



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE ESTATÍSTICA

ANÁLISE DA FECUNDIDADE ADOLESCENTE NO SEMIÁRIDO BRASILEIRO POR
REGRESSÃO LOGÍSTICA BINÁRIA

Tânia Barbosa Tomaz

João Pessoa – PB

Maio, 2025.

Tânia Barbosa Tomaz

ANÁLISE DA FECUNDIDADE ADOLESCENTE NO SEMIÁRIDO BRASILEIRO POR
REGRESSÃO LOGÍSTICA BINÁRIA

Monografia apresentada ao Curso de Bacharelado em Estatística da Universidade Federal da Paraíba como requisito parcial para obtenção do Grau de Bacharel. Área de Concentração: Estatística Aplicada.

Orientadora: Prof^ª. Dra. Everlane Suane de Araújo da Silva

João Pessoa - PB
Maio, 2025.

Catálogo na publicação
Seção de Catalogação e Classificação

T655a Tomaz, Tânia Barbosa.

Análise da fecundidade adolescente no semiárido brasileiro por regressão logística binária / Tânia Barbosa Tomaz. - João Pessoa, 2025.

67 p. : il.

Orientação: Everlane Suane de Araújo da Silva.

TCC (Curso de Bacharelado em Estatística) -
UFPB/CCEN.

1. Fecundidade. 2. Adolescente. 3. Semiárido brasileiro. I. Silva, Everlane Suane de Araújo da. II. Título.

UFPB/CCEN

CDU 311(043.2)



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
COORDENAÇÃO DO CURSO DE ESTATÍSTICA



ATA DE DEFESA DO TRABALHO DE CONCLUSÃO DE CURSO

“Análise da Fecundidade Adolescente no Semiárido Brasileiro por Regressão Logística Binária” Tânia Barbosa Tomaz

Ao sexto dia do mês de Maio de 2025 às 09h00m, de modo presencial, no Auditório do CCEN/UFPB, realizou-se a sessão pública de defesa do Trabalho de Conclusão de Curso da discente Tânia Barbosa Tomaz, matrícula 20180089379, com a Banca Examinadora composta pelos professores: Dra. Everlane Suane de Araújo da Silva, Presidente/Orientadora (Departamento de Estatística - UFPB), Dr. Neir Antunes Paes, Examinador (Departamento de Estatística - UFPB), Dra. Maria Lídia Coco Terra, Examinadora (Departamento de Estatística - UFPB) e Dra. Ana Flávia Uzeda dos Santos Macambira, Examinadora Suplente (Departamento de Estatística - UFPB). Iniciando-se os trabalhos, a presidente da Banca Examinadora cumprimentou os presentes, comunicou aos mesmos a finalidade da reunião e passou a palavra a candidata para que se fizesse, oralmente, a exposição do Trabalho de Conclusão de Curso intitulado *“Análise da Fecundidade Adolescente no Semiárido Brasileiro por Regressão Logística Binária”*. Concluída a apresentação, a Banca Examinadora iniciou a arguição da candidata. Encerrados os trabalhos de arguição os examinadores reuniram-se para avaliação e deram o parecer final sobre a apresentação e defesa oral da candidata, tendo sido atribuída à sua apresentação a nota Nove inteiros e cinco décimos (9,5), na disciplina de TCC II, resultante da média aritmética das notas atribuídas pelos membros da Banca Examinadora. A aprovação da discente está condicionada a entrega da versão final do Trabalho de Conclusão de Curso com a inserção da ficha catalográfica e, as alterações sugeridas pelos examinadores, à Coordenação do Curso de Bacharelado em Estatística no prazo de 7 dias corridos após a defesa

Documento assinado digitalmente
gov.br EVERLANE SUANE DE ARAUJO DA SILVA
Data: 07/05/2025 08:42:59-0300
Verifique em <https://validar.iti.gov.br>

Dra. Everlane Suane de Araujo da Silva
(Professora Orientadora)

Documento assinado digitalmente
gov.br MARIA LIDIA COCO TERRA
Data: 07/05/2025 08:56:53-0300
Verifique em <https://validar.iti.gov.br>

Dra. Maria Lidia Coco Terra
(Professora Examinadora)

Documento assinado digitalmente
gov.br NEIR ANTUNES PAES
Data: 08/05/2025 16:12:50-0300
Verifique em <https://validar.iti.gov.br>

Dr. Neir Antunes Paes
(Professor Examinador)

Documento assinado digitalmente
gov.br TANIA BARBOSA TOMAZ
Data: 07/05/2025 22:27:54-0300
Verifique em <https://validar.iti.gov.br>

Tânia Barbosa Tomaz
(Discente)

João Pessoa, 06 de Maio de 2025.

*Dedico à minha família,
cuja dedicação e apoio incansáveis foram
fundamentais em cada passo desta jornada.*

Agradecimentos

A Deus, por me fazer enxergar, mesmo nos momentos de exaustão, que sempre valeria a pena seguir. A Ele, minha fé, força e gratidão.

À minha família, pelo amor que nunca se ausenta, pelo carinho que acolhe e pelo amparo silencioso que sustentou cada conquista. Cheguei até aqui porque vocês sempre estiveram comigo.

À minha orientadora, Dr^a Everlane Suane de Araújo da Silva, que, com paciência, me guiou no universo da demografia e me fez enxergar as oportunidades de investigar a fecundidade no Semiárido brasileiro. Foi uma jornada feita de passos firmes, cuidadosamente construídos com dedicação e afeto. Ao me acolher no Laboratório de Estudos Demográficos, ofereceu não apenas um espaço de aprendizagem, mas também um ambiente acolhedor. Ter ali um lugar para pesquisar e estar foi o que suavizou os dias mais intensos da universidade.

À banca avaliadora, na pessoa do professor Dr. Neir Antunes Paes e da professora Dr^a Maria Lídia Coco Terra, expresse minha sincera gratidão. A presença de vocês neste momento é motivo de honra e alegria.

Aos meus professores do Departamento de Estatística, pelo incentivo constante e pelo apoio ao longo de toda a minha jornada formativa.

À professora Dr^a Bernadete de Lourdes Figueiredo de Almeida, que, ainda na graduação em Serviço Social, despertou em mim a paixão pela análise de dados e o compromisso com a qualidade na produção acadêmica. Sua influência foi decisiva em minha trajetória.

À minha amiga Janilce Magalhães, que há muitos anos me apresentou aos Sistemas de Informação em Saúde, por meio da Coordenação Epidemiológica do município de Conde-PB. Nunca sabemos ao certo onde nossos passos nos levarão, mas sou profundamente grata por aquele ter sido um dos caminhos que me trouxeram até aqui.

Com ternura, deixo meu agradecimento aos amigos de curso, especialmente aos "Veteranos", com quem compartilhei listas, risos, prazos e esperanças. Juntos, trilhamos um caminho que ficará guardado na memória com muito afeto.

Ao amigo de trabalho Flávio Nery, pelo companheirismo constante e pelo apoio generoso, que se fez presente de tantas formas: com escuta, incentivo e gestos que tornaram o caminho mais leve.

Aos meus alunos e minhas alunas, com quem aprendi tanto, enquanto também era aluna. Em cada troca de saberes, em cada partilha de dúvidas e angústias, construímos juntos um caminho de crescimento mútuo.

Por fim, às crianças e adolescentes, razão maior do meu fazer acadêmico. Foram elas que me ensinaram sobre a vida muito além das teorias.

RESUMO

A fecundidade na adolescência é um fenômeno demográfico relevante para a compreensão das desigualdades sociais e territoriais no Brasil e constitui um dos indicadores do ODS 3 da Agenda 2030. Suas implicações vão além do campo reprodutivo, afetando a saúde materno-infantil, o percurso educacional, a inserção no mercado de trabalho e a perpetuação da pobreza, especialmente em contextos vulneráveis. Apesar da tendência nacional de queda na fecundidade, essa redução não tem sido homogênea entre faixas etárias e regiões. Este estudo teve como objetivo geral: analisar os fatores sociodemográficos associados aos níveis da fecundidade das adolescentes entre 15 e 19 anos no Semiárido do Brasil em 2023. Com abordagem quantitativa, utilizou-se microdados do SINASC (2019–2023) e estimativas do IBGE, abrangendo 1.473 municípios. O estudo envolveu quatro etapas: tratamento dos dados, avaliação da qualidade das informações, análise do comportamento da fecundidade e aplicação de Regressão Logística Binária. A variável dependente foi a TEF adolescente, codificada como 1 para municípios abaixo da média nacional (0,0391) e 0 para os demais. As variáveis explicativas incluíram escolaridade, estado civil, filhos anteriores, cor/raça e tipologia territorial. Os resultados indicaram que a presença de adolescentes com ensino fundamental 2 aumentou em 3,75% a chance de o município ter uma TEF abaixo da média nacional. Além disso, municípios classificados como rurais adjacentes, intermediários adjacentes e rurais remotos também apresentaram maior probabilidade de baixa fecundidade adolescente em comparação aos urbanos, sugerindo que fatores como migração, mudanças culturais ou subnotificação podem ter influenciado os resultados, especialmente nas áreas rurais remotas. A presença de adolescentes casadas elevou em apenas 1,7% a chance do município estar abaixo da TEF nacional, indicando que, embora a nupcialidade ainda esteja associada à fecundidade, seu impacto foi limitado. Já as variáveis cor/raça e número de filhos anteriores não apresentaram associação estatisticamente significativa, refletindo a complexidade das interações entre os fatores sociodemográficos e a fecundidade no Semiárido brasileiro. O modelo mostrou bom desempenho, com AUC de aproximadamente 73%, sensibilidade de 54,2% e especificidade de 80,8%, validando sua aplicabilidade. Concluiu-se que, embora a fecundidade adolescente no Semiárido do Brasil esteja em queda, persistem desigualdades territoriais que demandam políticas públicas específicas e sensíveis ao contexto local.

Palavras-chave: Fecundidade; Adolescente; Semiárido brasileiro; Regressão Logística Binária.

ABSTRACT

Adolescent fertility is a significant demographic phenomenon for understanding social and territorial inequalities in Brazil and constitutes one of the indicators of SDG 3 of the 2030 Agenda. Its implications go beyond reproductive aspects, impacting maternal and child health, educational trajectories, access to the labor market, and the perpetuation of intergenerational poverty, especially in vulnerable contexts. Despite the national trend of declining fertility rates, this reduction has not occurred uniformly across age groups and regions. This study aimed to analyze the sociodemographic factors associated with fertility levels among adolescents aged 15 to 19 in Brazil's Semi-Arid region in 2023. A quantitative approach was adopted using microdata from SINASC (2019–2023) and population estimates from IBGE, covering 1,473 municipalities. The methodology involved four main steps: data processing, assessment of data quality, analysis of adolescent fertility trends, and application of a Binary Logistic Regression model. The dependent variable was the adolescent-specific fertility rate (ASFR), coded as 1 for municipalities below the national average (0.0391) and 0 otherwise. Explanatory variables included maternal education, marital status, number of previous children, race/skin color, and territorial typology. Results indicated that the presence of adolescents with lower secondary education increased the likelihood of a municipality having an ASFR below the national average by 3.75%. Additionally, municipalities classified as adjacent rural, adjacent intermediate, and remote rural areas showed a higher probability of low adolescent fertility compared to urban areas, suggesting that factors such as migration, cultural changes, or possible underreporting may have influenced these outcomes, particularly in remote rural areas. The presence of married adolescents increased the likelihood by only 1.7%, indicating that while marriage remains associated with fertility, its effect was limited. Variables such as race/skin color and previous childbirths showed no statistically significant association, reflecting the complexity of sociodemographic interactions with adolescent fertility in the Semi-Arid region. The model demonstrated good performance, with an AUC of approximately 73%, sensitivity of 54.2%, and specificity of 80.8%, supporting its applicability. It is concluded that, although adolescent fertility is declining in the Brazil's Semi-Arid, territorial disparities persist and require context-sensitive and targeted public policies.

Keywords: Fertility; Adolescents; Brazilian Semi-arid; Binary Logistic Regression.

LISTA DE FIGURAS

Figura 1: Composição do Semiárido brasileiro, 2025.	21
Figura 2: Taxa Específica de Fecundidade adolescente (15 – 19 anos) dos municípios do Semiárido brasileiro, 2023.....	51

LISTA DE GRÁFICOS

Gráfico 1: Taxa de Fecundidade Total para o Brasil e grandes regiões: 1940-2010.....	18
Gráfico 2: Taxa de Fecundidade Total para as Unidades da Federação: 1960-2010.	19
Gráfico 3: Taxa Específica de Fecundidade das Adolescentes do Brasil, Regiões e Semiárido Brasileiro, 2019-2023.	49
Gráfico 4: Análise dos pontos influentes no Modelo ao considerar os fatores sociodemográficos associados à fecundidade entre adolescentes no Semiárido brasileiro, 2023.....	55
Gráfico 5: Distância de Cook do Modelo ao considerar os fatores sociodemográficos associados à fecundidade entre adolescentes no Semiárido brasileiro, 2023.	55
Gráfico 6: Análise Combinada dos Pontos Influentes do Modelo ao considerar os fatores sociodemográficos associados à fecundidade entre adolescentes no Semiárido brasileiro, 2023.	56
Gráfico 7: Curva ROC do Modelo ao considerar os fatores sociodemográficos associados à fecundidade entre adolescentes no Semiárido brasileiro, 2023.....	57

LISTA DE QUADRO

Quadro 1: Identificação da variável sociodemográfica da Declaração de Nascido Vivo e respectivas categorias.	30
---	----

LISTA DE TABELAS

Tabela 1: Fator de correção calculados para os estados do Semiárido brasileiro, considerando o período de 2019 – 2023.	34
Tabela 2: Exemplo de Matriz de Confusão	41
Tabela 3: Percentual da cobertura de Nascidos Vivos (NV), entre mulheres de 10 a 49 anos de idade, dos Estados do Semiárido brasileiro, 2018 a 2022.	44
Tabela 4: Comparativo do percentual de cobertura de Nascidos Vivos entre adolescentes (15 a 19 anos, 2019-2023) e mulheres em idade reprodutiva (15 a 49 anos, 2018-2022) nos Estados do Semiárido brasileiro.....	45
Tabela 5: Número de vazios e percentual de completude das variáveis, relacionadas às mães de todos os grupos etários e recém-nascidos, selecionadas da Declaração de Nascido Vivo no Semiárido brasileiro, 2019 a 2023.....	47
Tabela 6: Número de vazios e percentual de completude das variáveis sociodemográficas de mães adolescentes (15 a 19 anos) selecionadas da Declaração de Nascido Vivo no Semiárido Brasileiro, 2019–2023.	48
Tabela 7: Estimativas dos Parâmetros do Modelo de Regressão Logística Binária ao considerar os fatores sociodemográficos associados à fecundidade entre adolescentes no Semiárido brasileiro, 2023.	53
Tabela 8: Avaliação da Capacidade Preditiva do Modelo ao considerar os fatores sociodemográficos associados à fecundidade entre adolescentes no Semiárido brasileiro, 2023.	57
Tabela 9: Estimativas das Razões de Chance (Odds Ratio) do Modelo ao considerar os fatores sociodemográficos associados à fecundidade entre adolescentes no Semiárido brasileiro, 2023.	58

LISTA DE SIGLAS

AUC – *Area Under the Curve* (*Área Sob a Curva*)

CEP – Comitê de Ética em Pesquisa

CONEP – Comissão Nacional de Ética em Pesquisa

CONDEL – Conselho Deliberativo da SUDENE

DATASUS – Departamento de Informática do Sistema Único de Saúde

DNV – Declaração de Nascido Vivo

IBGE – Instituto Brasileiro de Geografia e Estatística

IPEA - Instituto de Pesquisa Econômica Aplicada

GLM – Modelos Lineares Generalizados

LED – Laboratório de Estudos Demográficos

ODS – Objetivos de Desenvolvimento Sustentável

ONU – Organização das Nações Unidas

OR – *Odds Ratio* (Razão de Chances)

PIBIC – Programa Institucional de Bolsas de Iniciação Científica

PIVIC – Programa Voluntário de Iniciação Científica

PROPESQ – Pró-Reitoria de Pesquisa

ROC – *Receiver Operating Characteristic* (Característica de Operação do Receptor)

R^2 de Nagelkerke – Coeficiente de Determinação Ajustado para Modelos Logísticos

RIPSA – Rede Interagencial de Informações para a Saúde

SINASC – Sistema de Informações sobre Nascidos Vivos

SUDENE – Superintendência do Desenvolvimento do Nordeste

TEF – Taxa Específica de Fecundidade

TFT – Taxa de Fecundidade Total

UFPB – Universidade Federal da Paraíba

SUMÁRIO

1 INTRODUÇÃO	15
1.1 Objetivos do estudo	17
1.1.1 Objetivo Geral	17
1.1.2 Objetivos Específicos	17
2 FUNDAMENTAÇÃO TEÓRICA	17
2.1 Transição da Fecundidade no Brasil	18
2.2 Transição da Fecundidade no Semiárido brasileiro	21
2.3 Qualidade dos Registros Vitais de Nascimento no Brasil	23
2.4 Aplicação de Modelos de Regressão Logística Binária	26
3 PROCEDIMENTOS METODOLÓGICOS	27
3.1 Tipologia e Desenvolvimento do Estudo	27
3.2 Fontes dos Dados e Variáveis do Estudo	29
3.3 Análise da qualidade dos dados de nascimentos	31
3.4 Análise da fecundidade entre adolescentes	32
3.5 Aplicação das técnicas indiretas de correção das Taxas Específicas de Fecundidade	33
3.6 Modelagem Estatística dos Dados	35
3.6.1 Especificações do Modelo de Regressão Logística	35
3.6.2 R^2 de Nagelkerke	38
3.6.3 Análise de Diagnóstico	38
3.6.4 Análise do Desempenho do Modelo de Regressão Logística Binária	40
4 RESULTADOS E DISCUSSÕES	43
4.1 Cobertura e completude dos registros de nascimentos	43
4.2 Comportamento da Fecundidade entre Adolescentes	49
4.3 Modelagem dos fatores sociodemográficos associados à fecundidade	53
4.3.1 Análise de Diagnóstico	54
4.3.2 Análise dos Pontos Influentes	55
4.3.3 Avaliação do Desempenho do Modelo	56
4.3.4 Análise das Razões de Chances (<i>Odds Ratio</i>)	58
5 CONSIDERAÇÕES FINAIS	60
REFERÊNCIAS	62
PRODUÇÃO ACADÊMICA	66
ANEXO	67

1 INTRODUÇÃO

O estudo da fecundidade na adolescência representa um importante instrumento para compreender as dinâmicas da transição demográfica e os múltiplos impactos sociais, econômicos e de saúde pública associados à maternidade precoce. Conforme apontaram Sanhueza et al. (2023), a gravidez na adolescência está relacionada a riscos elevados tanto para as mães quanto para os bebês, sendo as complicações gestacionais e no parto uma das principais causas de mortalidade entre meninas adolescentes.

Além das consequências para a saúde, a maternidade precoce acarreta prejuízos significativos na trajetória educacional e profissional das jovens. Essa realidade contribui para a evasão escolar, reduz as oportunidades de qualificação e limita as perspectivas de inserção no mercado de trabalho, perpetuando o ciclo da pobreza, especialmente entre mulheres em contextos de vulnerabilidade, como áreas rurais e com baixos níveis de escolaridade (Azevedo et al., 2012).

Apesar de ações promovidas por governos e pela sociedade civil, a maternidade na adolescência permanece como um desafio global. Dados do Fundo de População das Nações Unidas (2023) indicaram que, em 2021, aproximadamente 13,3 milhões de bebês, cerca de 10% dos nascimentos mundiais, foram gerados por mães com até 20 anos. A distribuição desse fenômeno varia entre os países, sendo mais expressiva em regiões marcadas por desigualdades sociais e econômicas.

No caso brasileiro, desde a década de 1960 observou-se uma tendência de queda nas Taxas de Fecundidade Total (TFT), que passaram a se aproximar do nível de reposição populacional, em torno de 2,1 filhos por mulher, em todas as Grandes Regiões. Contudo, como apontou o relatório “Fecundidade e Dinâmica da População Brasileira” do Fundo de População das Nações Unidas (2018), esse declínio não se dá de forma homogênea entre os diferentes grupos etários e contextos regionais, estando fortemente associado às condições socioeconômicas das mulheres.

Entre as adolescentes, especialmente aquelas com idades entre 15 e 19 anos, o comportamento reprodutivo continua a exercer um papel relevante na configuração dos padrões de fecundidade. Embora o ritmo de rejuvenescimento da fecundidade tenha desacelerado entre os anos 2000 e 2010 (Cavenaghi; Berquó, 2014), os dados recentes apontaram que essa faixa etária mantém taxas expressivas de fecundidade, sobretudo em territórios marcados pela pobreza e vulnerabilidade social como na região Semiárida do Brasil (Pereira Júnior, 2020).

Neste cenário, o presente estudo se dedica à análise da fecundidade na adolescência no Semiárido brasileiro, região que tem sido objeto de diversas investigações desenvolvidas no âmbito do Laboratório de Estudos Demográficos (LED)/Universidade Federal da Paraíba (UFPB), ao qual esta pesquisa está vinculada. O recorte geográfico escolhido visa aprofundar a compreensão sobre os determinantes sociodemográficos da fecundidade juvenil em um contexto historicamente marcado por desigualdades estruturais.

Dessa forma, o estudo busca responder às seguintes questões centrais:

- Houve alterações relevantes nas taxas de fecundidade entre adolescentes no Semiárido brasileiro entre 2019 e 2023?
- Caso haja mudanças nos níveis e padrões de fecundidade, quais fatores sociodemográficos podem estar relacionados a essa variação?
- O ritmo de decaimento da fecundidade entre adolescentes no Semiárido brasileiro acompanha a tendência nacional?

Além da análise dos níveis de fecundidade, o estudo também destacou a importância de se avaliar a qualidade dos registros de nascidos vivos, uma vez que inconsistências ou lacunas nesses dados podem comprometer a precisão das estatísticas, dificultando a identificação de tendências reais e a formulação de políticas públicas eficazes.

Assim, tornou-se essencial aplicar técnicas de correção dos dados de nascimento e realizar modelagens que possibilitem identificar os fatores associados à fecundidade na adolescência. Para isso, este trabalho adotou como referência teórico-metodológica a dissertação de mestrado intitulada *Modelagem da Fecundidade Materno-Adolescente no Semiárido Brasileiro*, de Basílio Henrique Pereira Júnior, sob orientação dos professores Dr. Neir Antunes Paes e Dr^a. Caliandra Maria Bezerra Luna Lima, apresentada ao Programa de Pós-Graduação em Modelos de Decisão e Saúde da Universidade Federal da Paraíba.

Cabe destacar que este estudo teve como ponto de partida experiências anteriores desenvolvidas no âmbito do LED/UFPB, com o apoio da Pró-Reitoria de Pesquisa (PROPESQ) da UFPB. O primeiro projeto, intitulado “Estudo da Fecundidade no Semiárido Brasileiro, foi executado no período de setembro de 2020 a agosto de 2021”, por meio do Programa Voluntário de Iniciação Científica (PIVIC). O segundo, denominado “O Comportamento da Fecundidade no Semiárido Brasileiro na Pandemia da COVID-19”, foi desenvolvido entre setembro de 2022 e agosto de 2023, no âmbito do Programa Institucional de Bolsas de Iniciação Científica

(PIBIC). Ambos os projetos foram coordenados pela professora Dra. Everlane Suane de Araújo da Silva.

Para este estudo, delineou-se, a seguir, os seguintes objetivos:

1.1 Objetivos do estudo

No intuito de proceder com este estudo, os seguintes objetivos foram traçados:

1.1.1 Objetivo Geral

Analisar os fatores sociodemográficos associados aos níveis da fecundidade das adolescentes com idade entre 15 a 19 anos no Semiárido brasileiro, em 2023.

1.1.2 Objetivos Específicos

- Averiguar a qualidade dos dados de nascidos vivos do Semiárido brasileiro disponíveis no Sistema de Informação sobre Nascidos Vivos (SINASC);
- Estudar os níveis da fecundidade das adolescentes do Semiárido brasileiro no período de 2019 a 2023;
- Aplicar um Modelo de Regressão Logística Binária para analisar a associação de fatores sociodemográficos e a fecundidade das adolescentes.

Por fim, espera-se que os resultados apresentados por este estudo contribuam para o aprofundamento das discussões acadêmicas sobre o tema, bem como subsidie gestores públicos no planejamento e implementação de políticas voltadas à promoção de direitos das adolescentes.

2 FUNDAMENTAÇÃO TEÓRICA

A composição populacional tem uma relação direta com o número médio de filhos tidos por mulheres em um período de tempo. Assim, a fecundidade se apresenta como uma das componentes demográficas fundamentais para se compreender as transformações do padrão demográfico em um país, o que aponta para necessidade deste estudo partir da contextualização sócio-histórica do comportamento da fecundidade no Brasil e no Semiárido brasileiro.

2.1 Transição da Fecundidade no Brasil

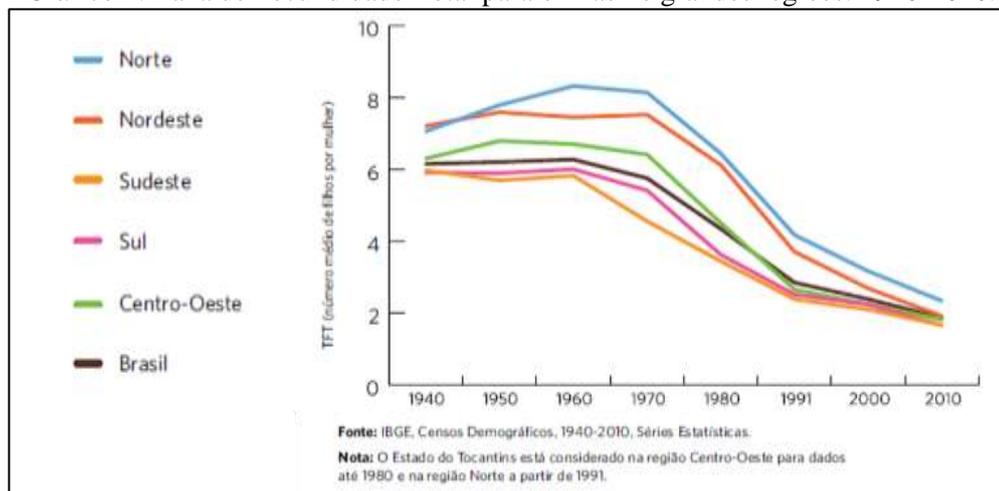
Até meados do século XX, o padrão demográfico brasileiro era relativamente estável com pequenas oscilações nos níveis de mortalidade e de natalidade. No entanto, a partir de 1960, o Brasil passou a ter uma queda expressiva dos níveis de fecundidade. Segundo Simões (2006), entre 1940 e 1960, a média de filhos tidos por mulheres era em torno de 6,2; nos anos 2000, essa média caiu para 2,4 filhos com tendência a convergir nas décadas subsequentes para valores abaixo do nível de reposição, ou seja, 2,1 filhos.

Ainda de acordo com Simões (2006, p. 15),

Essa radical transformação do padrão demográfico corresponde a uma das mais importantes modificações estruturais verificadas na sociedade brasileira, com reduções na taxa de crescimento populacional e alterações na estrutura etária, com crescimento mais lento no número de crianças e adolescentes, paralelamente a um aumento da população em idade ativa de pessoas idosas.

No entanto, no Brasil, a transformação de altas para baixas taxas de fecundidade, processo denominado de transição da fecundidade, não ocorreu no mesmo ritmo em todas as regiões do país (UNFPA, 2018), como demonstrado no Gráfico 1.

Gráfico 1: Taxa de Fecundidade Total para o Brasil e grandes regiões: 1940-2010.



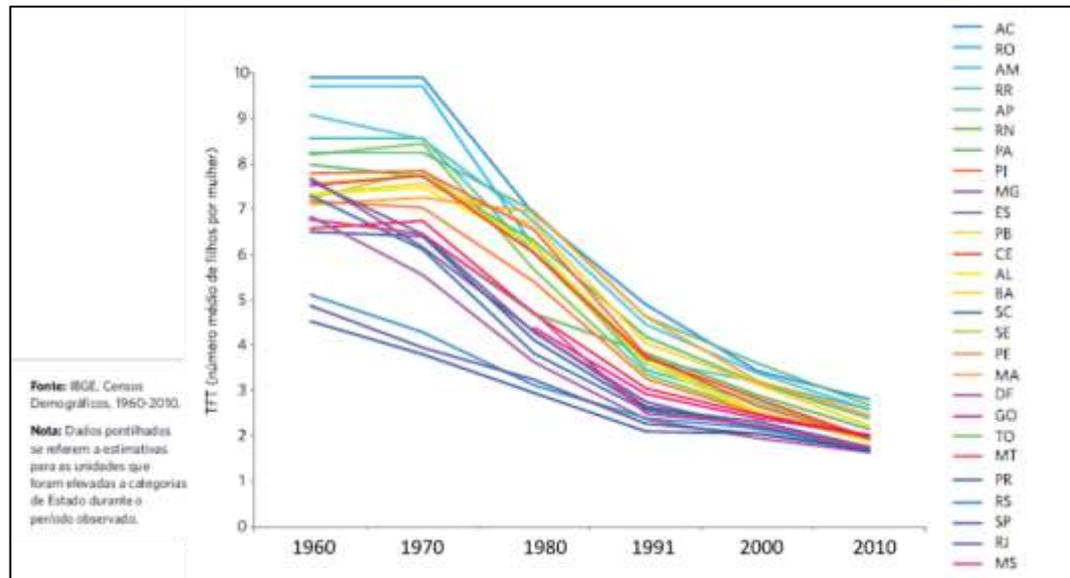
Fonte: Fundo de População das Nações Unidas, 2018, p. 14.

As Regiões Sul e Sudeste sempre apresentaram as menores Taxas de Fecundidade Total (TFT), seguido pela Região Centro-Oeste. Ao contrário, as Regiões Norte e Nordeste tinham taxas superiores à média nacional. Mesmo que a partir dos anos 2000, as TFTs nacional e de todas as regiões tenham se aproximado do nível de reposição, o número de filhos tidos pelas

mulheres na Região Norte era de 3,01 filhos e na Região Nordeste era de 2,54 filhos (RIPSA, 2012).

Ao analisar os níveis de fecundidade por Unidade Federativa, observou-se que as diferenças do ritmo de declínio foram ainda mais discrepantes, como se pode constatar no Gráfico 2.

Gráfico 2: Taxa de Fecundidade Total para as Unidades da Federação: 1960-2010.



Fonte: Fundo de População das Nações Unidas, 2018, p. 15.

O processo de declínio da fecundidade foi amplo e acentuado em todas as Unidades Federativas do país, apesar das diferentes trajetórias iniciais. Em 1960, os estados das Regiões Sudeste e do Sul, como São Paulo (SP), Rio Grande do Sul (RS) e Rio de Janeiro (RJ), que aparecem nas linhas azul-escuras e roxas, já apresentavam taxas mais baixas, em torno de 4 filhos por mulher, em relação aos demais estados. E os estados da Região Norte, Acre (AC), Rondônia (RO) e Amazonas (AM), representados pelas linhas azuis, estão entre os que apresentavam as taxas mais altas, com valores próximos a 10 filhos por mulher.

A partir da década de 1990, percebeu-se um declínio da fecundidade em todas as Unidades Federativas ainda que os ritmos variem significativamente. Em 2010, 18 das 27 Unidades Federativas e o Distrito Federal já apresentavam TFT abaixo do nível de reposição. Esse grupo incluía, além do Distrito Federal, todos os estados das Regiões Sudeste, Sul e Centro-Oeste, no Norte, Rondônia, e no Nordeste, Piauí, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Sergipe e Bahia.

Segundo Simões (2006), quando comparado com outros países, pode-se dizer que a transição da fecundidade no Brasil foi uma das mais rápidas do mundo. Nos países

desenvolvidos, o declínio da fecundidade se deu em um período de 80 a 100 anos, no Brasil, esse processo ocorreu em aproximadamente 30 anos, um ritmo além do que os especialistas esperavam.

De acordo com Wong (2018, p.69),

A previsão, hoje vista como conservadora, de se atingir o nível de substituição populacional, mediante o qual, os casais teriam filhos na quantidade suficiente para repor as gerações – algo em torno de 2,1 filhos por mulher – foi ultrapassado, igualmente, antes do esperado no Brasil.

Além disso, apesar das TFTs das regiões e, conseqüentemente, dos estados brasileiros, a partir de 2010, apontaram uma tendência à convergência para valores abaixo da taxa de reposição, observava-se diferenças no número médio de filhos tidos por mulheres ao se fazer uma análise do padrão etário da fecundidade.

Até a década de 1980, o Brasil apresentava um padrão de fecundidade considerado tardio, com maior concentração de nascimentos entre mulheres de 25 a 34 anos. A partir desse período, esse padrão começou a se modificar, com um aumento proporcional dos nascimentos entre mulheres mais jovens, especialmente na faixa de 20 a 24 anos. Entre os anos de 1990 e 2000, observou-se um declínio contínuo das taxas de fecundidade, abrangendo praticamente todas as faixas etárias. A única exceção foi o grupo de 15 a 19 anos, que apresentou leves aumentos nas taxas, apontando para uma tendência de rejuvenescimento da fecundidade (Simões, 2006).

Por outro lado, a redução dos nascimentos em idades mais avançadas esteve diretamente associada às mudanças na composição da fecundidade segundo a ordem de nascimento. Como destacaram Miranda-Ribeiro *et al.* (2017 *apud* Pereira Júnior 2020), a paridade, entendida como o número de filhos por mulher, passou por transformações significativas. Em outras palavras, ao longo do tempo, as mulheres passaram a ter menos filhos, com maior concentração nos primeiros nascimentos (1º e 2º filhos), enquanto a ocorrência de terceiros filhos ou mais foi progressivamente reduzida.

Em 2010, com base nos dados do Censo Demográfico, observaram-se novas alterações nos padrões de fecundidade entre as mulheres jovens e adolescentes. Embora esses grupos ainda apresentassem taxas relativamente elevadas em comparação com outras faixas etárias, já não era possível falar em uma tendência de rejuvenescimento da fecundidade (Cavenaghi; Berquó, 2014).

Paralelamente, a análise comparativa dos Censos Demográficos de 2000 e 2010 revelou que a queda da fecundidade foi mais acentuada entre mulheres com menor escolaridade e renda mais baixa. Apesar dessa redução, persistiam marcantes desigualdades regionais. A Região

Nordeste, por exemplo, continuava apresentando taxas de fecundidade superiores às registradas nas Regiões Sul e Sudeste do Brasil. Além disso, permaneciam elevados os contrastes por local de residência, sobretudo entre áreas urbanas e rurais, e entre diferentes regiões do país (Idem).

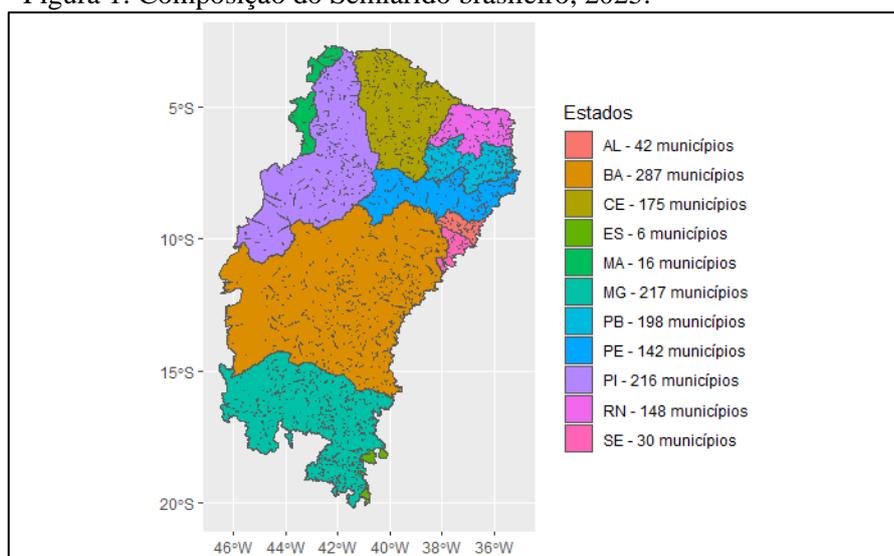
Diante das persistentes desigualdades regionais e socioeconômicas observadas no Brasil, torna-se, portanto, fundamental estudar a transição da fecundidade no Semiárido brasileiro. Essa região, marcada por vulnerabilidades históricas e limitações no acesso a serviços e oportunidades, apresenta dinâmicas demográficas que podem divergir significativamente das observadas em contextos mais desenvolvidos.

A análise da fecundidade nessa Região poderá revelar especificidades do comportamento reprodutivo de populações socialmente mais vulneráveis, além de oferecer subsídios valiosos para a formulação de políticas públicas mais sensíveis às particularidades regionais e à redução das desigualdades estruturais.

2.2 Transição da Fecundidade no Semiárido brasileiro

O Semiárido brasileiro é uma região marcada por intensas particularidades ambientais e sociais; caracteriza-se por um clima seco, chuvas irregulares, escassez hídrica e solos pobres em matéria orgânica. Abrange parte das Regiões Nordeste e Sudeste do país, ocupando cerca de 15% do território nacional. De acordo com a delimitação de 2024 da Superintendência de Desenvolvimento do Nordeste (SUDENE), o Semiárido compreende 1.477 municípios, como mostrado na Figura 1.

Figura 1: Composição do Semiárido brasileiro, 2025.



Fonte de dados básicos: CONDEL/SUDENE – Resolução nº 176, de 3 de janeiro de 2024.

Historicamente, o Semiárido brasileiro, além das adversidades climáticas, enfrenta desigualdades sociais, refletidas nos baixos indicadores de renda, na limitação dos serviços essenciais e na alta vulnerabilidade social. Dados mais recentes do Censo Demográfico de 2022 indicaram que 90% dos municípios com maior número de residências permanentemente fechadas estão no Semiárido brasileiro (Tiemi, 2023), o que sinalizou um esvaziamento demográfico da Região.

Quanto à transição da fecundidade, nas últimas décadas, o Semiárido brasileiro tem acompanhado a tendência nacional de declínio das taxas de fecundidade, ainda que com particularidades regionais. A transição dos níveis da Taxa de Fecundidade Total (TFT) nessa Região tem sido marcada por uma redução acentuada, refletindo transformações sociais, econômicas e culturais significativas.

De acordo com Silva (2019 *apud* Pereira Júnior 2021), esse declínio pode ser considerado acelerado, especialmente quando se observa a redução dos valores máximos registrados: em 1991, o pico da TFT foi de 7,84 filhos por mulher; em 2000, 6,09; e, em 2010, esse limite não ultrapassava 3,68.

O declínio da fecundidade nessa Região pode ser explicado por diversos fatores. Entre os mais relevantes estão o processo de migração das áreas rurais para os centros urbanos, muitas vezes impulsionado pela busca por melhores condições de vida; a elevação dos níveis de escolaridade, particularmente entre as mulheres, também pode ter contribuído significativamente, ao ampliar o acesso à informação e promover mudanças nas expectativas em relação à maternidade.

A partir do estudo realizado por Pereira Júnior (2021), mesmo com oscilações, entre 2010 e 2018, o declínio da fecundidade entre adolescentes se manteve como uma tendência consolidada no Semiárido brasileiro, em consonância com o observado nas demais regiões do país, aproximando-se inclusive da média nacional.

Ao analisar as Taxas Específicas de Fecundidade entre adolescentes para o ano de 2018, ele constatou que, dentre os 822 municípios do Semiárido brasileiro que registraram nascimentos entre adolescentes de 15 a 19 anos, 617 municípios (75,1%) apresentaram TEF abaixo da média nacional, que naquele ano foi de 0,0598, enquanto 205 municípios (24,9%) estiveram acima desse patamar, indicando uma tendência de redução da fecundidade adolescente na Região, alinhando-se ao padrão observado em outras áreas do país.

Quando a análise foi aprofundada em nível estadual, observou que as Unidades da Federação com maior proporção de municípios acima da TEF adolescente nacional foram Minas Gerais (43%), Sergipe (33,3%), Paraíba (30,8%) e Piauí (29,9%). Em contrapartida, os

estados com maior percentual de municípios abaixo da TEF nacional foram Pernambuco (85,8%), Ceará (84,5%), Rio Grande do Norte (75,5%) e Bahia (75,5%). Esses dados evidenciaram a heterogeneidade dos níveis de fecundidade adolescente no Semiárido brasileiro, tanto no plano regional quanto estadual, refletindo desigualdades estruturais e contextuais que influenciam os comportamentos reprodutivos das adolescentes.

Diante da complexidade e diversidade que marcam o processo de declínio da fecundidade entre adolescentes na região Semiárida do Brasil, tornou-se evidente a importância de aprofundar os estudos sobre as especificidades desse fenômeno. A transição da fecundidade nessa Região, embora acompanhe a tendência nacional, apresenta dinâmicas próprias que refletem desigualdades históricas, culturais e socioeconômicas.

2.3 Qualidade dos Registros Vitais de Nascimento no Brasil

No âmbito da Agenda 2030 da Organização das Nações Unidas (ONU), a meta 3.7 do Objetivo de Desenvolvimento Sustentável (ODS) 3, busca “Assegurar uma vida saudável e promover o bem-estar para todos, em todas as idades”, estabelecendo que até 2030 seja garantido o acesso universal aos serviços de saúde sexual e reprodutiva, incluindo planejamento familiar, informação e educação, além da integração da saúde reprodutiva em estratégias e programas nacionais (IPEA, 2024).

Para monitorar o progresso dessa meta um dos principais indicadores utilizados é a taxa de fecundidade entre adolescentes (indicador 3.7.2), o que requer registros vitais confiáveis para o acompanhamento das desigualdades regionais e o direcionamento de políticas públicas voltadas à juventude. A qualidade dos dados vitais, portanto, é um elemento essencial para o monitoramento desse indicador.

Nessa perspectiva, torna-se imprescindível a realização de estudos avaliativos que verifiquem a confiabilidade desses dados, sobretudo em regiões brasileiras economicamente menos desenvolvidas e em pequenos municípios, onde historicamente se observa menor qualidade nos registros. Essa verificação é crucial para garantir que as políticas públicas voltadas à saúde reprodutiva de adolescentes sejam baseadas em informações consistentes e representativas da realidade local.

No Brasil, os dados sobre nascimentos são provenientes de duas fontes oficiais: as Estatísticas do Registro Civil, produzidas pelo Instituto Brasileiro de Geografia e Estatística (IBGE), a partir dos registros lavrados em cartórios; e o Sistema de Informações sobre Nascidos Vivos (SINASC), coordenado pelo Ministério da Saúde. Desde 1999, ambas as bases utilizam

como fonte primária a Declaração de Nascido Vivo (DNV), emitida em três vias padronizadas. No entanto, a via processada pelo SINASC contempla um conjunto maior de variáveis do que aquela destinada ao Registro Civil, permitindo o detalhamento das informações.

Quanto à completude dos campos da DNV, o Ministério da Saúde estabelece procedimentos padronizados e de cumprimento obrigatório em todo o território nacional. Conforme preconiza o Manual de preenchimento da Declaração de Nascido Vivo (Brasil, 2022, p. 19), a responsabilidade pelo preenchimento da declaração recai sobre: (i) os profissionais de saúde do estabelecimento onde ocorreu o parto ou o primeiro atendimento pós-parto; (ii) profissionais de saúde ou parteiras tradicionais, em casos de partos domiciliares; ou (iii) cartórios autorizados pela Corregedoria de Justiça do Estado.

Entre as orientações, enfatiza-se a importância de evitar campos em branco. Caso a informação não esteja disponível, deve-se assinalar a opção “ignorado”; e, quando não aplicável, utilizar um traço (-). Um exemplo importante é o campo “sexo do recém-nascido”: a marcação como “ignorado” não implica ausência de informação, devendo ser utilizada apenas em casos específicos, como genitália ambígua ou hermafroditismo (Idem, p. 21).

Além disso, recomenda-se privilegiar informações fornecidas diretamente pela puérpera, pelos profissionais de saúde presentes no parto e por registros documentais, como prontuários. Deve-se também respeitar a autodeclaração da raça/cor e considerar que o responsável legal pelo recém-nascido nem sempre será o pai biológico, observando-se as normas legais vigentes.

Outro aspecto relevante refere-se à logística de distribuição e utilização da DNV. Trata-se de formulários impressos, distribuídos pelas Secretarias Municipais de Saúde às unidades de saúde. Após o preenchimento, uma via é encaminhada à Secretaria de Saúde para digitalização e arquivamento; outra é enviada ao cartório para o registro civil de nascimento; e a terceira permanece no estabelecimento de saúde onde ocorreu o parto ou atendimento pós-parto.

Quanto à composição da DNV, o formulário dispõe de 52 campos, o que possibilita ao SINASC fornecer dados detalhados sobre o nascimento, gestação, parto e características maternas, organizados em oito blocos principais: identificação do recém-nascido; local de ocorrência do parto; dados da parturiente; informações do responsável legal; gestação e parto; anomalias congênitas; campos de identificação e preenchimento.

Com relação à cobertura, desde a implantação do SINASC, o Ministério da Saúde tem promovido investimentos significativos para ampliar e qualificar a captação dos registros. Como resultado, “a partir de 1998, observou-se uma melhora na cobertura do SINASC, alcançando pelo menos 90% dos registrados em cartório. Ressalta-se que nesse período, o

SINASC já contava com mais de uma década de implantação em todos os municípios brasileiros” (Oliveira, 2015, p. 637).

No Semiárido brasileiro, uma análise realizada por Simões (2024), referente ao período de 2010 a 2021, apontou que todos os estados que compõem a Região, exceto o Maranhão, apresentavam cobertura superior a 90% desde 2010, de acordo com as estimativas da Secretaria de Vigilância em Saúde, utilizando a metodologia de Busca Ativa. Esses resultados demonstraram avanços significativos na cobertura dos registros de nascimentos, mesmo em áreas historicamente marcadas por fragilidades institucionais.

Sobre a completude das variáveis disponíveis no SINASC, o estudo de Silva e Tomaz (2023), avaliou 19 (dezenove) variáveis selecionadas da Declaração de Nascido Vivo, referente ao período de 2010 a 2019, utilizando como critério a proporção da mediana de preenchimento: igual ou superior a 90% para indicar completude adequada e até 5% para dados assinalados como “ignorados”.

Em 2010, foi observado que 16 (dezesesseis) das 19 (dezenove) variáveis da DNV analisadas apresentaram mediana de preenchimento acima de 90%, indicando nível satisfatório de completude. Entre essas variáveis, destacaram-se: “anomalia congênita”, “índices de APGAR ao 1º e 5º minuto de vida”, “consulta pré-natal”, “duração da gestação”, “escolarização da mãe”, “esfera administrativa onde ocorreu o parto”, “estado civil”, “idade da mãe”, “local de ocorrência”, “peso ao nascer”, “raça/cor”, “sexo”, “tipo de gravidez”, “tipo de parto” e “unidade federativa de ocorrência”.

As demais variáveis, “filhos tidos vivos”, “filhos tidos mortos” e “gestação anterior” apresentaram medianas de 87,5%, 79% e 2,5%, respectivamente, evidenciando completude considerada inadequada.

Quanto aos registros com informações ignoradas, apenas as variáveis “APGAR ao 1º e 5º minuto”, “filhos tidos vivos”, “filhos tidos mortos” e “gestação anterior” ultrapassaram o limite de 5%. No caso de “gestação anterior”, a mediana de proporção de dados ignorados atingiu 97,5%, indicando sério comprometimento da qualidade da informação. As demais variáveis mantiveram proporções inferiores a 5% para dados ignorados em todos os estados, o que indicou preenchimento adequado.

No ano de 2019, todas as 19 (dezenove) variáveis analisadas apresentaram medianas acima de 90%, o que representa melhora geral na completude das informações. Em relação à proporção de dados ignorados em 2019, observou-se que apenas a variável “filhos tidos mortos” ultrapassou o limite de 5%, ainda indicando fragilidade nesse aspecto.

Outro achado relevante para este estudo, refere-se à variável “ocupação da mãe”, cuja mediana de completude passou de 84,26%, em 2010, para 89,82%, em 2019, o que, embora represente avanço, ainda sinalizava necessidade de melhoria no preenchimento.

Portanto, a qualidade dos registros de nascimento no Brasil tem apresentado avanços significativos ao longo dos anos, especialmente com o fortalecimento e consolidação do Sistema de Informações sobre Nascidos Vivos (SINASC). Esse sistema tem se mostrado uma ferramenta eficiente na captação, sistematização e disseminação de informações vitais, contribuindo diretamente para o aprimoramento da vigilância em saúde e a formulação de políticas públicas.

Na região Semiárida do Brasil, ainda que persistam desafios pontuais, os resultados apontaram para uma evolução positiva na cobertura e completude dos dados, refletindo os esforços institucionais na padronização dos procedimentos de registro e no fortalecimento da rede de atenção à saúde materno-infantil.

2.4 Aplicação de Modelos de Regressão Logística Binária

A análise da fecundidade demanda o uso de métodos estatísticos capazes de captar relações complexas entre variáveis sociodemográficas, culturais e territoriais. Nesse contexto, a Regressão Logística Binária tem se consolidado como uma ferramenta metodológica robusta e adequada para investigar desfechos dicotômicos, como, por exemplo, a probabilidade de um município apresentar uma taxa de fecundidade adolescente abaixo ou acima de determinado valor de referência, como a TEF nacional.

Diferente da Regressão Linear clássica, que assume continuidade e normalidade dos dados, a Regressão Logística é voltada para situações em que a variável dependente é categórica binária (com dois possíveis resultados: 0 ou 1). O modelo estima a probabilidade de ocorrência de um evento a partir de um conjunto de variáveis explicativas, transformando essas probabilidades em uma função *logit*, ou seja, o logaritmo da razão de chances (*odds ratio*). Essa abordagem é especialmente útil em estudos populacionais e de saúde pública, em que muitos desfechos relevantes (como presença/ausência de gravidez precoce, evasão escolar, acesso a serviços etc.) assumem estrutura binária.

Nos estudos da fecundidade, o uso da Regressão Logística permite identificar quais características individuais ou contextuais aumentam ou reduzem a chance de determinado comportamento reprodutivo se manifestar em uma população, controlando-se mutuamente os efeitos das variáveis. Essa capacidade de modelar a influência simultânea de múltiplos fatores

é central para compreender fenômenos multifacetados como a fecundidade adolescente, que não decorre apenas de decisões individuais, mas também de fatores estruturais, como escolaridade, estrutura familiar, inserção territorial, acesso a políticas públicas e dinâmicas culturais locais.

A aplicação do Modelo Logístico também permite calcular as razões de chances (*odds ratio*), medidas que facilitam a interpretação dos efeitos das variáveis explicativas sobre o desfecho de interesse. Por exemplo, ao identificar que a escolaridade aumenta significativamente as chances de um município apresentar baixa taxa de fecundidade, ou que o predomínio de mulheres casadas está associado a uma maior probabilidade de altas taxas, o modelo fornece subsídios objetivos para análise dos fatores que podem influenciar o comportamento da fecundidade de um determinado grupo etário ou contexto.

Além disso, a flexibilidade do Modelo Logístico Binário favorece sua aplicação em diferentes escalas de análise, seja individual (microdados) ou ecológica (como no caso de agregar municipais), respeitando a estrutura dos dados disponíveis.

No entanto, embora o Modelo de Regressão Logística possa ser utilizado como alternativa para a análise da fecundidade na adolescência, sua aplicação tende a limitar a inclusão de variáveis potencialmente associadas ao fenômeno nessa faixa etária. Diante dessa limitação, Pereira Júnior (2021) recomendou a incorporação de análises temporais, que permitam avaliar os efeitos da variabilidade de características sociodemográficas e materno-infantis sobre as mudanças observadas na fecundidade adolescente ao longo das últimas décadas.

Nesse contexto, abordagens espaço-temporais poderiam ser particularmente relevantes, sobretudo quando integradas à classificação do IBGE (2017) sobre os espaços urbanos e rurais, podendo oferecer subsídios importantes para a compreensão territorial do fenômeno, contribuindo de forma significativa para o planejamento e a formulação de políticas públicas de saúde voltadas a contextos regionais e locais específicos. (idem)

3 PROCEDIMENTOS METODOLÓGICOS

3.1 Tipologia e Desenvolvimento do Estudo

Esta análise se configura em um estudo ecológico e longitudinal sobre a fecundidade na adolescência no Semiárido brasileiro. A unidade de análise observacional do estudo consiste aos nascidos vivos de mães com idade entre 15 a 19 anos e a população feminina, no mesmo

quinquênio, abrangendo os 1.473¹ municípios que compõem o Semiárido brasileiro, considerando a delimitação estabelecida pela Superintendência do Desenvolvimento do Nordeste (SUDENE)², em 2024.

A abrangência temporal deste estudo compreendeu o período de 2019 a 2023, considerando os registros de nascimentos de mães adolescentes (15 a 19 anos) por município, conforme dados do SINASC. A série histórica de 2019 a 2023 foi contemplada na análise do comportamento da fecundidade na adolescência e somente o ano de 2023 foi utilizado para a modelagem estatística dos níveis de fecundidade, com base em variáveis sociodemográficas.

A modelagem estatística foi realizada exclusivamente com dados do ano de 2023 por três principais razões. Primeiramente, trata-se de um período pós-pandemia de COVID-19, o que permitiu observar os possíveis efeitos da retomada das dinâmicas sociais. Em segundo lugar, 2023 é o último ano disponível no SINASC com dados consolidados e, por fim, a terceira razão, nesse ano apenas quatro municípios do Semiárido brasileiro não registraram nascimentos entre as adolescentes, o que garantiu maior cobertura territorial na análise.

Para o desenvolvimento do estudo, foram seguidas quatro etapas: na primeira, realizou-se o *download* e o pré-processamento dos arquivos de microdados do SINASC, referentes aos anos de 2019 a 2023, agregando as variáveis disponíveis por município do Semiárido brasileiro. Em seguida, na segunda etapa, procedeu-se a uma análise exploratória dos dados com o objetivo de avaliar a cobertura dos registros de nascimento e a completude das variáveis de interesse constantes na Declaração de Nascido Vivo (DNV).

Na terceira etapa, foi analisado o comportamento da Taxa Específica de Fecundidade (TEF) para adolescentes de 15 a 19 anos, ao longo do período de referência. Nessa fase, aplicou-se um fator de correção para os dados inconsistentes. Por fim, na quarta etapa, foi utilizada a modelagem por Regressão Logística Binária para investigar a associação entre as Taxas Específicas de Fecundidade entre adolescentes e as variáveis sociodemográficas selecionadas da DNV.

Cabe ressaltar que por se analisar dados e informações de domínio público, com base no inciso III do parágrafo único do artigo 1º da Resolução Nº 510/16 do Conselho Nacional de

¹ Dos 1.477 municípios que compõem o Semiárido brasileiro, 4 não registraram nascimentos da faixa etária de 15 a 19 anos, em 2023.

² De acordo com as Resoluções CONDEL/SUDENE Nº 150, de 13 de dezembro de 2021 e Nº 176, de 3 de janeiro de 2024.

Saúde, este estudo é isento da obrigatoriedade de registro e avaliação junto ao Comitê de Ética em Pesquisa (CEP) e Comissão Nacional de Ética em Pesquisa (CONEP).

3.2 Fontes dos Dados e Variáveis do Estudo

Os dados utilizados neste estudo foram extraídos de fontes oficiais, disponibilizadas pelo Ministério da Saúde, por meio do Sistema de Informações sobre Nascidos Vivos (SINASC) do Departamento de Informática do SUS (DATASUS), do Instituto Brasileiro de Geografia e Estatística (IBGE) e da Coordenação-Geral de Informações e Análises Epidemiológicas vinculada à Secretaria de Vigilância em Saúde do Ministério da Saúde. O acesso às informações ocorreu entre dezembro de 2024 e fevereiro de 2025.

Com o objetivo de não apenas contabilizar os nascidos vivos por município, mas também analisar um conjunto de variáveis constantes na DNV, optou-se pela geração de uma base de microdados a partir dos dados disponíveis no SINASC. A descrição dos códigos, nomes das variáveis e suas respectivas categorias foram obtidas na documentação oficial do SINASC, disponível no site do DATASUS. As variáveis relacionadas aos nascimentos consideradas neste estudo estão organizadas conforme os blocos da DNV, sendo elas:

- Bloco I – Identificação do recém-nascido (RN): sexo, peso ao nascer, índice de APGAR 1º e 5º minutos;
- Bloco II – Local de ocorrência: local de ocorrência;
- Bloco III – Parturiente: idade, situação conjugal, escolaridade, município de residência, raça/cor;
- Bloco V – Gestação e parto: duração gestacional, tipo de gravidez, tipo de parto, número de consultas pré-natal, quantidade de filhos tidos vivos, quantidade de filhos tidos mortos;
- Bloco VI – Anomalias Congênicas: anomalias detectadas.

Contudo, para a modelagem estatística dos fatores que podem influenciar a fecundidade na adolescência, foram selecionadas as variáveis do Bloco III e V. Embora o declínio das taxas de fecundidade seja uma realidade no Brasil, em 2000 já era possível observar que havia diferenças no comportamento reprodutivo das mulheres ao se analisar fatores sociodemográficos (Berquó, Cavenaghi, 2004).

Para aplicação da modelagem estatística, foram feitas adaptações com a recategorização de algumas variáveis (Quadro 1), observando-se as proporções das observações em cada categoria e assegurando a representatividade dos grupos formados, bem como a viabilidade da análise em termos estatísticos.

Quadro 1: Identificação da variável sociodemográfica da Declaração de Nascido Vivo e respectivas categorias.

Variável da Declaração de Nascido Vivo	Categorias
Escolarização mãe	Nenhuma ou Fundamental 1 Fundamental 2 Médio Superior (completo ou não)
Estado civil	Casada Solteira Separada/divorciada e viúva
Município de residência	Rural remoto Rural adjacente Intermediário remoto Intermediário adjacente Urbano
Quantidade de filhos tidos vivos	Não tem outro filho Sim tem outro filho
Raça/cor da mãe	Branca Indígena Negra

Fonte básica dos dados: Sistema de Informações sobre Nascidos Vivos (SINASC), 2025.

A variável “Escolaridade da mãe” foi recategorizada em quatro grupos: Fundamental 1, englobando adolescentes sem escolaridade ou que concluíram apenas os anos iniciais da Educação Fundamental; Fundamental 2, correspondente aos anos finais dessa etapa; Ensino Médio; e Ensino Superior — independentemente da conclusão desse nível.

No caso da variável “Estado Civil”, as categorias separada, divorciada e viúva foram agrupadas em uma única classe, em razão do baixo número de registros individuais. A variável “Município de residência” foi categorizada conforme a tipologia dos espaços rurais e urbanos do IBGE (2017), com o objetivo de superar uma abordagem dicotômica entre o rural e o urbano.

Essa decisão se baseou na compreensão de que “a localização ou a acessibilidade aos centros urbanos mais estruturados também se apresenta como um elemento importante, já que a relação das cidades menores com os centros urbanos de maior hierarquia reflete diretamente no modo de vida e na configuração do espaço” (Medeiros, 2023, p. 52).

A variável “Quantidade de filhos tidos vivos” foi simplificada em duas categorias: com ou sem filhos de gestações anteriores. Essa categorização permite captar, de forma objetiva, a

presença de experiência reprodutiva anterior, o que pode influenciar comportamentos relacionados à fecundidade atual, especialmente em contextos de vulnerabilidade socioeconômica.

E, por fim, a variável “Raça/Cor da mãe” foi recategorizada em três grupos: branca, indígena e negra. As observações classificadas como amarelas foram incorporadas ao grupo branco, enquanto os dados de mães autodeclaradas pretas e pardas foram agrupados sob a categoria negra.

Quanto à contagem da população feminina na faixa etária de 15 a 19 anos, desagregada por idade simples e por ano (2019 a 2023), foi obtida a partir do Estudo de Estimativas Populacionais por Município, Idade e Sexo 2000 – 2024 (Brasil, 2025).

Para o *download* e o pré-processamento das informações individualizadas dos nascimentos, utilizou-se o pacote Microdatasus do *software R*. Esse pacote permite o acesso facilitado aos dados de diversos sistemas de informação em saúde, com funcionalidades para limpeza e rotulagem automática de variáveis categóricas (Saldanha, 2019).

A delimitação geográfica dos municípios pertencentes ao Semiárido brasileiro seguiu os critérios estabelecidos pela SUDENE em 2024. Para tanto, foi utilizado o pacote Geobr (Pereira *et. al*, 2019), o qual fornece dados geoespaciais oficiais do Brasil em diferentes escalas e períodos, permitindo a agregação precisa dos dados por município.

3.3 Análise da qualidade dos dados de nascimentos

Para dar início ao estudo sobre a fecundidade no Semiárido brasileiro, partiu-se da avaliação da qualidade dos dados oficiais relacionados aos nascimentos, um dos principais eventos vitais que, quando bem monitorado, pode fornecer informações relevantes sobre os determinantes sociais e econômicos que impactam as condições de vida e saúde das mulheres e de seus filhos.

A análise da qualidade dos dados dos municípios do Semiárido brasileiro foi realizada com base em duas dimensões: cobertura e completude. A cobertura refere-se à proporção de nascimentos efetivamente registrados no Sistema de Informações sobre Nascidos Vivos (SINASC). Para estimar a cobertura dos nascimentos entre adolescentes de 15 a 19 anos, utilizou-se a estatística $C = \frac{1}{k}$, em que k representa o fator de correção dos nascimentos, conforme descrito na Seção 3.5.

Além disso, adotou-se, para fins comparativos, os dados divulgados pela Secretaria de Vigilância em Saúde, por meio da Coordenação-Geral de Informações e Análises Epidemiológicas, em que a cobertura estadual foi utilizada como uma variável proxy representativa dos municípios pertencentes ao respectivo estado na área do Semiárido brasileiro.

Já para a completude, adotou-se como referência os critérios definidos no estudo de Oliveira (2015), que estabeleceu como padrão de qualidade uma mediana de preenchimento igual ou superior a 90% e um limite máximo de 5% de registros com informações ignoradas. Esses parâmetros possibilitaram aferir se os dados disponíveis têm boa qualidade. Para essa análise, observou-se os dados de nascimentos entre todas as mulheres em idade reprodutiva, 15 a 49 anos de idade, como também se fez uma análise isolando os nascimentos entre as adolescentes de 15 a 19 anos de idade.

3.4 Análise da fecundidade entre adolescentes

Com o objetivo de analisar a fecundidade entre adolescentes da região Semiárida brasileira, foi construída uma base de dados com as Taxas Específicas de Fecundidade (TEF) para o grupo etário de 15 a 19 anos, desagregadas por municípios e Unidades Federativas da Região, para o período de 2019 a 2023. Essas taxas permitiram avaliar a intensidade do fenômeno da fecundidade adolescente no território e possibilitaram comparações com os níveis nacionais e das demais Grandes Regiões brasileiras.

Para fins comparativos, foram utilizadas também as TEFs oficiais disponibilizadas pelo IBGE, referentes ao Brasil e as Grandes Regiões. Dessa forma, foi possível contextualizar os dados do Semiárido brasileiro em relação ao panorama nacional.

A Taxa Específica de Fecundidade por grupo etário é definida como o quociente entre o número de nascidos vivos de mães pertencentes a um determinado grupo etário, em um dado ano, e o número estimado de mulheres deste mesmo grupo etário na população geral, no mesmo período. A fórmula é expressa como:

$${}_nTEF_x = \frac{{}_nNVobs_x}{{}_nM_x}$$

Em que:

- ${}_nTEF_x$ = Taxa Específica de Fecundidade na faixa etária quinquenal $x, x+n$;
- ${}_nNVobs_x$ = Número de nascimentos observado na faixa etária quinquenal $x, x+n$;
- ${}_nM_x$ = Número de mulheres observado na faixa etária quinquenal $x, x+n$.

A fecundidade geralmente se refere à população feminina em idade reprodutiva, compreendida entre 15 a 49 anos, e costuma ser analisada por meio de grupos etários quinquenais. Neste estudo, o foco foi direcionado especificamente ao grupo etário do ciclo reprodutivo (15 – 19 anos), considerando sua relevância social e epidemiológica no contexto das políticas públicas de saúde sexual e reprodutiva.

Além da análise das taxas de fecundidade, realizou-se uma análise exploratória das variáveis sociodemográficas selecionadas para a modelagem. Após a avaliação da qualidade dessas variáveis e a exploração inicial dos dados, utilizou-se a Regressão Logística Binária como método estatístico para investigar os fatores sociodemográficos associados aos níveis de fecundidade entre adolescentes no Semiárido brasileiro.

3.5 Aplicação das técnicas indiretas de correção das Taxas Específicas de Fecundidade

Diante da possibilidade de subestimação das Taxas Específicas de Fecundidade em determinados municípios da região Semiárida do Brasil, adotou-se uma técnica indireta de correção das taxas baseada na aplicação de um fator multiplicativo proposto por Pereira Júnior (2021), com o objetivo de ajustar os valores do ano 2023, e garantir maior comparabilidade entre os municípios analisados.

O fator de correção, denominado f , foi calculado a partir da razão entre a média das TEF municipais do estado e a TEF global do estado no mesmo período (2019 – 2023). Essa abordagem partiu do pressuposto de que, mesmo diante de possíveis falhas na cobertura ou sub-registro, a proporção relativa entre os municípios dentro de cada estado tende a ser preservada.

Os passos metodológicos adotados para obtenção do fator de correção foram os seguintes:

1. Cálculo das TEF_adolescente³ acumuladas por município: para cada município do Semiárido brasileiro, somou-se o total de nascidos vivos de mães adolescentes (15 a 19 anos) no período de 2019 a 2023, e esse total foi dividido pela soma da população feminina da mesma faixa etária e período. Obteve-se, assim, a TEF acumulada para cada município;
2. Cálculo da TEF_adolescente média dos municípios por estado (denominada de TEF_adolescente_média): em seguida, foi calculada a média das TEFs acumuladas dos municípios pertencentes a cada estado, refletindo o comportamento médio da fecundidade adolescente nos municípios daquele estado;

³ Neste estudo se considerou como TEF_adolescente, a TEF relacionada ao grupo etário de 15 a 19 anos.

3. Cálculo da TEF_adolescente para o estado (denominada de TEF_adolescente estado): foi também calculada a TEF global de cada estado para o mesmo período (2019–2023), considerando o total de nascimentos de mães adolescentes dividido pela população feminina correspondente;
4. Determinação do fator de correção f : com esses valores, o fator foi definido como apresentado a seguir.

$$f = \frac{TEF_{adolescente\ média}}{TEF_{adolescente\ estado}}$$

5. Aplicação do fator de correção: para o ano de 2023, a TEF de cada município foi multiplicada pelo fator f correspondente ao seu estado, gerando uma estimativa ajustada da fecundidade adolescente.

Ao final do processo de cálculo, foi organizada uma tabela-síntese contendo, para cada Unidade Federativa, os seguintes elementos:

Tabela 1: Fator de correção calculados para os estados do Semiárido brasileiro, considerando o período de 2019 – 2023.

Estado	TEF_adolescente Média	TEF_adolescente estado	Fator de correção TEF_adolescente
Alagoas	0,0661	0,0621	1,0644
Bahia	0,0469	0,0490	0,9571
Ceará	0,0473	0,0467	1,0128
Espírito Santo	0,0433	0,0436	0,9931
Maranhão	0,0620	0,0658	0,9422
Minas Gerais	0,0452	0,0427	1,0585
Paraíba	0,0508	0,0493	1,0304
Pernambuco	0,0534	0,0529	1,0094
Piauí	0,0564	0,0513	1,0994
Rio Grande do Norte	0,0481	0,0451	1,0665
Sergipe	0,0567	0,0552	1,0271

Fonte básica dos dados: Sistema de Informações sobre Nascidos Vivos (SINASC), 2025.

Essa técnica apresentou como principal vantagem a preservação da proporção relativa entre os municípios, ou seja, mantém a distribuição territorial dos níveis de fecundidade, ao mesmo tempo em que minimiza possíveis distorções decorrentes de registros incompletos. Além disso, permite comparações mais equitativas e realistas entre os municípios e os estados

de uma região, especialmente em análises que envolvem desigualdades regionais na qualidade dos dados.

3.6 Modelagem Estatística dos Dados

Para compreender os fatores sociodemográficos associados aos níveis de fecundidade na adolescência no Semiárido brasileiro, optou-se pela aplicação do Modelo de Regressão Logística Binária, uma vez que se trata de uma abordagem estatística adequada para situações em que a variável dependente é dicotômica, ou seja, assume apenas dois valores: presença (1) ou ausência (0) do evento de interesse.

Neste estudo, a variável resposta foi definida com base na Taxa Específica de Fecundidade (TEF) entre adolescentes de 15 a 19 anos. Os municípios cuja TEF se encontra abaixo da TEF nacional foram codificados como 1, enquanto os demais foram codificados como 0.

A escolha desse ponto de corte se justificou pelo fato de que, embora a fecundidade adolescente venha apresentando tendência de queda no Semiárido do Brasil, ainda persistem desigualdades importantes entre os municípios. Utilizar a TEF nacional como referência permite compreender os fatores que contribuem para essas diferenças.

3.6.1 Especificações do Modelo de Regressão Logística

Como a variável dependente (Y_i) é uma variável aleatória de Bernoulli, assumindo probabilidade 1 quando Y_i possui uma determinada característica, e 0, quando não a possui, tem-se a seguinte distribuição de probabilidade (Montgomery; Peck; Viking, 2001):

$$Y_i = \begin{cases} 1, \text{ com probabilidade } P(Y_i = 1) = \pi_i \\ 0, \text{ com probabilidade } P(Y_i = 0) = 1 - \pi_i \end{cases}$$

Em que,

π_i representa a probabilidade de ocorrência do evento associado a Y_i ocorrer.

Ao incorporar um conjunto de variáveis explicativas x_1, x_2, \dots, x_p , a equação estrutural do Modelo Logístico pode ser expressa como:

$$\text{Logit} \left(\frac{\pi_i}{1 - \pi_i} \right) = \beta_0 + \beta_1 x_{i_1} \dots \beta_p x_{i_p} + \varepsilon_i = \eta_i + \varepsilon_i$$

Em que,

β_0 : é o intercepto;

β_i : são os coeficientes associados às variáveis independentes x_{i_p} ;

η_i : é o preditor linear;

ε_i : é o erro aleatório.

A estimação dos parâmetros β é realizada por meio do método da máxima verossimilhança, dado que a Regressão Logística pertence à família dos Modelos Lineares Generalizados (GLMs), cuja função de ligação é a *logit* e a distribuição da variável resposta é Bernoulli.

Sendo assim, considerando um Modelo de Regressão Logística simples com $Y_i \sim Ber(\pi_i)$ a função de probabilidade é:

$$f(y_i, \pi_i) = \pi_i^{y_i} (1 - \pi_i)^{1-y_i}$$

E a função de verossimilhança é dada por:

$$L(\beta) = \prod_{j=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i}, \beta \in R^{(2)}$$

Aplicando o logaritmo, obtém-se a log-verossimilhança:

$$\begin{aligned} L(\beta) &= \ln[L(\beta)] \\ &= \ln \left[\prod_{j=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \right] \\ &= \sum_{i=1}^n [y_i \ln(\pi_i) + (1 - y_i) \ln(1 - \pi_i)] \\ &= \sum_{i=1}^n [y_i \ln(\pi_i) + \ln(1 - \pi_i) - y_i \ln(1 - \pi_i)] \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n \left[y_i \ln \left(\frac{\pi_i}{1 - \pi_i} \right) + \ln(1 - \pi_i) \right] \\
&= \sum_{i=1}^n \left[y_i (\beta_0 + \beta_1 x_1) + \ln \left(\frac{1}{1 + \exp(\beta_0 + \beta_1 x_1)} \right) \right] \\
&= \sum_{i=1}^n [y_i (\beta_0 + \beta_1 x_1) - \ln(1 + \exp(\beta_0 + \beta_1 x_1))]
\end{aligned}$$

Para encontrar os valores de β que maximizam $\ell\beta$, deriva-se em relação a cada parâmetro. As derivadas parciais resultam em:

$$\begin{aligned}
\frac{\partial \ell\beta}{\partial (\beta_0)} &= \sum_{i=1}^n \left[y_i - \frac{1}{1 + \exp(\beta_0 + \beta_1 x_1)} \exp(\beta_0 + \beta_1 x_1) \right] \quad (i) \\
\frac{\partial \ell\beta}{\partial (\beta_1)} &= \sum_{i=1}^n \left[y_i x_i - x_i \frac{1}{1 + \exp(\beta_0 + \beta_1 x_1)} \exp(\beta_0 + \beta_1 x_1) \right] \quad (ii)
\end{aligned}$$

Igualando as expressões (i) e (ii), obtém-se um sistema de equações que não possui solução analítica fechada devido à natureza não linear em β_0 e β_1 .

$$\begin{cases} \sum_{i=1}^n = (y_i x_1) = 0 \\ \sum_{i=1}^n = x_i (y_i x_i) = 0 \end{cases}$$

Por esse motivo, a obtenção das estimativas exige o uso de métodos numéricos iterativos, como o algoritmo de *Newton-Raphson* ou métodos baseados em gradiente, amplamente disponíveis em diversos programas estatísticos e linguagens computacionais.

3.6.2 R² de Nagelkerke

O R² de *Nagelkerke*, também conhecido como pseudo-R² de *Nagelkerke*, é uma medida de qualidade do ajuste para Modelos de Regressão Logística. Trata-se de uma adaptação do coeficiente de determinação utilizado na Regressão Linear, ajustada para lidar com a natureza da Regressão Logística.

Essa métrica fornece uma estimativa da proporção da variabilidade da variável dependente explicada pelo modelo, com base nas probabilidades previstas para o evento de interesse. Seus valores variam entre 0 e 1, em que 0 indica ausência de poder explicativo e 1 representa um ajuste perfeito do modelo aos dados.

O cálculo do R² de *Nagelkerke* se baseia na comparação entre a log-verossimilhança do modelo ajustado e a do modelo nulo (ou seja, um modelo sem preditores), conforme proposto por *Cox e Snell* (1989), sendo posteriormente ajustado por *Nagelkerke* para atingir o limite superior igual a 1.

3.6.3 Análise de Diagnóstico

A análise de diagnóstico é uma etapa fundamental no processo de modelagem, pois permite avaliar a adequação do modelo ajustado aos dados observados, identificar observações influentes e verificar possíveis violações dos pressupostos do modelo. Na Regressão Logística, embora os pressupostos sejam menos restritivos do que na Regressão Linear, ainda é essencial garantir a validade dos resultados obtidos. Para este estudo foram utilizadas diferentes métricas e representações gráficas para a etapa de diagnóstico:

3.6.3.1. Análise de Resíduos

A análise dos resíduos é uma etapa fundamental para verificar a adequação do modelo e identificar possíveis discrepâncias entre os valores observados e os valores ajustados (Montgomery; Peck; Viking, 2001). Para isso, foram utilizados os resíduos padronizados, os quais permitem detectar observações discrepantes em relação ao comportamento esperado.

Nessa etapa, foi utilizado o Teste de *Durbin Watson*, o qual permite verificar a presença de autocorrelação nos resíduos de um modelo de regressão, especialmente a autocorrelação de primeira ordem, ou seja, a correlação entre resíduos consecutivos. O valor da estatística de Durbin-Watson (D), varia entre 0 e 4, indicando ausência ou presença de autocorrelação:

- $D \approx 2$: indica ausência de autocorrelação;
- $D < 2$: sugere autocorrelação positiva;
- $D > 2$: sugere autocorrelação negativa.

A estatística de Durbin-Watson parte da suposição de que os erros no modelo de regressão são gerados por um processo autorregressivo, sendo expressa como:

$$\varepsilon_i = \rho \varepsilon_{i-1} + \alpha_i$$

Generalizando, temos a fórmula geral do Teste:

$$DW = \frac{\sum_{i=2}^n (\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=3}^n \varepsilon_i^2}$$

3.6.3.2 Bondade de Ajuste

Para avaliar a qualidade do ajuste global do modelo, foi utilizada a estatística do desvio (*deviance*), ajustada pela dispersão. A comparação entre esse desvio e o quantil da distribuição Qui-quadrado permite verificar se há indícios de má adequação do modelo. Se a condição desvio < Qui-quadrado for satisfeita, considera-se que o modelo apresenta bom ajuste, ou seja, a variação explicada é compatível com a variabilidade esperada sob o modelo ajustado.

3.6.3.3 Distância de Cook

A Distância de *Cook* é utilizada para identificar observações que possuem grande influência sobre os coeficientes estimados do modelo, medindo o quanto os resultados da regressão podem mudar quando uma observação específica é removida do conjunto de dados (Montgomery; Peck; Viking, 2001).

Para este estudo, a análise foi realizada graficamente e um ponto foi considerado influente se sua Distância de *Cook* superou o valor de corte adotado. Sua fórmula é definida como:

$$DC_j = r_j^2 \frac{h_j}{(1 - h_j)}$$

Em que,

r_j^2 é o resíduo estandardizado de Pearson e h_j é o valor de alavancagem da observação j .

3.6.3.4 Análise dos Pontos Influentes (*Outliers*)

Os *Outliers* devem ser analisados com cautela, a fim de verificar possíveis justificativas para seu comportamento atípico. Em alguns casos, esses valores extremos podem ser resultado de falhas explicáveis, como erros de medição, registros incorretos ou problemas nos instrumentos utilizados para coleta dos dados. Nessas situações, é recomendável que o valor seja corrigido ou removido do conjunto de dados, caso haja possibilidade e justificativa para isso. Contudo, deve-se ressaltar que a exclusão de *outliers* só deve ocorrer quando houver evidências sólidas e externas ao modelo estatístico de que se trata, de fato, de um valor inválido (Montgomery; Peck; Viking, 2001).

3.6.4 Análise do Desempenho do Modelo de Regressão Logística Binária

A avaliação do desempenho de um Modelo de Regressão Logística é fundamental para verificar sua capacidade de classificar corretamente os casos observados. Entre as principais ferramentas utilizadas nesse processo, destacam-se a matriz de confusão e a curva ROC (*Receiver Operating Characteristic*).

3.6.4.1 Matriz de confusão

A matriz de confusão é uma tabela que compara os valores previstos pelo modelo com os valores reais da variável dependente. Ela permite identificar a quantidade de verdadeiros positivos, falsos positivos, verdadeiros negativos e falsos negativos.

Com base nessa matriz, podem ser calculadas métricas importantes, como:

- Acurácia: proporção de previsões corretas sobre o total de casos;
- Sensibilidade (ou *recall*): capacidade do modelo de identificar corretamente os casos positivos;
- Especificidade: capacidade de identificar corretamente os casos negativos;
- Precisão: proporção de verdadeiros positivos entre todas as previsões positivas.

Essas métricas oferecem uma visão detalhada do comportamento do modelo em diferentes aspectos da classificação. A matriz pode ser definida como descrita a seguir na Tabela 2:

Tabela 2: Exemplo de Matriz de Confusão

Dado observado	Valor predito	
	Sim	Não
Sim	Verdadeiro Positivo (VP)	Falso Negativo (FN)
Não	Falso Positivo (FP)	Verdadeiro Negativo (VN)

Em que temos,

- Verdadeiro Positivo (VP): O modelo classificou corretamente uma instância positiva como positiva. Ou seja, trata-se de um caso em que a condição realmente está presente e foi corretamente identificada pelo modelo;
- Falso Negativo (FN): O modelo classificou incorretamente uma instância positiva como negativa. Ou seja, a condição está presente, mas o modelo falhou em identificá-la;
- Falso Positivo (FP): O modelo classificou incorretamente uma instância negativa como positiva. Ou seja, a condição não está presente, mas o modelo indicou que sim;
- Verdadeiro Negativo (VN): O modelo classificou corretamente uma instância negativa como negativa. Ou seja, a condição realmente não está presente e o modelo acertou ao não identificá-la.

3.6.4.2 Curva ROC

A curva ROC (*Receiver Operating Characteristic*) é uma ferramenta gráfica amplamente utilizada para avaliar o poder discriminativo de modelos classificatórios, como a Regressão Logística. Ela é construída com base na relação entre a taxa de verdadeiros positivos (sensibilidade) e a taxa de falsos positivos (1 - especificidade), considerando diferentes limiares de classificação.

O desempenho do modelo é resumido por meio da área sob a curva ROC (*AUC – Area Under the Curve*), que varia entre 0 e 1. Quanto mais próxima de 1 for a *AUC*, melhor será a capacidade do modelo em distinguir entre as classes da variável resposta. Em contrapartida, valores próximos de 0,5 indicam que o modelo não possui capacidade discriminativa, ou seja, equivale a uma classificação aleatória (Fawcet, 2006).

A interpretação dos valores de *AUC* pode ser feita da seguinte forma:

- 0,5 – 0,6 → nenhum poder discriminativo (classificação aleatória);
- 0,7 – 0,8 → desempenho satisfatório (bom);
- 0,8 – 0,9 → desempenho elevado (muito bom);
- acima de 0,9 → desempenho excelente.

Dessa forma, a curva *ROC*, associada à *AUC*, constitui uma importante métrica para comparar e validar a capacidade preditiva de modelos classificatórios.

3.6.4.3 Razão de Chances (*Odds Ratio*)

A razão de chances, ou *odds ratio* (OR), é uma medida amplamente utilizada para interpretar os coeficientes estimados em Modelos de Regressão Logística. Ela representa a razão entre as chances de ocorrência de um evento em um grupo em comparação com outro grupo de referência, considerando o efeito de uma variável explicativa.

Na Regressão Logística, os coeficientes β_j são expressos na escala do *logit* (logaritmo da razão de chances). Assim, ao aplicar a exponenciação ao coeficiente estimado, obtém-se a razão de chances associada a uma unidade de aumento na variável x_j :

$$OR_j = e^{\beta_j}$$

Essa medida pode ser interpretada da seguinte forma:

- $OR > 1$: indica que o aumento de uma unidade em x_j está associado a um aumento nas chances do evento ocorrer;
- $OR < 1$: indica uma associação negativa, ou seja, o aumento de x_j está associado à redução nas chances do evento;
- $OR = 1$: sugere ausência de efeito, ou seja, x_j não influencia a ocorrência do evento.

Sendo assim, considerando o Modelo Logístico com apenas uma variável independente x . A razão de chances ($RC(x)$) é expressa por:

$$RC(x) = \frac{\pi(x)}{1 - \pi(x)} = e^{\beta_0} e^{\beta_1 x}$$

Ao incrementar uma unidade em x , obtém-se a nova razão de chances:

$$RC(x + 1) = \frac{\pi(x + 1)}{1 - \pi(x + 1)} = e^{\beta_0} e^{\beta_1(x+1)} = e^{\beta_0} e^{\beta_1 x} e^{\beta_1}$$

Dessa forma, a razão de chances entre os dois valores de x , ou seja, a odds ratio (OR) associada a um aumento unitário em x é dada por:

$$OR(X) = \text{logit} \frac{RC(x + 1)}{RC(x)} = \frac{e^{\beta_0} e^{\beta_1 x} e^{\beta_1}}{e^{\beta_0} e^{\beta_1 x}} = e^{\beta_1}$$

Aplicando o logaritmo natural em ambos os lados:

$$\log \frac{RC(x + 1)}{RC(x)} = \text{logit} \frac{\pi(x + 1)}{1 - \pi(x + 1)} - \text{logit} \left(\frac{\pi(x)}{1 - \pi(x)} \right) = \beta_1$$

Portanto, o coeficiente β_1 representa a variação no *logit* da probabilidade do evento de interesse para cada unidade adicional em x . Em termos práticos, a *odds ratio* e^{β_1} e β_1 permite quantificar o acréscimo (quando $OR > 1$) ou decréscimo (quando $OR < 1$) nas chances do evento ocorrer a cada aumento unitário na variável x .

4 RESULTADOS E DISCUSSÕES

4.1 Cobertura e completude dos registros de nascimentos

A análise da cobertura e da completude dos registros de nascimentos nos estados que compõem a região do Semiárido brasileiro revelou um padrão elevado de qualidade nas informações declaradas. Observou-se que a proporção de preenchimento das variáveis selecionadas se manteve, no período estudado, 2019 a 2023, acima da mediana de 90%, com baixos percentuais de informações ignoradas (inferiores a 5%).

Segundo Paes (2018), a manutenção de uma cobertura dos registros vitais igual ou superior a 90% por um período mínimo de cinco anos consecutivos pode ser adotada como um critério indicador de boa qualidade dos dados. No contexto deste estudo, esse parâmetro foi plenamente atendido, o que evidenciou o preenchimento adequado dos campos essenciais das Declarações de Nascido Vivo, o que reflete não apenas a efetividade da captação dos registros

de nascimentos, como também a qualidade das informações produzidas e disponibilizadas pelo Sistema de Informações sobre Nascidos Vivos (SINASC).

Ao se analisar o percentual de cobertura por estado (Tabela 3), tendo como parâmetro os dados divulgados pela Secretaria de Vigilância em Saúde, como já mencionado, os resultados deste estudo corroboraram com achados anteriores da literatura quanto à confiabilidade e à expansão da cobertura do SINASC, como já discutido na seção de Fundamentação Teórica deste trabalho.

A leitura da Tabela 3 permitiu constatar que todos os estados da região mantiveram cobertura superior a 90% ao longo do período em análise. Destaca-se, ainda, que mesmo nos anos de maior impacto da pandemia de COVID-19, 2020 e 2021, os níveis de cobertura permaneceram estáveis em comparação com 2019.

Tabela 3: Percentual da cobertura de Nascidos Vivos (NV), entre mulheres de 10 a 49 anos de idade, dos Estados do Semiárido brasileiro, 2018 a 2022⁴.

Estados	2018	2019	2020	2021	2022
Alagoas	93,3	93,6	96,8	98,2	98,7
Bahia	95,8	96,2	96,9	98,1	98,7
Ceará	95,1	95,5	96,2	98,3	99,0
Espírito Santo	100	100	100	100	100
Maranhão	91,0	91,4	94,8	97,9	98,9
Minas Gerais	97,2	97,6	98,0	99,1	99,3
Paraíba	98,0	98,2	97,7	99,6	99,7
Pernambuco	98,3	98,7	97,0	99,5	99,6
Piauí	96,9	97,1	96,2	98,1	98,7
Rio Grande do Norte	97,0	97,3	97,1	99,1	99,3
Sergipe	96,5	96,8	96,2	98,4	99,1

Fonte de dados básicos: Secretaria de Vigilância em Saúde - Coordenação-Geral de Informações e Análises Epidemiológicas, 2025.

Também se observou que o Estado do Espírito Santo apresentou cobertura de 100% em todos os anos analisados, o que se alinha a um padrão historicamente observado nos estados da Região Sudeste do Brasil. A longa trajetória de investimentos em saúde e tecnologia da informação contribui para a consolidação de bases de dados mais fidedignas, como é o caso do

⁴ Os percentuais de cobertura apresentados abrangem até o ano de 2022, conforme a última atualização disponibilizada pela Secretaria de Vigilância em Saúde, por meio da Coordenação-Geral de Informações e Análises Epidemiológicas. Optou-se, portanto, por destacar os percentuais referentes ao ano de 2018, com o objetivo de evidenciar a manutenção da cobertura acima de 90% por cinco anos consecutivos.

SINASC. Nesse sentido, o desempenho do Espírito Santo pode ser compreendido como resultado de um contexto regional historicamente mais favorável à organização e à efetividade dos sistemas de vigilância em saúde.

Ao analisar a cobertura dos registros de nascimentos entre adolescentes de 15 a 19 anos no período de 2019 a 2023, utilizando dados corrigidos conforme descrito na Seção 3.5, e compará-la com a cobertura média dos registros a partir dos dados disponibilizados pela metodologia Busca Ativa, para o período de 2018 a 2022, observou-se uma discrepância em alguns estados do Semiárido brasileiro. Em seis estados, Alagoas, Minas Gerais, Paraíba, Piauí, Rio Grande do Norte e Sergipe, a cobertura dos registros de nascimentos entre adolescentes foi inferior à cobertura média observada entre mulheres em idade reprodutiva geral.

Essa diferença sugere a existência de possíveis sub-registros específicos na faixa etária de 15 a 19 anos nos estados em que a cobertura foi inferior, o que pode indicar subnotificações associadas à maternidade precoce. Em contraste, os estados da Bahia, Maranhão, Pernambuco e Ceará apresentaram cobertura superior à média observada para o grupo de mulheres de 15 a 49 anos no período de 2018 a 2022, enquanto o Espírito Santo manteve a mesma proporção entre os dois grupos etários, evidenciando maior uniformidade na cobertura dos registros de nascimentos. No entanto, é esperado que a cobertura da metodologia da Busca Ativa seja maior porque inclui dados das capitais dos estados, os quais, em geral, tem mais qualidade.

Tabela 4: Comparativo do percentual de cobertura de Nascidos Vivos entre adolescentes (15 a 19 anos, 2019-2023) e mulheres em idade reprodutiva (15 a 49 anos, 2018-2022) nos Estados do Semiárido brasileiro.

Estados	Cobertura NV (15–19 anos, 2019–2023) ⁵	Cobertura NV (10-49 anos, 2018 a 2022) ⁶
Alagoas	93,94	96,12
Bahia	100,0	97,14
Ceará	98,73	96,81
Espírito Santo	100,0	100,0
Maranhão	100,0	94,80
Minhas Gerais	94,47	98,23
Paraíba	97,04	98,64
Pernambuco	99,06	98,62
Piauí	90,96	97,39
Rio Grande do Norte	93,77	97,95
Sergipe	97,35	97,39

Fonte de dados básicos: Sistema de Informações sobre Nascidos Vivos (SINASC), 2025.

⁵ Cobertura calculada neste estudo após a aplicação do fator de correção dos dados de nascimentos.

⁶ Cobertura calculada pela Secretaria de Vigilância em Saúde através da metodologia Busca Ativa.

Sobre a completude dos campos da DNV, optou-se, inicialmente, fazer uma análise do número absoluto de registros incompletos (valores vazios) e o percentual de completude de 17 variáveis relacionadas às mães e aos recém-nascidos, no período de 2019 a 2023, considerando todos os grupos etários. Posteriormente, observou-se apenas as variáveis de interesse deste estudo para a faixa etária de 15 a 19 anos de idade.

De modo geral, na Tabela 5, observou-se um elevado grau de completude das variáveis analisadas, com 13 (treze) delas atingindo percentuais iguais ou superiores a 98% ao longo de todos os anos. No entanto, algumas discrepâncias merecem destaque, a saber: enquanto variáveis como “Consulta pré-natal”, “Local de ocorrência”, “Peso ao nascer”, “Sexo”, “Tipo de parto”, “Tipo de gravidez”, “Idade da mãe” e “Município de residência” manteve percentuais de completude aproximadamente ou iguais a 100%, outras, como “Quantidade de filhos tidos mortos”, “Quantidade de filhos tidos vivos” e “Raça/cor da mãe”, registraram os níveis mais baixos de completude, variando entre 92% e 97%.

As variáveis “Anomalia congênita” e os índices de APGAR no 1º e 5º minutos de vida também concentraram números relevantes de valores ausentes, com completude oscilando entre 95% (em 2019) e 98% (em 2023). Apesar da melhora gradual, a recorrência de registros incompletos pode indicar dificuldades no preenchimento desses campos, possivelmente relacionadas à ausência de diagnóstico no momento do nascimento ou à baixa valorização dessas informações nos fluxos de registro.

As variáveis “Escolarização da mãe”, “Semanas de gestação” e “Estado Civil” também demonstraram leve oscilação ao longo dos anos, com percentuais de completude entre 96% e 98%. Embora esses valores permaneçam em um patamar elevado e tenham sido corrigidos nos anos seguintes, a variação pode sinalizar desafios pontuais na coleta ou registro dessas informações.

Assim, a análise permitiu concluir que os dados do SINASC referentes ao Semiárido brasileiro, no período de 2019 a 2023, apresentaram qualidade satisfatória em termos de completude, com destaque positivo para a maioria das variáveis essenciais à vigilância em saúde materno-infantil.

Tabela 5: Número de vazios e percentual de completude das variáveis, relacionadas às mães de todos os grupos etários e recém-nascidos, selecionadas da Declaração de Nascido Vivo no Semiárido brasileiro, 2019 a 2023.

Variável da Declaração de Nascido Vivo	2019		2020		2021		2022		2023	
	N	%	n	%	N	%	n	%	n	%
Anomalia Congênita	12.051	97	10.389	98	7.458	98	6.672	98	6.609	98
Consulta pré-natal	1.647	~ 100	1.438	~ 100	2.020	~ 100	14	~ 100	2.388	99
Escolarização mãe	18.036	96	10.321	98	11.119	97	8.045	98	8.064	98
Estado Civil	8.881	98	7.541	98	8.889	98	5.450	99	5.229	99
Idade da mãe	0	100	0	100	0	100	0	100	0	100
Índice de APGAR ao 1º minuto de vida	12.832	97	9.428	98	8.446	98	7.044	98	6.029	98
Índice de APGAR ao 5º minuto de vida	12.859	97	9.392	98	8.452	98	7.036	98	6.089	98
Local de ocorrência	11	~ 100	14	~ 100	17	~ 100	0	100	12	~ 100
Município de residência	0	100	0	100	0	100	0	100	0	100
Peso ao nascer	74	~ 100	68	~ 100	71	~ 100	55	~ 100	18	~ 100
Quantidade de filhos tidos mortos	36.155	92	27.969	93	25.129	94	17.621	96	17.515	96
Quantidade de filhos tidos vivos	23.265	95	17.919	96	16.555	96	11.530	97	11.571	97
Raça/cor da mãe	31.351	93	25.141	94	23.843	94	17.934	96	16.044	96
Semanas de gestação	11.448	97	9.958	98	10.124	98	6.717	98	6.316	98
Sexo	80	~ 100	76	~ 100	90	~ 100	75	~ 100	68	~ 100
Tipo de parto	456	~ 100	496	~ 100	461	~ 100	328	~ 100	207	~ 100
Tipo de gravidez	355	~ 100	586	~ 100	496	~ 100	372	~ 100	251	~ 100
Total de nascidos vivos – Semiárido	445.072		429.878		426.330		399.694		396.127	
– Brasil	2.849.146		2.730.145		2.677.101		2.561.922		2.537.576	

Fonte de dados básicos: Sistema de Informações sobre Nascidos Vivos (SINASC). Secretaria de Vigilância em Saúde, 2025.

Ao observar as variáveis selecionadas para a modelagem da fecundidade entre adolescentes também se constatou, como apresentadas na Tabela 6, que de um modo geral, as variáveis apresentaram bons níveis de completude, embora com algumas diferenças relevantes entre elas.

Tabela 6: Número de vazios e percentual de completude das variáveis sociodemográficas de mães adolescentes (15 a 19 anos) selecionadas da Declaração de Nascido Vivo no Semiárido Brasileiro, 2019–2023.

Variável da Declaração de Nascido Vivo	2019		2020		2021		2022		2023	
	N	%	N	%	n	%	N	%	n	%
Escolarização mãe	2.249	96	1.468	97	1.276	98	879	98	842	98
Estado civil	1.573	97	1.262	98	1.434	97	827	98	736	98
Município de residência	0	100	0	100	0	100	0	100	0	100
Quantidade de filhos tidos vivos	6.461	91	4.770	92	4.140	93	2.716	95	2.546	95
Raça/cor da mãe	5.140	93	3.873	94	3.604	94	2.501	95	2.190	95
Nascidos vivos - Semiárido	74.281		68.054		66.251		55.547		52.299	
Brasil	399.922		364.074		347.278		301.313		289.340	

Fonte de dados básicos: Sistema de Informações sobre Nascidos Vivos (SINASC). Secretaria de Vigilância em Saúde, 2025.

A variável "Município de residência" se destacou pela completude absoluta, com ausência total de registros vazios em todos os anos analisados. Esse resultado reflete a obrigatoriedade e padronização do preenchimento dessa informação, essencial para a localização geográfica dos eventos de nascimento.

As variáveis "Escolarização da mãe" e "Estado civil" também apresentaram completude elevada e estável, com percentuais variando entre 96% e 98% ao longo do período. Esses valores são considerados satisfatórios para análises estatísticas e sugerem que, mesmo em populações jovens, como o grupo de mães adolescentes, essas informações são relativamente bem registradas no sistema.

Por outro lado, a variável "Quantidade de filhos tidos vivos" apresentou os menores percentuais de completude da tabela, com destaque para o ano de 2019, em que apenas 91% dos registros estavam completos. Ainda que haja uma melhora progressiva até 2023 (95%), esse campo ainda apresenta limitações que podem comprometer a robustez de análises que considerem essa variável como preditora.

A variável "Raça/cor da mãe" também apresentou níveis mais baixos de completude em comparação às demais, oscilando entre 93% (2019) e 95% (2023). Embora os percentuais indiquem melhora ao longo do tempo, a presença de vazios nessa variável pode refletir dificuldades de categorização, resistência na autodeclaração ou falhas no processo de coleta de dados.

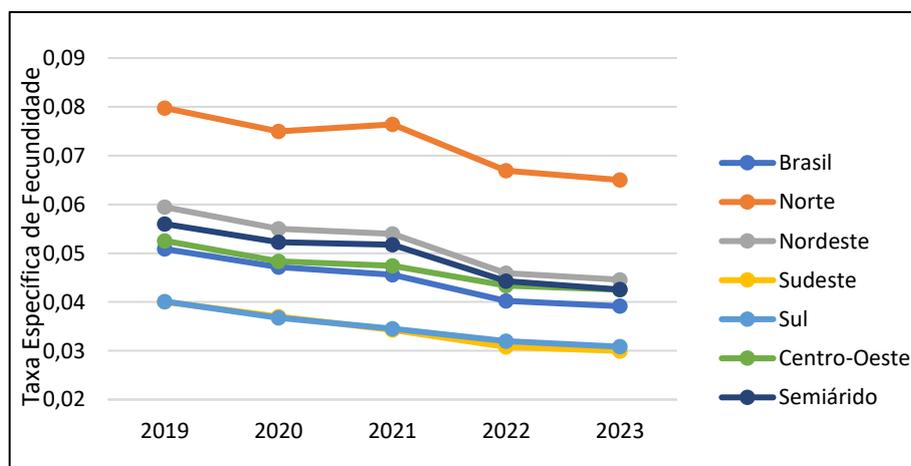
Em resumo, a análise da Tabela 5 revelou que, embora os dados do SINASC sobre mães adolescentes no Semiárido brasileiro apresentem boa qualidade geral, é importante considerar os limites de completude de algumas variáveis, especialmente “Quantidade de filhos tidos vivos” e “Raça/cor da mãe”, que exigem atenção para garantir a fidedignidade das análises inferenciais a serem realizadas.

4.2 Comportamento da Fecundidade entre Adolescentes

A partir da análise das Taxas Específica de Fecundidade entre Adolescentes (15 a 19 anos), observou-se que o Brasil e suas regiões, inclusive o Semiárido, apresentou uma tendência de decréscimo da fecundidade entre o grupo etário de 15 a 19 anos de idade, como demonstrado no decaimento das linhas entre os anos de 2019 a 2023 (Gráfico 3).

No entanto, o Gráfico 3 ilustrou que o ritmo de decaimento da taxa de fecundidade nessa faixa etária foi diferente entre as regiões. A Região Norte, por exemplo, apresentou, em todo o período, as maiores taxas de fecundidade, embora tenha registrado uma redução de aproximadamente 0,0798 para 0,0650 no intervalo analisado.

Gráfico 3: Taxa Específica de Fecundidade das Adolescentes do Brasil, Regiões e Semiárido Brasileiro, 2019-2023.



Fonte de dados básicos: Sistema de Informações sobre Nascidos Vivos (SINASC), 2025.

Em contraste, as Regiões Sudeste e Sul apresentaram os menores valores da TEF, com taxas iguais ou próximas a 0,0300, em 2023. As Regiões Centro-Oeste e o Nordeste apresentaram valores intermediários, 0,0425 e 0,0446 respectivamente, mas ainda superiores as Regiões Sudeste e Sul, indicando a persistência de desigualdades regionais no comportamento reprodutivo das adolescentes.

Apesar de aproximadamente 85% dos municípios do Semiárido brasileiro estarem situados na Região Nordeste, observou-se que os valores da Taxa Específica de Fecundidade para essa área (0,0425) são ligeiramente inferiores à TEF nordestina (0,0446). Ao final do período analisado, a curva do Semiárido se aproximou dos valores observados para o Brasil como um todo. Tal comportamento pode indicar uma possível subestimação dos níveis reais de fecundidade na região Semiárida do Brasil.

Diante desse possível viés de subestimação, impôs-se a necessidade de aplicar fatores de correção aos dados do Semiárido, conforme estabelecido nos procedimentos metodológicos deste estudo, a fim de assegurar maior fidedignidade às estimativas obtidas. Após os ajustes, a TEF do Semiárido em 2023 foi estimada em 0,0449, valor superior ao observado para a Região Nordeste (0,0446) e ainda distante da TEF nacional (0,0391), o que reflete valores mais realista e reforça a importância da correção para a adequada interpretação dos níveis de fecundidade na Região.

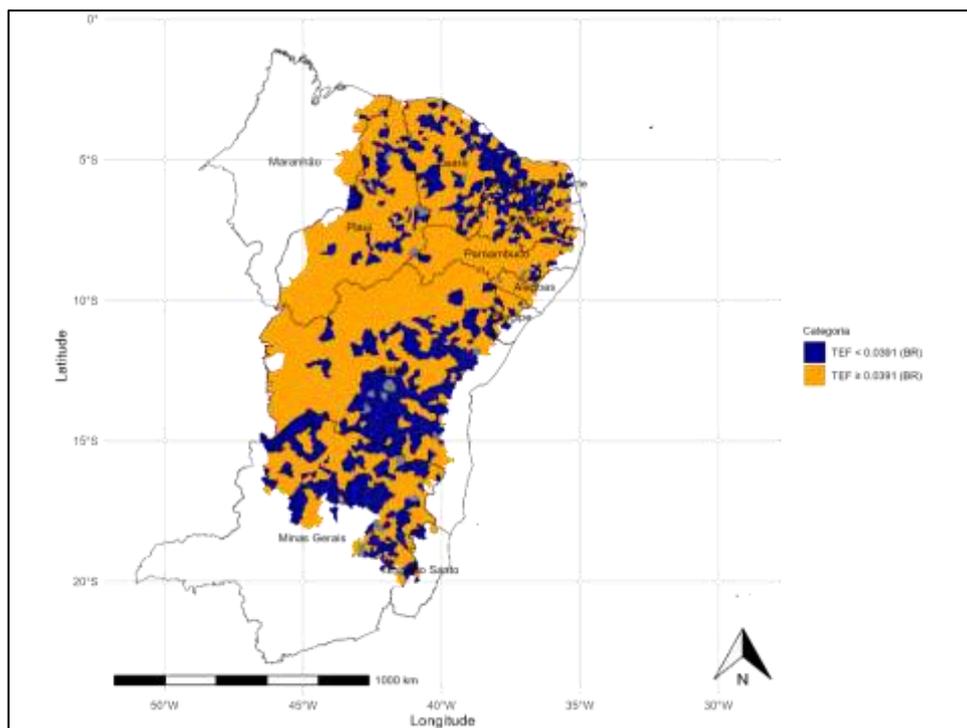
Ao examinar as TEF dos 1.473 municípios do Semiárido brasileiro com registro de nascidos vivos de mães adolescentes (15 a 19 anos), em 2023, observou-se que, em 2023, 563 (38,22 %) apresentavam TEF abaixo da TEF nacional (0,0391) e 61,77 % situam-se acima desse referencial. Esses dados, ao serem comparados com o estudo de Pereira Júnior (2020) para 2018, indicaram que houve alterações no comportamento reprodutivo nessa faixa etária no Semiárido brasileiro.

Primeiro, o número de municípios com registro de nascimento nessa faixa etária aumentou em quase toda Região. Dos 1.477 municípios do Semiárido brasileiro, 1.473 (99,73%) registraram nascimentos entre adolescentes. Em 2018, dos 1.262 municípios que compunha o Semiárido naquele ano, 822 (65,13%) apresentaram nascimentos entre o grupo etário de 15-19 anos. Por outro lado, a proporção de municípios com TEF adolescente abaixo da TEF nacional, em 2018, era de 75,1%, em 2023, somente 38,22% estavam abaixo da taxa nacional, como já mencionado na Fundamentação Teórica deste estudo.

Dentre as possíveis razões que podem ter influenciado o número de municípios abaixo da TEF nacional ser menor em 2023, destaca-se a inclusão de 215 municípios no Semiárido brasileiro (nova delimitação da SUDENE para 2024), dos quais 129, em 2018, já tinham a TEF adolescente maior que a TEF adolescente nacional daquele ano.

A análise da Figura 2 revelou que a distribuição espacial das TEF adolescentes no Semiárido, em 2023, é marcada por elevada variabilidade municipal, mesmo entre estados vizinhos, sem a formação de grandes aglomerados homogêneos. Notou-se, contudo, uma maior concentração de municípios com TEF abaixo da TEF nacional no sul da Bahia e em partes da Região Sudeste (Minas Gerais e Espírito Santo). Essa heterogeneidade intrarregional sugere que determinantes sociais locais, como por exemplo, nível de escolaridade, acesso a serviços de saúde, dentre outros exercem influência diferenciada sobre os padrões de fecundidade.

Figura 2: Taxa Específica de Fecundidade adolescente (15 – 19 anos) dos municípios do Semiárido brasileiro, 2023.



Fonte de dados básicos: Sistema de Informações sobre Nascidos Vivos (SINASC), 2025.

Outro aspecto sobre as TEF adolescente no Semiárido brasileiro é que entre 563 municípios que apresentaram uma TEF abaixo da TEF nacional, 531 (94,32%) estão em áreas urbanas ou são adjacentes dos espaços urbanos. Segundo Gonçalves (2019), a urbanização é uma variável intrínseca à fecundidade. No Brasil, de 1940 a 2010, à medida que o grau de urbanização no país aumentava, a fecundidade declinava. Talvez porque “os setores urbanos

criam novas oportunidades de trabalho, principalmente para as mulheres, aumentam o custo de se criar os filhos, do lazer e dos investimentos em capital humano” (Idem, p, 221-222).

Por outro lado, os municípios com a TEF adolescente abaixo de 0,0391 (TEF nacional), a proporção de mães adolescentes, em 2023, com ensino médio e superior, completo ou não, era de 31,87%, enquanto que 68,13% são de municípios com a TEF acima. No entanto, como alertou Martins (2016), o que vem ocorrendo é uma mudança na cobertura educacional do país, melhorando os níveis de escolaridade das adolescentes e não necessariamente um efeito negativo dos graus mais altos de escolaridade na fecundidade.

Quanto às observações sobre a situação conjugal, constatou-se que 24,16% das adolescentes dos municípios com TEF abaixo da TEF nacional foram registradas como casadas, 75,66% como solteiras e 0,18% como separadas ou divorciadas, apontando para a confirmação dos estudos de Pereira e Silva (2022) sobre a dissociação entre nupcialidade e gravidez na adolescência.

Segundo as autoras, no período de 2000 a 2020, foi possível constatar, no Brasil, a diminuição da proporção de mães adolescentes casadas ou em união estável, de 46,87% para 28%, enquanto que de solteiras um aumento de 49,7% para 70,8%. Embora a nupcialidade seja um determinante próximo ou intermediário da fecundidade “a maioria das gravidezes ocorrem em contexto de solteirice das jovens, o que reforça os achados sobre a queda de adolescentes unidas com filhos” (Idem, 2022, p. 4-5).

No tocante ao recorte de raça ou cor de pele, observou-se que 83,91% das mães adolescentes autodeclaradas pardas são residentes em municípios com TEF menor que a TEF nacional; 10,63% se autodeclararam brancas e 4,79% pretas. Os valores para as categorias amarela e indígena são inferiores a 0,5%. Essas proporções podem ser explicadas pela composição populacional do Semiárido brasileiro. Segundo Cavenaghi e Berquó (2014), a composição dos grupos populacionais determina o comportamento da fecundidade por raça ou cor da pele. “O comportamento das mulheres declaradas brancas espelha o que ocorre na Região Sudeste e o das mulheres declaradas negras assemelha-se à média do Nordeste” (Idem, p.79).

No entanto, as autoras salientaram que o ritmo de declínio da fecundidade, no Brasil, foi mais lento entre as mulheres negras, quando comparada as mulheres brancas. Entre as mulheres autodeclaradas brancas, a fecundidade, em 2000, já tinha alçado o nível de reposição (2,1) e entre as mulheres negras isso veio ocorrer apenas em 2010, com 50% dessas mulheres nessa condição e as demais com 2,1 a 4,9 filhos.

Esses resultados, portanto, ressaltaram a importância de investigações em escalas geográficas mais restritas, tanto para verificar a qualidade e a completude dos registros quanto para conhecer as singularidades do comportamento da fecundidade entre os diferentes grupos.

4.3 Modelagem dos fatores sociodemográficos associados à fecundidade

Para a aplicação do Modelo de Regressão Logística Binária, a variável TEF (Taxa Específica de Fecundidade) foi definida como variável dependente, utilizando como ponto de corte o valor da TEF nacional (0,0391) no ano de 2023. Os municípios com TEF inferior à TEF nacional foram codificados como “1” e os demais como “0”. As variáveis independentes incluíram os fatores sociodemográficos extraídos da DNV, previamente categorizados conforme descrito nos procedimentos metodológicos deste estudo.

Tabela 7: Estimativas dos Parâmetros do Modelo de Regressão Logística Binária ao considerar os fatores sociodemográficos associados à fecundidade entre adolescentes no Semiárido brasileiro, 2023.

Parâmetros	Estimativa β	Erro padrão	Estatística Wald	p-valor
Intercepto	-0,4401	0,0962	-4,5736	0,0001***
Estado Civil:				
Solteira (R)	-	-	-	-
Casada	0,0166	0,0054	3,0548	0,0022**
Separada	-0,2561	0,2700	-0,9488	0,342730
Escolaridade da mãe:				
Superior (completo ou não)	-	-	-	-
Nenhuma/Fundamental 1	0,0297	0,0748	0,3972	0,69120
Fundamental 2	0,0367	0,0144	2,5405	0,0110*
Médio	0,0028	0,0142	0,2034	0,83880
Município de residência:				
Urbano (R)	-	-	-	-
Rural adjacente	0,0484	0,0051	9,4763	0,001***
Intermediário adjacente	0,0331	0,0050	6,5419	0,001***
Rural remoto	0,0711	0,0166	4,2871	0,0001***
Intermediário remoto	0,0926	0,0795	1,1647	0,244150
Raça/cor da mãe:				
Branca(R)	-	-	-	-
Indígena	0,0977	0,0675	1,4466	0,1480
Negra	-00066	0,0112	-0,5898	0,5553
Quantidade de filhos tidos vivos:				
Sim tem outro filho (R)	-	-	-	-
Não tem outro filho	-0,0154	0,0095	-16228	0,1046

Fonte de dados básicos: Sistema de Informações sobre Nascidos Vivos (SINASC), 2025.

(R): correspondem às categorias de referência; (*): indica significância estatística; p-valor < 0,05 indica significância estatística entre a TEF adolescente e a categoria da variável.

Para a análise da significância estatística das variáveis e suas respectivas categorias, adotou-se como critério um *p-valor* inferior a 0,05, ou seja, a probabilidade de se obter os

resultados observados, assumindo que a hipótese nula (de que a taxa do município é igual à taxa nacional) seja verdadeira. Valores abaixo desse limiar indicam evidências estatísticas suficientes para rejeitar a hipótese nula.

A partir da Tabela 7, observou-se que, ao nível de significância de 5%, apenas as categorias “Casada” (estado civil da mãe), “Fundamental 2” (escolaridade da mãe), “Rural remoto”, “Rural adjacente” e “Intermediário adjacente” (Município de residência) apresentaram significância estatística no Modelo, uma vez que seus respectivos p-valores foram inferiores a 0,05. Após a análise da significância estatística, fez-se uma Análise de Diagnóstico a fim de avaliar a qualidade de ajuste do Modelo.

4.3.1 Análise de Diagnóstico

Nessa etapa, constatou-se que o Modelo apresentou um pseudo R^2 de Nagelkerke de 0,1881, sugerindo que aproximadamente 18,81% da variação na Taxa Específica de Fecundidade (TEF) entre adolescentes, superior ou inferior à TEF nacional do Semiárido brasileiro em 2023, pode ser explicada pelas variáveis sociodemográficas incluídas no Modelo.

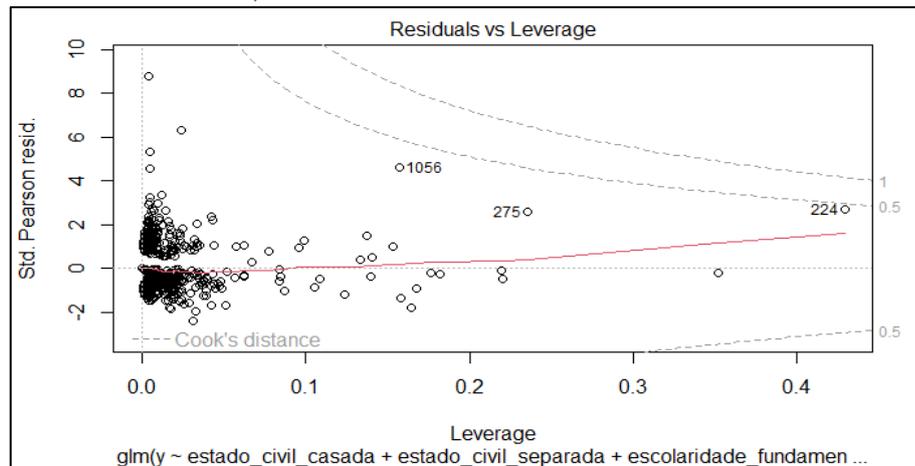
Também foi realizada uma análise dos resíduos para identificar possíveis discrepâncias entre os valores observados e os valores ajustados. Inicialmente, foi aplicado o Teste de *Durbin-Watson*, concluindo que como o valor da estatística (1,9623) está próximo de 2 e o p-valor (0,4607) é superior ao nível de significância de 5%, não há evidências de autocorrelação entre os resíduos do Modelo, o que indica que os dados estão distribuídos de forma aleatória e tem um padrão de variabilidade constante.

E, por fim, para avaliar a bondade de ajuste do Modelo, foi realizada a comparação entre o desvio do modelo nulo (1959,5) e o desvio do modelo ajustado (1740,1). A diferença entre esses desvios (219,4), com 12 graus de liberdade, foi avaliada utilizando o teste do qui-quadrado, resultando em um valor-p inferior a 0,05. Dessa forma, concluiu-se que o Modelo apresentou um ajuste significativamente melhor que o modelo nulo, sendo, portanto, adequado aos dados.

4.3.2 Análise dos Pontos Influentes

Para analisar a influência dos pontos sobre os coeficientes do Modelo, foram realizadas três verificações. A primeira consistiu na geração do Gráfico 4 dos resíduos padronizados, por meio do qual foi possível identificar observações que se desviam do comportamento esperado, ou seja, aquelas que ultrapassam o limite de $\pm 0,5$.

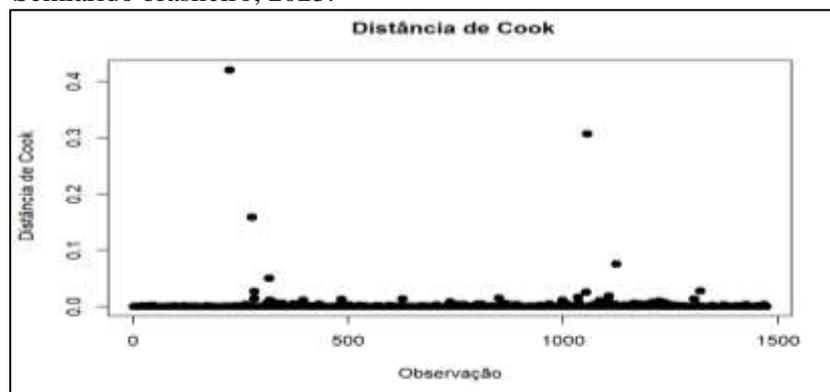
Gráfico 4: Análise dos pontos influentes no Modelo ao considerar os fatores sociodemográficos associados à fecundidade entre adolescentes no Semiárido brasileiro, 2023.



Fonte básica dos dados: Sistema de Informações sobre Nascidos Vivos (SINASC), 2025

Para checagem dos pontos de alavancagem apresentados no Gráfico 4, observou-se a Distância de *Cook*, através do Gráfico 5, o que permite avaliar o quanto a estimativa dos coeficientes de regressão seria afetada se uma observação específica fosse removida do conjunto de dados.

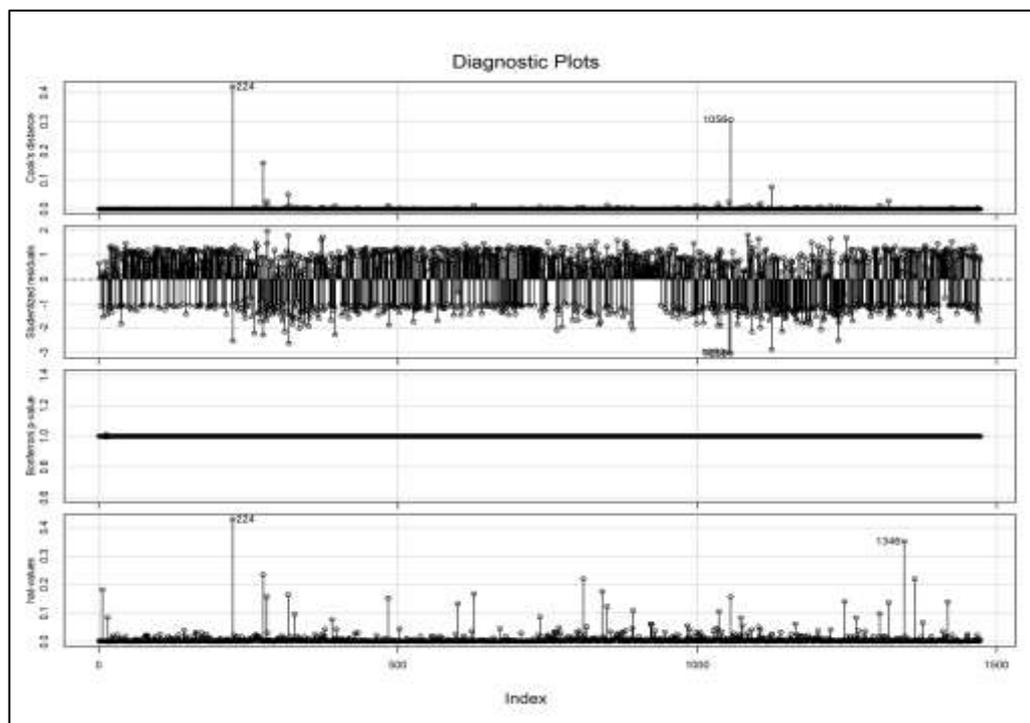
Gráfico 5: Distância de Cook do Modelo ao considerar os fatores sociodemográficos associados à fecundidade entre adolescentes no Semiárido brasileiro, 2023.



Fonte básica dos dados: Sistema de Informações sobre Nascidos Vivos (SINASC), 2025

No intuito de aprofundar a análise dos pontos discrepantes, optou-se por fazer uma análise combinada de medidas de diagnóstico através da função *IndexPlot* do R, como demonstrado no Gráfico 6, o que permitiu visualizar que o Modelo continha muitos valores extremos.

Gráfico 6: Análise Combinada dos Pontos Influentes do Modelo ao considerar os fatores sociodemográficos associados à fecundidade entre adolescentes no Semiárido brasileiro, 2023.



Fonte básica dos dados: Sistema de Informações sobre Nascidos Vivos (SINASC), 2025

Diante da leitura do Gráfico 6, fez-se uma varredura no Modelo para identificar todos os pontos de influência, identificando 179 pontos. Após a análise, decidiu-se por não excluí-los por não comprometerem o desempenho do Modelo.

4.3.3 Avaliação do Desempenho do Modelo

Para avaliação do desempenho do Modelo, primeiro, construiu-se uma matriz de confusão entre os valores previstos e os valores observados. A matriz de confusão indicou que o Modelo teve desempenho satisfatório na classificação dos municípios quanto à TEF adolescente em relação à taxa nacional. Dos municípios que apresentavam TEF acima da taxa nacional, 71,1% foram corretamente identificados. Já entre os que apresentavam TEF abaixo da taxa nacional, a taxa de acerto foi de 59,4%. A proporção de acerto global (acurácia geral) do modelo foi de 67,6%.

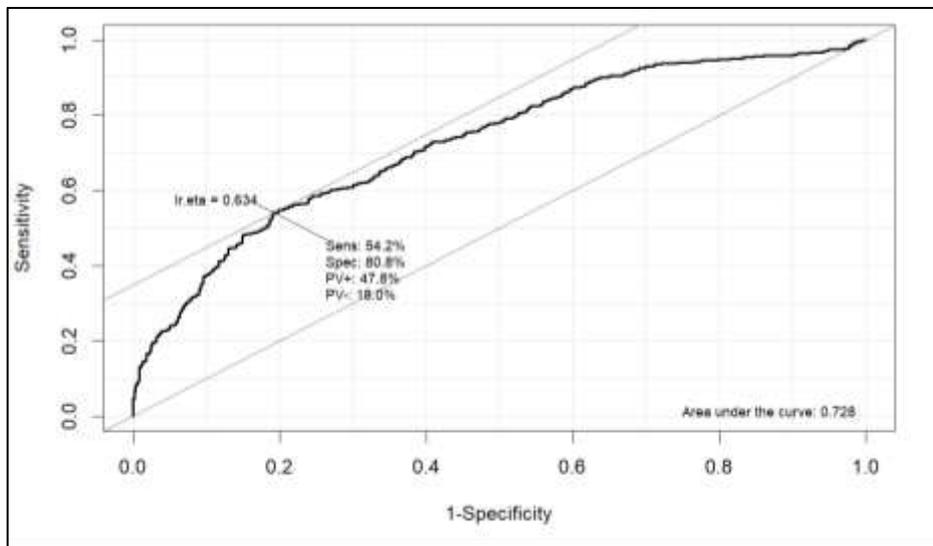
Tabela 8: Avaliação da Capacidade Preditiva do Modelo ao considerar os fatores sociodemográficos associados à fecundidade entre adolescentes no Semiárido brasileiro, 2023.

TEF adolescente observada	TEF adolescente prevista		% correta	% Acerto global
	Acima da TEF nacional	Abaixo da TEF nacional		
Acima da TEF nacional	727 (correto)	295 (erro)	71,1%	67,6%
Abaixo da TEF nacional	183 (erro)	268 (correto)	59,4%	
TOTAL	910	451		

Fonte básica dos dados: Sistema de Informações sobre Nascidos Vivos (SINASC), 2025.

Além da matriz de confusão, foi construído a curva ROC (*Receiver Operating Characteristic*) para avaliar o desempenho do modelo de classificação binária em diferentes limiares de decisão (Gráfico 7). Nela, a sensibilidade (taxa de verdadeiros positivos) é plotada no eixo y e em função da taxa de falsos positivos (1 – especificidade) no eixo x. Quanto mais próxima do canto superior esquerdo, maior a capacidade do modelo de distinguir entre as classes.

Gráfico 7: Curva ROC do Modelo ao considerar os fatores sociodemográficos associados à fecundidade entre adolescentes no Semiárido brasileiro, 2023.



Fonte básica dos dados: Sistema de Informações sobre Nascidos Vivos (SINASC), 2025.

No Gráfico 7, a área sob a curva (*AUC*) foi de 0,728, revelando bom poder discriminatório. Para o limiar ótimo ($lr.eta = 0,634$), obtivemos:

- Sensibilidade: 54,2%;
- Especificidade: 80,8%;
- Valor Preditivo Positivo (PV+): 47,8%;
- Valor Preditivo Negativo (PV-): 18%.

Esses resultados sugeriram que, embora o Modelo tenha apresentado uma boa especificidade (80,8%) e um *AUC* adequado (72,8%), ele possui limitações importantes quanto à precisão preditiva, principalmente no que se refere ao valor preditivo negativo. Em outras palavras, o Modelo tem maior segurança ao afirmar que um município está acima da TEF nacional (especificidade alta), do que ao prever corretamente os que estão abaixo (sensibilidade moderada e PV- muito baixo).

4.3.4 Análise das Razões de Chances (*Odds Ratio*)

O *odds ratio* (OR) ou razões de chances, indicou o quanto a chance de pertencer à categoria 1, aumenta ou diminui com o aumento de uma unidade na variável explicativa, mantendo as demais constantes. No caso deste estudo, a chance de um município ter uma TEF adolescente abaixo da TEF adolescente nacional (0,0391).

Para interpretar o efeito das variáveis independentes sobre a variável resposta, utilizou-se a razão de chances (*Odds Ratio*), obtida por meio da exponenciação dos coeficientes estimados pelo Modelo. Além disso, foram calculados os respectivos intervalos de confiança de 95% ($IC_{95\%}$) para indicar a precisão dessas estimativas.

Tabela 9: Estimativas das Razões de Chance (*Odds Ratio*) do Modelo ao considerar os fatores sociodemográficos associados à fecundidade entre adolescentes no Semiárido brasileiro, 2023.

Parâmetros	Odds Ratio	IC _{95%} Odds		p-valor
		LI	LS	
Intercepto	-	-	-	-
Estado Civil:	1	-	-	-
Solteira (R)				
Casada	1,0168	1,0060	1,0277	0,0022
Escolaridade da mãe:				
Superior (completo ou não) (R)	1	-	-	-
Fundamental 2	1,0375	1,0084	1,0673	0,0110
Município de residência:				
Urbano (R)	1	-	-	-
Rural adjacente	1,0497	1,0392	1,0603	0,0001
Intermediário adjacente	1,0337	1,0235	1,0440	0,0001
Rural remoto	1,0738	1,0394	1,1093	0,0001

Fonte básica dos dados: Sistema de Informações sobre Nascidos Vivos (SINASC), 2025.

Na Tabela 9, foram consideradas apenas as variáveis, e suas respectivas categorias, com significância estatística para o Modelo. Ademais, as categorias identificadas com (R) correspondem às categorias de referência, fixadas com razão de chances (*OR*) igual a 1 para fins de comparação. Pode-se verificar que os fatores sociodemográficos que estão significativamente associados à chance de um município registrar TEF adolescente abaixo da nacional são:

Estado civil – casada

A presença de adolescentes casadas, em comparação às solteiras, está associada a uma maior probabilidade do município apresentar uma Taxa Específica de Fecundidade (TEF) na adolescência inferior à TEF adolescente nacional. Especificamente, municípios com maior proporção de adolescentes casadas apresentaram somente 1,7% de mais chance de estarem abaixo da TEF nacional, em relação àqueles com maior proporção de adolescentes solteiras.

Com esses resultados, concluiu-se que embora, entre as adolescentes do Semiárido brasileiro a nupcialidade ainda seja um fator associado à fecundidade, os dados sinalizam para o início do movimento, segundo Pereira e Silva (2022), de dissociação do casamento ou união estável com a gravidez na adolescência.

Escolaridade – Ensino Fundamental 2

Com relação a escolaridade, os dados indicaram que a proporção de adolescentes com escolaridade até o Ensino Fundamental II está positivamente associada à chance de o município apresentar uma TEF adolescente inferior à média nacional, quando comparada à proporção de adolescentes com ensino superior.

Em termos quantitativos, os municípios com maior presença de adolescentes com essa escolaridade apresentaram uma chance 3,75% maior de estarem abaixo da média nacional de fecundidade. No entanto, conforme Martins (2016), ressalta-se que o grau de escolaridade mais alto não necessariamente é um fator associado ao aumento da fecundidade, o que tem ocorrido é a democratização do acesso à Educação.

Tipologia do município de residência

Os resultados revelaram que, em comparação com os municípios urbanos (*odds* = 1), os municípios classificados como rurais adjacentes, intermediários adjacentes e rurais remotos apresentaram maiores chances de possuir uma Taxa Específica de Fecundidade (TEF) adolescente abaixo da média nacional, com associações estatisticamente significativas ($p < 0,0001$). Isso indicou que, mesmo em áreas tradicionalmente associadas a maiores índices de

fecundidade precoce, há padrões locais que podem estar influenciando uma redução relativa desses indicadores.

O maior efeito observado nos municípios rurais remotos sugeriu a presença de fatores específicos, tais como migração das adolescentes, mudanças culturais ou possíveis limitações na notificação de nascimentos, que merecem investigações mais aprofundada. Esses achados reforçaram a importância de considerar o recorte territorial nas análises de fecundidade adolescente, especialmente em regiões como o Semiárido brasileiro.

5 CONSIDERAÇÕES FINAIS

O presente trabalho investigou os fatores sociodemográficos associados à fecundidade na adolescência no Semiárido brasileiro, com foco nos municípios que apresentaram Taxa Específica de Fecundidade (TEF) entre adolescentes de 15 a 19 anos inferior à TEF nacional em 2023. Para isso, foi aplicada uma modelagem estatística por Regressão Logística Binária, utilizando microdados do SINASC e informações populacionais do IBGE.

O modelo proposto permitiu identificar quais características dos municípios estão mais fortemente associadas à maior ou menor chance de apresentarem TEF adolescente abaixo da TEF adolescente nacional (0,0391). Os principais achados foram:

A fecundidade adolescente no Semiárido brasileiro está associada a múltiplos fatores contextuais, entre eles o estado civil, a escolaridade e a tipologia territorial. Municípios com maior proporção de adolescentes casadas apresentaram uma chance apenas 1,7% maior de estarem abaixo da TEF nacional, o que sugere que, embora a nupcialidade ainda tenha relação com padrões reprodutivos, seu impacto é modesto nesse contexto.

A escolaridade, especialmente o ensino fundamental II, mostrou-se levemente associada à menor fecundidade, com aumento de 3,75% na chance de o município estar abaixo da média nacional. Contudo, é necessário cautela ao interpretar essa associação, pois a expansão do acesso à educação pode estar mais relacionada à democratização do ensino do que à mudança direta nos comportamentos reprodutivos.

Por fim, os municípios classificados como rurais adjacentes, intermediários adjacentes e, sobretudo, rurais remotos, apresentaram maior probabilidade de ter TEF adolescente abaixo da média nacional. O destaque para os rurais remotos levanta hipóteses sobre possíveis migrações, transformações culturais ou subnotificações, evidenciando a necessidade de aprofundar os

estudos sobre as especificidades territoriais que influenciam a fecundidade entre adolescentes no Semiárido brasileiro.

O Modelo de Regressão Logística Binária apresentou bom desempenho estatístico com resultados satisfatórios na análise de diagnóstico (resíduos, influência, ajuste) e área sob a curva *ROC*, aproximadamente 73%, indicando boa capacidade discriminativa.

Além dos aspectos analíticos, o estudo também evidenciou a melhoria na qualidade dos dados de nascimentos no Semiárido brasileiro, tanto em cobertura quanto em completude, o que confere maior robustez às conclusões obtidas.

Os resultados mostraram que, embora a região Semiárida do Brasil siga a tendência nacional de queda da fecundidade, persistem desigualdades significativas entre os municípios, fortemente influenciadas por fatores estruturais como escolaridade e urbanização. A utilização de Modelos de Regressão Logística Binária se mostrou adequada para capturar essas nuances e oferecer evidências quantitativas que podem subsidiar o planejamento de políticas públicas voltadas à saúde sexual e reprodutiva das adolescentes.

No entanto, este estudo apresentou algumas limitações que restringiram a realização de análises mais específicas sobre a realidade do Semiárido brasileiro. Embora tenha sido aplicado um fator de correção, os registros permanecem frágeis nos municípios menores, o que indica a necessidade de investigações mais aprofundadas e contextualizadas nessas localidades.

A ausência de dados completos do Censo Demográfico de 2022 também impossibilitou a incorporação de variáveis socioeconômicas fundamentais para uma compreensão mais ampla e detalhada da fecundidade adolescente nos diferentes contextos municipais. Além disso, algumas variáveis disponibilizadas no SINASC, como a ocupação da mãe, ainda apresentam baixa confiabilidade, o que impediu de utilizá-la na modelagem estatística.

Os resultados, assim como as limitações, deste estudo apontaram para a necessidade de análises sobre as peculiaridades do comportamento da fecundidade no Semiárido brasileiro e para políticas públicas integradas que promovam o acesso à educação e à saúde sexual e reprodutiva, especialmente, em áreas mais vulneráveis. Espera-se, portanto, que os achados deste estudo contribuam para o monitoramento do alcance da meta 3.7 do ODS 3 da Agenda 2030, subsidiando planejamentos territoriais de ações voltadas à promoção de oportunidades para as adolescentes da Região.

Como recomendação para estudos futuros, sugere-se a ampliação das abordagens analíticas por meio da aplicação de modelos pertencentes a outras famílias de distribuições além

da binomial, como as famílias Poisson, quase-Poisson, binomial negativa ou gaussiana, a depender da natureza da variável resposta. Além disso, recomenda-se a utilização de dados individualizados, quando disponíveis, para modelagem, possibilitando maior precisão nas estimativas, melhor controle de variáveis e a aplicação de modelos hierárquicos ou mistos que considerem a estrutura dos dados.

REFERÊNCIAS

- ADELMAN, S. *et al.* **State-specific fertility rate changes across the USA following the first two waves of COVID-19**. *Human Reproduction*, v. 38, n. 6, p. 1202–1212, abr. 2023. Disponível em: <https://shre.ink/aX8w>. Acesso em: 17 ago. 2023.
- AMADEI, J. R. P.; FERRAZ, V. C. T. **Guia para elaboração de referências: ABNT NBR 6023:2018**. Bauru, fev. 2023. 53 p. Disponível em: <https://usp.br/sddarquivos/arquivos/abnt6023.pdf>. Acesso em: 17 ago. 2023.
- AZEVEDO, J. P.; LOPEZ-CALVA, L. F.; PEROVA, E. **Is the Baby to Blame? An Inquiry into the Consequences of Early Childbearing**. Washington, DC: The World Bank, 2012. Disponível em: <https://encurtador.com.br/tupdC>. Acesso em: 10 de dez. 2024.
- BERQUÓ, E. S.; CAVENAGHI, S. M. **Mapeamento socioeconômico e demográfico dos regimes de fecundidade no Brasil e sua variação entre 1991 e 2000**. In: Encontro Nacional de Estudos Populacionais, Caxambu. Anais [...]. Caxambu: ABEP, 2004. p. 1–18. Disponível em: <https://encurtador.com.br/iT5BH>. Acesso em: 21 abr. 2025.
- BERQUÓ, E. S.; CAVENAGHI, S. M. **Notas sobre os diferenciais educacionais e econômicos da fecundidade no Brasil**. *Revista Brasileira de Estudos de População*, v. 31, n. 2, p. 471–482, 2014. Disponível em: <https://encurtador.com.br/atIgA>. Acesso em: 21 abr. 2025.
- BRASIL. Instituto Brasileiro de Geografia e Estatística – IBGE. **Estudo de estimativas populacionais para os municípios brasileiros, desagregados por sexo e idade, 2000 - 2024**. Rio de Janeiro: IBGE, 2025. Disponível em: <http://tabnet.datasus.gov.br/>. Acesso em: 10 mar. 2025.
- BRASIL. Instituto Brasileiro de Geografia e Estatística – IBGE. **Proposta Metodológica para Classificação dos Espaços do Rural, do Urbano e da Natureza no Brasil**. Rio de Janeiro: IBGE, 2023. Disponível em: <https://encurtador.com.br/A1QvJ>. Acesso em: 10 de abr. 2025.
- BRASIL. Instituto Brasileiro de Geografia e Estatística – IBGE. **Sistemas de estatísticas vitais no Brasil: avanços, perspectivas e desafios**. Rio de Janeiro: IBGE, 2018. Disponível em: <https://shre.ink/azDz>. Acesso em: 4 ago. 2021.
- BRASIL. Instituto Brasileiro de Geografia e Estatística – IBGE. **Síntese de indicadores sociais: uma análise das condições de vida da população brasileira**. Rio de Janeiro: IBGE, 2020. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv101678.pdf>. Acesso em: 20 nov. 2020.

BRASIL. Ministério da Saúde. **Conselho Nacional de Saúde. Resolução nº 510, de 7 de abril de 2016.** Dispõe sobre as normas aplicáveis a pesquisas em Ciências Humanas e Sociais. Diário Oficial da União: seção 1, Brasília, DF, n. 98, p. 44–46, 24 maio 2016.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Análise Epidemiológica e Vigilância de Doenças Não Transmissíveis. **Declaração de Nascido Vivo: manual de instruções para preenchimento.** 4. ed. Brasília: Ministério da Saúde, 2022. Disponível em: <https://shre.ink/HRQE>. Acesso em: 10 jan. 2023.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Análise Epidemiológica e Vigilância de Doenças Não Transmissíveis. **Indicadores de cobertura que utilizam a metodologia do Busca Ativa.** Brasília: Ministério da Saúde, 2022. Disponível em: <https://encurtador.com.br/2hPRE>. Acesso em: 10 mar. 2025.

BRASIL. Ministério da Saúde. **SINASC – Sistema de Informações sobre Nascidos Vivos.** Disponível em: <https://encurtador.com.br/hpsTI>. Acesso em: 16 dez. 2024.

BRASIL. Superintendência do Desenvolvimento do Nordeste. **Resolução CONDEL/SUDENE nº 176, de 3 de janeiro de 2024.** Altera a Resolução nº 150, de 13 de dezembro de 2021, que aprova a nova delimitação do Semiárido brasileiro. Diário Oficial da União: seção 1, Brasília, DF, n. 3, p. 57, 4 jan. 2024. Disponível em: <https://encurtador.com.br/D8Qn9>. Acesso em: 15 de abr. 2025.

BRASIL. Instituto de Pesquisa Econômica Aplicada – IPEA. **Agenda 2030:** objetivos de desenvolvimento sustentável. Avaliação do progresso das principais metas globais para o Brasil – ODS 3: assegurar uma vida saudável e promover o bem-estar para todas e todos, em todas as idades. (Cadernos ODS, 3). Brasília: Ipea, 2024. Disponível em: https://repositorio.ipea.gov.br/bitstream/11058/14124/3/Agenda_2030_ODS_3_Assegurar_uma_vida_saudavel.pdf Acesso em 20 de abr. 2025.

CAVENAGHI, S. M.; BERQUÓ, E. S. **Perfil socioeconômico e demográfico da fecundidade no Brasil de 2000 a 2010.** Cap. 3. In: CAVENAGHI, S.; CABELLA, W. (org.). Comportamiento reproductivo y fecundidad en América Latina: una agenda inconclusa. Serie e-Investigaciones n. 3. Rio de Janeiro: ALAP, 2014. Disponível em: <https://encurtador.com.br/zHXVp> Acesso em: 10 mar. 2025.

COX, D. R.; SNELL, E. J. **The analysis of binary data.** 2. ed. London: Chapman and Hall, 1989.

FAWCETT, T. **Introduction to ROC analysis.** Pattern Recognition Letters, v. 27, p. 861–874, 2006.

FIGUEIREDO, A. M. R. **Mapas em R com geobr.** Campo Grande, MS: RStudio/Rpubs, 2020. Disponível em: http://rpubs.com/amrofi/maps_geobr. Acesso em: 28 ago. 2024.

FRIAS, P. G. **Adequação das informações sobre nascidos vivos e óbitos e a estimação da mortalidade infantil no Brasil.** Dissertação de Mestrado. Universidade Federal de Pernambuco, Recife, 2013. Disponível em: <https://repositorio.ufpe.br/handle/123456789/13400>. Acesso em: 24 abr. 2021.

FUNDO DE POPULAÇÃO DAS NAÇÕES UNIDAS. **Fecundidade e dinâmica da população brasileira**. Brasília: UNFPA, 2018. Disponível em: <https://shre.ink/azDQ>. Acesso em: 4 ago. 2021.

FUNDO DE POPULAÇÃO DAS NAÇÕES UNIDAS. **UNFPA: Brasil segue com índices elevados de gravidez na adolescência**. Brasília: UNFPA, 2023. Disponível em: <https://brasil.un.org/pt-br/199938-unfpa-brasil-segue-com-%C3%ADndices-elevados-de-gravidez-na-adolesc%C3%AAncia>. Acesso em: 21 abr. 2025.

GONÇALVES, G. Q. **Tendências regionais da transição da fecundidade brasileira corrente e de coorte ao longo do século XX. 2019**. Tese de Doutorado. Faculdade de Ciências Econômicas, Centro de Desenvolvimento e Planejamento Regional, Programa de Pós-graduação em Demografia, Universidade Federal de Minas Gerais, Belo Horizonte, 2019. Disponível em: <https://repositorio.ufmg.br/handle/1843/30344> Acesso em: 10 de abr. 2025.

MARINELLI, N. P. *et al.* **Evolução de indicadores e capacidade de atendimento no início da epidemia de COVID-19 no Nordeste do Brasil, 2020**. Epidemiologia e Serviços de Saúde, v. 29, n. 3, p. 1–10, 2020. Disponível em: <https://shre.ink/azI7>. Acesso em: 10 jan. 2023.

MARTINS, P. H. V. **Mudanças recentes na fecundidade adolescente no Brasil: a associação com a escolaridade continua a mesma?**. Dissertação de Mestrado. Universidade Federal de Minas Gerais, 2016. Disponível em: <https://repositorio.ufmg.br/handle/1843/FACE-AFXMSC> Acesso em: 10 abr. 2025.

MEDEIROS, J. B. **Análise da qualidade dos dados populacionais e registros de óbitos dos centenários do Semiárido brasileiro no período de 2000 a 2020**. Tese de Doutorado. Programa de Pós-Graduação em Modelos de Decisão em Saúde. Universidade Federal da Paraíba, João Pessoa, 2023. Disponível em: https://repositorio.ufpb.br/jspui/handle/123456789/27990?locale=pt_BR Acesso em: 5 de abr. 2025.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to linear regression analysis**. New York: Wiley, 2001.

OLIVEIRA, M. M. de *et al.* **Avaliação do Sistema de Informações sobre Nascidos Vivos – Brasil, 2006 a 2010**. Epidemiologia e Serviços de Saúde, Brasília, v. 24, n. 4, p. 629–640, 2015. Disponível em: <https://shre.ink/az0X>. Acesso em: 10 set. 2022.

PAES, M. A. **Demografia estatística dos eventos vitais: com exemplos baseados na experiência brasileira**. 2. ed. João Pessoa: CCTA, 2018.

PEREIRA JÚNIOR, B. H. **Modelagem da fecundidade materno-adolescente no Semiárido brasileiro**. Dissertação de Mestrado. Programa de Pós-Graduação em Modelos de Decisão em Saúde. Universidade Federal da Paraíba, João Pessoa, 2020. Disponível em: <https://repositorio.ufpb.br/jspui/handle/123456789/21166>. Acesso em: 1 set. 2022.

PEREIRA, L. M.; SILVA, I. A. da. **Formação familiar na adolescência: quais as imbricações entre a fecundidade e a nupcialidade?** In: Encontro Nacional de Estudos Populacionais, Anais [...]. Belo Horizonte: ABEP, 2022. Disponível em: <https://www.abep.org.br>. Acesso em: 21 abr. 2025.

PEREIRA, R *et al.* **Geobr: Download Official Spatial Data Sets of Brazil**. R package version 1.9.1. 2019. Disponível em: <https://github.com/ipeaGIT/geobr>. Acesso em: 1 abr. 2025.

SALDANHA, R. F.; BASTOS, R. R.; BARCELLOS, C. **Microdatasus: pacote para download e pré-processamento de microdados do Departamento de Informática do SUS (DATASUS)**. Cadernos de Saúde Pública, Rio de Janeiro, v. 35, n. 9, e00032419, 2019. Disponível em: <https://rfsaldanha.github.io/microdatasus/>. Acesso em: 1 set. 2022.

SANHUEZA, A. *et al.* **Adolescent pregnancy and associated risks in Latin America: a systematic review**. Pan American Journal of Public Health, v. 47, e2023-120, 2023. Disponível em: <https://encurtador.com.br/635hg>. Acesso em: 10 de jan. 2025.

SILVA, E. S. de A. da. **Impacto do Programa Bolsa Família na mortalidade infantil do semiárido brasileiro**. Tese de Doutorado. Programa de Pós-Graduação em Modelos de Decisão em Saúde. Universidade Federal da Paraíba, João Pessoa, 2017. Disponível em: https://repositorio.ufpb.br/jspui/handle/tede/9673?mode=full&locale=pt_BR. Acesso em: 20 de abr. 2025.

SILVA, E. S. de A. da. TOMAZ, T.B. **O Comportamento da Fecundidade no Semiárido Brasileiro na Pandemia da COVID-19**. PIBIC, Setembro de 2022 – Agosto de 2023. PROPESQ/UFPB, João Pessoa, 2023.

SIMÕES, C. C. S. **Revisão das estimativas de população nos anos intercensitários entre 1980 e 2000: uma avaliação da consistência demográfica dos resultados**. Revista Brasileira de Estudos de População, v. 23, n. 1, p. 5–18, 2006. Disponível em: <https://rebep.org.br/revista/issue/view/38>. Acesso em: 10 de mar. 2025.

SIMÕES, N. G. da C. **Análise da associação da fecundidade com fatores materno-infantis no semiárido brasileiro na pandemia da COVID-19 em 2021**. Trabalho de Conclusão de Curso. Departamento de Estatística. Centro de Ciências Exatas e da Natureza. Universidade Federal da Paraíba, João Pessoa, 2023. Disponível em: <https://encurtador.com.br/znkqf>. Acesso em: 10 de abr. 2025.

SOARES, R. A. S. **Modelo decisório espacial para a redução da mortalidade infantil: uma discussão no contexto da ruralidade na Paraíba**. 2019. 185 f. Tese de Doutorado. Universidade Federal da Paraíba, João Pessoa, 2019. Disponível em: https://repositorio.ufpb.br/jspui/handle/123456789/15499?locale=pt_BR. Acesso em: 10 de abr. 2025.

SZWARCWALD, C. L. *et al.* **Busca ativa de óbitos e nascimentos no Nordeste e na Amazônia Legal: estimacão da mortalidade infantil nos municípios brasileiros**. Brasília: Ministério da Saúde, 2010. Disponível em: <https://www.paho.org/publicaciones-relacsis/file>. Acesso em: 23 abr. 2021.

TERRA, M.L.C. **Regressão II**. Slides disponibilizados em sala de aula. Fevereiro a junho de 2023. João Pessoa: DE/UFPB.

TIEMI, R. **Censo mostra que 90% dos municípios com maior número de residências fechadas estão no semiárido**. Jornal da USP, 3 ago. 2023. Disponível em:

<https://jornal.usp.br/ciencias/censo-mostra-que-90-dos-municipios-com-maior-numero-de-residencias-fechadas-estao-no-semiarido/>. Acesso em: 21 abr. 2025.

WONG, L. L. R. **População: A mutante dinâmica da população brasileira** In: BRASIL. Instituto Brasileiro de Geografia e Estatística – IBGE. Brasil em números. Centro de Documentação e Disseminação de Informações. - Vol. 1 (1992-). - Rio de Janeiro: IBGE, 2018. Disponível em: <https://repositorio.ufmg.br/handle/1843/37029> Acesso em: 17 de abr. 2025.

PRODUÇÃO ACADÊMICA

SILVA, E. S. de A. da; TOMAZ, T. B. **A fecundidade no semiárido brasileiro nos cenários pré e pandêmico da COVID-19**. In: SINAPE – SIMPÓSIO BRASILEIRO DE PROBABILIDADE E ESTATÍSTICA, 25., 2024, Fortaleza. Anais [...]. Fortaleza: SINAPE, 2024.

SILVA, E. S. de A. da; TOMAZ, T. B. **Impactos socioeconômicos e da pandemia da COVID-19 na fecundidade da região semiárida brasileira**. In: ENCONTRO NACIONAL DE ESTUDOS POPULACIONAIS, 23., 2024, Brasília, Anais [...] Brasília: ABEP, 2024.

TOMAZ, T. B.; SILVA, E. S. de A. da; LIMA, M. A. M. **Desigualdades regionais na fecundidade adolescente: um estudo longitudinal do Semiárido brasileiro (2019–2023)**. In: REUNIÃO ANUAL DA RBRAS, 69; SEAGRO – SEMINÁRIO DE AGROECOLOGIA, 21. Vitória: RBRAS; SEAGRO, 2025. (Trabalho aprovado).

ANEXO

Implementação do código no *software* R

```

> # Pacotes necessários
> library(MASS)
> library(car)
> library(ROCR)
> library(Epi)
> # Importação e preparação da base
> base_tef <- read.csv2("D:/Documentos/Demografia/TCC/Base/2023/....csv")
> base_tef$tef <- as.factor(base_tef$tef)
> base_tef <- na.omit(base_tef)
> # Define a categoria de referência
> base_tef$tef <- relevel(base_tef$tef, ref = "1")
> # Ajuste do modelo
> modelo1 <- glm(
>   tef ~ (variáveis de interesse)
>   family = binomial(link = "logit"),
>   data = base_tef)
> )summary(modelo1)
> # Bondade de ajuste
> desvio1 <- modelo1$deviance
> graus_liberdade1 <- modelo1$df.residual
> q.quadr1 <- qchisq(0.95, graus_liberdade1)
> # Diagnóstico de influências
> cook_d <- cooks.distance(modelo1)
> limite_cook <- 4 / nrow(base_tef)
> out_cook <- which(cook_d > limite_cook)
> residuos <- rstandard(modelo1)
> out_residuos <- which(abs(residuos) > 2)
> leverage <- hatvalues(modelo1)
> limite_lev <- 2 * length(coef(modelo1)) / nrow(base_tef)
> out_leverage <- which(leverage > limite_lev)
> cat("Observações influentes (Cook > lim):", out_cook, "\n")
> cat("Resíduos padronizados > 2:", out_residuos, "\n")
> cat("Alavancagem alta:", out_leverage, "\n")
> # Avaliação com curva ROC (ROCR)
> fit <- fitted(modelo1)
> pred <- prediction(fit, base_tef$tef)
> perf <- performance(pred, "tpr", "fpr")
> plot(perf, main = "Curva ROC (pacote ROCR)")
> auc_valor <- performance(pred, "auc")@y.values[[1]]
> cat("AUC:", round(auc_valor, 4), "\n")
> # Curva ROC com pacote Epi
> ROC(
>   form = tef ~ (variáveis de interesse)
>   data = base_tef, plot = "ROC", MX = TRUE)

```