com as contagens históricas do CSV SINAN/DATASUS (Brasil, [s. d.]). Nota: os valores observados de 2023–2025 não integraram o treinamento; IC 95% da média condicional, não de predição.

Na Tabela 3, as contagens masculinas de 2023–2025 (77.515, 79.485 e 81.633) superaram as estimativas (65.153,22, 65.583,67 e 66.014,13), resultando em erros de −12.361,78, −13.901,33 e −15.618,87. Na série feminina, os observados (32.071, 33.775 e 34.582) também superaram as estimativas (26.308,30, 26.082,99 e 25.857,69), com erros de −5.762,70, −7.692,01 e −8.724,31. A contagem de 2025 é considerada definitiva no CSV disponibilizado. Em ambas as séries, os erros de 2023–2025 mostram que o prolongamento mecânico das inclinações de 2001–2022 não reproduziu adequadamente os registros posteriores, independentemente da consolidação da observação final.

A Figura 1 apresenta os 22 valores masculinos históricos, a regressão crescente ajustada e sua extrapolação até 2050. Os três pontos posteriores ao ajuste situam-se acima das respectivas estimativas e reforçam a diferença entre tendência estatisticamente significativa no período de treinamento e desempenho extrapolativo satisfatório.

A Figura 2 mostra as 22 observações femininas históricas e a reta ajustada com inclinação negativa. Os registros de 2023–2025 apresentaram valores superiores às extrapolações do modelo, contrariando o prolongamento simples da tendência decrescente. A curva não capta necessariamente alterações nos critérios de registro, na detecção ou na dinâmica epidemiológica.

Na série masculina, a correlação entre ano e contagens foi positiva ($r = 0,6752$; $t(20) = 4,09$; $p = 0,0006$) e, na feminina, negativa ($r = −0,6846$; $t(20) = −4,20$; $p = 0,0004$). Os testes globais foram $F(1,20) = 16,76$ ($R^2 = 0,4559$) para a série masculina e $F(1,20) = 17,64$ ($R^2 =$

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

0,4687) para a feminina. Para o total nacional, r = 0,2798, R^2 = 0,0783 e p = 0,2072, sem evidência de tendência linear significativa ao nível de 5%.

As inclinações OLS de 2001–2022 foram +430,45 casos/ano (EP = 105,14) para a série masculina, −225,31 casos/ano (EP = 53,64) para a feminina e +205,15 casos/ano (EP = 157,38) para o total. Os interceptos estimados para 2001 foram 55.683,23, 31.265,03 e 86.948,26, respectivamente. Esses parâmetros foram estimados diretamente do CSV, sem reconstituição algébrica a partir de resultados arredondados.

A Tabela 4 reúne os coeficientes, erros-padrão e indicadores de ajuste das três regressões. Os parâmetros do total foram estimados diretamente com essa coluna do CSV; sua incerteza utiliza os resíduos do modelo do total, e não a soma dos limites dos intervalos individuais por sexo.

Tabela 4. Parâmetros OLS, erros-padrão e indicadores de ajuste das regressões de tuberculose, Brasil, 2001–2022

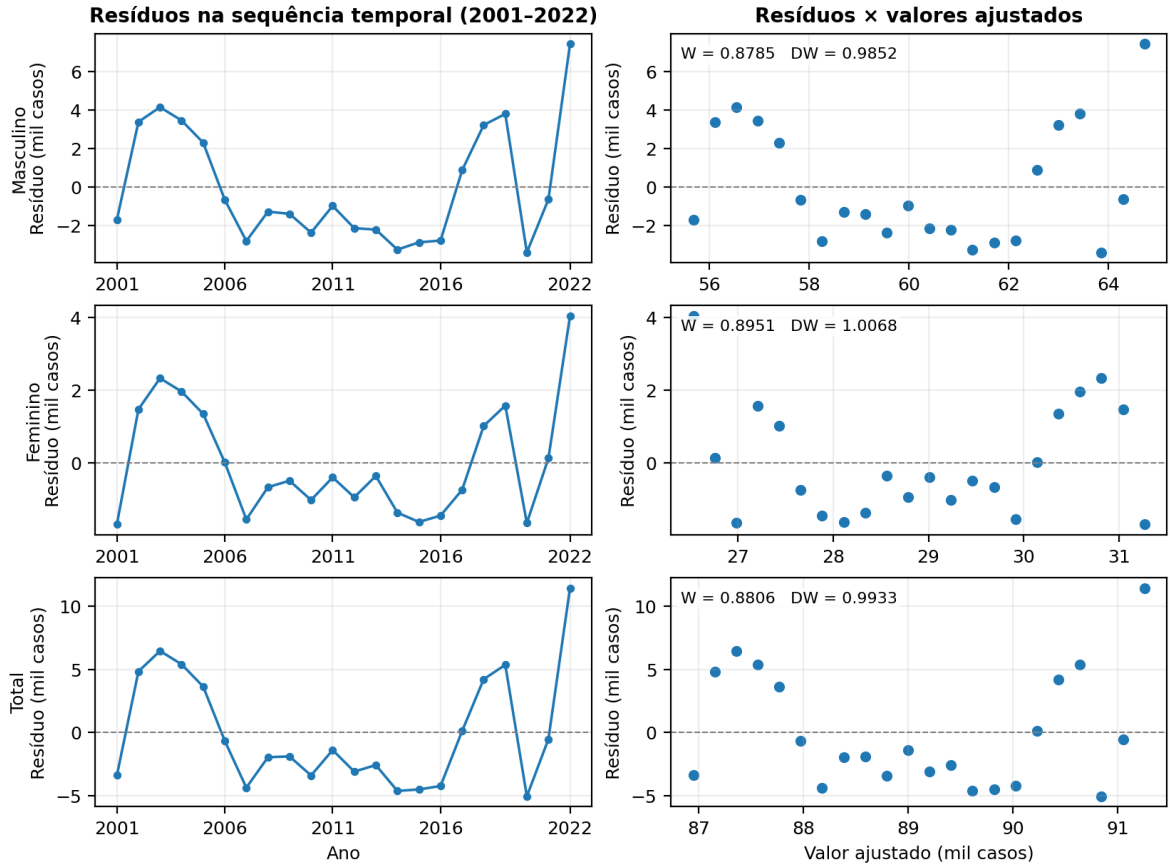
Série	Intercepto	EP intercepto	Inclinação anual	EP inclinação	R^2	F(1,20)
Masculino	55.683,23	1.289,90	430,45	105,14	0,456	16,76
Feminino	31.265,03	658,06	−225,31	53,64	0,469	17,64
Total	86.948,26	1.930,76	205,15	157,38	0,078	1,70

Fonte: cálculos dos autores a partir das 22 linhas de treinamento do CSV associado ao SINAN/DATASUS.

A Tabela 4 confirma tendências de sentidos opostos por sexo e uma associação linear fraca no total nacional. Esses resultados descrevem o comportamento das contagens no período de ajuste.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Figura 4. Diagnóstico gráfico dos resíduos das regressões lineares: sequência temporal e resíduos versus valores ajustados, Brasil, 2001–2022



Fonte: elaboração dos autores a partir dos resíduos das regressões OLS ajustadas com as 22 observações anuais de 2001–2022 do CSV. Linhas horizontais tracejadas representam resíduo igual a zero; gráficos descritivos, não substitutos de testes de hipóteses.

Para complementar a avaliação dos pressupostos das regressões, a Tabela 5 apresenta os resultados dos testes de Shapiro–Wilk e das estatísticas de Durbin–Watson, calculados com os resíduos das três séries no período de ajuste (2001–2022). A Figura 4 apresenta os mesmos resíduos em ordem cronológica e em relação aos valores ajustados, auxiliando a identificação visual de padrões e possível dependência temporal.

Nos três modelos, o teste de Shapiro–Wilk rejeitou a normalidade residual ao nível de 5%: masculino (W = 0,8785; p = 0,0113), feminino (W = 0,8951; p = 0,0237) e total (W = 0,8806; p = 0,0124). As estatísticas Durbin–Watson foram 0,9852, 1,0068 e 0,9933, respectivamente; as correlações residuais de primeira defasagem foram 0,4350, 0,3985 e 0,4221. Esses valores sugerem dependência serial positiva, sem conclusão inferencial específica para DW, cujo valor de p não foi calculado. Os achados recomendam cautela na utilização dos erros-padrão e dos

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

intervalos convencionais, mas não testam diretamente a estabilidade das tendências em 2023–2025.

Tabela 5. Diagnósticos dos resíduos das regressões lineares de tuberculose, Brasil, 2001–2022

Série	Shapiro–Wilk (W)	p (Shapiro–Wilk)	Durbin–Watson (DW)	r resíduos (lag 1)
Masculino	0,8785	0,0113	0,9852	0,4350
Feminino	0,8951	0,0237	1,0068	0,3985
Total	0,8806	0,0124	0,9933	0,4221

Fonte: cálculos dos autores a partir das 22 observações anuais de 2001–2022 do CSV associado ao SINAN/DATASUS. Nota: n = 22 em cada série; W = estatística de Shapiro–Wilk; DW = estatística de Durbin–Watson; lag 1 = correlação entre resíduos de anos consecutivos. O valor de p aplica-se somente ao teste de Shapiro–Wilk.

A análise de sensibilidade realizada com o estimador HAC de Newey-West, considerando duas defasagens temporais e correção para amostra pequena, apresentou erros-padrão dos coeficientes angulares de 133,93 para a série masculina, 71,82 para a feminina e 203,95 para o total nacional. Os respectivos valores de p foram de aproximadamente 0,0044, 0,0052 e 0,3265. Em comparação com a inferência convencional por mínimos quadrados ordinários, os erros-padrão robustos foram maiores nas três séries, indicando redução da precisão inferencial quando considerada a possível dependência temporal dos resíduos. Apesar desse ajuste, as tendências temporais masculina e feminina permaneceram estatisticamente significativas ao nível de 5%, mantendo, respectivamente, os sentidos crescente e decrescente. Para o total nacional, a tendência linear continuou sem significância estatística. Esses resultados indicam que as conclusões sobre a direção das tendências por sexo foram preservadas nessa análise de sensibilidade. Entretanto, a correção HAC não elimina possíveis inadequações da especificação linear, não resolve a ausência de normalidade residual e não demonstra capacidade preditiva para os anos de 2023 a 2025 ou para as extrapolações até 2050.

Para 2023–2025, os MAE/RMSE foram de 21.353,67/21.504,72 registros no total nacional, 13.960,66/14.023,90 na série masculina e 7.393,01/7.494,21 na feminina. Os valores de MASE foram 6,51, 6,39 e 6,44, respectivamente, com escala calculada sobre as diferenças absolutas entre anos consecutivos do período de ajuste. As medidas descrevem os erros dos três anos não utilizados no treinamento.

Na validação retrospectiva por origem móvel (2018–2022), os valores de MAE/RMSE foram 7.374,17/8.518,14 para o total, 4.910,32/5.695,64 para masculino e 2.463,85/2.844,75 para feminino. Esses resultados e a subestimação observada em 2023–2025 mostram limitações do ajuste linear para previsões fora da amostra.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

A fim de verificar se a extrapolação linear acrescentou precisão em relação a uma referência simples, confrontaram-se as previsões OLS com as estimativas ingênuas baseadas nos valores de 2022, iguais a 102.708 registros no total, 72.141 na série masculina e 30.567 na feminina. A Tabela 6 apresenta MAE, RMSE e MASE de ambos os métodos para 2023–2025. Os valores observados desses três anos foram utilizados somente para calcular os erros de avaliação.

Tabela 6. Comparação do desempenho preditivo da regressão linear e da previsão ingênua, Brasil, 2023–2025

Série	Método	MAE (casos)	RMSE (casos)	MASE
Total nacional	OLS	21.353,67	21.504,72	6,51
Total nacional	Ingênuo	10.312,33	10.662,87	3,14
Masculina	OLS	13.960,66	14.023,90	6,39
Masculina	Ingênuo	7.403,33	7.591,93	3,39
Feminina	OLS	7.393,01	7.494,21	6,44
Feminina	Ingênuo	2.909,00	3.091,57	2,53

Fonte: elaboração dos autores a partir do CSV associado ao SINAN/DATASUS e das regressões ajustadas a 2001–2022. Nota: OLS = mínimos quadrados ordinários; previsão ingênua = repetição da contagem de 2022 nos três anos avaliados; MASE calculado com denominador comum de 2001–2022. Menores valores indicam menor erro.

Nos três grupos, a previsão ingênua apresentou erros menores que a regressão linear. No total nacional, o MAE caiu de 21.353,67 para 10.312,33 registros e o RMSE de 21.504,72 para 10.662,87; nas séries masculina e feminina, os MAE da referência foram 7.403,33 e 2.909,00, ante 13.960,66 e 7.393,01 nas regressões. Os resultados mostram que a extrapolação linear não superou a persistência do nível observado em 2022 no período avaliado. Essa comparação utiliza apenas três anos e não permite concluir que o modelo ingênuo seja superior em outros horizontes ou contextos epidemiológicos.

As flutuações de 2020 e 2021 devem ser analisadas à luz da pandemia de COVID-19, período no qual alterações na oferta de serviços, na busca por diagnóstico e na notificação de tuberculose podem ter perturbado as séries. Chitwood *et al.* (2025) discutiram interrupções na detecção no Brasil. No presente estudo, a pandemia não foi incluída como variável explicativa e, portanto, não é possível atribuir causalmente a ela os erros de extrapolação. As quedas e retomadas identificadas nos anos finais do treinamento limitam a adequação de uma única reta ao comportamento epidemiológico.

As desigualdades socioeconômicas e territoriais influenciam a ocorrência, o diagnóstico e o registro de tuberculose (Jesus *et al.*, 2022; Li *et al.*, 2022; Mesquita *et al.*, 2023; Sales *et al.*, 2018). A análise agregada nacional, sem taxas padronizadas pela população, não permite quantificar essas influências nem distinguir mudanças de incidência de alterações no volume

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

absoluto de registros. A relação de soma entre as séries por sexo deve ser interpretada de modo descritivo, sem equivaler à comparação de taxas específicas por sexo.

O CSV não contém informações sobre acesso aos serviços, métodos diagnósticos ou volume de exames. Assim, embora mudanças nesses fatores possam alterar as notificações, este estudo não permite atribuir a eles as diferenças entre as estimativas e as contagens de 2023–2025.

Os histogramas da Figura 3 sintetizam a distribuição das contagens de 2001–2022 por sexo, sem preservar a sequência temporal. Por isso, complementam os gráficos de evolução anual, mas não substituem a análise dos resíduos.

Para 2023, as estimativas por sexo foram 65.153,22 registros masculinos (IC 95% da média: 62.272,59–68.033,86) e 26.308,30 femininos (24.838,70–27.777,90). Para 2050, a extrapolação dos modelos históricos apresentou 76.775,48 (IC 95%: 68.217,48–85.333,48) registros masculinos, 20.225,04 (IC 95%: 15.859,04–24.591,03) femininos e 97.000,51 (IC 95%: 84.190,64–109.810,39) no total. Esses limites refletem apenas a precisão da média estimada sob a hipótese de tendência linear constante, não representando intervalos de predição nem mudanças epidemiológicas futuras previstas.

5. CONSIDERAÇÕES FINAIS

Foram ajustadas três regressões lineares simples às contagens anuais de 2001–2022, reservando-se 2023–2025 para avaliar as estimativas. No período de ajuste, a tendência masculina foi crescente e a feminina decrescente, enquanto a tendência do total não foi estatisticamente significativa. Essas relações não pressupõem, por si só, continuidade nos anos seguintes.

A avaliação externa evidenciou subestimação do total nacional pelas regressões em 2023, 2024 e 2025, com diferenças absolutas de 18.124,48, 21.593,33 e 24.343,18 registros. O MAE da regressão foi de 21.353,67, enquanto a referência ingênua baseada na contagem de 2022 apresentou MAE de 10.312,33. O mesmo padrão, favorável à referência simples, ocorreu nas séries masculina e feminina (Tabela 6). Dessa forma, a significância das tendências históricas por sexo não correspondeu a melhor desempenho de previsão nos três anos posteriores ao ajuste.

A interpretação dos resultados deve considerar as limitações da base agregada e dos pressupostos estatísticos das regressões. Os testes de Shapiro–Wilk indicaram desvios da normalidade residual, enquanto as estatísticas de Durbin–Watson sugeriram autocorrelação

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

temporal positiva. A análise de sensibilidade com erros-padrão robustos HAC de Newey-West, considerando duas defasagens, preservou a significância estatística das tendências masculina e feminina, enquanto a tendência do total nacional permaneceu não significativa ao nível de 5%. Embora esses resultados ofereçam maior sustentação à interpretação das associações temporais históricas, não asseguram estabilidade dos parâmetros nem precisão das extrapolações. As discrepâncias identificadas entre as estimativas e as contagens observadas de 2023 a 2025 reforçam essa limitação. Além disso, a ausência de informações individuais e de variáveis explicativas adicionais impede atribuir os desvios observados a fatores epidemiológicos específicos, como alterações diagnósticas, condições socioeconômicas ou efeitos da pandemia de COVID-19.

A contribuição do estudo consiste em demonstrar, com a mesma série de dados, a distinção entre associação temporal, avaliação dos pressupostos e desempenho preditivo. A comparação fora da amostra e a validação retrospectiva tornam visíveis os limites das extrapolações lineares, enquanto o modelo ingênuo oferece uma referência direta para interpretar seus erros. A base agregada permite reproduzir os cálculos, mas não examinar os fatores epidemiológicos que podem alterar as contagens. Estudos futuros poderão comparar métodos de séries temporais, incluir mudanças estruturais e indicadores regionais e populacionais, além de avaliar horizontes adicionais de previsão. As extrapolações até 2050 permanecem cenários condicionais, sem precisão validada para uso isolado no planejamento epidemiológico.

ASPECTOS ÉTICOS E DISPONIBILIDADE DOS DADOS

A pesquisa utilizou exclusivamente dados secundários agregados, disponibilizados em arquivo CSV associado ao SINAN/DATASUS, sem acesso a registros individualizados ou informações pessoais identificáveis. O conjunto de dados utilizado na modelagem e na avaliação das estimativas está disponível no endereço eletrônico informado na Metodologia, permitindo a conferência das observações anuais e a reprodução dos cálculos estatísticos apresentados.

AGRADECIMENTOS

Este estudo foi desenvolvido no âmbito do Programa de Pós-Graduação em Governança e Transformação Digital da Universidade Federal do Tocantins, em parceria com o Ministério da Saúde, por meio de um Termo de Execução Descentralizada (TED).

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

APÊNDICE A – CONTAGENS ANUAIS UTILIZADAS NO ESTUDO (2001–2025)

Tabela A1. Série anual do CSV associado ao SINAN/DATASUS, Brasil, 2001–2025. Os anos 2001–2022 compõem a amostra de treinamento; 2023–2025 são reservados à avaliação das estimativas

Ano	Masculino	Feminino	Total
2001	53.978	29.582	83.560
2002	59.490	32.515	92.005
2003	60.685	33.141	93.826
2004	60.422	32.550	92.972
2005	59.704	31.711	91.415
2006	57.167	30.156	87.323
2007	55.456	28.365	83.821
2008	57.412	29.025	86.437
2009	57.730	28.977	86.707
2010	57.182	28.221	85.403
2011	59.002	28.619	87.621
2012	58.274	27.845	86.119
2013	58.633	28.207	86.840
2014	58.023	26.976	84.999
2015	58.830	26.492	85.322
2016	59.355	26.443	85.798
2017	63.435	26.924	90.359
2018	66.214	28.453	94.667
2019	67.223	28.785	96.008
2020	60.462	25.339	85.801
2021	63.648	26.892	90.540
2022	72.141	30.567	102.708
2023	77.515	32.071	109.586
2024	79.485	33.775	113.260
2025	81.633	34.582	116.215

Fonte: arquivo CSV 'tuberculose-2001 a 2025.csv' associado à base 'Casos de tuberculose: desde 2001 (SINAN)' (Brasil, [s. d.]); acesso informado: 20 out. 2026. Nota: o arquivo tem quatro colunas e 25 linhas anuais; Total = Masculino + Feminino em todas as linhas.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

REFERÊNCIAS

BRASIL. Ministério da Saúde. Departamento de Informática do Sistema Único de Saúde (DATASUS). **Casos de tuberculose: desde 2001 (SINAN)**. Brasília, DF: Ministério da Saúde, [s. d.]. Disponível em: https://datasus.saude.gov.br/acesso-a-informacao/casos-de-tuberculose-desde-2001-sinan/. Acesso em: 20 out. 2026.

CHITWOOD, Melanie H. *et al.* Quantifying disruptions to tuberculosis case detection in Brazilian states during the COVID-19 pandemic. *International Journal of Epidemiology*, Oxford, v. 54, n. 5, art. dyaf146, 2025. DOI: https://doi.org/10.1093/ije/dyaf146.

COSTA, Fernanda Bruzadelli Paulino da *et al.* Spatial distribution and temporal trend of childhood tuberculosis in Brazil. **Tropical Medicine and Infectious Disease**, Basel, v. 8, n. 1, art. 12, 2023. DOI: https://doi.org/10.3390/tropicalmed8010012.

DURBIN, J.; WATSON, G. S. Testing for serial correlation in least squares regression. I. **Biometrika**, Oxford, v. 37, n. 3-4, p. 409-428, 1950. DOI: https://doi.org/10.1093/biomet/37.3-4.409.

GASPAR, Roberta S. *et al.* Temporal analysis of reported cases of tuberculosis and of tuberculosis-HIV co-infection in Brazil between 2002 and 2012. **Jornal Brasileiro de Pneumologia**, Brasília, DF, v. 42, n. 6, p. 416-422, 2016. DOI: https://doi.org/10.1590/S1806-37562016000000054.

GIACOMET, Clóvis Luciano *et al.* Temporal trend of tuberculosis incidence and its spatial distribution in Macapá, Amapá. **Revista de Saúde Pública**, São Paulo, v. 55, art. 96, 2021. DOI: https://doi.org/10.11606/s1518-8787.2021055003431.

GOMES, Thiéberson; FERRACIOLI, Laércio. A investigação da construção de modelos no estudo de um tópico de Física utilizando um ambiente de modelagem computacional qualitativo. **Revista Brasileira de Ensino de Física**, São Paulo, v. 28, n. 4, p. 453-461, 2006.

HYNDMAN, Rob J.; KOEHLER, Anne B. Another look at measures of forecast accuracy. **International Journal of Forecasting**, Amsterdam, v. 22, n. 4, p. 679-688, 2006. DOI: https://doi.org/10.1016/j.ijforecast.2006.03.001.

JESUS, Gabriela S. *et al.* The effect of primary health care on tuberculosis in a nationwide cohort of 7.3 million Brazilian people: a quasi-experimental study. **The Lancet Global Health**, London, v. 10, n. 3, p. e390-e397, 2022. DOI: https://doi.org/10.1016/S2214-109X(21)00550-7.

LI, Yunfei *et al.* Excess tuberculosis cases and deaths following an economic recession in Brazil: an analysis of nationally representative disease registry data. **The Lancet Global Health**, London, v. 10, n. 10, p. e1463-e1472, 2022. DOI: https://doi.org/10.1016/S2214-109X(22)00320-5.

LIMA, M. T. **Técnica estatística**: teoria e prática (R programming). Londrina: Editora e Distribuidora Educacional, 2009.

MACIEL, Ethel Leonor Noia; SALES, Carolina Maia Martins. Epidemiological surveillance of tuberculosis in Brazil: how can more progress be made? **Epidemiologia e Serviços de Saúde**,

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Brasília, DF, v. 25, n. 1, p. 175-178, 2016. DOI: https://doi.org/10.5123/S1679-49742016000100018.

MESQUITA, Cristal Ribeiro *et al.* Spatial analysis of tuberculosis patient flow in a neglected region of Northern Brazil. **Tropical Medicine and Infectious Disease**, Basel, v. 8, n. 8, art. 397, 2023. DOI: https://doi.org/10.3390/tropicalmed8080397.

MORETTIN, Pedro Alberto; BUSSAB, Wilton de Oliveira. **Estatística básica**. 9. ed. São Paulo: Saraiva, 2017.

NEWEY, Whitney K.; WEST, Kenneth D. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. **Econometrica**, v. 55, n. 3, p. 703-708, 1987. DOI: https://doi.org/10.2307/1913610.

PAIVA JÚNIOR, Francisco Pessoa de; MARCOLINO, Raimundo dos Santos; FALCÃO, Danilo Lima; LEITE, Jefferson Cruz dos Santos. **Modelagem matemática e epidemiologia com aplicações**. 1. ed. Curitiba: Appris, 2021. ISBN 978-65-250-0689-5.

PRADO, Thiago Nascimento do *et al.* Clinical and epidemiological characteristics associated with unfavorable tuberculosis treatment outcomes in TB-HIV co-infected patients in Brazil: a hierarchical polytomous analysis. **Brazilian Journal of Infectious Diseases**, Salvador, v. 21, n. 2, p. 162-170, 2017. DOI: https://doi.org/10.1016/j.bjid.2016.11.006.

QUEIROZ, Ana Angélica Rêgo de *et al.* Spatial pattern and temporal trend of mortality due to tuberculosis. **Revista Latino-Americana de Enfermagem**, Ribeirão Preto, v. 26, e2992, 2018. DOI: https://doi.org/10.1590/1518-8345.2049.2992.

SALES, Carolina Maia Martins *et al.* Social determinants of tuberculosis via a zero-inflated model in small areas of a city in Southeastern Brazil. **Revista da Sociedade Brasileira de Medicina Tropical**, Uberaba, v. 51, n. 5, p. 638-643, 2018. DOI: https://doi.org/10.1590/0037-8682-0015-2018.

SANTANA, Verônica. **STAT 501: método de regressão**. Pennsylvania: Penn State, 2020. Disponível em: https://online.stat.psu.edu/stat501/. Acesso em: 31 out. 2026.

SCAVETTA, Rick J. **Python e R para cientistas de dados modernos**: o melhor dos dois mundos. São Paulo: Novatec, 2022.

SCHOBER, Patrick; BOER, Christa; SCHWARTE, Lothar A. Correlation coefficients: appropriate use and interpretation. **Anesthesia & Analgesia**, Philadelphia, v. 126, n. 5, p. 1763-1768, 2018. DOI: https://doi.org/10.1213/ANE.0000000000002864.

SCHMULLER, Joseph. **Análise estatística com R**. Tradução de Samantha Batista. Rio de Janeiro: Alta Books, 2019.

SHAPIRO, Samuel S.; WILK, Martin B. An analysis of variance test for normality (complete samples). Biometrika, Oxford, v. 52, n. 3-4, p. 591-611, 1965. DOI: https://doi.org/10.1093/biomet/52.3-4.591.

SILVA, Amanda Souza da. **Análise de dados**. Londrina: Editora e Distribuidora Educacional, 2020.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

SILVA, Gabriela Drummond Marques da; BARTHOLOMAY, Patrícia; CRUZ, Oswaldo Gonçalves; GARCIA, Leila Posenato. Evaluation of data quality, timeliness and acceptability of the tuberculosis surveillance system in Brazil's micro-regions. **Ciência & Saúde Coletiva**, Rio de Janeiro, v. 22, n. 10, p. 3307-3319, 2017. DOI: https://doi.org/10.1590/1413-812320172210.18032017.

SILVA, Marcus Tolentino; GALVÃO, Taís Freire. Incidência de tuberculose no Brasil: análise de série temporal entre 2001 e 2021 e projeção até 2030. **Revista Brasileira de Epidemiologia**, Rio de Janeiro, v. 27, e240027, 2024. DOI: https://doi.org/10.1590/1980-549720240027.2.

VILLALVA-SERRA, Klauss *et al.* Impact of Xpert MTB/RIF implementation in tuberculosis case detection and control in Brazil: a nationwide intervention time-series analysis (2011-2022). **The Lancet Regional Health – Americas**, London, v. 36, art. 100804, 2024. DOI: https://doi.org/10.1016/j.lana.2024.100804.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.