

---

Universidade Federal da Paraíba  
Centro de Ciências Exatas e da Natureza  
Departamento de Estatística

**ANÁLISE DA ASSOCIAÇÃO DA FECUNDIDADE  
COM FATORES MATERNO-INFANTIS NO  
SEMIÁRIDO BRASILEIRO NA PANDEMIA DA  
COVID-19 EM 2021**

Nícollas Gabriel da Costa Simões

Junho/2023

---

Nícollas Gabriel da Costa Simões

**ANÁLISE DA ASSOCIAÇÃO DA FECUNDIDADE  
COM FATORES MATERNO-INFANTIS NO  
SEMIÁRIDO BRASILEIRO NA PANDEMIA DA  
COVID-19 EM 2021**

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<sup>ª</sup>. Dra. Everlane Suane de Araújo da Silva

**João Pessoa  
Junho/2023**

**Catálogo na publicação**  
**Seção de Catálogo e Classificação**

S581a Simoes, Nicollas Gabriel da Costa.

Análise da associação da fecundidade com fatores materno-infantis no semiárido brasileiro na pandemia da COVID-19 em 2021 / Nicollas Gabriel da Costa Simoes. - João Pessoa, 2023.  
98 p. : il.

Orientação: Everlane Suane de Araújo da Silva.  
TCC (Curso de Bacharelado em Estatística) - UFPB/CCEN.

1. Semiárido brasileiro. 2. COVID-19. 3. Fecundidade. 4. Declaração de Nascidos Vivos. 5. Modelo de Regressão Logística. I. Silva, Everlane Suane de Araújo da. II. Título.

UFPB/CCEN

CDU 311(043.2)



UNIVERSIDADE FEDERAL DA PARAÍBA  
DEPARTAMENTO DE ESTATÍSTICA

Trabalho de Conclusão de Curso de Estatística intitulado *Análise da Associação da Fecundidade com Fatores Materno-infantis no Semiárido Brasileiro na Pandemia da COVID-19 em 2021* de autoria de Nícollas Gabriel da Costa Simões, aprovado pela banca examinadora constituída pelos seguintes professores:

Documento assinado digitalmente  
**gov.br** EVERLANE SUANE DE ARAÚJO DA SILVA  
Data: 18/06/2023 12:36:35-0300  
Verifique em <https://validar.iti.gov.br>

---

Prof<sup>ª</sup>. Dra. Everlane Suane de Araújo da Silva  
Universidade Federal da Paraíba

Documento assinado digitalmente  
**gov.br** NEIR ANTUNES PAES  
Data: 19/06/2023 09:55:17-0300  
Verifique em <https://validar.iti.gov.br>

---

Prof. Dr. Neir Antunes Paes  
Universidade Federal da Paraíba

Documento assinado digitalmente  
**gov.br** MARIA LIDIA COCO TERRA  
Data: 18/06/2023 14:29:52-0300  
Verifique em <https://validar.iti.gov.br>

---

Prof<sup>ª</sup>. Dra. Maria Lídia Coco Terra  
Universidade Federal da Paraíba

João Pessoa, 12 Junho de 2023

Centro de Ciências Exatas e da Natureza, Universidade Federal da  
Paraíba Cidade Universitária s/n CEP 58.051-900, João Pessoa,  
Paraíba, Brasil Fone: +55 83 3216-7785

*Dedico esse estudo à todos que estiveram ao meu lado  
durante a graduação, em especial, minha família.*

## AGRADECIMENTOS

Em primeiro lugar, gostaria de agradecer aos meus maravilhosos pais. Agradeço toda a confiança e esforço que tiveram em me criar, permitindo que eu pudesse desfrutar desta conquista que estou vivendo. Não posso deixar de agradecer à minha namorada, Letícia Garcez. Quero expressar meu profundo agradecimento pelo apoio incondicional, encorajamento e compreensão que ela me proporcionou ao longo desta jornada acadêmica. Suas palavras de incentivo e confiança foram essenciais para superar os desafios e perseverar até o fim.

Gostaria de expressar meus sinceros agradecimentos a todas as pessoas que contribuíram para a realização deste Trabalho de Conclusão de Curso (TCC). Sem o apoio e suporte dessas pessoas, este trabalho não teria sido possível.

Agradeço também aos membros da banca examinadora por dedicarem seu tempo e expertise na avaliação do meu trabalho. Suas contribuições e *feedbacks* foram extremamente valiosos para aprimorar a qualidade e consistência do TCC.

Sou grato aos professores e profissionais do curso de Estatística, que gentilmente compartilharam seus conhecimentos e experiências durante o desenvolvimento deste estudo. Suas orientações e *insights* enriqueceram minha compreensão e me forneceram uma base sólida para a pesquisa.

Por fim, gostaria de estender meu agradecimento a todos aqueles que, direta ou indiretamente, contribuíram para este trabalho, mesmo que não tenham sido mencionados nominalmente. Seu apoio e encorajamento foram fundamentais para o sucesso deste estudo.

Novamente, expresso minha profunda gratidão a todos que estiveram ao meu lado durante esta jornada. O apoio e companheirismo de vocês foram essenciais, e sou profundamente grato por cada contribuição. Muito obrigado.

*“Tenha coragem de seguir, seu coração  
e intuição. Eles, de certa forma,  
já sabem o que você realmente quer ser”*

*(Steve Jobs)*

A fecundidade é um dos fenômenos demográficos mais importantes da dinâmica populacional, onde faz referência direta a relação entre nascimentos vivos e mulheres em idade reprodutiva, que varia dos 15 aos 49 anos. Após o início da pandemia da COVID-19 muito se discutiu sobre os efeitos que essa apresentou sobre a mortalidade. Porém, estudos relacionados ao tema da natalidade e fecundidade ainda são muito escassos para algumas áreas, tal como para o Semiárido brasileiro. Ao considerar esse cenário, esse estudo teve como objetivo identificar as variáveis materno-infantis que estão associadas aos níveis de fecundidade do Semiárido brasileiro na pandemia da COVID-19 em 2021, e avaliar também os impactos da pandemia sobre a qualidade dos dados dos nascidos vivos da região durante o referente ano. Foi utilizada como ferramenta estatística a regressão logística binária, em conjunto com microdados dos 1.262 municípios do Semiárido brasileiro, trazendo resultados significativos para a região. A priori, o estudo buscou avaliar os impactos da COVID-19 sobre a cobertura e completude das informações dos microdados das declarações de nascidos vivos do ano de 2021 para a região. O estudo como esperado, mostrou avanços dos Sistemas de Informação em Saúde no Brasil, apesar do período atípico do estudo que foi durante a COVID-19. Porém, ainda foram detectadas algumas dificuldades em relação a incompletude de alguns municípios. Em seguida, foi aplicado a modelagem com base na regressão logística, onde foi possível observar a associação entre a Taxa de Fecundidade Total, e o Índice de Apgar 1º minuto, Escolaridade, Estado Civil e Idade das Mães. Assim, identificando os possíveis fatores influentes na saúde reprodutiva das mulheres durante todo seu período fértil da região semiárida brasileira.

**Palavras-chave:** Semiárido brasileiro; COVID-19; Fecundidade; Declaração de Nascidos Vivos; Modelo de Regressão Logística.

## ABSTRACT

Fertility is one of the most important demographic phenomena in population dynamics, referring directly to the relationship between live births and women of reproductive age, ranging from 15 to 49 years old. After the onset of the COVID-19 pandemic, much discussion has been centered around its effects on mortality. However, studies related to the topics of birth rates and fertility are still scarce for certain areas, such as the Brazilian Semiarid region. Considering this scenario, this study aimed to identify the maternal and child variables associated with fertility levels in the Brazilian Semi-Arid region during the COVID-19 pandemic in 2021, and also evaluate the impacts of the pandemic on the quality of data on live births in the region during that year. Binary logistic regression was used as a statistical tool, along with microdata from the 1,262 municipalities in the Brazilian Semiarid region, yielding significant results for the region. Initially, the study sought to assess the impacts of COVID-19 on the coverage and completeness of microdata from live birth records for the year 2021 in the region. As expected, the study demonstrated advancements in Health Information Systems in Brazil, despite the atypical study period during the COVID-19 pandemic. However, some difficulties regarding the incompleteness of data from certain municipalities were still detected. Subsequently, modeling was applied based on logistic regression, allowing for the observation of the association between the Total Fertility Rate and the 1-minute Apgar Score, Education, Marital Status, and Maternal Age. Thus, identifying possible influential factors on women's reproductive health throughout their fertile period in the Brazilian Semiarid region.

**Keywords:** Brazilian Semiarid region; COVID-19; Fertility; Live Birth Certificate; Logistic Regression Model.

LISTA DE QUADROS

3.1 Descrição das variáveis do estudo . . . . . 48

## LISTA DE FIGURAS

2.1	Espaço geográfico do Semiárido brasileiro . . . . .	21
2.2	Taxa de Fecundidade Total segundo as regiões brasileiras e para o Brasil 1940-2010 . . . . .	25
3.1	Curva ROC . . . . .	47
4.1	Cobertura do registro de nascidos vivos do Nordeste brasileiro, 2010-2021 .	50
4.2	Preenchimento da variável anomalias congênitas dos bebês das maes de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021. . . . .	54
4.3	Preenchimento da variável apgar 1 ° minuto dos bebês das mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021. . . . .	56
4.4	Preenchimento da variável apgar 5° minuto dos bebês das mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021. . . . .	57
4.5	Preenchimento da variável quantidade de consultas pré-natal das mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021. . . . .	58
4.6	Preenchimento da variável duração da gestação das mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021. . . . .	60
4.7	Preenchimento da variável escolaridade das mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021. . . . .	61
4.8	Preenchimento da variável estado civil das mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021. . . . .	62
4.9	Preenchimento da variável idade das mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021. . . . .	63
4.10	Preenchimento da variável local de ocorrência do parto das mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021. . . . .	64
4.11	Preenchimento da variável peso dos bebês das mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021. . . . .	65

4.12	Preenchimento da variável raça das mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021. . . . .	67
4.13	Preenchimento da variável sexo dos bebês das mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021. . . . .	68
4.14	Preenchimento da variável tipo de gestação das mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021. . . . .	69
4.15	Preenchimento da variável tipo de parto realizado pelas mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021. . . . .	70
4.16	Taxas de Fecundidade Total da região Nordeste e do estado de Minas Gerais, e Taxa de Fecundidade Total da região Semiárida brasileira, em 2021. . . .	73
4.17	Gráficos dos resíduos padronizados para o modelo de regressão logístico para os municípios do Semiárido brasileiro, 2021. . . . .	77
4.18	Distância de Cook do modelo de regressão logístico binário, 2021. . . . .	80
4.19	Avaliação de pontos influentes, 2021. . . . .	81
4.20	Avaliação de pontos influentes II, 2021. . . . .	82
4.21	Curva ROC, 2021. . . . .	83

## LISTA DE TABELAS

2.1	Taxa de Fecundidade Total segundo as regiões brasileiras e para o Brasil, 1940 - 2010 . . . . .	25
3.1	Exemplo de matriz de confusão . . . . .	46
4.1	Porcentagem da cobertura do registro do nascidos vivos dos estados que compõem o Semiárido brasileiro, 2010-2020. . . . .	50
4.2	Número de registros vazios e percentual de incompletude das variáveis das declarações de nascidos vivos do Semiárido brasileiro, 2019 e 2021 . . . . .	52
4.3	Preenchimento da variável anomalias congênitas para o Semiárido brasileiro, 2021. . . . .	54
4.4	Preenchimento da variável apgar 1º minuto para o Semiárido brasileiro, 2021. . . . .	55
4.5	Preenchimento da variável apgar 5º minuto para o Semiárido brasileiro, 2021. . . . .	57
4.6	Preenchimento da variável quantidade de consultas pré-natal para o Semiárido brasileiro, 2021. . . . .	58
4.7	Preenchimento da variável duração da gestação para o Semiárido brasileiro, 2021. . . . .	59
4.8	Preenchimento da variável escolaridade da mãe para o Semiárido brasileiro, 2021. . . . .	61
4.9	Preenchimento da variável estado civil da mãe para o Semiárido brasileiro, 2021. . . . .	62
4.10	Preenchimento da variável idade da mãe para o Semiárido brasileiro, 2021. . . . .	63
4.11	Preenchimento da variável local de ocorrência do parto para o Semiárido brasileiro, 2021. . . . .	64
4.12	Preenchimento da variável peso do bebê para o Semiárido brasileiro, 2021. . . . .	65
4.13	Preenchimento da variável raça da mãe para o Semiárido brasileiro, 2021. . . . .	66
4.14	Preenchimento da variável sexo do bebê para o Semiárido brasileiro, 2021. . . . .	67
4.15	Preenchimento da variável tipo de gestação para o Semiárido brasileiro, 2021. . . . .	68

4.16	Preenchimento da variável tipo de parto para o Semiárido brasileiro, 2021.	70
4.17	Taxa de Fecundidade Total da região do Nordeste brasileiro, 2021. . . . .	73
4.18	Taxa de Fecundidade Total do Semiárido brasileiro, 2021. . . . .	74
4.19	Classificação dos municípios segundo o nível de reposição da Taxa de Fecundidade Total do Semiárido brasileiro, 2021. . . . .	74
4.20	Classificação dos municípios considerados na modelagem segundo o nível de reposição da Taxa de Fecundidade Total do Semiárido brasileiro, 2021. . .	75
4.21	Variáveis selecionados para compor o modelo de regressão logística binária.	76
4.22	Bondade de ajuste . . . . .	79
4.23	Matriz confusão do modelo em termos do percentual de classificação correta para os municípios do Semiárido brasileiro, 2021 . . . . .	80

<b>1</b>	<b>Introdução</b>	<b>17</b>
1.1	Objetivo Geral . . . . .	19
1.1.1	Objetivos Específicos . . . . .	19
<b>2</b>	<b>Fundamentação Teórica</b>	<b>20</b>
2.1	Semiárido brasileiro . . . . .	20
2.2	Fecundidade . . . . .	22
2.3	A Fecundidade no Brasil, na região do Nordeste e Semiárido brasileiro . . .	23
2.4	Transição demográfica . . . . .	26
2.5	A pandemia da COVID-19 e a fecundidade no Brasil . . . . .	27
2.6	Sistemas de informações de estatísticas vitais no Brasil . . . . .	28
2.6.1	Registro Civil - IBGE . . . . .	29
2.6.2	Sistema de Informações sobre Nascidos Vivos - SINASC . . . . .	30
2.7	Qualidade dos dados . . . . .	31
2.7.1	Cobertura . . . . .	31
2.7.2	Incompletude . . . . .	32
2.8	Modelo de Regressão Logística . . . . .	33
<b>3</b>	<b>Procedimento Metodológico</b>	<b>34</b>
3.1	Cenário do Estudo . . . . .	34
3.2	Fontes dos dados, variáveis do estudo e softwares utilizados . . . . .	35
3.3	Procedimentos para ajuste das Taxas Específicas de Fecundidade . . . . .	36
3.4	Qualidade dos registros dos nascidos vivos . . . . .	38
3.4.1	Cobertura e incompletude dos registros de nascidos vivos . . . . .	38
3.5	Análise exploratória das variáveis de saúde materno-infantil e o estudo do impacto da COVID-19 . . . . .	39
3.6	Modelagem estatística dos dados . . . . .	39

3.6.1	Modelo de Regressão Logística . . . . .	40
3.6.2	Razão de Chances (Odds Ratio) . . . . .	42
3.6.3	R <sup>2</sup> de Nagelkerke . . . . .	42
3.6.4	Análise de diagnósticos . . . . .	43
3.6.5	Análise de pontos influentes ( <i>Outlier</i> ) . . . . .	46
3.6.6	Curva ROC . . . . .	46
<b>4</b>	<b>Resultados e discussões</b>	<b>49</b>
4.1	Qualidade dos registros dos nascidos vivos e avaliação dos impactos da pandemia da COVID-19 . . . . .	49
4.1.1	Cobertura dos registros de nascimentos . . . . .	49
4.1.2	Incompletude dos registros de nascimentos . . . . .	51
4.2	Caracterização das variáveis das Declarações de Nascidos Vivos . . . . .	53
4.2.1	Anomalias congênitas . . . . .	53
4.2.2	Apgar 1º minuto . . . . .	55
4.2.3	Apgar 5º minuto . . . . .	56
4.2.4	Quantidade de consultas de pré-natal . . . . .	57
4.2.5	Duração da gestação em semanas . . . . .	59
4.2.6	Escolaridade . . . . .	60
4.2.7	Estado civil . . . . .	61
4.2.8	Idade das mães . . . . .	62
4.2.9	Local de ocorrência do parto . . . . .	63
4.2.10	Peso do bebê . . . . .	64
4.2.11	Raça/cor da mãe . . . . .	66
4.2.12	Sexo do bebê . . . . .	67
4.2.13	Tipo de gestação . . . . .	68
4.2.14	Tipo de parto . . . . .	69
4.3	Comportamento dos níveis da fecundidade do Semiárido brasileiro . . . . .	70
4.4	Modelagem da Fecundidade . . . . .	75
4.4.1	Análise de diagnóstico . . . . .	77
4.4.2	Avaliação de pontos influentes . . . . .	81
4.4.3	Curva ROC . . . . .	82
<b>5</b>	<b>Considerações finais</b>	<b>84</b>
	<b>Referências bibliográficas</b>	<b>88</b>
	<b>Anexo</b>	<b>93</b>
	<b>Implementação computacional com software estatístico - R</b>	<b>94</b>

# CAPÍTULO 1

## INTRODUÇÃO

O Semiárido brasileiro é uma região que se estende pelos nove estados da região do Nordeste, além da região setentrional do estado de Minas Gerais. Ocupa cerca de 12% do território nacional e abriga cerca de 27 milhões de habitantes, correspondendo a 12% de toda população nacional (ASA, 2022). Trata-se de uma região rica sob vários aspectos sociais, culturais e econômicos, porém, é uma região de extrema vulnerabilidade social. Segundo o Instituto Brasileiro de Geografia e Estatística (IBGE), mais da metade (59,1%) dos brasileiros em situação de extrema pobreza estão na região do Nordeste. Destes, mais da metade (52,5%) vivem em áreas rurais da região, como por exemplo o Semiárido que compõe boa parte da região rural nordestina (ASA, 2022).

Além da vulnerabilidade social, outro grande problema da região do Semiárido brasileiro está também nos componentes demográficos, em específico a fecundidade. Um dos componentes demográficos mais importantes da dinâmica populacional, onde faz referência direta a relação entre nascimentos vivos e mulheres em idade reprodutiva, segundo definições internacionais, compreende as mulheres que estão na faixa etária de 15 a 49 anos (MARTINS et al., 2014). Sendo assim, é de extrema importância acompanhar e estudar esse componente, pois através deles conseguimos acompanhar e analisar o crescimento e o comportamento de uma determinada população. Logo, a fecundidade mostra-se como um agente resolutivo tanto para a questão da estrutura etária da população, quanto em questões econômicas, sociais e sanitárias. (ARBACHE, 2011).

Em relação as questões sanitárias, a COVID-19 é uma infecção respiratória causada pelo corona vírus SARS-COV-2, potencialmente grave e de elevado grau de transmissibilidade. Os primeiros casos surgiram na região de Wuhan, na China, em dezembro de 2019 e em menos de 3 meses já tinha se alastrado para diversos países, entre eles o Brasil (Ministério

da Saúde, 2022). Muito se discutiu a respeito dos impactos que uma pandemia desse grau poderia causar nos componentes demográficos da população mundial, mais especificamente na mortalidade. Porém, estudo a respeito da fecundidade e os impactos que esse poderia trazer para as mulheres ainda é bastante escasso.

A diretora da Organização Pan-Americana da Saúde (OPAS), Carissa F. Etienne, advertiu que a interrupção contínua dos serviços de saúde da mulher devido à COVID-19 poderia “destruir” mais de 20 anos de progresso na redução da mortalidade materna e o aumento do acesso ao planejamento familiar na América Latina e no Caribe (OPAS, 2022). Durante boa parte da pandemia as mulheres foram extremamente afetadas pelas interrupções dos serviços de saúde reprodutiva e materna, onde estima-se que cerca de 20 milhões de mulheres americanas tiveram seu controle de natalidade interrompido durante o período da pandemia, motivadas tanto por questões econômicas, políticas e sanitárias (ONU, 2022).

De acordo com a representante do Fundo de População das Nações Unidas (UNFPA), Astrid Brant, no caso do Brasil a principal preocupação em relação ao impacto no acesso a serviços de saúde reprodutiva tem sido em regiões afastadas, de difícil acesso, localizadas no Norte e o Nordeste do país (BANT, 2022). Como por exemplo a região do Semiárido brasileiro. Essa que é uma região caracterizada por condições climáticas adversas e baixos índices de desenvolvimento humano. O que acaba auxiliando na falta de infraestrutura e vulnerabilidade socioeconômica da população local.

No Brasil, os Sistemas de Informações de Saúde que auxilia nos estudos dos componentes demográficos é o Sistema de Informações sobre Nascidos Vivos (SINASC), que é alimentado pelo formulário da Declaração de Nascidos Vivos (DNV) em todo o território nacional. E também o Instituto Brasileiro de Geografia e Estatística (IBGE). Esses auxiliam principalmente no planejamento de políticas sanitárias e de saúde da população. Além da formulação de indicadores epidemiológicos estratégico de auxílio de diversas ações voltadas à gestão em saúde.

Sendo assim, vemos que as variáveis contidas nas DNVs como por exemplo: Ano de nascimento, tipo de parto, semanas de gestação, instrução da mãe e entre outros, são importantes indicadores para detectar as condições de saúde materna das mulheres. Fazendo-se necessário que esses dados possuam alta qualidade estatística, visto que são de extrema importância para indicadores na saúde materna em todo território nacional e em específico no Semiárido (PAES, 2018).

Alem disso, uma boa qualidade de coleta e preenchimento dessas informações é de extrema relevância para a população, pois o preenchimento incompleto ou ausente da

DNV pode ocasionar falsos resultados, dificultando o planejamento de políticas públicas e epidemiológicas essenciais para o desenvolvimento da região. Sendo assim, essa monografia buscou identificar as variáveis materno-infantis que estão associadas aos níveis de fecundidade das mulheres em idade fértil dos municípios do Semiárido brasileiro. Além de trazer um estudo do impacto da COVID-19 sobre a completude e cobertura das informações fornecidas na declaração de nascidos vivos do ano de 2021.

## 1.1 Objetivo Geral

- Identificar as variáveis materno-infantis que estão associadas aos níveis de fecundidade do Semiárido brasileiro na pandemia da COVID-19 em 2021.

### 1.1.1 Objetivos Específicos

- Avaliar os impactos da pandemia da COVID-19 sobre a qualidade dos dados dos nascidos vivos através da cobertura e da completude das variáveis das declarações de nascidos vivos;
- Identificar os fatores materno-infantis que implicaram na fecundidade por meio da regressão logística binária.

## CAPÍTULO 2

# FUNDAMENTAÇÃO TEÓRICA

### 2.1 Semiárido brasileiro

O Semiárido brasileiro é composto por parte da região do Nordeste brasileiro e do estado do Minas Gerais. É constituída por 1.262 municípios e ocupa cerca de 12% do território nacional, abrigando cerca de 28 milhões de habitantes divididos entre zona urbana (62%) e rural (38%). Segundo o IBGE (Instituto Brasileiro de Geografia e Estatística) é um dos Semiáridos mais povoados do mundo. Destaca-se como sendo a maior em relação a sua extensão e densidade demográfica. Correspondendo a uma área de 1,03 milhão de km<sup>2</sup>, conforme observada na Figura 2.1 (IBGE, 2022).

Seu clima é basicamente marcado por grandes irregularidades das chuvas e por altas taxas de evapotranspiração, que juntos contribuem para o risco constante da escassez hídrica, ocasionando assim uma insuficiência climática para sustentar rios caudalosos que se mantenham contínuos por longos períodos do ano. Com exceção do Rio São Francisco. Além disso, desde a sua primeira delimitação, o Semiárido sempre foi uma região caracterizada socialmente por baixos níveis educacionais e econômicos, quando comparada com as demais regiões do país. Estima-se que 60,9% dos municípios do Semiárido possui um IDH (Índice de Desenvolvimento Humano) variando entre muito baixo a baixo, mostrando assim as dificuldades enfrentadas diariamente por essa região (ASA, 2022).

As contradições e injustiças sociais que rodeiam a região podem ser percebidas inclusive no acesso à renda, que conseqüentemente reflete também uma forte desigualdade de gênero.

Figura 2.1: Espaço geográfico do Semiárido brasileiro



Fonte: Superintendência do Desenvolvimento do Nordeste (SUDENE), 2022.

Segundo dados do IBGE (Censo demográfico de 2000), metade da população do Semiárido não tinha renda monetária, ou tem como única fonte de renda os benefícios oferecidos pelo governo, a maioria desses sendo mulheres equivalendo a cerca de (59,5%) (AQUINO, 2010).

Devido a esses diversos problemas enfrentados pela região, em 2007 foi estabelecida a competência para fixar critérios técnicos e científicos para delimitação do Semiárido ao Conselho Deliberativo (Condel) da Superintendência do Desenvolvimento do Nordeste (Sudene) pela Lei Complementar n° 125, de 3 de janeiro de 2007. Que garante a população desses municípios acessar políticas públicas voltadas especificamente para o Semiárido, como por exemplo, o tratamento diferente das políticas de crédito, benefícios fiscais, carro pipa que auxilia bastante no combate a sede enfrentadas por bastantes famílias e entre outros (SUDENE, 2021).

Sendo assim, essa vulnerabilidade enfrentada pelo Semiárido brasileiro, faz com o que os programas de assistência social, financeira e de saúde seja de extrema importância para os cidadãos dessa região (NERI et al., 2013). Esses programas se mostraram essenciais durante o período mais crítico da pandemia da COVID-19, apesar da pandemia ter acentuado ainda mais as dificuldades estruturais e desigualdades que já existiam no período pré-pandêmico. Questões como os fechamentos de feiras, comércio e interrupções dos serviços de saúde reprodutiva e materna, fez com o que o progresso na redução da mortalidade materna ficasse comprometida no Semiárido, afetando principalmente a população rural da respectiva região(OPAS, 2022).

## 2.2 Fecundidade

A fecundidade é um dos componentes da dinâmica demográfica, que faz referência ao desempenho efetivo das mulheres durante toda a idade reprodutiva. Desde o início dos anos de 1960, muito vem se debatendo a respeito do crescimento populacional e da fecundidade. Nos anos de 1960 e 1970, o mundo passou por um período de aumento explosivo de suas populações para um cenário de estabilização do crescimento populacional, ou até mesmo de decréscimo. A mudança no contexto demográfico global, vem preocupando cada vez mais (SOUSA, 2022).

Ao decorrer dos anos foi observado uma queda nas Taxas de Fecundidade Total (TFT), que se aproximam cada vez mais, ou já estão em diversas áreas abaixo do nível de reposição – 2,1 filhos por mulher, que é o necessário para que cada geração se reponha em sua totalidade. O Canadá e alguns países da Europa, já vem apresentando taxas de

fecundidade total bem abaixo do nível de reposição, levando a um envelhecimento da população, acarretando diversos problemas econômicos e sociais para os países (G1, 2022).

Essa queda observada em todo o mundo está associada a diversos motivos, um deles foi a evolução dos métodos contraceptivos e das políticas voltadas para a perspectiva do controle da natalidade. Porém, alguns autores associam o declínio das taxas de fecundidade à modernização econômica e social contínua (EASTERLIN, 1975; CALDWELL, 1977).

## **2.3 A Fecundidade no Brasil, na região do Nordeste e Semiárido brasileiro**

Nas últimas décadas houveram diversas discussões a respeito da fecundidade e do declínio que essa vem sofrendo em todo o mundo. No Brasil, o comportamento da fecundidade não foi diferente. Estudos apontaram que a fecundidade no Brasil está passando por uma fase de transição (CAVENAGHI, 2016; WONG, 2016).

Em meados da década de 1930, o Brasil teve um grande marco que foi a industrialização e a urbanização no país. Desde essa década até a década de 1960, o Brasil apresentou uma leve queda nas suas taxas de fecundidade, porém um declínio mais acentuado só ocorreu durante o período do governo de Juscelino Kubitschek, que permitiu a vinda de várias empresas internacionais para o Brasil, auxiliando ainda mais na industrialização do país e consequentemente influenciando os níveis de fecundidade (GOBBI, 2022).

Contudo, as transformações da fecundidade no país não foram homogêneas, essas apresentaram variações regionais, socioeconômicas e culturais, apontando assim que o seu controle e evolução foi diferente em todo o território nacional, principalmente entre os espaços urbanos e rurais (GONÇALVES et al., 2019).

Na Tabela 2.1 e na Figura 2.2 é possível analisar melhor a evolução da fecundidade de 1940 a 2010 para o Brasil e as suas regiões. Ficou evidente o comentário anteriormente trazido sobre a transição que a fecundidade do Brasil passou nos últimos anos. Observou-se que em 1991, a TFT no parâmetro nacional era de 2,89 filhos por mulher, já em 2010, esse número reduziu para 1,90, ficando abaixo do nível de reposição. Essa queda também foi refletida nas regiões mais pobres, como é o caso da população rural, de baixa escolaridade e negra da região do Nordeste, ainda que de forma relativamente mais lenta, quando comparada com o que aconteceu com o restante da população (GIANNINI, 2022).

A década de 1990 foi marcada por um verdadeiro rejuvenescimento da fecundidade

em todo o país. Em 1980, cabia as mulheres entre 25 e 29 anos de idade a maior frequência na quantidade de filhos. Já nos anos seguintes houve um deslocamento cada vez mais frequente da concentração de mães da idade dos 20 aos 24 anos, além também do aumento da fecundidade do grupo de 15 a 19 anos de idade. Alterando a caracterização da paridade ao decorrer desses anos. Além disso, o comportamento da TFT também está relacionado as mudanças que a sociedade brasileira vem sofrendo com o processo da transição demográfica (SILVA; PAES; CESAR, 2018).

Em relação as regiões do Brasil, ficou evidente que a região Nordeste sempre mostrou taxas de fecundidade mais elevadas que as demais, tendo assim como a região Norte uma queda mais tardia que as demais regiões do Centro-Sul. Observamos que no Nordeste um declínio mais acentuado da fecundidade se deu no período de 1980 a 1991, enquanto nas demais regiões do Centro-sul esse declínio começou na década de 1970.

O Nordeste brasileiro sempre foi uma região marcada por diversos problemas sociais e demográficos históricos. Na metade do século XX sofreu por uma intensa transição demográfica de sua população, marcada principalmente pelas constantes migrações para outras regiões do país. Essas foram motivadas principalmente pelas grandes ofertas de empregos que estavam surgindo em outras regiões e também a fuga constantes da população rural da seca que sempre assolou a região rural e também semiárida do Nordeste (OJIMA; FUSCO, 2015).

Esse fator faz que a região do Nordeste cada dia mais perdesse participação no total da população nacional, declinando de 34,6% na década de 1950 para 27,8% no ano de 2010, segundo o censo demográfico. Em 1980, a região tinha na sua totalidade 35,5 milhões de nordestinos divididos em 50,2% nas áreas urbanas e 49,8% nas rurais, e uma Taxa de Fecundidade Total, da ordem de 6,2 filhos por mulher como mostrado na Tabela 2.1. Em 2010, foi visto um cenário bem diferente da década de 1950, o Nordeste apresentava uma população de 53,3 milhões de pessoas, dos quais 73,1% eram residentes de áreas urbanas, e 27,9% de áreas rurais. Além disso, os níveis de fecundidade era de 2,06 filhos por mulher, algo bem próximo da média nacional que era de 1,9 filhos por mulher. Evidenciando as consequências da transição demográfica (FUSCO; MORVAN, 2017).

Ao citar a região nordestina, é fácil lembrar da fecundidade no Semiárido brasileiro. Visto que 90% do Semiárido está localizado no nordeste. A fecundidade no Semiárido tem apresentado uma redução significativa nas últimas décadas. Segundo os dados do Instituto Brasileiro de Geografia e Estatística (IBGE), a taxa de fecundidade total (TFT) no Semiárido brasileiro em 2010 era de 2,19 filhos por mulher, o que é próximo da taxa de reposição populacional (2,1 filhos por mulher).

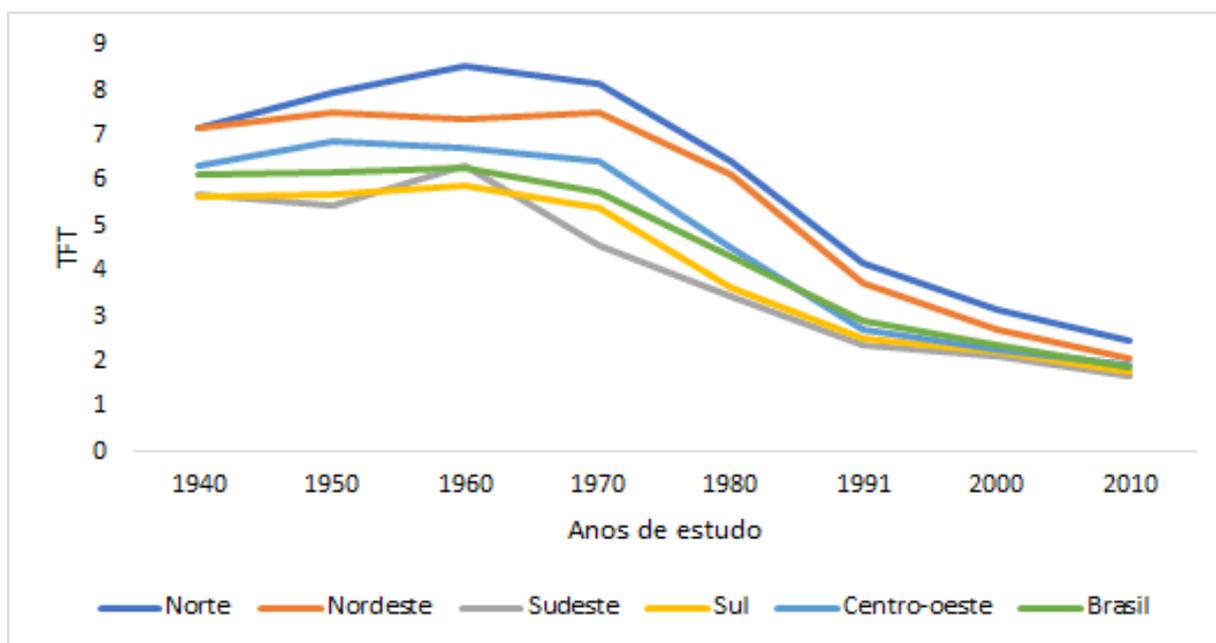
Essa redução na fecundidade pode ser atribuída a diversos fatores, como a transição demográfica e epidemiológica, a melhoria no acesso à educação e aos serviços de saúde, a disseminação de métodos contraceptivos e mudanças culturais e socioeconômicas que têm levado a uma menor ênfase no papel reprodutivo das mulheres (UNFPA , 2022). Trazendo também impactos significativos na dinâmica social e econômica da região, afetando questões como a distribuição de renda, o mercado de trabalho, o acesso aos serviços públicos e a estrutura etária da população.

Tabela 2.1: Taxa de Fecundidade Total segundo as regiões brasileiras e para o Brasil, 1940 - 2010

<b>Grandes Regiões</b>	<b>1940</b>	<b>1950</b>	<b>1960</b>	<b>1970</b>	<b>1980</b>	<b>1991</b>	<b>2000</b>	<b>2010</b>
<b>Norte</b>	7,17	7,97	8,56	8,15	6,45	4,20	3,16	2,47
<b>Nordeste</b>	7,15	7,50	7,39	7,53	6,13	3,75	2,69	2,06
<b>Sudeste</b>	5,69	5,45	6,34	4,56	3,45	2,36	2,10	1,70
<b>Sul</b>	5,65	5,70	5,89	5,42	3,63	2,51	2,24	1,78
<b>Centro-Oeste</b>	6,36	6,86	6,74	6,42	4,51	2,69	2,25	1,92
<b>Brasil</b>	6,16	6,21	6,28	5,76	4,35	2,89	2,38	1,90

Fonte: Instituto Brasileiro de Geografia e Estatística - IBGE, 2022.

Figura 2.2: Taxa de Fecundidade Total segundo as regiões brasileiras e para o Brasil 1940-2010



Fonte: Instituto Brasileiro de Geografia e Estatística - IBGE, 2022.

## 2.4 Transição demográfica

A transição demográfica trata-se de uma mudança no comportamento da mortalidade, fecundidade e natalidade de um país, o que pode provocar diversas mudanças na população como, por exemplo, o envelhecimento da população. As primeiras teorias a respeito da transição passou a surgir no início do século XX, segundo essas teorias o desenvolvimento econômico e o processo de modernização de uma nação estariam na origem das mudanças das taxas demográficas (VASCONCELOS; GOMES, 2012).

Essa transição ocorre em quatro estágios principais, o primeiro a mortalidade é concentrada nos primeiros anos de vida (grande percentagem da população é jovem); no segundo a mortalidade decresce e as taxas de crescimento aumentam ou se mantêm elevadas (percentagem de jovens na população aumenta); no terceiro momento, as taxas de fertilidade decrescem e a mortalidade continua a cair (aumento da percentagem de adultos jovens, e progressivamente de pessoas idosas) e no último estágio, a mortalidade em todos os grupos etários continua a cair (aumento contínuo na percentagem de idosos na população)(ELÍDIO, 2019).

Estima-se que a população brasileira começou a passar por esse processo de transição por volta de 1970, quando a fecundidade passou a decrescer junto com a mortalidade, resultando assim na diminuição das taxas de fecundidade total generalizada para todas as regiões como mostrado na Figura 2.1 (GOBBI, 2022).

O processo de urbanização acompanhou-se de importantes alterações sociais na população, a inserção da mulher na sociedade, rearranjos familiares, incrementos tecnológicos fez com o que esse padrão tecnológico fosse alterado ao decorrer dos anos. A forte queda na fecundidade e o aumento da longevidade impulsionaram um envelhecimento acelerado da população brasileira (VASCONCELOS; GOMES, 2012). Em 2010, o país apresentou taxa de fecundidade de 1,9 filhos por mulher, inferior à média que era observada na América Latina que era de em média 2,1 filhos por mulher, obtendo uma taxa de 2,06 na região no Nordeste, que era um das regiões com maiores taxas de fecundidade total do país (IBGE, 2022).

Em 2020, a pandemia da COVID-19 teve um impacto profundo na dinâmica demográfica global, acelerando e modificando a transição demográfica em muitos países ao redor do mundo. Com a chegada da pandemia, medidas de contenção, como o distanciamento social, lockdowns e restrições de viagens, afetaram diretamente os padrões reprodutivos das populações. Em muitos países, o isolamento social e a incerteza econômica resultaram em uma redução significativa do número de nascimentos. Casais adiaram os planos de ter

filhos devido às preocupações com a saúde e às dificuldades financeiras, impactando assim as taxas de fecundidade (COUTINHO et al., 2020).

## 2.5 A pandemia da COVID-19 e a fecundidade no Brasil

A Organização Mundial da Saúde (OMS) declarou a infecção pelo novo coronavírus como uma emergência global em 31 de janeiro de 2020. A rápida propagação a nível mundial levou a OMS a declarar no dia 11 de Março de 2020, a infecção do COVID-19, uma pandemia mundial.

Desde o paciente zero, identificado na província de Wuhan, China, em dezembro de 2019, até 20 de julho de 2020, 15 milhões de pessoas foram infectadas e 610 mil morreram por COVID-19 em todo o mundo (ALVES, 2020). Menos de um ano depois, em 29 de maio de 2021 esses números aumentaram para 169.118.995 de casos confirmados e 3.519.175 óbitos, mostrando assim o alto poder de propagação e contaminação do vírus (OMS, 2022).

No Brasil não foi diferente, até 20 de julho de 2020, existiam 2,1 milhões de infectados e 80 mil óbitos, sem considerar as subnotificações (ALVES, 2020), já no dia 19 de outubro de 2022, segundo a OMS, existiam 34 milhões de casos confirmados e 687 mil óbitos, mostrando que o Brasil vinha seguindo o mesmo modelo de aumentos de casos e óbitos ditas anteriormente em relação ao mundo.

Irrefutavelmente, as populações do mundo todo e em específico a população brasileira, sofreram diversas modificações sociais e econômica que perpetuou durante o período da pandemia. A crise sanitária e a diminuição das interações sociais devido à necessidade do distanciamento social, fez com o que as pessoas ficassem cada dia mais trancadas dentro de suas casas, gerando o aumento das atividades domésticas e de cuidado com pessoas idosas e crianças (COUTINHO et al., 2020).

Durante a pandemia foi possível pensar que quanto maior é o tempo de convivência no lar, mais relações sexuais os casais tendiam a manter e conseqüentemente a reprodução nesse período aumentasse. Porém, as conseqüências da COVID-19 foram basicamente ao contrário do que logicamente se pensava. A perda de empregos, fechamentos de empresas, incertezas em relação à renda e à saúde, falecimento de parentes, afetaram o desejo e a escolha dos casais de terem filhos (COUTINHO et al., 2020).

Além disso, o estresse e a ansiedade causados pela pandemia podem afetar a saúde reprodutiva das pessoas, afetando a ovulação e a qualidade do sêmen. O estresse também

pode levar a mudanças no comportamento sexual e afetar o libido. Portanto, embora a infecção pelo SARS-CoV-2 não pareça ter impactos significativos na fecundidade, a pandemia da COVID-19 impactou significativamente a saúde sexual e reprodutiva de muitas pessoas em todo o mundo, tornando ainda mais importante o acesso aos serviços de saúde reprodutiva e ao suporte emocional necessário para lidar com o estresse e a ansiedade causados pela pandemia (SANTOS, 2022).

Os primeiros impactos foram sentidos em 2020, onde houve um grande adiamento de gestações nesse ano (IDOETA, 2022). Enquanto no ano de 2019, segundo o Instituto Brasileiro de Geografia Estatística (IBGE), havia tido cerca de 2,8 milhões de bebês, no ano de 2020 esse número caiu para um pouco mais de 2,6 milhões. Sendo assim, considerando a projeção feita pelo IBGE, o Brasil no ano de 2020 teria cerca de 2,9 milhões de bebês nascidos vivos, porém, na prática como mostrado anteriormente observamos uma redução de 300 mil bebês do que era esperado para o ano.

A literatura sobre a COVID-19 é recente e ainda está evoluindo, portanto, falar sobre os verdadeiros impactos que a pandemia causou na fecundidade do Brasil, e em específico nos municípios do Semiárido brasileiro é de extrema importância (AASSVE et al., 2020). Principalmente, relacionado aos impactos na coleta e interpretação desses dados, que desempenham um papel crucial na compreensão da propagação do vírus.

## **2.6 Sistemas de informações de estatísticas vitais no Brasil**

As estatísticas vitais e de saúde são componentes essenciais para um estudo a respeito de uma determinada população. Essas buscam entender a evolução e comportamento dos nascimentos, óbitos e casamentos, trazendo assim uma resposta a respeito da dinâmica demográfica.

Os sistemas de saúde de informação estão em constante evolução. Desde o final do século XIX, o Brasil dispõe de dados de nascimentos vivos e de mortes. Porém, essas informações no passado eram extremamente viesadas visto que só se referiam as capitais dos estados. A forma de coleta, armazenamento, tratamento e disseminação das informações já estão bem mais evoluídas desde a criação do registro civil (SANTOS, 2006).

Sendo assim, desde a sua criação o registro imediatamente após o nascimento é um direito humano, confere identidade e facilita o acesso aos serviços essenciais como saúde, educação e benefícios sociais. Além disso, auxilia no planejamento de políticas públicas

essenciais para cada população (UNICEF, 1948).

### 2.6.1 Registro Civil - IBGE

O registro civil brasileiro teve sua origem ainda durante o período colonial e imperial, onde na época cabia a igreja católica, os registros de casamento, batismos e óbitos ocorridos no território brasileiro. No entanto, esse sistema ao decorrer dos anos deixou de atender satisfatoriamente as necessidades da sociedade brasileira, especialmente com as intensificações das imigrações e da abolição da escravatura. Justamente pelo fato de muitos dos imigrantes e escravos libertos não seguirem a religião católica, viasando assim os registros da época. Sendo assim, na época houve uma forte demanda social por um sistema de registro com menos influência da igreja católica e mais do estado. Aliás, a secularização do registro civil é apontada como um marco na transição para o Estado laico brasileiro que temos nos dias atuais (SANTOS, 2006).

Primeiramente, foram previstos os registros dos nascimentos, casamentos e óbitos das pessoas não católicas, feito em livros próprios dos Escrivães dos Juízos de Paz (Lei 1.144, de 11 de setembro de 1861, e Regulamento 3.069, de 17 de abril de 1863). Em seguida, no dia 25 de abril de 1874 foi editado o Decreto nº 5.604 para a execução do art. 2º da Lei nº 1.829, de 9 de setembro de 1870, onde pela referida alteração foram encarregados dos assentos de registro civil, em cada Juizado de Paz, o Escrivão respectivo, sob a imediata direção e inspeção do Juiz de Paz. Surgindo assim o Registro Civil das Pessoas Naturais, sob responsabilidade do Escrivão do Juizado da Paz em cada freguesia do Império. Onde segundo esses livros, foi capaz de identificar na época uma enorme quantidade de imigrantes, indígenas, libertos, alienados e até mesmo condenados (SANTOS, 2006).

Logo em seguida já no período da república, houve um novo regulamento que foi o decreto nº 9.886, de 7 de março de 1888, que entrou em vigor em 1º de janeiro de 1889, que já impôs certa obrigatoriedade da emissão do registro civil. Atualmente o processo de coleta das informações do Registro Civil passou por algumas alterações. Em uma primeira etapa os cartórios enviavam diretamente a Diretoria Geral de Estatísticas (DGE) os dados referentes aos nascimentos, óbitos e casamento registrados. Em seguida essa função ficou sobre responsabilidade dos departamentos de estatísticas dos estados, que recebiam essas informações e passavam ao DGE (SANTOS, 2006).

Sendo assim, surgiu a necessidade de unificar o processo de pesquisa e isso passou a ser atribuída ao IBGE. Onde está sobre sua responsabilidade reunir as informações sobre os registros de nascimentos, óbitos e óbitos fetais no Brasil, além da construção das estatísticas

e da publicação dessas informações. Para os cartórios de Registro Civil de Pessoas Naturais, segundo a Lei Federal nº 6.015, de 31 de dezembro de 1973, determina que os oficiais do registro civil devem remeter ao IBGE as informações dentro dos primeiros oito dias dos meses de janeiro, abril, junho e outubro de cada ano, os registros de nascimentos e mortes ocorridos no trimestre anterior (IBGE, 2022).

Além disso, após o decreto nº 70.210, de 28.02.1972, foi transferido também para o IBGE a responsabilidade pela coleta e apuração das informações ditas anteriormente, que até então era levantada pelo Serviço de Estatísticas Demográfica, Moral e Política do Ministério da Justiça.

## **2.6.2 Sistema de Informações sobre Nascidos Vivos - SINASC**

O Sistema de Informações sobre Nascidos Vivos (SINASC) foi implementado oficialmente a partir de 1990, tem como fonte de dados a declaração de nascidos vivos (DNVs) composta por 41 variáveis em sua totalidade. No qual é um sistema informatizado desenvolvido pelo Ministério da Saúde do Brasil para registro e acompanhamento dos nascimentos ocorridos no país (SINASC, 2022).

Seus principais objetivos são a coleta de dados sobre nascidos vivos em todo o território nacional, formando um banco de dados nacional sobre nascimentos, mediante a agregação dos dados estaduais; fornecer dados sobre Nascidos Vivos a todos os níveis do Sistema de Saúde; permitir um acompanhamento das estatísticas de nascimentos, com variáveis que são de grande importância para a saúde pública, como por exemplo, peso ao nascer, apgar 1 e 5 minutos, escolaridade da mãe, consultas de pré-natal, presença e descrição de anomalia congênita, etc; avaliar os riscos na gravidez, no parto e ao recém-nascido; subsidiar a execução das ações básicas na área materno-infantil, e permitir uma maior confiabilidade na elaboração dos coeficientes de mortalidade infantil (MINISTÉRIO DA SAÚDE, 2021).

Sua implantação ocorreu de forma lenta e gradual em todas as unidades da Federação, em contrapartida, já apresentam um número de registros maior do que o publicado pelo IBGE, com base nos dados de Cartório de Registro Civil. Além disso, o SINASC oferece auxílio para o desenvolvimento de ações com intuito de melhorar o atendimento as gestantes e aos recém-nascidos de todo o Brasil, além do monitoramento e avaliação das políticas públicas de saúde relacionadas à saúde materna e infantil (SINASC, 2022).

## 2.7 Qualidade dos dados

A Qualidade de dados também é conhecida pelo termo em inglês *Data Quality*. Pode ser definida como o valor que determinada informação possui. Este valor é averiguado a partir de alguns fatores. Dizemos que um dado tem qualidade quando é possível observar a origem da informação, a forma como ela foi produzida, armazenada, tratada e manipulada. Segundo Paes (2018), quatro dimensões são consideradas no estudo da qualidade dos dados de estatísticas vitais, que são: cobertura, completude, fidedignidade e regularidade.

De acordo com a constituição brasileira, segundo a Lei de nº 6.015/73, quando uma criança nasce no Brasil o registro de nascimento deve ser feito em até 15 dias a partir da data de nascimento, prorrogando-se por mais 45 dias se o requerente for a mãe, ampliando-se por três meses para aquelas situações em que os lugares de residência do responsável seja a uma distancia maior que 30 km do cartório. Com exceção dos índios que esse registro pode ser tanto feito no cartório como no livro próprio do órgão federal de assistência ao indígena, no caso, a Fundação Nacional do Índio - Funai, que emitirá o Registro Administrativo de Nascimento de Indígena - RANI (OLIVEIRA et al., 2018).

Apesar de todos os esforços do poder público, o sub-registro ainda é uma realidade do mundo e brasileira. Desde o final da década de 1990 e ao longo dos anos 2000, inúmeras campanhas e iniciativas governamentais para a erradicação do sub-registro foram implementadas, sendo a mais importante delas a determinação da gratuidade do registro de nascimento, que passou a vigorar em 1997 (art. 30 da Lei n. 9.534, de 10.12.1997). Essas ações tiveram impacto real no aumento da cobertura de tais registros, auxiliando assim na melhora da qualidade dos registros ao decorrer dos anos (OLIVEIRA et al., 2018).

### 2.7.1 Cobertura

O sub-registro de nascimento refere-se a um conjunto de nascimento não registrados no mesmo ano de sua ocorrência ou no primeiro trimestre do ano subsequente. Além disso, a má qualidade dos dados faz com que levantamentos a esse respeito sejam mais difíceis de serem implementados.

Para minorar esses problemas, os demógrafos desenvolveram um conjunto de procedimentos para estimação da cobertura dos nascimentos e análise de dados que é dado pela razão entre número de nascimento observados e nascimentos estimados como mostrado na Formula 2.1.

$$C = \frac{NV_{obs}}{NV_{est}} \quad (2.1)$$

Em que, o NVest é calculado pela soma das Taxas Específicas de Fecundidade (TEF.) em intervalos quinquenais entre as faixas etárias das mães de 15-49 anos de idade, multiplicado pelo total de mulheres dos mesmos intervalos etários, como mostrando na Formula 2.2. (PAES, 2018).

$$NV_{est(i)} = \sum_{j=15}^{49} TEF_j(i) * TM_j \quad (2.2)$$

Onde, segundo informações fornecidas por Szwarcwald et al.(2019) a cobertura das informações de nascidos vivos é adequada para a maior parte dos municípios brasileiros, embora ainda seja encontrado municípios com níveis inferiores a 60% no território nacional, como por exemplo, alguns municípios do estado do Amazonas, Pará e Maranhão. Outro problema detectado que também se faz necessário considerar é a incompletude das informações.

## 2.7.2 Incompletude

A incompletude de uma variável é caracterizado quando temos algum dado faltante ou preenchido como ignorando nos campos de resposta dos formulários de pesquisas ou avaliações. No SINASC e no Registro Civil diferentes fatores podem estar relacionados a ausência das informações no ato do preenchimento. Em “branco” pode estar relacionado por não preenchimento do campo pelo profissional; “ignorado” pode ser pela indisponibilidade da informação (SILVESTRIN et al., 2018).

A completude das variáveis é de extrema importância para uma boa avaliação da qualidade das informações da Declaração de Nascido Vivo (DNV). Através dela, conseguimos identificar e rastrear o descuido e a importância dada pelos profissionais de saúde, a ausência de dados nos prontuários médicos e de certa forma conseguimos identificar a falta de conhecimento de certas informações no acompanhamento da assistência materno-infantil (MAIA et al., 2017).

Segundo a alguns autores, os dados da SINASC nos últimos anos vêm obtendo uma melhora significativa em relação a completude. Porém algumas variáveis ainda continuam sendo um problema no preenchimento, como por exemplo; a idade gestacional, escolaridade

da mãe, raça/cor da pele e ocupação. A identificação dos problemas nessas variáveis é de extrema importância pois elas dificultam as análises de vulnerabilidade e desigualdades sociais, além de dificultar um melhor acompanhamento dos nascidos vivos em todo o território nacional. (SZWARCOWALD et al., 2019; LINO et al., 2019).

Sendo assim, para melhor investigarmos a magnitude das associações e relações da fecundidade materna no Semiárido brasileiro com as variáveis de saúde materno-infantis expressas na DNVs, optaremos pelo uso da Regressão, em específico a Regressão logística, visto que essa é uma ótima ferramenta estatística para o estudo da relação entre duas ou mais variáveis de interesse.

## 2.8 Modelo de Regressão Logística

O estudo da regressão teve sua origem em meados dos anos de 1885, por Sir Francis Galton. Onde buscou demonstrar que não existia relação entre as alturas dos pais nas alturas dos filhos, mas sim que esse regride para uma média da população. Atualmente, o termo “Análise de Regressão” define um conjunto de técnicas que busca estudar a relação entre duas ou mais variáveis de interesses, divididas entre variáveis dependentes (ou de respostas) e variáveis independentes (ou preditoras).

Com o passar do tempo, surgiu a necessidade de se criar um modelo que permitisse fazer predições de valores tomados por uma variável categórica, surgindo assim a regressão logística. Tem como principal objetivo modelar a partir de um conjunto de dados, a relação entre uma variável resposta dicotômica, e variáveis explicativas numéricas ou categóricas. Um exemplo é o modelo proposto por Cox e Shell (1989), onde tem sido bastante utilizado para casos que se procura ajustar um modelo linear do tipo regressão.

Na regressão logística, a variável resposta é dicotômica, onde geralmente atribuímos o valor 1 ao acontecimento de interesse (sucesso) e 0 ao acontecimento complementar (insucesso). Sendo assim, como o objetivo do estudo foi investigar e classificar a variável resposta com base na Taxa de Fecundidade Total está acima ou abaixo do nível de reposição (2,1), a regressão logística se mostra como um bom modelo resolutivo para avaliarmos a associação da fecundidade materna com as variáveis materno-infantis da DNV.

## CAPÍTULO 3

# PROCEDIMENTO METODOLÓGICO

### 3.1 Cenário do Estudo

O Semiárido brasileiro é uma região que compreende 1.262 municípios de dez estados diferentes. Além disso, nessa área vivem atualmente 27 milhões de habitantes divididos entre zonas urbanas correspondendo a 62% do total de habitantes e rurais com 38%. É uma região que apresenta diversas dificuldades sociais, climáticas e sanitárias, porém, ao mesmo tempo, é uma região extremamente rica e importante para o Brasil pelas diversas iniciativas econômicas que se desenvolvem nessa região (ASA, 2022).

Esse estudo abrange o ano de 2021, por município, passando por um dos períodos mais críticos da pandemia da COVID-19. No qual todos os dados foram obtidos através do do Sistema de Informações sobre Natalidade (SINASC) e da declaração de nascidos vivos (DNV), onde por motivos da ausência dos dados oficiais, e como o objetivo foi retratar reflexos do período da pandemia, foi utilizado os dados preliminares do referente ano.

Segundo a Secretaria de Vigilância em Saúde, esses dados preliminares disponibilizados, passam por rigorosos processos de aprimoramento da qualificação dos dados de natalidade, onde esses são realizados em 3 etapas junto aos respectivos Estados e Municípios. A primeira é o tratamento prévio dos dados, onde são indicadas críticas como inconsistência do dado e possíveis duplicidades; a segunda é o tratamento preliminar dos dados, onde são revistas as indicadas do tratamento prévio e reenviado aos Estados um relatório de qualidade indicando as possíveis novas e antigas inconsistências; e a terceira onde se encontrado os dados coletados nessa monografia, é um tratamento final dos dados, onde

são revisados e tratados os dados para a publicação final nos meios de comunicação do Ministério da Saúde (SINASC, 2023). Os dados definitivos só foram disponibilizados em maio de 2023, dessa maneira, a elaboração do Trabalho de Conclusão de Curso já estava concluído

Sendo assim, foi utilizado o ano de 2021 para a modelagem estatística dos níveis de fecundidade com as variáveis materno-infantis, além da avaliação os impactos da COVID-19 sobre a incompletude e cobertura das informações das DNVs das mães do Semiárido brasileiro.

## 3.2 Fontes dos dados, variáveis do estudo e softwares utilizados

Foram utilizadas duas fontes principais para as coletas das variáveis do estudo, a primeira foi o Instituto Brasileiro de Geografia e Estatística (IBGE), onde esse foi utilizado para a coleta de dados populacionais desagregados por municípios pertencentes ao Semiárido brasileiro, focando no sexo feminino e nas faixas etárias correspondente ao período fértil das mulheres.

A segunda fonte utilizada foi o Sistema de Informações sobre Nascidos Vivos (SINASC), onde através deste foram coletados os dados presentes nas declarações de nascidos vivos (DNVs) para o ano de 2021, além das variáveis materno-infantis. Importante ressaltar que assim como na obtenção dos dados populacionais, também foram utilizados o filtro por município, sexo e faixas etárias quinquenais dos 15 aos 49 anos. Sendo assim, foi construída uma base de dados relativa ao ano de 2021 para o Semiárido brasileiro, totalizando 1.262 municípios. Destes foram selecionados as variáveis das mães consideradas no estudo mostrado no Quadro 3.1.

Todavia, foram realizadas adaptações para a recategorização de 5 variáveis. A primeira e segunda foram as variáveis “Apgar no primeiro minuto” e “Apgar no quinto minuto” foram categorizadas segundo os índices propostos pela Sociedade Brasileira de Pediatria (APGAR, 1986), a terceira foi a variável “semanas de gestação” foi categorizada entre prematuro ( $< 37$  semanas) e não-prematuro ( $\geq 37$  semanas) e a quarta foi a variável “peso ao nascer” foi categorizada entre o peso maior ou igual a 2.500g (peso normal) e menor que 2.500g (baixo peso) segundo a Sociedade Brasileira de Pediatria (2019).

Além disso, a última variável foi a “instrução da mãe” que foi categorizada como:  $\leq 7$  anos de estudo (fora da idade escolar) e  $\geq 8$  anos de estudos (dentro da idade escolar),

conforme classificação proposta no Artigo 4º da Portaria nº 1.035 de 5 de outubro de 2018, do Ministério da Educação, através do Conselho Nacional de Educação e Câmara de Educação Básica, que diz: “O Ensino Fundamental, com duração de 9 (nove) anos, abrange a população na faixa etária dos 6 (seis) aos 14 (quatorze) anos de idade”.

Para a manipulação e tratamento dos bancos de dados, foi utilizado o software estatístico de programação R (versão 4.2.2), além do Microsoft Excel para o desenvolvimento e análise das estatísticas descritivas que foram apresentadas.

### 3.3 Procedimentos para ajuste das Taxas Específicas de Fecundidade

Foram utilizadas duas fontes principais para a coleta dos dados como mencionado anteriormente, a primeira foi o SINASC, com os dados referentes aos nascidos vivos e as informações das declarações de nascidos vivos, e a segunda o IBGE com os dados populacionais. Além disso, vale ressaltar que foi utilizada a classificação oficial proposta pela SUDENE para seleção dos municípios, constituindo assim a região de estudo que é o Semiárido brasileiro (SUDENE, 2021).

O primeiro banco de dados fez referência direta a declaração de nascidos vivos, onde contém todas as informações principais utilizadas na modelagem, informações tanto da mãe como do bebê que foi registrado em determinado município na declaração de nascidos vivos.

O segundo banco, foi utilizado para a obtenção das taxas específicas de fecundidade, obtido pela divisão da quantidade de nascidos vivos ( $N_x$ ) das mães em uma determinada idade, pelo número de mulheres nesta mesma idade ou grupo etário ( $M_x$ ). Além da taxa de fecundidade total das mães de 15 a 49 anos de cada município, que consiste na soma de todas as específicas multiplicado pela escala de idade que está sendo levado em consideração. Ambas mostradas nas fórmulas 3.1 e 3.2. Esse banco além de trazer informações sobre as taxas citadas, também será utilizado para classificar nossa variável independente na modelagem que será empregada.

$$TEF = \frac{N_x}{M_x} \quad (3.1)$$

$$TFT = \sum_{i=1}^n (TEF_n) * 5 \quad (3.2)$$

Além disso, foi necessário fazer uma correção dos dados de nascidos vivos do ano de 2021, apesar do nível de cobertura dos referentes dados ter sido relativamente alto. Sendo assim, para o problema em questão, foi utilizado o método relacional de Gompertz no ajuste da fecundidade, que busca corrigir a Taxa Específica de Fecundidade (TEF) em uma população, levando em consideração a fecundidade de diferentes idades. O método em si consiste em ajustar uma curva exponencial de Gompertz aos dados de fecundidade da população nas respectivas idades. Uma vez que a equação do modelo de fecundidade relacional Gompertz foi ajustada aos dados de fecundidade, foi possível calcular a TEF corrigida para qualquer idade e período de tempo. Para o problema em questão, foram consideradas as Taxas Específicas de Fecundidade por município das mulheres do Semiárido brasileiro, nas idades quinquenais de 15 a 49 anos.

A função de Gompertz foi expressa abaixo:

$$Gompit(p) = -\ln(-\ln(p))$$

Onde é importante destacar que o modelo não necessita do suposto de fecundidade constante no período mais recente, assim como fornece uma maneira adequada de interpolar os dados em idades simples, permitindo o reagrupamento em grupos etários não convencionais para grupos convencionais (MOULTRIE et al., 2013).

E o modelo é definido por:

$$Gompit(x) = \alpha + \beta.Gompit_s(x)$$

- $\alpha$  = nível da fecundidade  $-0,5 < \alpha < 0,5$ .
- $\beta$  = padrão da fecundidade  $0,65 < \beta < 1,5$ .
- $Gompit_s(x)$  = Gompit padrão para respectiva idade .

Se  $\alpha=0$  e  $\beta=1$ , padrão = observado.

Uma vez que a equação do modelo de fecundidade relacional Gompertz foi ajustada

aos dados de fecundidade, foi possível calcular a TEF é corrigida para qualquer idade e período de tempo.

## **3.4 Qualidade dos registros dos nascidos vivos**

A avaliação da qualidade dos dados antes da implementação de toda a metodologia foi de extrema importância para garantir estudos mais próximos da realidade. Dentro das quatro dimensões propostas por Paes para avaliação dos dados, serão abordadas as duas principais: a cobertura dos registros de nascidos vivos e a completude de preenchimento de todas as variáveis da Declaração de Nascido Vivo (DNV)(PAES, 2018).

### **3.4.1 Cobertura e incompletude dos registros de nascidos vivos**

A cobertura dos registros de nascidos vivos, como dito anteriormente, é uma das 4 principais dimensões de avaliação da qualidade de uma pesquisa. Sendo assim, para a avaliação do nível de cobertura das informações de nascidos vivos no Semiárido brasileiro em 2021, foram consideradas os resultados disponibilizados pela Secretaria de Vigilância em Saúde, através do sítio eletrônico: <https://svs.aids.gov.br/daent/acesso-a-informacao/acoes-e-programas/busca-ativa/indicadores-de-saude/cobertura/>.

Além disso, outra dimensão extremamente importante é a completude das variáveis de estudo. Sendo assim, foram utilizados os seguintes critérios de avaliação do grau de completude das informações: uma variável é considerada "excelente" quando possui menos de 5% de incompletude, "boa" quando apresenta uma incompletude entre 5% e 9%, "regular" quando a incompletude está entre 10% e 19%, "ruim" quando a incompletude varia de 20% a 49%, e "muito ruim" quando a incompletude é superior a 50% (ROMERO; CUNHA, 2007).

Vale a pena ressaltar que foi investigado o percentual de incompletude para todas as variáveis da Tabela 3.1, por município. E aqueles municípios que obtiveram em alguma de suas variáveis um percentual de completude acima de 10%, foi descartado do estudo.

### 3.5 Análise exploratória das variáveis de saúde materno-infantil e o estudo do impacto da COVID-19

Assim como a avaliação das dimensões de qualidade, a análise exploratória de dados foi de extrema importância para o estudo. Foram feitos gráficos e tabelas para as respectivas variáveis das declarações de nascidos vivos. Essa etapa teve como principal finalidade fazer uma prévia análise antes da aplicação de outros métodos estatísticos, que, no caso, será a modelagem.

Para o estudo do impacto da COVID-19, foram calculadas as frequências e as porcentagens de cada variável, além de ser realizado um estudo comparativo entre a quantidade de nascidos vivos registrada pelo SINASC para o ano de 2021 e a quantidade de nascidos vivos do ano de 2019. Isso auxiliou na obtenção de respostas sobre os impactos trazidos pela pandemia da COVID-19 na fecundidade e na incompletude do preenchimento das informações das Declarações de Nascidos Vivos (DNVs).

No que diz respeito à cobertura e impactos da pandemia da COVID-19, o estudo foi realizado por meio de análise gráfica e tabular das coberturas dos estados do Semiárido brasileiro no período de 2010 a 2021. Isso proporcionou uma compreensão mais abrangente da evolução da cobertura ao longo do tempo e permitiu identificar possíveis tendências ou variações entre os estados.

### 3.6 Modelagem estatística dos dados

Com o objetivo de trazer resultados a respeito das mães do Semiárido brasileiro durante o período fértil, e por conta da natureza dos dados, iremos optar por realizar uma regressão logística binária para o ano de 2021. Auxiliando no estudo da associação da fecundidade materna com as variáveis maternas infantis da DNV.

Considerou-se uma modelagem dicotômica tendo como variável dependente a Taxa de Fecundidade Total (TFT) das mães de 15 a 49 anos, cujo ponto de corte foi de está acima do nível de reposição que é de 2,1 filhos por mulher. Os municípios foram divididos em dois grupos: grupo 0 (municípios com a TEF abaixo do nível de reposição) e grupo 1 (municípios com a TEF acima do nível de reposição). As variáveis independentes foram as descritas na Tabela 3.1, oriundas da Declaração de Nascido Vivo.

### 3.6.1 Modelo de Regressão Logística

O modelo de regressão logística, também conhecido como *logit*, representa um caso particular dos MLGs (Modelos Lineares Generalizados) (NELDER; WEDDERBURN, 1972). Tem como principal diferença a composição da variável resposta com dados qualitativos, como por exemplo: sexo, raça/cor, classe social e entre outros. Além disso, possui como componente aleatória a distribuição Binomial e a função de ligação logística (MONTGOMERY; PECK; VIKING, 2001).

Considerando  $Y$  uma variável aleatória Bernoulli, onde 1 é a probabilidade do indivíduo possuir uma determinada característica, e 0 de não possuir, temos a seguinte distribuição de probabilidade:

$$Y = \begin{cases} 1 = P(Y = 1) = \pi \\ 0 = P(Y = 0) = 1 - \pi \end{cases}$$

Agora, consideraremos um conjunto de variáveis explicativas  $X_1, X_2, \dots, X_p$ . Temos o modelo de regressão logística definido pela Formula 3.3:

$$\text{logit}\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i = \eta_i + \varepsilon_i \quad (3.3)$$

Fazendo a estimação dos parâmetros por Máxima-Verossimilhança, considere novamente um modelo de regressão logística simples com  $Y_i \text{ Ber}(\pi_i)$ , a função de distribuição de probabilidade é dada por:

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

Sendo assim, a função de verossimilhança será:

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

Fazendo a aplicação do logaritmo, a expressão será definida por:

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

Sendo assim, para encontrarmos o valores de  $\beta$  que maximiza  $l(\beta)$ , deriva-se  $l(\beta)$  em relação a cada parâmetro  $(\beta_0, \beta_1)$ , obtendo-se as duas equações 3.4 e 3.5:

$$\frac{\partial(\beta)}{\partial(\beta_1)} = \sum_{i=1}^n \left[ y_i - \frac{1}{1 + \exp(\beta_0 + \beta_1 x_i)} \exp(\beta_0 + \beta_1 x_i) \right] \quad (3.4)$$

$$\frac{\partial(\beta)}{\partial(\beta_1)} = \sum_{i=1}^n \left[ y_i x_i - \frac{1}{1 + \exp(\beta_0 + \beta_1 x_i)} \exp(\beta_0 + \beta_1 x_i) \right] \quad (3.5)$$

que igualadas a zero, geram o seguinte sistema de equações:

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

Como as equações não são lineares em  $\beta_0$  e  $\beta_1$ , são necessários métodos iterativos para resolução, estes métodos encontram-se disponíveis em vários programas computacionais com os respectivos auxílios.

### 3.6.2 Razão de Chances (Odds Ratio)

No modelo de regressão logística, as estimativas dos parâmetros do modelo tem uma interpretação particular, chamada de OR (Odds Ratio) ou razão de chances (RC), onde é a razão entre a probabilidade de que um evento ocorra dividido pela probabilidade de que ele não ocorra (MONTGOMERY; PECK; VIKING, 2001). É obtido pela comparação de indivíduos que diferem apenas na variável de interesse.

Sendo assim, seja o modelo abaixo definido com apenas uma variável independente X, temos que:

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

Incrementando uma unidade em X, teríamos o seguinte efeito na razão de chances:

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

Assim,

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

Logo,

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

Sendo assim, através da OR(X) é possível mensurar o acréscimo (ou decréscimo) na P(Y=1) em relação a P(Y=0), tomando como base a oscilação da variável X. Ou seja, mensuramos a probabilidade de que determinado evento ocorra em nosso modelo.

### 3.6.3 R<sup>2</sup> de Nagelkerke

O R<sup>2</sup> Nagelkerke, também conhecido como pseudo-R<sup>2</sup> de Nagelkerke, é uma medida de ajuste em modelos de regressão logística. Trata-se de uma extensão do coeficiente

de determinação, comumente utilizado em regressão linear. Na regressão logística, o  $R^2$  Nagelkerke fornece uma indicação de quão bem o modelo explica a variação na variável dependente, considerando a probabilidade do evento sendo modelado. Ele varia de 0 a 1, em que 0 indica que o modelo não explica nenhuma variação e 1 indica um ajuste perfeito. Sua fórmula é calculada comparando a verossimilhança do modelo com a verossimilhança de um modelo nulo (um modelo sem preditores) (COX; SNELL, 1989).

### 3.6.4 Análise de diagnósticos

Após a estimação dos coeficientes é necessário avaliar a significância das variáveis para o modelo. Onde a análise de diagnóstico consiste basicamente em avaliarmos alguns pressupostos estabelecidos pela literatura para a obtenção de um bom modelo. Para esta pesquisa foi verificado os seguintes pontos;

#### **Análise de resíduos (Durbin Watson).**

A análise de resíduos é uma etapa importante na avaliação de um modelo de regressão logística. Os resíduos consistem na diferença entre os valores observados e os valores previstos pelo modelo. A análise de resíduos na regressão logística visa verificar a adequação do modelo aos dados e identificar possíveis problemas, como violação de pressupostos ou padrões sistemáticos nos erros de previsão. Existem diferentes métodos para analisar os resíduos em um modelo de regressão logística, mas no estudo foi avaliado os gráficos de resíduos parciais e o teste de Durbin Watson.

A análise gráfica dos resíduos padronizados foi realizada com base no diagrama de dispersão entre o resíduo e o predito, onde é esperado que eles estejam concentrados em pelo menos 95% entre -2 e 2 e poucos pontos acima de -3 e 3 (PEREIRA, 2020). Além do teste de Durbin Watson a fim de detectar a presença de autocorrelação nos resíduos.

O teste de Durbin-Watson é um teste estatístico utilizado para verificar a presença de autocorrelação nos resíduos de um modelo de regressão. Especificamente, ele avalia se existe autocorrelação de primeira ordem, ou seja, se há correlação entre os resíduos adjacentes. É aplicável a diversos tipos de modelos de regressão, incluindo a regressão linear simples, múltipla e também modelos de regressão logística. O valor do teste de Durbin-Watson varia entre 0 e 4 (SOBRAL; ROCHA, 2020), sendo que:

- Um valor próximo a 2 indica ausência de autocorrelação nos resíduos;

- Um valor menor que 2 indica autocorrelação positiva (resíduos adjacentes correlacionados positivamente);
- Um valor maior que 2 indica autocorrelação negativa (resíduos adjacentes correlacionados negativamente).

Sua estatística é baseada na suposição de que os erros no modelo de regressão são gerados por um processo autorregressivo conforme a mostrado abaixo:

$$\xi_i = \rho\xi_{i-1} + \alpha_i$$

Generalizando, tem-se a estatística de teste de Durbin Watson, mostrado na Formula 3.6:

$$Dw = \frac{\sum_{i=2}^n (\xi_i - \xi_{i-1})^2}{\sum_{i=3}^n \xi_i^2} \quad (3.6)$$

### **Bondade de Ajuste.**

A bondade de ajuste é uma medida que avalia o quão bem um modelo se ajusta aos dados observados. É importante verificar a bondade de ajuste para garantir que o modelo seja adequado e capaz de representar os padrões e as relações presentes nos dados, existem diferentes métodos e métricas para avaliar a bondade de ajuste, dependendo do tipo de modelo estatístico utilizado. Para esse estudo a bondade do ajuste será avaliado por meio do desvio escalonado e do teste qui-quadrado (OLIVEIRA, 2018).

O Desvio escalonado (*Scaled Deviance*), é uma medida da discrepância entre as frequências observadas e as frequências esperadas sob o modelo. É calculado como a diferença entre o desvio nulo (*deviance null*) e o desvio do modelo ajustado (*deviance*).

Já o teste qui-quadrado é uma medida estatística que compara as frequências observadas e esperadas sob o modelo. Ele quantifica a discrepância entre as frequências esperadas sob o modelo ajustado e as frequências observadas. O teste é calculado como a soma dos desvios ao quadrado divididos pelas frequências esperadas.

Sendo assim, nosso modelo estará bem ajustado quando o desvio escalonado for menor

que nossa qui-quadrado de referência, além do desvio escalonado ser próximo de zero e o valor do teste qui-quadrado não está estatisticamente significativo (OLIVEIRA, 2018).

### **Distância de Cook.**

A distância de Cook é uma medida estatística utilizada na análise de regressão para avaliar a influência de observações individuais nos resultados do modelo. Ela mede o quanto os resultados da regressão podem mudar quando uma observação específica é removida do conjunto de dados. É calculada para cada observação e leva em consideração tanto os resíduos (diferença entre os valores observados e os valores preditos pelo modelo) quanto a influência da observação nos coeficientes de regressão. Valores maiores de distância de Cook indicam que a observação em questão tem uma influência maior nos resultados do modelo (PREGIBON, 1981). Sua fórmula é definido com base na Fórmula 3.7 mostrado abaixo:

$$DC_j = r_j^2 \frac{h_j}{(1 - h_j)} \quad (3.7)$$

Onde,  $r_j^2$  é o resíduo estandardizado de Pearson e  $h_j$  é o leverage. Com esta medida, é possível plotar um gráfico e verificar a sua distância. Ao interpretar a distância de Cook, geralmente se considera que valores abaixo de 1 indicam uma influência mínima ou baixa, enquanto valores acima de 1 podem ser considerados como indicativos de uma influência mais significativa. No entanto, a interpretação da distância de Cook depende do tamanho do conjunto de dados, do número de variáveis independentes e do contexto específico da análise (AGRANONIK, 2005).

### **Matriz de confusão para as estimativas do modelo.**

A matriz de confusão é uma representação tabular usada para avaliar o desempenho de um modelo de classificação ou regressão em relação às classes verdadeiras dos dados. Ela fornece uma visão detalhada dos resultados das previsões do modelo, permitindo a análise de erros e acertos em cada classe. É especialmente útil quando trabalhamos com problemas de classificação binária, ou seja, quando temos duas classes distintas, como "positivo" e "negativo". No entanto, também pode ser adaptada para problemas de classificação multiclasse (CASTRO; BRAGA, 2011).

A matriz é definida com base na Tabela 3.1:

Tabela 3.1: Exemplo de matriz de confusão

		Valor Predito	
		Sim	Não
Real	Sim	Verdadeiro Positivo (VP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (VN)

Fonte: Elaboração própria.

- Verdadeiro Positivo (VP): O modelo previu corretamente uma instância como positiva. Essas são as instâncias que são verdadeiramente positivas e também foram classificadas corretamente como positivas pelo modelo;
- Falso Negativo (FN): O modelo previu incorretamente uma instância como negativa quando, na verdade, era positiva. Essas são as instâncias que são verdadeiramente positivas, mas foram classificadas erroneamente como negativas pelo modelo;
- Falso Positivo (FP): O modelo previu incorretamente uma instância como positiva quando, na verdade, era negativa. Essas são as instâncias que são verdadeiramente negativas, mas foram classificadas erroneamente como positivas pelo modelo;
- Verdadeiro Negativo (VN): O modelo previu corretamente uma instância como negativa. Essas são as instâncias que são verdadeiramente negativas e também foram classificadas corretamente como negativas pelo modelo.

### 3.6.5 Análise de pontos influentes (*Outlier*)

A análise de *outliers* é uma técnica utilizada para identificar observações que são inconsistentes ou atípicas em relação ao restante dos dados, ou seja, é um valor valor que está significativamente afastado dos demais valores da amostra afetando as estatísticas do modelo. Existe várias formas de tratar esses pontos influentes, mas nesse trabalho utilizaremos o gráfico dos resíduos padronizados versus alavancagem. Onde um ponto vai ser considerado possível ponto discrepante, se ficar fora do intervalo padronizado  $[-0,5;0,5]$ .

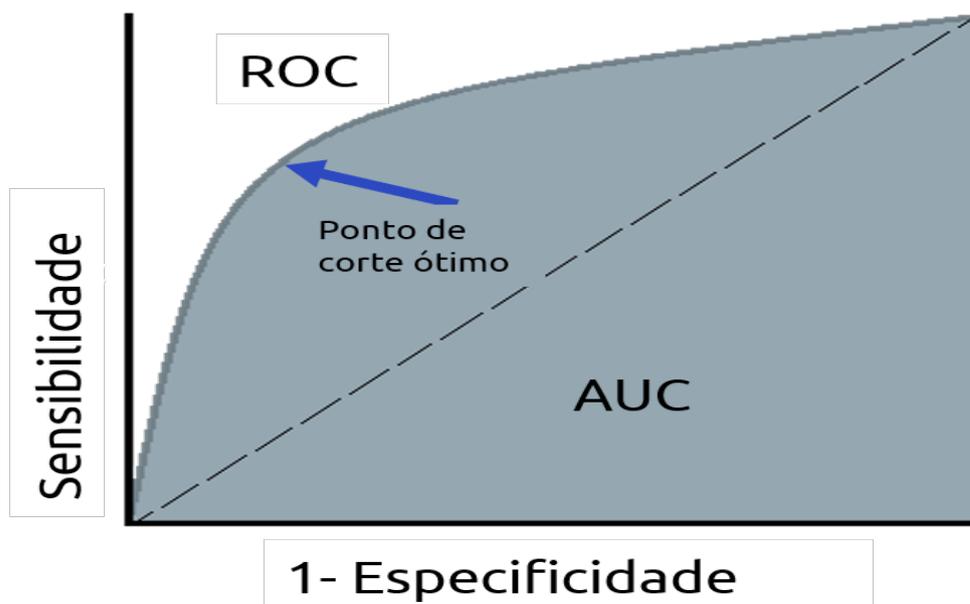
### 3.6.6 Curva ROC

A curva ROC (*Receiver Operating Characteristics*) mede a capacidade da predição do modelo, onde é uma forma de observarmos a qualidade do ajuste para o modelo obtido com base nas respectivas taxas de erro e acerto do modelo, se dando por meio da análise da matriz de contingência, que nos possibilita obter um ponto de corte que maximiza a relação

entre a taxa de acerto e de falso positivo para o respectivo evento de interesse do estudo. A área abaixo da curva ROC (Figura 3.1), é chamada de AUC (*Area Under the ROC Curve*), basicamente compara os classificadores da curva em um único valor, indicando a probabilidade do modelo realizar previsões corretas (FAWCET, 2006). Onde é importante destacar que esse valor da AUC sempre vai variar de 0 a 1.

Frequentemente o valor mais utilizado é 0,5, porém nesse trabalho o ponto de corte será o que maximiza ambas as curvas de sensibilidade e especificidade mostrado na Figura 3.1. A sensibilidade é definida como a probabilidade do teste fornecer um resultado positivo, e a especificidade é definida como a probabilidade do teste fornecer um resultado negativo. O ponto de corte ideal, é aquele ponto que maximiza a probabilidade de sucesso e ao mesmo tempo diminua a probabilidade de erro do modelo que será abordado.

Figura 3.1: Curva ROC



Quadro 3.1: Descrição das variáveis do estudo

V1	Apgar no primeiro minuto	0-3 Asfixia grave 4-6 Asfixia moderada 6-7 Asfixia leve 8-10 Boa vitalidade
V2	Apgar no quinto minuto	0-3 Asfixia grave 4-6 Asfixia moderada 6-7 Asfixia leve 8-10 Boa vitalidade
V3	Sexo do bebê	Masculino Feminino
V4	Estado civil	Situação conjugal Solteira Casada/União estável Viúva/separada judicialmente/divorciada
V5	Semanas de gestação	Prematuro (<37 semanas) Não prematuro ( $\geq 37$ semanas)
V6	Instrução da mãe	$\leq 7$ anos de estudo (fora da idade escolar) $\geq 8$ anos de estudos (dentro da idade escolar)
V7	Município de residência	Código IBGE
V8	Número de consultas pré-natal	Nenhuma 1 a 3 consultas 4 a 6 consultas 7 e mais consultas
V9	Tipo de gravidez	Única Dupla Tripla +
V10	Tipo de parto	Vaginal Cesáreo
V11	Raça/cor	Branca Preta Amarela Parda Indígena
V12	Anomalia congênita	Sim Não
V13	Idade da mãe (anos)	15-19 20-24 25-29 30-34 35-39 40-44 45-49
V14	Local de ocorrência do nascimento	Hospital Domicílio Outro estabelecimento de Saúde Outros
V15	Peso ao nascer	< 2500 gramas (baixo peso) $\geq 2500$ gramas (peso normal)

Nesse capítulo foram apresentados os resultados referentes as discussões abordadas anteriormente sobre o Semiárido brasileiro. Está subdividido nas seguintes sessões, respectivamente: análise da qualidade dos registros dos nascimentos oriundos das mulheres em idade fértil do Semiárido brasileiro, e avaliação do impacto da COVID-19 sobre ele; discussão sobre as principais variáveis materno-infantis que constam nas Declarações de Nascidos Vivos; análise do comportamento dos níveis da fecundidade; e modelagem da fecundidade.

### **4.1 Qualidade dos registros dos nascidos vivos e avaliação dos impactos da pandemia da COVID-19**

#### **4.1.1 Cobertura dos registros de nascimentos**

A avaliação da cobertura das informações é de extrema importância em uma análise estatística, especialmente ao lidar com bancos de dados provenientes de questionários. Assim, com base nas estimativas feitas pela Secretaria de Vigilância em Saúde, na Tabela 4.1, podemos observar as estimativas de cobertura para os estados que compõem o Semiárido brasileiro no período de 2010 a 2021, as quais utilizaremos como referência em nosso estudo. É importante ressaltar que essas estimativas de cobertura são referentes aos níveis estaduais e não municipais. Muitos municípios ainda apresentam problemas em sua cobertura, mesmo que os níveis de cobertura sejam satisfatórios em nível estadual,

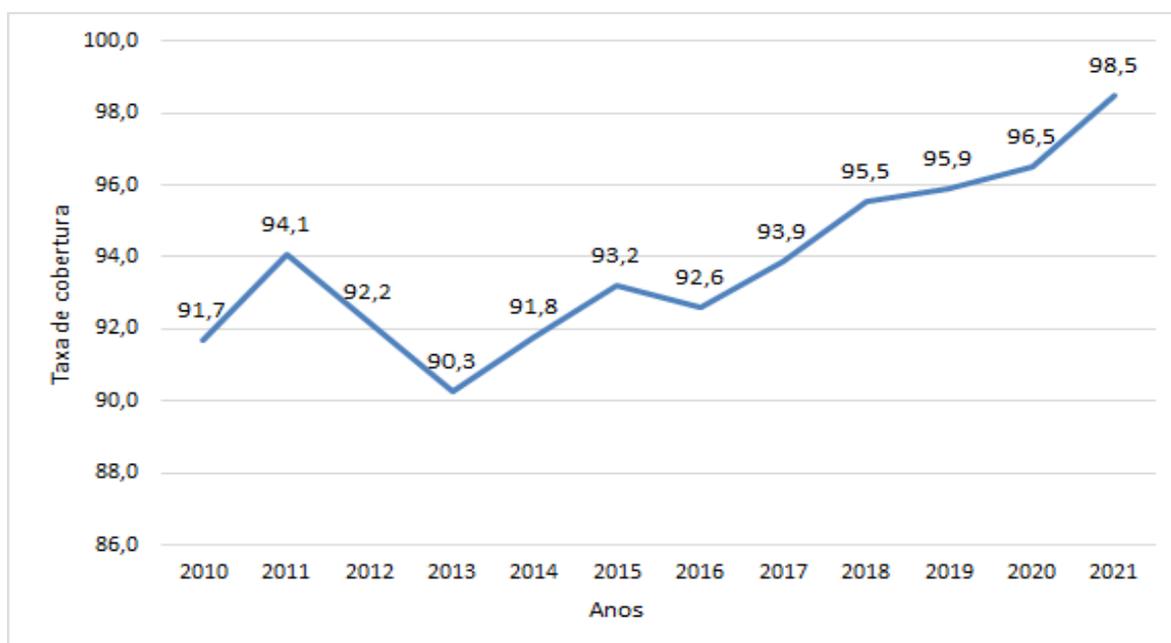
Tabela 4.1: Porcentagem da cobertura do registro do nascidos vivos dos estados que compõem o Semiárido brasileiro, 2010-2020.

Unidades da Federação	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Maranhão	87,5	90,2	86,3	83,4	84,5	86,6	86,2	88,0	91,0	91,4	94,8	97,9
Piauí	92,3	95,8	91,8	90,0	93,1	92,7	94,9	95,0	96,9	97,1	96,2	98,1
Ceará	92,9	93,8	91,6	90,7	93,1	93,5	95,1	94,6	95,1	95,5	96,2	98,3
Rio Grande do Norte	92,8	95,4	93,6	91,8	93,6	94,0	96,3	95,1	97,0	97,3	97,1	99,1
Paraíba	93,5	95,0	92,8	92,8	94,1	99,7	95,0	96,0	98,0	98,2	97,7	99,6
Pernambuco	92,6	95,2	95,2	94,4	95,8	98,1	97,0	98,0	98,3	98,7	97,0	99,5
Alagoas	91,7	93,3	91,8	91,3	91,8	92,4	91,2	93,0	93,3	93,6	96,8	98,2
Sergipe	93,6	97,4	93,9	95,2	96,3	96,4	96,1	96,3	96,5	96,8	96,2	98,4
Bahia	91,7	94,5	93,4	89,6	90,9	91,7	90,0	93,0	95,8	96,2	96,9	98,1
Minas Gerais	96,2	97,8	97,2	95,0	99,2	99,1	97,0	97,0	97,2	97,6	98,0	99,1
<b>Brasil</b>	<b>95,1</b>	<b>97,1</b>	<b>96,2</b>	<b>95,0</b>	<b>96,6</b>	<b>97,0</b>	<b>96,5</b>	<b>97,0</b>	<b>97,6</b>	<b>97,8</b>	<b>98,2</b>	<b>99,0</b>

Fonte: Sistema de Informações sobre Nascidos Vivos (SINASC), 2022.

conforme demonstrado na Tabela 4.1.

Figura 4.1: Cobertura do registro de nascidos vivos do Nordeste brasileiro, 2010-2021



Fonte: Sistema de Informações sobre Nascidos Vivos (SINASC). Secretaria de Vigilância em Saúde, 2022.

Observando a Tabela 4.1 e o Gráfico 4.1, verifica-se que, desde 2010, as coberturas de nascidos vivos eram relativamente boas, com a maioria dos estados apresentando taxas acima de 90%, exceto o estado do Maranhão. No caso do Maranhão, pode-se observar que o aumento da cobertura de nascidos vivos ocorreu de forma mais lenta em comparação com outros estados. Em 2010, a cobertura era de aproximadamente 87,5%, e em 2011 houve um aumento de 2,7% em relação ao ano anterior. No entanto, após 2011, houve

uma queda na cobertura até iniciar uma recuperação gradual a partir de 2014. Entretanto, verifica-se novamente um leve decaimento da cobertura do estado no ano de 2016, onde um dos fatores que podem está associado a essa queda de 0,4% foi epidemia de dengue que o Brasil vinha passando nesse ano (PAGNO, 2022). Em 2020, que foi o início da pandemia da COVID-19, não foi observado um impacto negativo sobre o estado, isso pode está associado aos poucos meses de pandemia que esse ano vinha sofrendo. Porém, em 2021, constata-se a permanência desse cenário sobre a cobertura dos dados da DNV do referente ano, mostrando a constante evolução dos índices de cobertura que o Brasil vem sofrendo ao decorrer dos anos.

Em relação aos outros estados, observamos que desde 2010, todos vinham apresentando avanços na cobertura dos registros dos nascidos vivos, sendo notório a evolução dos estados no qual o Semiárido está presente. Em 2021, observamos que os estados do Rio Grande do Norte, Paraíba, Pernambuco e Minas Gerais foram os estados que obtiveram os melhores percentuais da cobertura do registro dos nascidos vivos, destacando a Paraíba e Pernambuco, que apresentaram uma porcentagem de cobertura de 99,6% e 99,5%. Comprovando ausência dos impactos da COVID-19 ainda nesse ano.

Falando agora em um parâmetro mais geral, na Figura 4.1, foi possível observar a cobertura no Nordeste brasileiro, onde o padrão se mantém. Em 2010, a região apresentava uma cobertura de 91,7%, com um leve aumento em 2011, seguido por uma queda que pode estar relacionada à epidemia da dengue mencionada anteriormente, assim como em 2016. No entanto, em 2021, a região apresenta uma boa cobertura, como constatado anteriormente na maioria dos estados que compõem o Nordeste e o Semiárido brasileiro.

Os avanços na cobertura dos registros dos nascidos vivos no Semiárido foram notórios, contudo, ainda é preciso aprimorar para que cada dia mais tenhamos a nosso dispor informações menos viesadas e mais próximas da realidade de cada região.

#### **4.1.2 Incompletude dos registros de nascimentos**

A identificação das fragilidades na completude dos registros de nascidos vivos nos auxiliam a detectar os principais obstáculos que afetam a coleta e a consolidação das informações. Alguns exemplos desses obstáculos incluem dificuldades no preenchimento das informações por parte dos profissionais de saúde, veracidade das informações preenchidas, entre outros.

Sendo assim, procurando identificar a situação da completude das variáveis da Declaração de Nascidos Vivos do SINASC referente ao Semiárido brasileiro, a Tabela 4.2 mostra o

Tabela 4.2: Número de registros vazios e percentual de incompletude das variáveis das declarações de nascidos vivos do Semiárido brasileiro, 2019 e 2021

Variáveis	2019		2021		Comparativo
	n	%	n	%	
Município de residência	0	0,00	0	0,00	0
Local de ocor. Do nascimento	8	0,00	13	0,00	5
Idade da mae	0	0,00	0	0,00	0
Estado civil da mãe	11.576	2,99	7.899	2,14	-3.677
Semanas de gestação	9.878	2,55	9.738	2,64	-140
Tipo de gravidez	300	0,08	515	0,14	215
Tipo de parto	386	0,10	524	0,14	138
Numero de consultas pré-natal	1.526	0,39	1.930	0,52	404
Apgar 1° minuto	11.222	2,90	7.239	1,96	-3.986
Apgar 5° minuto	11.252	2,90	7.238	1,96	-4.014
Raça/cor da mãe	18.910	4,88	15.520	4,20	-3.390
Peso ao nascer	46	0,01	36	0,01	-10
Anomalia congênita	9.237	2,38	6.763	1,83	-2.474
Escolaridade da mãe	11.576	2,99	7.428	2,01	-4.148
Total de Declarações Vazias	85.917		64.843		-21.074
Total de Declarações	387.576		369.534		-18.042

Fonte: Sistema de Informações sobre Nascidos Vivos (SINASC).

número de caselas vazias e o percentual de incompletude de quatorze variáveis das quinze selecionadas para este estudo, levando em consideração as informações tidas como vazias, em branco, não informado e ignorados. É importante ressaltar que a variável "sexo" não foi levada em consideração por conta de na hora do preenchimento ser atribuído a resposta ignorado para bebês que não houveram como identificar o sexo, ou que nasceram com casos de hermafroditismo (MINISTÉRIO DA SAÚDE, 2023). Além disso, a Tabela 4.2 apresentará o total geral de declarações vazias, bem como a quantidade total de declarações para cada ano correspondente.

Conforme a Tabela 4.2, foi possível observar o número de vazios e o percentual de incompletude das variáveis das declarações de nascidos vivos. No ano de 2019, período pré pandêmico, notamos que as variáveis com maiores problemas na sua completude foi a raça/cor da mãe, com um total de 18.910 ignorados, representando 4,88% do total geral, além do estado civil da mãe e das semanas de gestação, ambas com 11.576 ignorados, com um percentual de 2,99%. Além disso, olhando para um parâmetro geral das variáveis, observamos que nesse ano houve um total de 85.917 caselas vazias nas demais declarações.

Em relação ao ano de 2021, período da pandemia da COVID-19, notamos algumas mudanças em relação à completude das variáveis. Primeiramente, vemos um aumento de vazios nas variáveis tipo de gravidez, parto e dos números de consultas pré-natal. Isso se deve principalmente pela diminuição dos acompanhamentos pré-natal motivada

pelas diversas interrupções dos serviços de saúde da mulher, provocada pela pandemia da COVID-19 (OPAS, 2022).

Por outro lado, observou-se uma diminuição dos vazios nas variáveis que apresentaram maiores problemas no ano de 2019, como por exemplo, a raça/cor da mãe, com uma queda de 3.390 ignorados, estado civil da mãe com uma queda de 3.677 ignorados, semanas de gestação com uma leve queda de 140 ignorados. Mostrando assim uma melhora na completude dessas informações e também no total geral de vazios no comparativo dos dois períodos. Observamos que no ano de 2019 tínhamos um total de 85.917 vazios, e no ano pandêmico de 2021 um total de 64.843 vazios, totalizando uma redução de 21.074 vazios. Dessa maneira, os dados foram na do que era previsto acontecer para esse ano.

Em relação à quantidade geral de declarações de nascidos vivos do Semiárido brasileiro, vemos que no ano de 2019 foram registrados 387.576 declarações, enquanto no ano de 2021 foram registrados 369.534 declarações, mostrando assim uma queda de 2,38%, totalizando 18.042 declarações.

## 4.2 Caracterização das variáveis das Declarações de Nascidos Vivos

Nessa serão apresentadas as frequências absolutas e a porcentagem das respectivas categorias das variáveis da DNVs consideradas para esse estudo. Onde serão apresentadas duas perspectivas de análises, uma considerando todo o Semiárido, e a outra irá trazer uma perspectiva ao nível estadual do Semiárido brasileiro.

### 4.2.1 Anomalias congênitas

A variável anomalia congênita refere-se a uma condição médica presente em um indivíduo desde o nascimento, causada por fatores genéticos ou ambientais durante o desenvolvimento pré-natal. Na Tabela 4.3, vemos o preenchimento dessa variável para o todo do Semiárido brasileiro, e sua respectiva categorização. De 369.534 declarações, cerca de 359.941 não tiveram nenhum tipo de anomalia detectado, correspondendo a 97% do total geral. Além disso, foram apresentadas 2.830 declarações com anomalias, e 6.763 ignorados, resultado respectivamente em 1% e 2% do total geral para o ano de 2021.

Observando a Figura 4.2, temos uma perspectiva da região semiárida ao nível estadual.

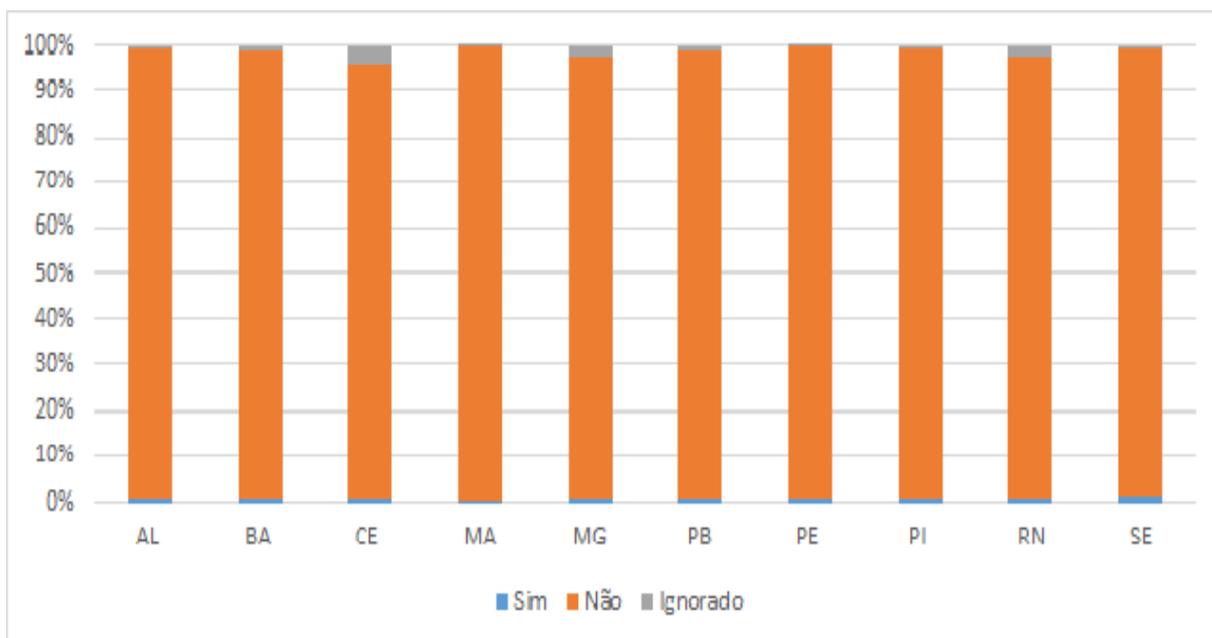
Para quase todos os estados, mais de 90% das declarações coletadas, não apresentaram anomalias, mostrando um bom cenário para o Semiárido. Além disso, foi constatado que os estados que mais apresentaram ignorados em suas declarações, foi o estado do Ceará (5%), Minas Gerais (3%) e o Rio Grande do Norte (3%) do total de declarações gerais de cada estado.

Tabela 4.3: Preenchimento da variável anomalias congênicas para o Semiárido brasileiro, 2021.

Categoria	n	%
Sim	2.830	1
Não	359.941	97
Ignorado	6.763	2
Total	369.534	100

Fonte: Sistema de Informações sobre Nascidos Vivos, 2022.

Figura 4.2: Preenchimento da variável anomalias congênicas dos bebês das maes de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021.



Fonte: Sistema de Informações sobre Nascidos Vivos (SINASC). Secretaria de Vigilância em Saúde, 2022.

## 4.2.2 Apgar 1° minuto

A pontuação de Apgar é um teste realizado em recém-nascidos para avaliar sua condição física e determinar se eles necessitam de cuidados médicos imediatos para diagnóstico de possíveis complicações. Os reflexos do recém-nascido, a respiração, a frequência cardíaca, entre outros aspectos, são avaliados (LESLIE V, 2023). O teste é realizado no primeiro minuto de vida do recém-nascido e novamente cinco minutos após o nascimento.

Na Tabela 4.4, observamos o preenchimento da variável Apgar 1 ° minuto para todo o Semiárido brasileiro e suas respectivas categorias. Das 369.534 DNVs, observamos que 315.231 foram enquadradas como bebês que tiveram ótimas condições após parto, ou seja, tiveram uma pontuação de 8 a 10. Além disso, observamos um total de 34.125 bebês com dificuldades leves, enquadradas na pontuação de 6 a 7 na escala Apgar. Por outro lado, nas dificuldades moderada e grave, temos um total de 10.300 e 2.639 bebês nas respectivas categorias, totalizando 4% do total geral. Além disso, observamos um total de 7.239 DNV com preenchimento incompleto, representando 2% do total geral para todo o Semiárido brasileiro.

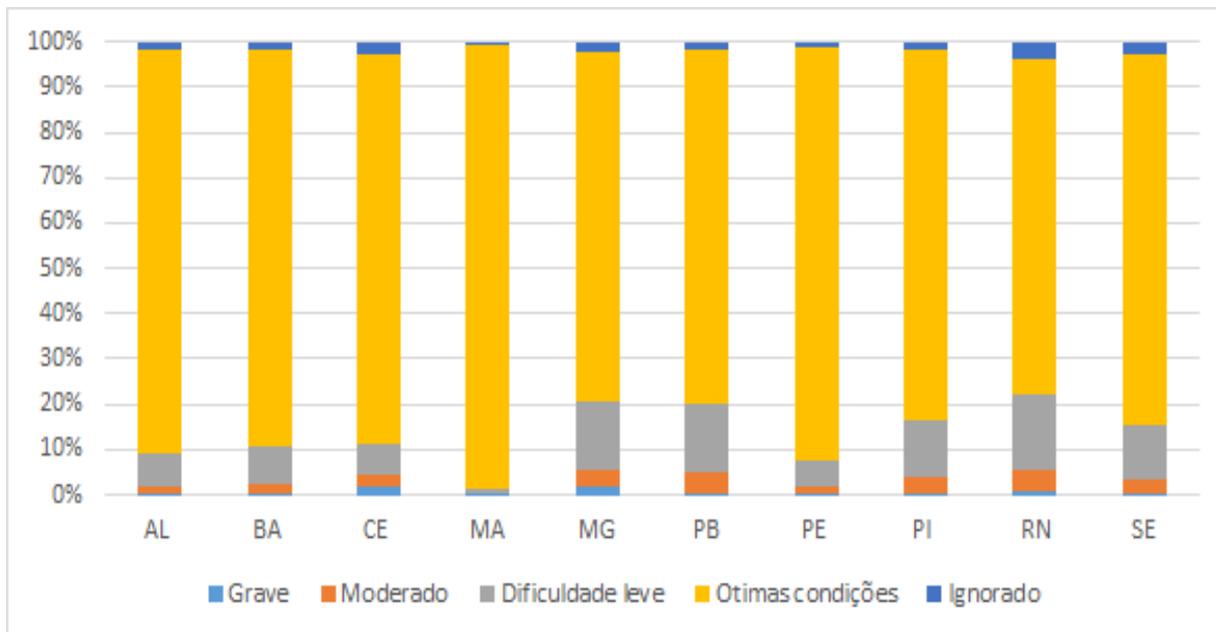
Ao analisar a Figura 4.3, foi possível confirmar o que foi mencionado anteriormente sobre a distribuição das categorias. Em todos os 10 estados analisados, foi observado que, no mínimo, 75% das declarações de nascidos vivos foram classificadas como bebês nascidos em ótimas condições. Além disso, os estados do Rio Grande do Norte, Paraíba e Minas Gerais apresentaram as maiores porcentagens na categoria de dificuldades leves, com 17%, 15% e 15% das declarações, respectivamente. No entanto, é importante destacar que as porcentagens das demais categorias foram baixas, o que indica que os bebês nascidos de mães do Semiárido estão, em sua maioria, saudáveis e com poucas complicações pós-parto.

Tabela 4.4: Preenchimento da variável apgar 1° minuto para o Semiárido brasileiro, 2021.

Categoria	n	%
Grave	2.639	1
Moderado	10.300	3
Dificuldade leve	34.125	9
Ótimas condições	315.231	85
Ignorado	7.239	2
Total	369.534	100

Fonte: Sistema de Informações sobre Nascidos Vivos, 2022.

Figura 4.3: Preenchimento da variável apgar 1° minuto dos bebês das mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021.



Fonte: Sistema de Informações sobre Nascidos Vivos (SINASC). Secretaria de Vigilância em Saúde, 2022.

### 4.2.3 Apgar 5° minuto

Assim como a variável apgar 1° minuto, o apgar 5° minuto busca reavaliar a condição física do recém-nascido e verificar se houve melhora ou piora desde o primeiro exame. Reavaliando os 5 aspectos do teste do apgar, que é a frequência cardíaca, respiração, reflexos, tônus muscular e cor da pele do bebê. Sendo assim, na Tabela 4.5 podemos observar o preenchimento do apgar 5° minuto para o Semiárido brasileiro para mães durante todo o seu período reprodutivo do ano de 2021. De uma forma geral, observamos que na avaliação do apgar 5° minuto para os bebês do Semiárido brasileiro, cerca de 96% das declarações foram classificadas na categoria de 8-10 no qual foi atribuído como ótimas condições, esse representando 353.801 declarações.

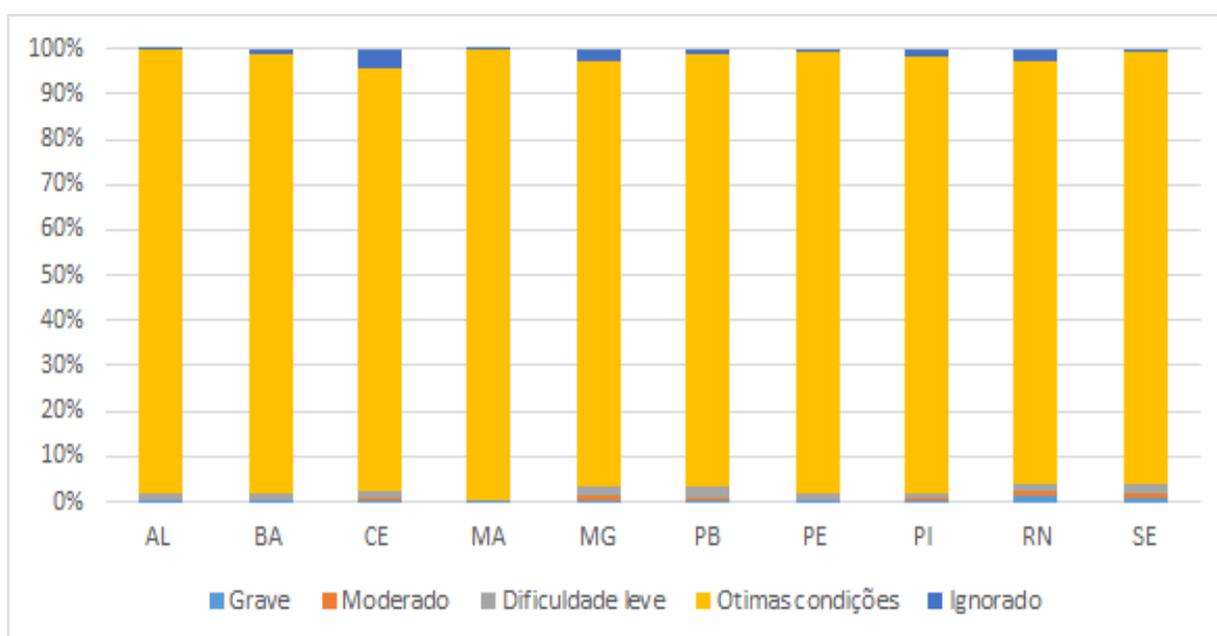
Olhando a Figura 4.4, podemos observar o gráfico com a distribuição do preenchimento por estado do Semiárido brasileiro. Assim como no apgar 1° minuto, observamos uma melhora nas condições dos recém-nascidos, visto que pra todos os estados, esses obtiveram no mínimo 95% do preenchimento das suas DNVs classificadas como ótimas condições, mostrando assim, ótimas condições de saúde após o parto para toda a região.

Tabela 4.5: Preenchimento da variável apgar 5° minuto para o Semiárido brasileiro, 2021.

Categoria	n	%
Grave	1.144	0
Moderado	1.585	0
Dificuldade leve	5.766	2
Ótimas condições	353.801	96
Ignorado	7.238	2
Total	369.534	100

Fonte: Sistema de Informações sobre Nascidos Vivos, 2022.

Figura 4.4: Preenchimento da variável apgar 5° minuto dos bebês das mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021.



Fonte: Sistema de Informações sobre Nascidos Vivos (SINASC). Secretaria de Vigilância em Saúde, 2022.

#### 4.2.4 Quantidade de consultas de pré-natal

Em relação a variável consultas de pré-natal, é indispensável para mãe e para o bebê um acompanhamento adequado. Pois é através desse acompanhamento que os médicos conseguem identificar a saúde do bebê e da mãe durante todo o período da gestação, evitando possíveis complicações e problemas futuros. Sendo assim, de acordo com a tabela 4.6, podemos ver o preenchimento da variável quantidade de consultas pré-natal para o Semiárido brasileiro, onde cerca de 75% das mães tiveram 7 ou mais consultas de pré-natais, totalizando 275.474 declarações com essa categoria. Mostrando assim, um bom acompanhamento dos bebês e das mães durante o período da pandemia da COVID-19 em 2021. Por outro lado, vemos que cerca de 6% das mães tiveram menos de 3 consultas

durante a gestação, totalizando 21.284 declarações.

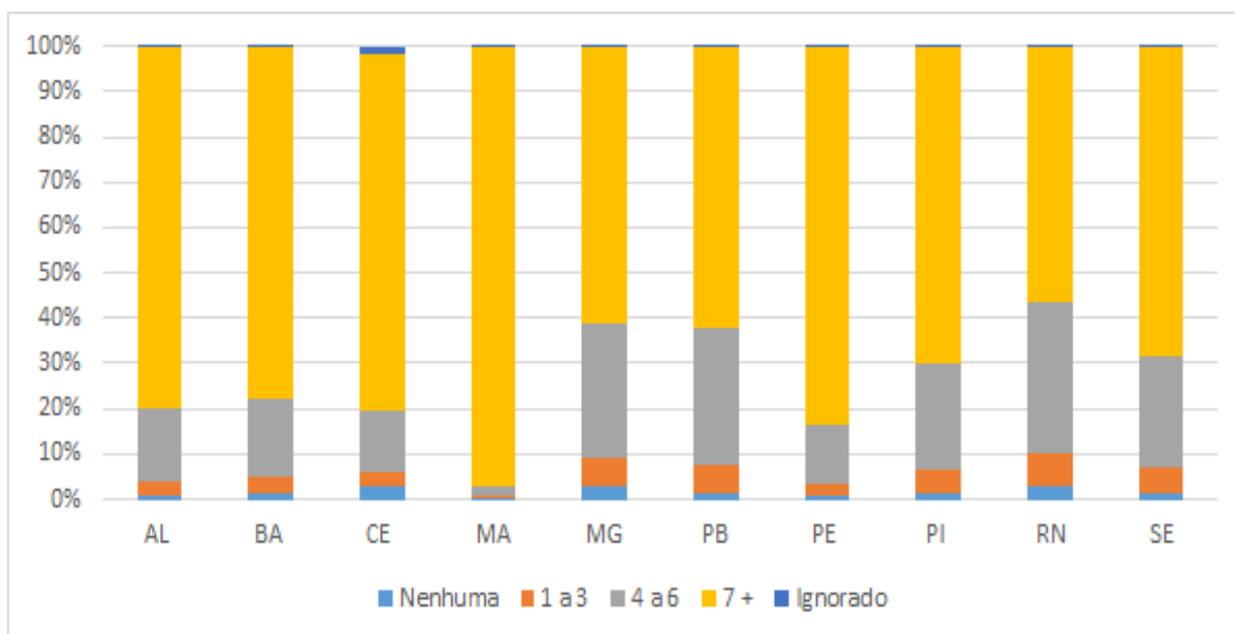
Na Figura 4.5, observou-se a distribuição das consultas pré-natal das mulheres do Semiárido, por seu respectivo estado. E olhando o Gráfico 4.5, observou-se que os estados que apresentaram uma menor quantidade de consultas pré-natal, foi o estado de Minas Gerais e do Rio Grande do Norte, apresentando cerca de 10% das declarações na categoria de até 3 consultas durante todo o período da gestação. Por outro lado, foi visto um bom exemplo do acompanhamento pré-natal das mulheres do Maranhão, onde mais de 90% de todas as declarações, mostrou que as mulheres tiveram mais de 7 consultas durante toda a gestação, mostrando um ótimo cenário para os municípios.

Tabela 4.6: Preenchimento da variável quantidade de consultas pré-natal para o Semiárido brasileiro, 2021.

Categoria	n	%
Nenhuma	6.539	2
1 a 3	14.745	4
4 a 6	70.846	19
7 +	275.474	75
Ignorado	1.930	1
Total	369.534	100

Fonte: Sistema de Informações sobre Nascidos Vivos, 2022.

Figura 4.5: Preenchimento da variável quantidade de consultas pré-natal das mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021.



Fonte: Sistema de Informações sobre Nascidos Vivos (SINASC). Secretaria de Vigilância em Saúde, 2022.

## 4.2.5 Duração da gestação em semanas

Segundo o Ministério da Saúde, o tempo de gestação ideal tem sido definido em 9 meses ou respectivamente 40 semanas. De acordo com o órgão, se o bebê nascer antes das 37 semanas, é considerado prematuro, e fora disso, não prematuro. Sendo assim, na Tabela 4.7 observo-se o preenchimento da duração da gestação para o Semiárido, que foi classificado conforme as recomendações do Ministério da Saúde falado anteriormente. Para todo o semiárido brasileiro, cerca de 86% dos nascimentos foram não prematuro, correspondendo a 317.531 declarações. Por outro lado, vemos 11% dos nascimentos enquadrados como prematuro, e 3% como ignorados, correspondendo respectivamente a 42.265 e 9.738 declarações.

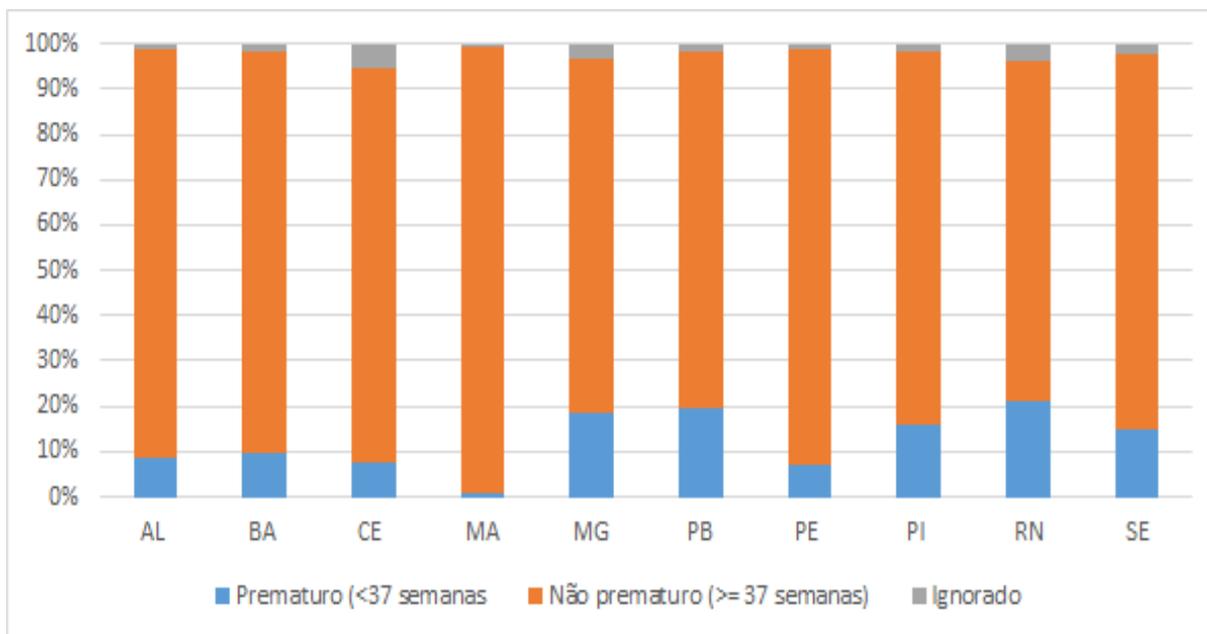
Em relação aos estados, observo-se na Figura 4.6, o gráfico correspondendo a distribuição da gestação das mães do Semiárido por estado. Sendo assim, o estado do Maranhão foi o que apresentou a menor quantidade de gestações prematuras e de declarações ignoradas, ficando abaixo de 5% a representação dos dois. Por outro lado, os estados de Minas Gerais, Paraíba e Rio Grande do Norte, foram os que apresentaram uma maior quantidade de gestações prematuras, chegando a aproximadamente 20% das declarações de cada estado.

Tabela 4.7: Preenchimento da variável duração da gestação para o Semiárido brasileiro, 2021.

Categoria	n	%
Prematuro (<37 semanas )	42.265	11
Não prematuro ( $\geq$ 37 semanas)	317.531	86
Ignorado	9.738	3
Total	369.534	100

Fonte: Sistema de Informações sobre Nascidos Vivos, 2022.

Figura 4.6: Preenchimento da variável duração da gestação das mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021.



Fonte: Sistema de Informações sobre Nascidos Vivos (SINASC). Secretaria de Vigilância em Saúde, 2022.

#### 4.2.6 Escolaridade

Em relação à variável escolaridade das mães dos municípios do Semiárido brasileiro, com base na Tabela 4.9, observamos que 78% das mães estão dentro da idade escolar, que corresponde a mais de 7 anos de estudo, representando 287.751 mulheres. Por outro lado, 20% das mães do Semiárido estão fora da idade escolar, correspondendo a menos de 7 anos de estudo, representando 74.355 mulheres que foram mães durante o ano de referência.

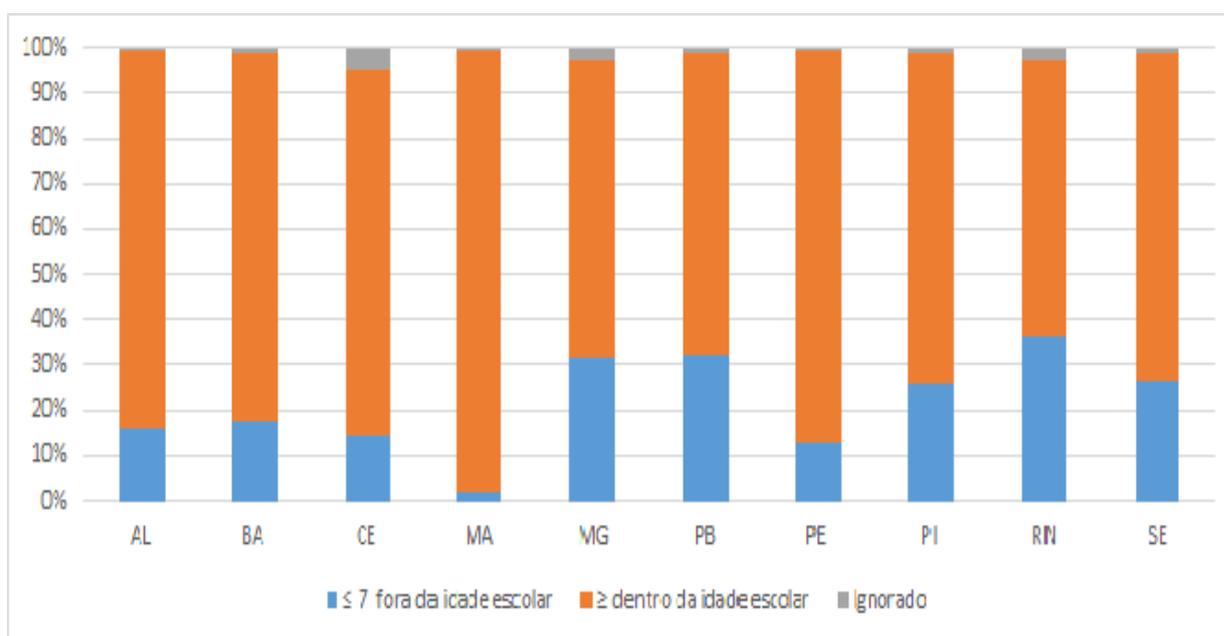
a Figura 4.7, foi possível visualizar o gráfico relacionado ao preenchimento da escolaridade das mães do Semiárido, distribuído por estado. Fica evidente que, em todos os estados, mais de 50% das mulheres se encontravam dentro da faixa etária escolar. Por outro lado, os estados que apresentaram uma maior porcentagem de mulheres fora da idade escolar foram Minas Gerais, Paraíba e Rio Grande do Norte, representando aproximadamente 35% do total de mulheres em seus respectivos estados.

Tabela 4.8: Preenchimento da variável escolaridade da mãe para o Semiárido brasileiro, 2021.

Categoria	n	%
≤ 7 fora da idade escolar	74.355	20
≥ 8 dentro da idade escolar	287.751	78
Ignorado	7.428	2
Total	369.534	100

Fonte: Sistema de Informações sobre Nascidos Vivos, 2022.

Figura 4.7: Preenchimento da variável escolaridade das mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021.



Fonte: Sistema de Informações sobre Nascidos Vivos (SINASC). Secretaria de Vigilância em Saúde, 2022.

#### 4.2.7 Estado civil

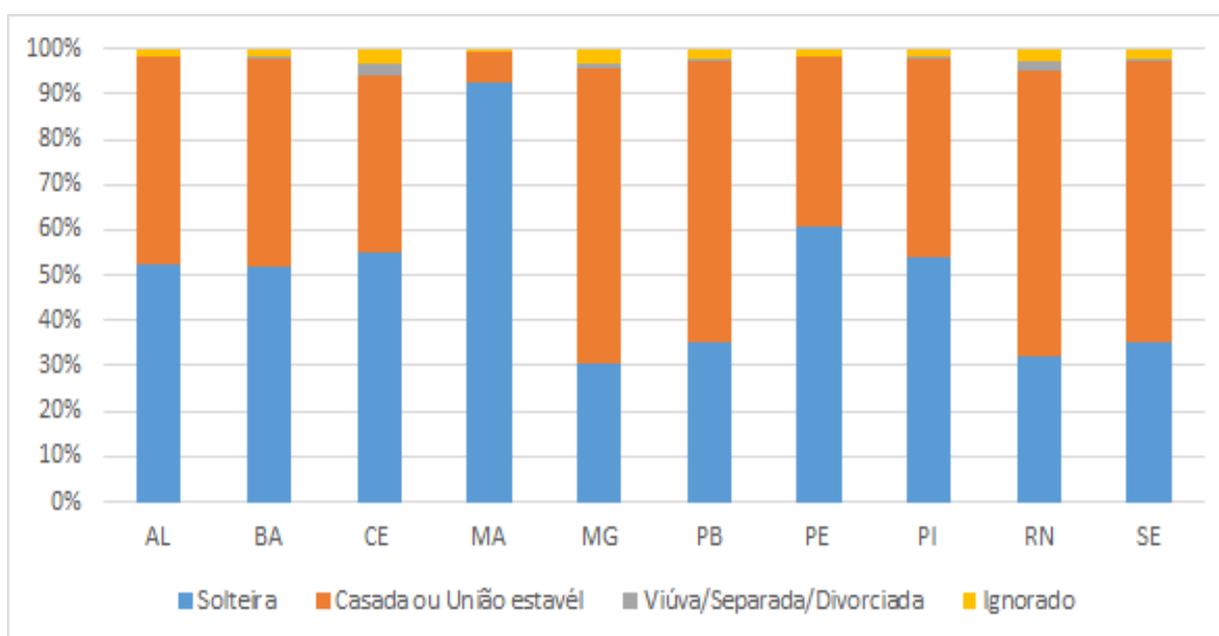
Com base na Tabela 4.9, foi possível observar o preenchimento do estado civil das mães nas declarações. Mais da metade das mães são solteiras, representando 51% do total de declarações, o que equivale a 187.149 declarações. Além disso, 46% das declarações, correspondendo a 170.317 mulheres, são casadas ou estão em união estável. Na Figura 4.8, pode-se observar o gráfico que mostra o preenchimento da variável do estado civil das mulheres do Semiárido, discriminado por estado. É evidente que os estados apresentam uma divisão significativa entre mães solteiras e aquelas que estão casadas ou em união estável.

Tabela 4.9: Preenchimento da variável estado civil da mãe para o Semiárido brasileiro, 2021.

Categoria	n	%
Solteira	187.149	51
Casada ou União estável	170.317	46
Viúva/Separada/Divorciada	4.169	1
Ignorado	7.899	2
Total	369.534	100

Fonte: Sistema de Informações sobre Nascidos Vivos, 2022.

Figura 4.8: Preenchimento da variável estado civil das mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021.



Fonte: Sistema de Informações sobre Nascidos Vivos (SINASC). Secretaria de Vigilância em Saúde, 2022.

#### 4.2.8 Idade das mães

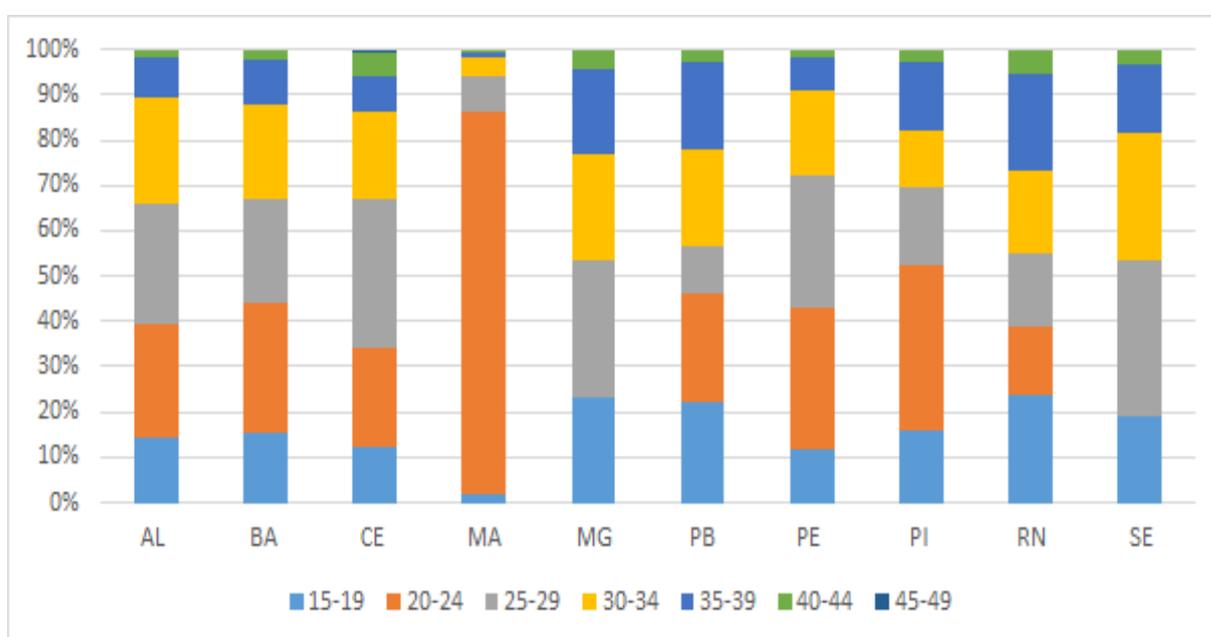
Em relação à variável idade das mães, com baseando-se na Tabela 4.10 e na figura 4.9, observa-se uma maior prevalência de mães de 15 a 29 anos, tendo uma maior frequência de mães de 20 a 24 anos, totalizando 95.219 mulheres, equivalente a 26% do total geral de mulheres. Além disso, logo em seguida temos com um total de 90.371 mães de 25 a 29 anos, e 57.625 mães de 15 a 19 anos, representando 24% e 16% do total geral. Por outro lado, vemos que a partir dos 30 anos, já conseguimos observar uma queda na quantidade de mães, sendo intensificada na faixa etária de 35 a 39 anos conforme observamos na tabela. Dessa maneira, foi possível afirmar que há uma grande prevalência de mães jovens na região.

Tabela 4.10: Preenchimento da variável idade da mãe para o Semiárido brasileiro, 2021.

Categoria	n	%
15-19	57.625	16
20-24	95.219	26
25-29	90.371	24
30-34	71.434	19
35-39	42.436	11
40-44	11.736	3
45-49	713	0
Total	369.534	100

Fonte: Sistema de Informações sobre Nascidos Vivos, 2022.

Figura 4.9: Preenchimento da variável idade das mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021.



Fonte: Sistema de Informações sobre Nascidos Vivos (SINASC). Secretaria de Vigilância em Saúde, 2022.

#### 4.2.9 Local de ocorrência do parto

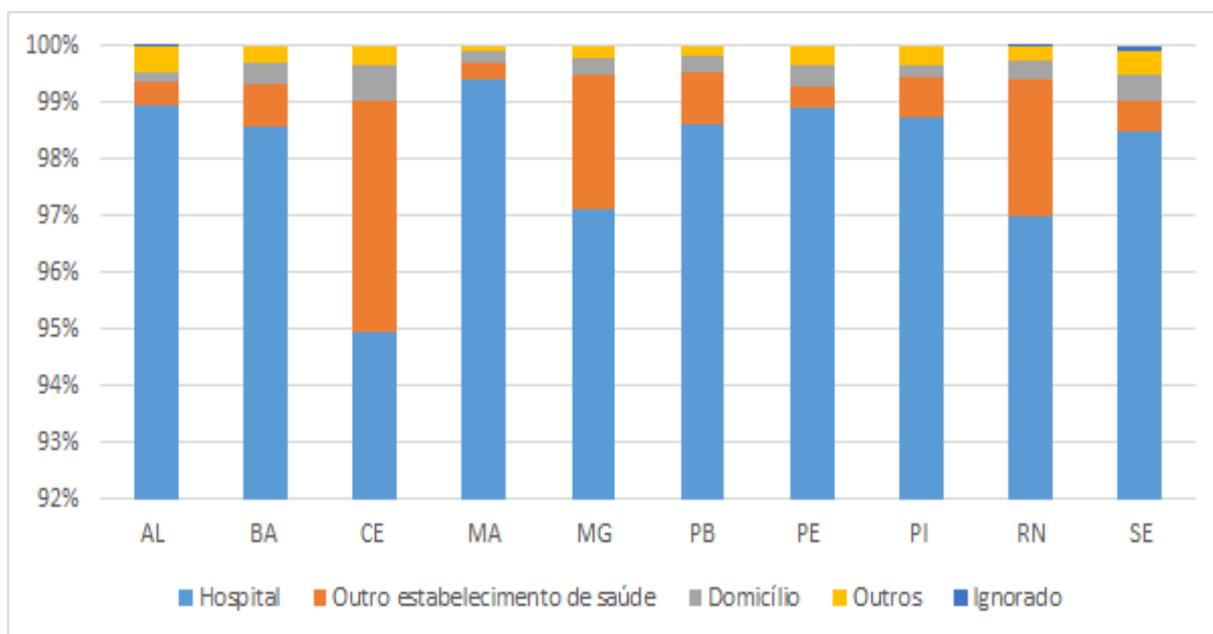
Em relação ao local de ocorrência do parto, a partir de 1940, houve uma grande mudança no local de ocorrência dos partos mundialmente, com grande parte dos partos passando a ser realizados em hospitais. Na Tabela 4.11, observe-se que para as mães do Semiárido brasileiro não foi diferente, 98% dos partos foram realizados em hospitais, representando 361.241 partos. Obtendo uma quantidade bem pequena de partos fora de hospitais e de outros estabelecimentos de saúde. Isso também sendo refletido nos estados como observamos na Figura 4.10 o gráfico referente ao preenchimento do local de ocorrência por seu respectivo estado.

Tabela 4.11: Preenchimento da variável local de ocorrência do parto para o Semiárido brasileiro, 2021.

Categoria	n	%
Hospital	361.241	98
Outro estabelecimento de saúde	5.698	2
Domicílio	1.417	0
Outros	1.165	0
Ignorado	13	0
Total	369.534	100

Fonte: Sistema de Informações sobre Nascidos Vivos, 2022.

Figura 4.10: Preenchimento da variável local de ocorrência do parto das mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021.



Fonte: Sistema de Informações sobre Nascidos Vivos (SINASC). Secretaria de Vigilância em Saúde, 2022.

#### 4.2.10 Peso do bebê

De acordo com o Ministério da Saúde, um bebê é considerado tendo baixo peso quando ao nascer ele apresente um peso inferior a 2.500 gramas. Sendo assim, utilizando essa classificação proposta, na Tabela 4.12, observou-se o preenchimento da variável do peso dos bebês das mães do Semiárido brasileiro durante o ano de 2021. Com base nisso, 92% dos bebês nascidos durante o período de referência nasceram com o peso normal, representando 338.252 bebês, e apenas 31.246 bebês nasceram abaixo do peso.

Além disso, com base na Figura 4.11 observamos o preenchimento do peso do bebê por seu respectivo estado. Sendo assim, para os 10 estados, observamos que no mínimo

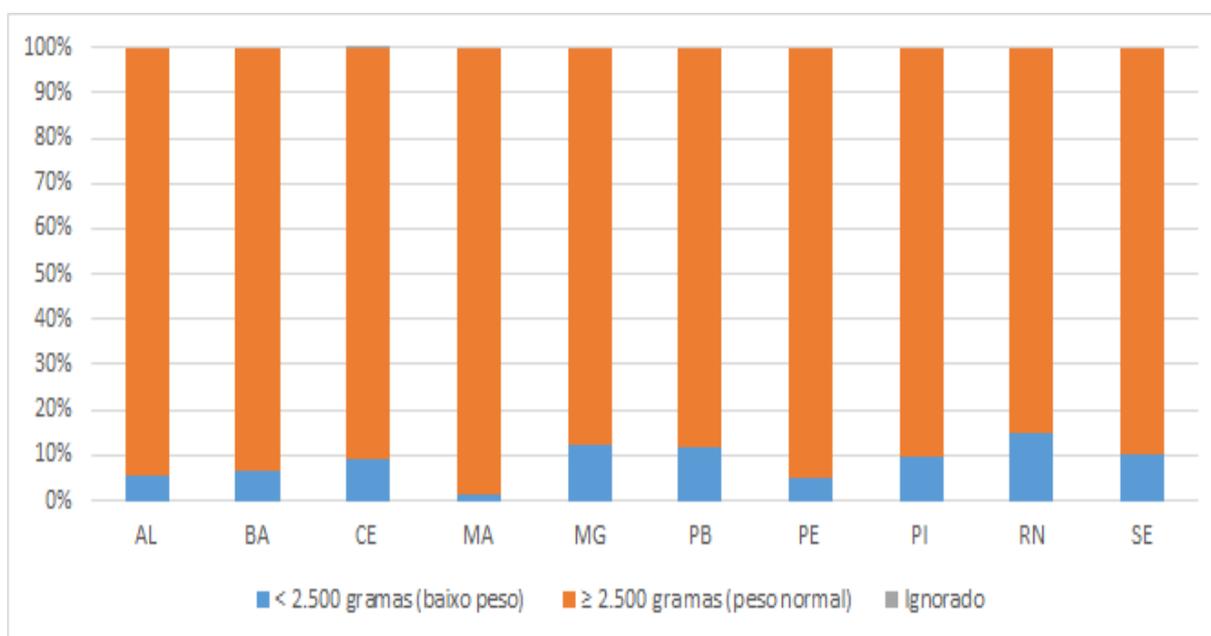
80% dos bebês nasceram como peso normal. Desses estados, o estado do Rio Grande do Norte foi o que apresentou uma maior porcentagem de bebês nascidos abaixo do peso, apresentando uma porcentagem próxima dos 15% dos nascidos.

Tabela 4.12: Preenchimento da variável peso do bebê para o Semiárido brasileiro, 2021.

Categoria	n	%
<2.500 gramas (baixo peso)	31.246	8
≥ 2.500 gramas (peso normal)	338.252	92
Ignorado	36	0
Total	369.534	100

Fonte: Sistema de Informações sobre Nascidos Vivos, 2022.

Figura 4.11: Preenchimento da variável peso dos bebês das mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021.



Fonte: Sistema de Informações sobre Nascidos Vivos (SINASC). Secretaria de Vigilância em Saúde, 2022.

### 4.2.11 Raça/cor da mãe

Em relação ao preenchimento da variável raça/cor das mães do Semiárido brasileiro do ano de 2021, com base na Tabela 4.13 e na Figura 4.12 observamos uma maior prevalência de mulheres na cor parda, essas representando 80% das declarações de nascidos vivos do ano de referência, totalizando 296.597 mães da respectiva cor. Além disso, 11% das mulheres se declararam da cor branca e 4% da cor preta, totalizando 40.836 mulheres brancas e 12.270 mulheres pretas como visto na tabela.

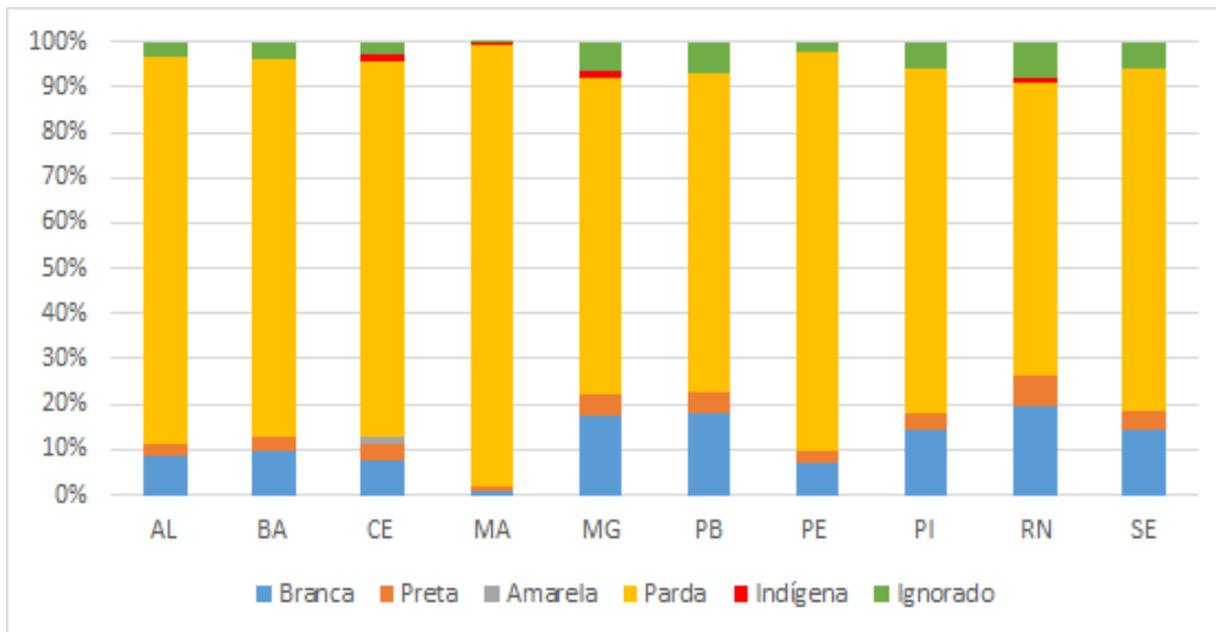
Na Figura 4.12, foi possível constatar o que foi dito através do gráfico apresentado da distribuição desses preenchimentos por estado, observamos uma maior prevalência da cor parda, essa sendo de no mínimo 70% em todos os estados, seguidos pela cor branca que tem uma maior porcentagem no estado do Rio Grande do Norte, Paraíba e Minas Gerais.

Tabela 4.13: Preenchimento da variável raça da mãe para o Semiárido brasileiro, 2021.

Categoria	n	%
Branca	40.836	11
Preta	12.270	4
Amarela	1.189	0
Parda	296.597	80
Indígena	2.122	1
Ignorado	15.520	4
Total	369.534	100

Fonte: Sistema de Informações sobre Nascidos Vivos, 2022.

Figura 4.12: Preenchimento da variável raça das mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021.



Fonte: Sistema de Informações sobre Nascidos Vivos (SINASC). Secretaria de Vigilância em Saúde, 2022.

#### 4.2.12 Sexo do bebê

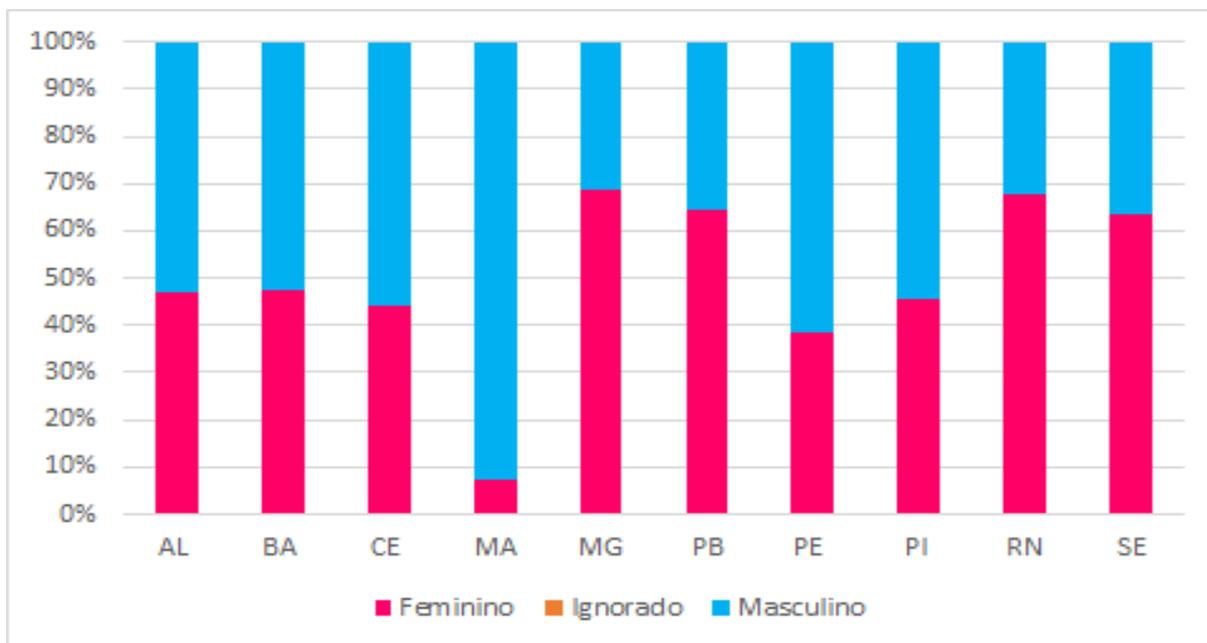
Com base na Tabela 4.14 e na Figura 4.13, foram observadas informações sobre o preenchimento do sexo dos bebês das mães no Semiárido brasileiro durante o ano de 2021. Verificou-se que 51% das declarações informaram que os bebês eram do sexo masculino, totalizando 189.300 bebês do respectivo sexo. Além disso, o restante das declarações correspondeu a bebês do sexo feminino, totalizando 180.155 bebês. Em relação aos estados, observou-se que Alagoas, Bahia, Ceará, Maranhão, Pernambuco e Piauí apresentaram uma maior prevalência de bebês do sexo masculino, enquanto os demais estados tiveram uma predominância de bebês do sexo feminino.

Tabela 4.14: Preenchimento da variável sexo do bebê para o Semiárido brasileiro, 2021.

Categoria	n	%
Masculino	189.300	51
Feminino	180.155	49
Ignorado	79	0
Total	369.534	100

Fonte: Sistema de Informações sobre Nascidos Vivos, 2022.

Figura 4.13: Preenchimento da variável sexo dos bebês das mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021.



Fonte: Sistema de Informações sobre Nascidos Vivos (SINASC). Secretaria de Vigilância em Saúde, 2022.

#### 4.2.13 Tipo de gestação

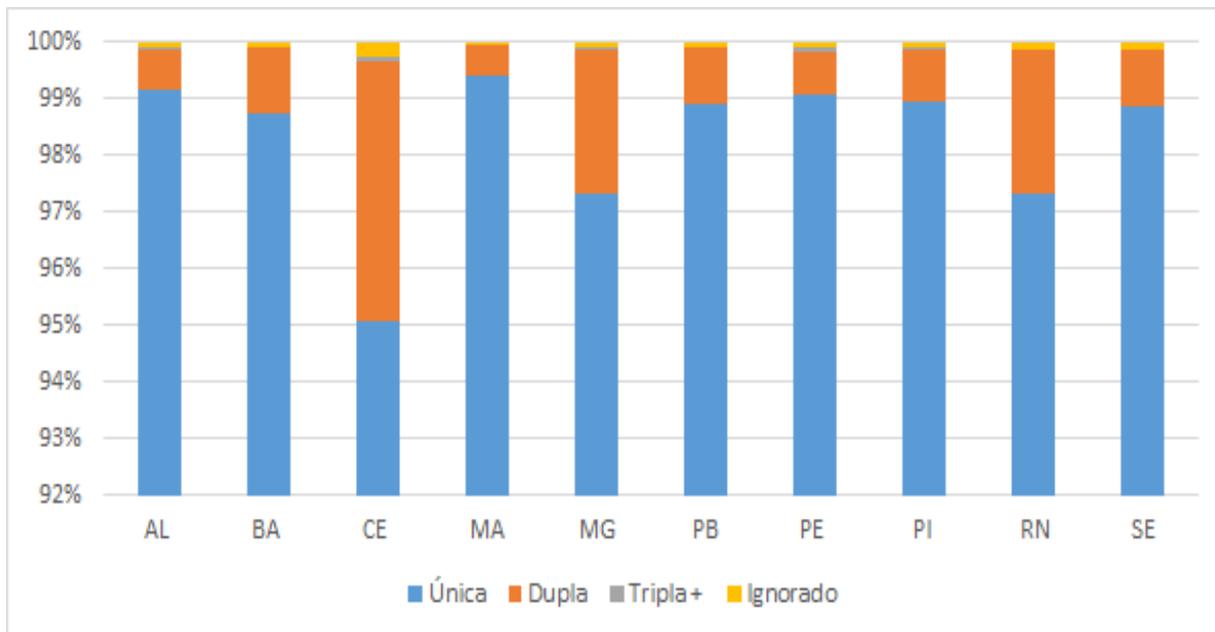
Com base na Tabela 4.15, pode-se observar o preenchimento das declarações referentes ao tipo de gestação das mães no Semiárido brasileiro. Cerca de 98% do total de declarações de nascidos vivos do ano em questão foram marcadas como gestação única, o que corresponde a 361.928 mulheres. Além disso, foram registradas 6.958 gestações duplas e 133 gestações triplas ou mais. Ao analisar a Figura 4.14, verifica-se o preenchimento dessas variáveis por estado. Para todos os estados, observa-se uma maior prevalência de gestação única, confirmando o que foi constatado anteriormente. Além disso, observa-se que o estado do Ceará apresentou a maior prevalência de gestações duplas e triplas, seguido pelos estados de Minas Gerais e Rio Grande do Norte.

Tabela 4.15: Preenchimento da variável tipo de gestação para o Semiárido brasileiro, 2021.

Categoria	n	%
Única	361.928	98
Dupla	6.958	2
Tripla ou mais	133	0
Ignorado	515	0
Total	369.534	100

Fonte: Sistema de Informações sobre Nascidos Vivos, 2022.

Figura 4.14: Preenchimento da variável tipo de gestação das mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021.



Fonte: Sistema de Informações sobre Nascidos Vivos (SINASC). Secretaria de Vigilância em Saúde, 2022.

#### 4.2.14 Tipo de parto

De acordo com a Organização Mundial da Saúde, o parto vaginal é o mais comum em todo mundo e também é a forma mais segura de ter um bebê. Porém, no Brasil esse tipo de parto vem ficando um pouco abaixo do parto Cesário, isso sendo justificado por diversos motivos, como por exemplo; medo da dor, a ideia de que a cesariana é mais segura para o bebê, medo de não conseguir vaga nos hospitais e entre outros (BARBOSA et al., 2003). Sendo assim, analisando um pouco as mães do Semiárido brasileiro, na Tabela 4.16 observamos o preenchimento das declarações de nascidos vivos referente ao tipo de parto realizado. Conforme os resultados mostram, 56% das mulheres tiveram parto Cesário, totalizando 208.079 partos. Além disso, 44% das mulheres tiveram parto Vaginal, totalizando 161.931 partos.

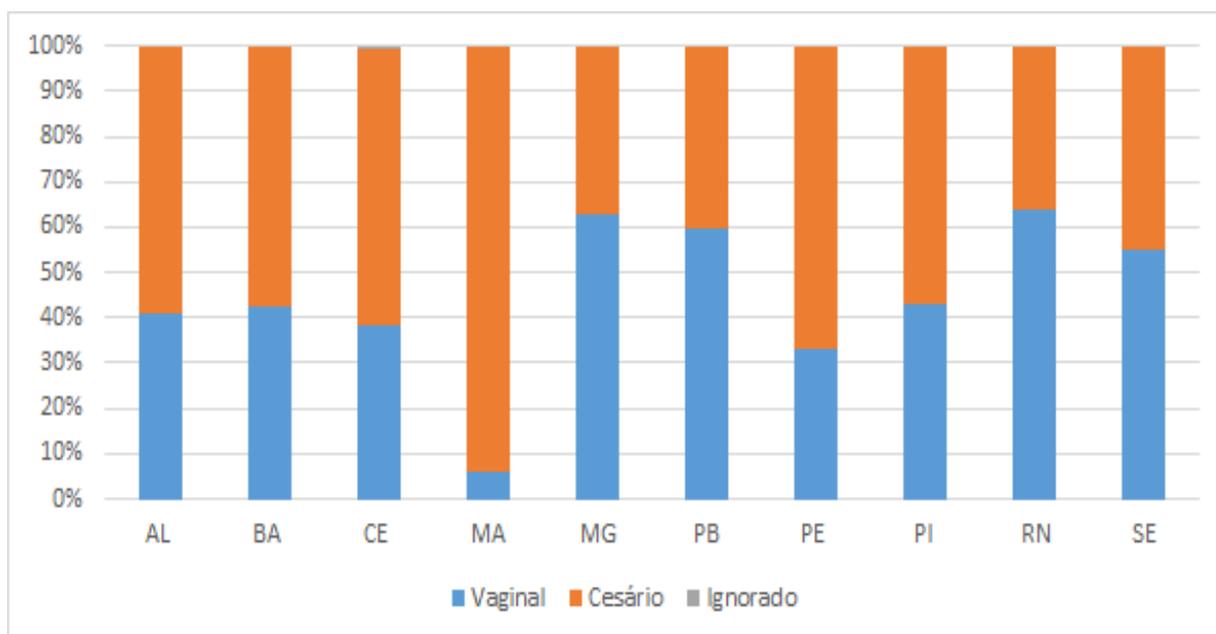
Na Figura 4.15, podemos ver o preenchimento do tipo de parto das mães do Semiárido por seu respectivo estado. Sendo assim, com base no gráfico apresentado, os estados de Minas Gerais, Paraíba, Rio Grande do Norte e Sergipe, foram os que apresentaram uma maior porcentagem de parto do tipo vaginal. Por outro lado, os demais estados apresentaram uma maior prevalência de partos do tipo cesário, como por exemplo, o estado do Maranhão que teve mais de 95% dos seus partos no tipo cesário.

Tabela 4.16: Preenchimento da variável tipo de parto para o Semiárido brasileiro, 2021.

Categoria	n	%
Vaginal	161.931	44
Cesário	207.079	56
Ignorado	524	0
Total	369.534	100

Fonte: Sistema de Informações sobre Nascidos Vivos, 2022.

Figura 4.15: Preenchimento da variável tipo de parto realizado pelas mães de 15 a 49 anos, dos estados do Semiárido brasileiro, 2021.



Fonte: Sistema de Informações sobre Nascidos Vivos (SINASC). Secretaria de Vigilância em Saúde, 2022.

### 4.3 Comportamento dos níveis da fecundidade do Semiárido brasileiro

Nos últimos anos, houveram algumas transformações na fecundidade das mulheres do Brasil. De acordo com dados do Instituto Brasileiro de Geografia e Estatística (IBGE) em 2022, a Taxa de Fecundidade Total (TFT) do país apresentou uma queda significativa nas últimas décadas. Em 1980, a TFT do Brasil era de 4,3 filhos por mulher em idade fértil. Em 2010, essa taxa havia caído para 1,9 filhos por mulher, e em 2020, a TFT ficou em 1,67 filhos por mulher.

Além disso, houve uma mudança no padrão de fecundidade, com uma queda na taxa de natalidade entre mulheres mais jovens e um aumento na taxa de natalidade entre mulheres

mais velhas. De acordo com o IBGE, em 1980 a taxa de fecundidade das mulheres de 15 a 19 anos era de 82,3 nascimentos por mil mulheres, enquanto em 2020 essa taxa diminuiu para 43,4 nascimentos por mil mulheres. Por outro lado, a taxa de fecundidade das mulheres de 30 a 34 anos era de 75,1 nascimentos por mil mulheres em 1980, e em 2020, essa taxa aumentou para 107,7 nascimentos por mil mulheres. Essa mudança no padrão de fecundidade está relacionada a vários fatores, como o aumento da escolaridade das mulheres, o acesso a métodos contraceptivos e informações sobre planejamento familiar, o adiamento da maternidade e outros fatores (IBGE, 2022).

Essas transformações na fecundidade das mulheres têm impacto em diversos aspectos da sociedade, como o envelhecimento da população, a estrutura familiar e a economia. Por isso, é importante compreender as tendências e fatores que influenciam a fecundidade para planejar políticas públicas e programas de saúde adequados às necessidades das mulheres e da sociedade na sua totalidade.

No Semiárido brasileiro, a fecundidade tem sido marcada por uma série de mudanças nas últimas décadas. Historicamente, essa região apresentou altas taxas de fecundidade, principalmente em razão do papel central da reprodução na organização social e econômica das famílias rurais. No entanto, a partir da década de 1960, houve uma queda gradual nas taxas de fecundidade, em grande parte devido à ampliação do acesso aos métodos contraceptivos e à melhoria das condições de vida e saúde da população (GOBBI L, 2022). Porém, entre as décadas de 1980 e 1990, houve uma desaceleração acentuada na queda da fecundidade no Semiárido, em contraste com o que ocorreu em outras regiões do país. Isso pode ser explicado pelo aumento da migração de famílias rurais para áreas urbanas, onde as condições de vida eram mais precárias e os serviços de saúde e educação eram mais limitados.

No entanto, a partir dos anos 2000, houve uma retomada da queda da fecundidade no Semiárido, em linha com o que ocorreu em outras regiões do país. Isso se deve em grande parte aos avanços nas políticas públicas de saúde e educação, que ampliaram o acesso aos serviços e melhoraram as condições de vida da população (NERI et al., 2013).

A Taxa de Fecundidade Total (TFT) no Semiárido brasileiro tem apresentado uma queda significativa nas últimas décadas. Segundo dados do Instituto Brasileiro de Geografia e Estatística (IBGE), a TFT na região passou de 5,2 filhos por mulher em 1980 para 2,5 em 2020. No entanto, é importante destacar que essa taxa ainda é superior à média nacional, que foi de 1,7 filhos por mulher em 2020. Além disso, existem variações significativas na TFT entre os estados que compõem o semiárido, sendo que alguns apresentam taxas mais elevadas do que outros.

Na Figura 4.16, observamos a diferença das TFTs dos estados com a dos seus respectivos Semiárido. Fica evidente as maiores taxas de fecundidade total no Semiárido quando comparada com a de todo estado. Os estados que obtiveram maiores TFTs foram os estados de Alagoas, Maranhão e Piauí, em contrapartida, os que tiveram maiores TFTs no semiárido foram os estados de Alagoas, Maranhão e Pernambuco. Mostrando assim a forte influência que a capital e alguns outros municípios que não estão no Semiárido pernambucano tem sobre o estado. Pois apesar dele apresentar uma das menores TFTs do nordeste, apresenta ainda uma das maiores do Semiárido.

Além disso, aqueles que tiveram menores taxas no Semiárido, foram os mesmos que tiveram nos índices estaduais, que foram os estados do Rio Grande do Norte, Minas Gerais e Ceará. Isso podendo ser explicado por fatores socioeconômicos, culturais e educacionais, que influenciam diretamente as escolhas reprodutivas das mulheres.

Além das interpretações citadas, observamos também que todas as TFTs dos estados e do Semiárido no ano de 2021 ficaram abaixo da média nacional, refletindo assim a tendência de redução da fecundidade no Semiárido brasileiro e no Brasil. Isso pode trazer diversos prejuízos para a população, visto que com a diminuição da fecundidade a população tende a envelhecer mais rapidamente, o que pode gerar desafios para o sistema previdenciário e de saúde, além de impactar a dinâmica econômica do país.

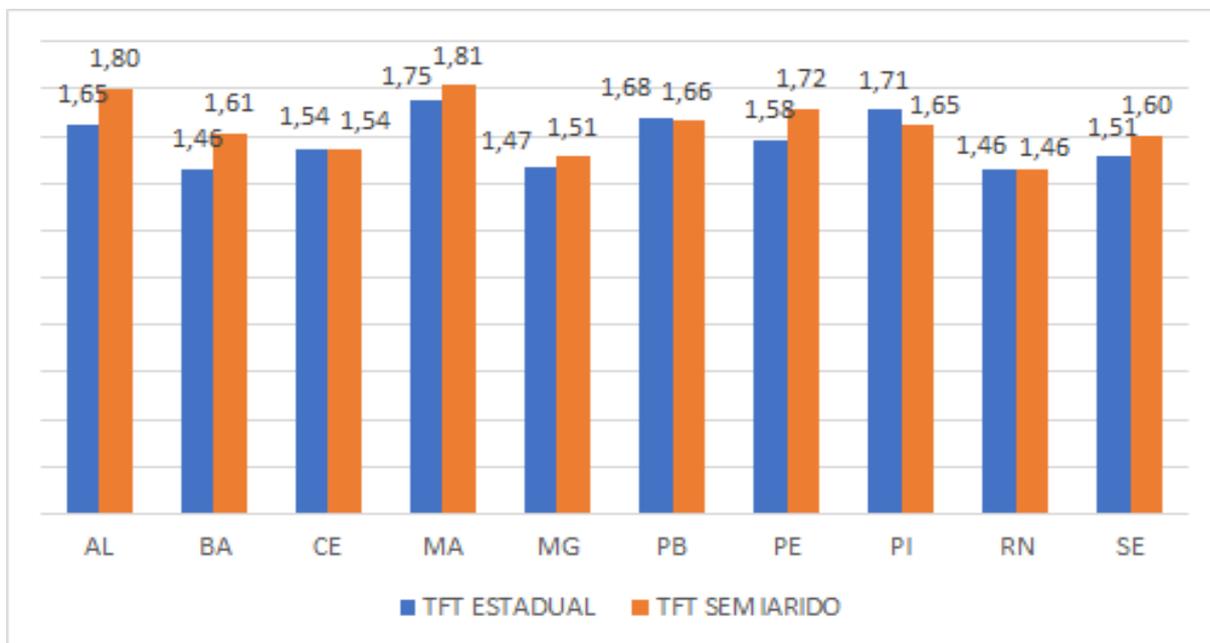
Outro ponto do impacto dessa queda são os impactos nas mudanças culturais e sociais, como a diminuição da importância da família como unidade social, a redução do tamanho médio das famílias e a diminuição da importância da maternidade como papel social das mulheres. Essas mudanças podem gerar impactos positivos ou negativos, dependendo do contexto social e cultural onde está se empregando (SUDENE, 2021).

Por conta da fecundidade está sendo calculada com base em uma projeção do número de mulheres, foi realizada a correção proposta no item 3.3 para ajuste dos indicadores da TEF e TFT dos respectivos grupos de idades das mulheres do Semiárido. Auxiliando na diminuição do viés dos nossos indicadores.

Nas Tabelas 4.17 e 4.18, temos as Taxas de Fecundidade Específica e Total para o nordeste e o Semiárido. Como 90% do território no Semiárido está localizado no nordeste, trazer esse estudo relacionando a TFT das duas regiões é de extrema importância, pois além das interpretações plausíveis, esses nos auxiliam a saber se nossos dados estão sofrendo com problemas de subestimação. Como se espera que a TFT do Nordeste seja menor que a do Semiárido, observamos que para o ano de 2021 isso se concretizou, mostrando uma boa qualidade dos dados em relação a subestimação dos mesmos. Além disso, a TFT nordestina foi de 1,57 filhos por mulher, enquanto a do Semiárido foi de 1,61 filhos por

mulher. Apresentando TFTs bem próximas, mas ainda assim vemos um número médio de filhos maior na região do Semiárido.

Figura 4.16: Taxas de Fecundidade Total da região Nordeste e do estado de Minas Gerais, e Taxa de Fecundidade Total da região Semiárida brasileira, em 2021.



Fonte: Sistema de Informações sobre Nascidos Vivos (SINASC). Secretaria de Vigilância em Saúde, 2022.

Tabela 4.17: Taxa de Fecundidade Total da região do Nordeste brasileiro, 2021.

Variáveis	Nordeste						
	15-19	20-24	25-29	30-34	35-39	40-44	45-49
Nascimentos	119.329	197.184	182.886	144.807	87.724	24.153	1.450
Mulheres	2.452.523	2.397.198	2.333.690	2.429.150	2.537.294	2.254.481	1.881.307
TEF	0,0487	0,0823	0,0784	0,0596	0,0346	0,0107	0,0008
TEF Gompertz	0,0396	0,0927	0,0827	0,0566	0,0320	0,0105	-
TFT	1,5747						
TFT Gompertz	1,5707						

Fonte: Sistema de Informações sobre Nascidos Vivos (SINASC). Secretaria de Vigilância em Saúde, 2022.

Tabela 4.18: Taxa de Fecundidade Total do Semiárido brasileiro, 2021.

Variáveis	Semiárido						
	15-19	20-24	25-29	30-34	35-39	40-44	45-49
<b>Idade</b>							
<b>Nascimentos</b>	57.625	95.219	90.371	71.434	42.436	11.736	713
<b>Mulheres</b>	1.076.885	1.170.175	1.176.534	1.162.668	1.113.443	1.020.925	886.593
<b>TEF</b>	0,0535	0,0814	0,0768	0,0614	0,0381	0,0115	0,0008
<b>TEF Gompertz</b>	0,0404	0,0945	0,0848	0,0585	0,0334	0,0111	-
<b>TFT</b>	1,6177						
<b>TFT Gompertz</b>	1,6134						

Fonte: Sistema de Informações sobre Nascidos Vivos (SINASC). Secretaria de Vigilância em Saúde, 2022.

Olhando mais para os municípios do Semiárido, com base na tabela 4.19, dos 1.262 municípios, apenas 76 apresentaram uma Taxa de Fecundidade Total maior que a do nível de reposição populacional, e 1.186 municípios apresentaram uma TFT menor que o nível de reposição. Os maiores sendo de sua maioria do estado da Paraíba, com 19 municípios, seguidos pela Bahia e Pernambuco, ambos com 12 municípios, como mostrado na Tabela 4.19.

Sendo assim, foi possível observar também no Semiárido brasileiro uma relativa queda das taxas de fecundidade total das mulheres durante toda sua idade reprodutiva. Isso, como citado anteriormente, pode ser reflexo do aumento do acesso à educação e à informação, a maior participação das mulheres no mercado de trabalho, a um aumento da disseminação dos métodos contraceptivos e até mesmo relativos reflexos da pandemia da COVID-19 sobre a fecundidade no Semiárido no ano de 2021.

Tabela 4.19: Classificação dos municípios segundo o nível de reposição da Taxa de Fecundidade Total do Semiárido brasileiro, 2021.

Estados	Municípios		Total Geral
	≤ 2,1	>2,1	
<b>AL</b>	32	6	38
<b>BA</b>	266	12	278
<b>CE</b>	167	8	175
<b>MA</b>	2	0	2
<b>MG</b>	89	2	91
<b>PB</b>	175	19	194
<b>PE</b>	111	12	123
<b>PI</b>	178	7	185
<b>RN</b>	138	9	147
<b>SE</b>	28	1	29
<b>Total Geral</b>	<b>1.186</b>	<b>76</b>	<b>1.262</b>

Fonte: Sistema de Informações sobre Nascidos Vivos (SINASC), 2022.

## 4.4 Modelagem da Fecundidade

Anteriormente foi analisada a incompletude das variáveis da DNV dos municípios do Semiárido brasileiro. Como citado no item 3.4.1, os municípios que apresentaram mais de 9% do total de ignorados em uma das 15 variáveis apresentado na Tabela 3.1, foram retiradas do estudo. Sendo assim, dos 1.262 municípios do Semiárido brasileiro, 296 apresentaram problemas em relação à incompletude dos dados, levando a exclusão. Para a modelagem, foram consideradas 966 municípios como mostrando na Tabela 4.20.

Tabela 4.20: Classificação dos municípios considerados na modelagem segundo o nível de reposição da Taxa de Fecundidade Total do Semiárido brasileiro, 2021.

Estado	Municípios		Total Geral
	$\leq 2,1$	$> 2,1$	
<b>AL</b>	25	5	30
<b>BA</b>	202	8	210
<b>CE</b>	128	8	136
<b>MA</b>	2	0	2
<b>MG</b>	55	0	55
<b>PB</b>	126	16	142
<b>PE</b>	106	11	117
<b>PI</b>	126	3	129
<b>RN</b>	112	7	119
<b>SE</b>	26	0	26
<b>Total Geral</b>	<b>908</b>	<b>58</b>	<b>966</b>

Fonte: Sistema de Informações sobre Nascidos Vivos (SINASC), 2022.

Dessa forma, após o ajuste citado, foi aplicado modelo de regressão logística binário, entre a Taxa de Fecundidade Total das mulheres em idade fértil do Semiárido brasileiro, e as variáveis materno-infantis presente na Tabela 3.1. Antes da obtenção do modelo final, foi realizado uma seleção de variáveis com base no p-valor, onde aquelas que apresentaram uma significância estatística entre a TFT e a categoria da respectiva variável menor que 0,10, foram incluídas no modelo, e as que apresentaram valores maiores foram excluídas. Obtendo assim, o modelo final mostrado na Tabela 4.21.

Tabela 4.21: Variáveis selecionados para compor o modelo de regressão logística binária.

Parâmetros	Estimativa (B)	Erro Padrão	Estatística de Wald	g.l	OR (B)	P-valor
Constante	-2,753	0,177	241,284	1	0,064	<0,001
Apgar 1m 0-2	-0,029	0,096	0,088	1	0,972	0,766
Apgar 1m 3-5	-0,12	0,051	5,653	1	0,887	0,017
Apgar 1m 6-7	-0,065	0,027	5,868	1	0,937	0,015
Apgar 1m 8-10	-0,067	0,026	6,534	1	0,935	0,011
Escolaridade fora da idade E	-0,061	0,028	4,673	1	0,941	0,031
Escolaridade dentro da idade E	-0,05	0,027	3,599	1	0,951	0,058
Estado Civil casada união e	0,149	0,042	12,678	1	1,161	<0,001
Estado Civil solteira	0,145	0,042	12,27	1	1,157	<0,001
Estado Civil viuva separada	0,287	0,079	13,137	1	1,332	<0,001
Idade 15-19	-0,015	0,025	0,355	1	0,985	0,551
Idade 20-24	-0,043	0,021	4,075	1	0,958	0,044
Idade 25-29	-0,024	0,018	1,763	1	0,976	0,184
Idade 30-34	-0,055	0,028	3,855	1	0,947	0,050
Idade 35-39	-0,121	0,058	4,35	1	0,886	0,037
Idade 40-44	-0,093	0,228	0,167	1	0,911	0,682

Fonte: Sistema de Informações sobre Nascidos Vivos (SINASC), 2022/ Instituto Brasileiro de Geografia e Estatística (IBGE).

Logo, o modelo apontou que a probabilidade dos recém nascidos das mães do Semiárido brasileiro que apresentou um Apgar 1 ° minuto de 3 a 5, 6 a 7 ou 8 a 10, tem probabilidades menores de pertencer a municípios com taxa de fecundidade total acima do nível de reposição quando comparado com o Apgar 1 ° minuto de 0-2. Além disso, em relação à escolaridade, vemos que as mães dos recém nascidos que estão dentro da idade escolar tem menos chance de pertencerem a municípios com uma TFT acima do nível de reposição de 2,1, quando comparado a mães fora da idade escolar.

Em relação a variável estado civil das mães, através das odds ratio (OD), foi constatado que mães com um estado civil viúva ou separada, tem mais chances de pertencer os municípios acima do nível de reposição quando comparado com as demais. Já em relação as idades, vemos que mães de 35 a 44 anos tem menos chances de pertencer a municípios acima do nível de reposição quando comparado com as demais idades. A idade com uma maior probabilidade foi as mulheres de 15 a 19 como mostrado na Tabela 4.21.

Além disso, também foi calculado um pseudo R<sup>2</sup> de Nagelkerke, de 11,7%, o que significa que cerca de 11,7% das categorias significativas dessas variáveis materno-infantis explicaram a TFT das mães do Semiárido brasileiro em 2021.

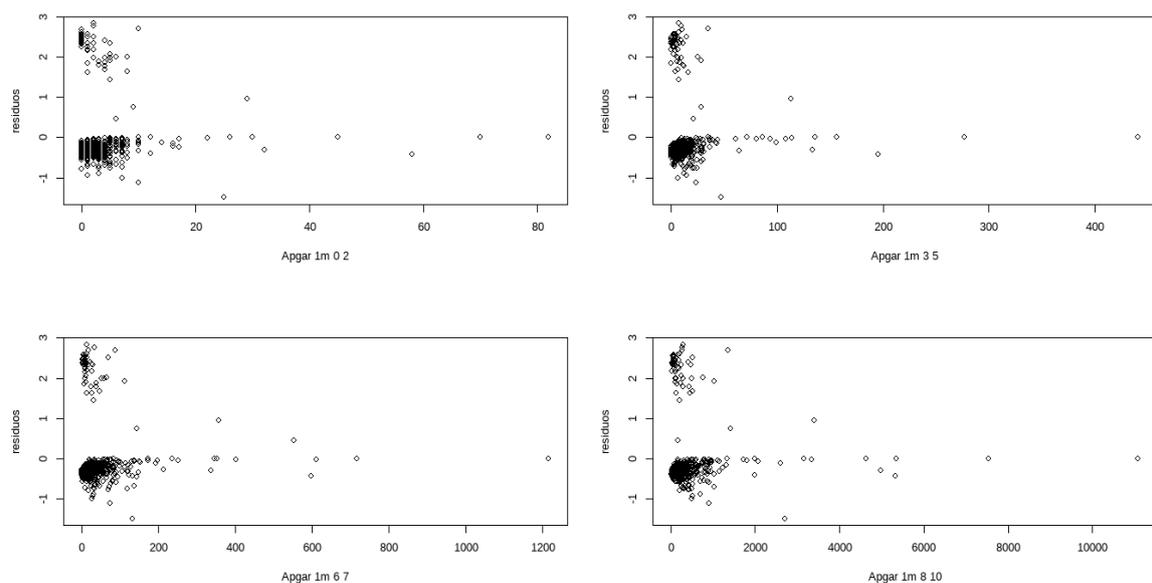
### 4.4.1 Análise de diagnóstico

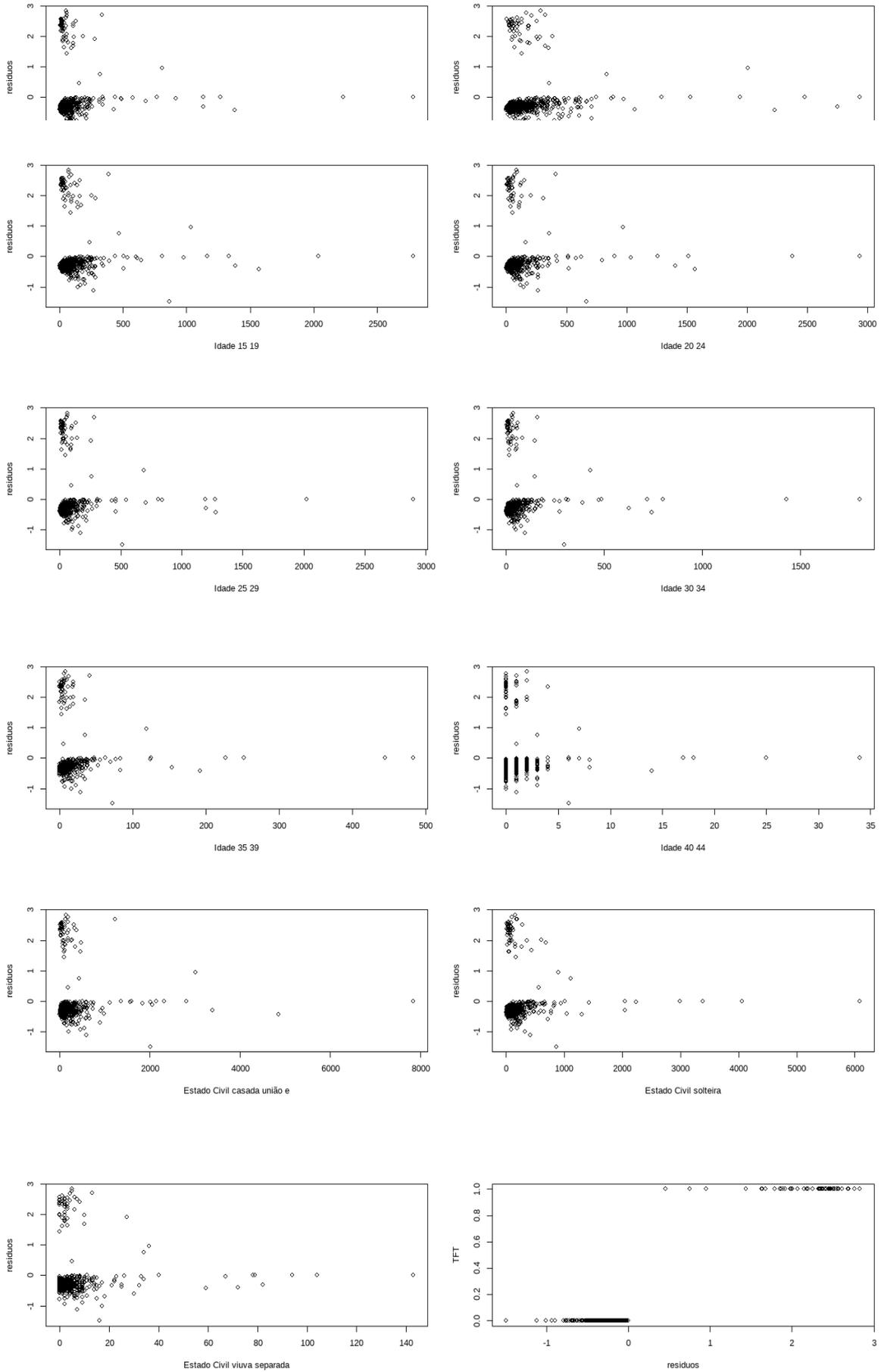
Fazer uma análise de diagnóstico em um modelo de regressão é extremamente importante para avaliar a qualidade do modelo e identificar possíveis problemas ou áreas de melhoria. Essa análise ajuda a verificar se os pressupostos do modelo de regressão estão sendo atendidos e se os resultados são confiáveis. Sendo assim, para o modelo avaliaremos se o pressupostos abaixo estão sendo atendidas;

#### Análise de resíduos

A análise de resíduos desempenha um papel importante no nosso estudo de diagnóstico, pois é através dela que avaliamos a adequação do modelo. O valor esperado dos resíduos padronizados, em um modelo de regressão, geralmente varia de -2 a 2. Mas dependendo do contexto e da variável de estudo, esse intervalo poderá ser um pouco maior. Sendo assim, observando a Figura 4.17, como os resíduos do modelo proposto varia em sua maioria entre  $[-1;3]$ , considera-se que não existe um “padrão” que afeta nossos dados, esses estão distribuídos de forma aleatória e com um padrão de variabilidade constante. Além disso, também foi aplicado o teste de Durbin Watson para comprovar o que foi dito anteriormente, e foi obtido uma estatística de 2,048 bem próximo a 2, o que sugere uma ausência de autocorrelação dos resíduos.

Figura 4.17: Gráficos dos resíduos padronizados para o modelo de regressão logístico para os municípios do Semiárido brasileiro, 2021.





## Bondade de ajuste

A bondade de ajuste em modelos de regressão logística pode ser avaliada por meio de medidas como o qui-quadrado e o desvio escalonado (*deviance*). A qui-quadrado é um teste estatístico usado para avaliar a adequação do modelo aos dados observados. No contexto da regressão logística, o teste qui-quadrado é aplicado comparando a *deviance* do modelo ajustado com a *deviance* de um modelo nulo (modelo com apenas o intercepto, sem nenhuma variável independente). A diferença entre essas duas *deviances* segue uma distribuição qui-quadrado, com graus de liberdade igual ao número de variáveis independentes do modelo.

Um valor baixo de qui-quadrado indica uma boa adequação do modelo aos dados, enquanto um valor alto sugere que o modelo não se ajusta bem aos dados. Portanto, um valor de qui-quadrado significativo pode indicar a necessidade de revisão ou melhoria do modelo. Para o modelo de estudo, os respectivos valores do estudo obtido foram apresentados na Tabela 4.22:

Tabela 4.22: Bondade de ajuste

Estatística	Valores calculados
Deviance	396,068
Qquadrado	1022,81

Fonte: Elaboração própria.

Sendo assim, concluímos que; como o valor obtido pela *deviance* foi de 396 e menor que 1.022 da estatística da qui quadrado, a diferença do modelo saturado para o modelo em investigação é menor que a qui-quadrado de referência, o modelo ajustado é adequado.

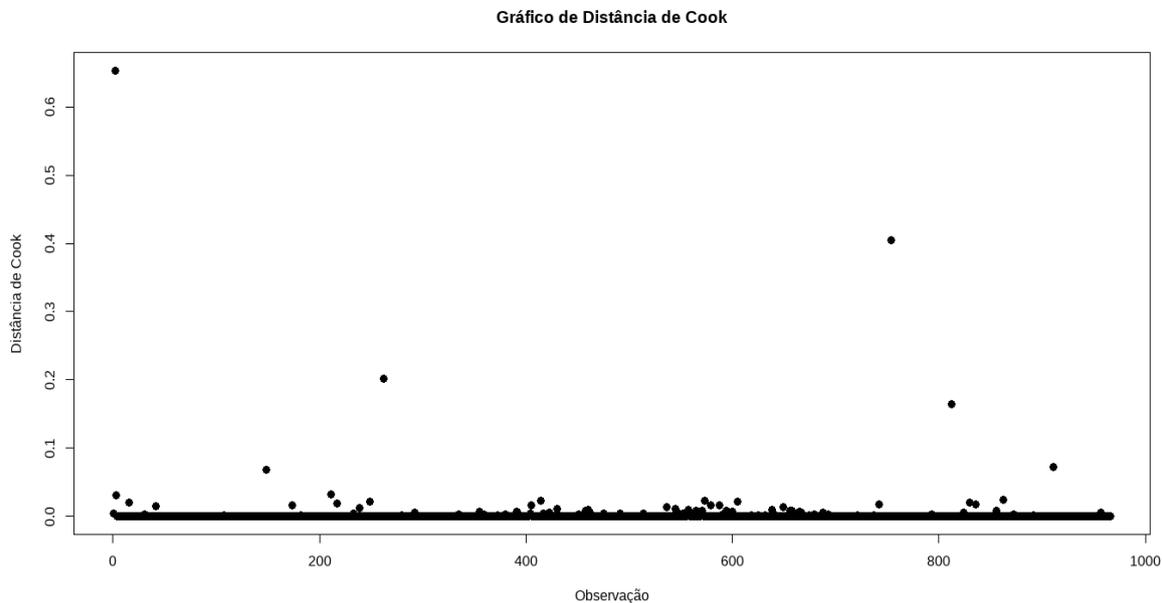
## Distância de Cook

Usaremos a distância de Cook para medir o quanto a estimativa dos coeficientes de regressão seria afetada se uma observação específica fosse removida do conjunto de dados. Ela leva em consideração tanto a magnitude dos resíduos da observação quanto a sua influência nas estimativas dos coeficientes de regressão. Onde valores maiores de distância de Cook indicam que a observação correspondente tem uma influência maior no modelo, ou seja, a remoção dessa observação afetaria consideravelmente as estimativas dos coeficientes. Geralmente uma observação é considerada influente quando apresenta valores maiores que 1 (LEMONTE, 2008).

Com base na Figura 4.18, observamos o gráfico da distância de Cook para o modelo final levado em consideração. Observo-se que o gráfico não apresentou nenhum ponto maior

que 1, mostrando assim uma ausência de pontos influentes.

Figura 4.18: Distância de Cook do modelo de regressão logístico binário, 2021.



Fonte: Elaboração própria.

## Desempenho do modelo

Para avaliar o desempenho do modelo, foi construído uma matriz de confusão entre os valores estimados e os valores observados e foi obtido os resultados descritos na Tabela 4.23:

Tabela 4.23: Matriz confusão do modelo em termos do percentual de classificação correta para os municípios do Semiárido brasileiro, 2021

TFT observada	TFT prevista			% de acerto global
	Abaixo da taxa de reposição (n)	Acima da taxa de reposição (n)	% correta	
Abaixo da taxa de reposição	907	1	99,9	94,2
Acima da taxa de reposição	55	3	5,2	
<b>Total</b>	962	4	-	-

Fonte: Elaboração própria

Sendo assim, os resultados mostraram que dos 908 municípios que apresentaram taxa de fecundidade total abaixo do nível de reposição, 907 (99,9%) foi classificado corretamente, enquanto 1 não. Por outro lado, dos municípios que apresentaram taxa de fecundidade total acima do nível de reposição, o modelo não apresentou uma porcentagem de acerto

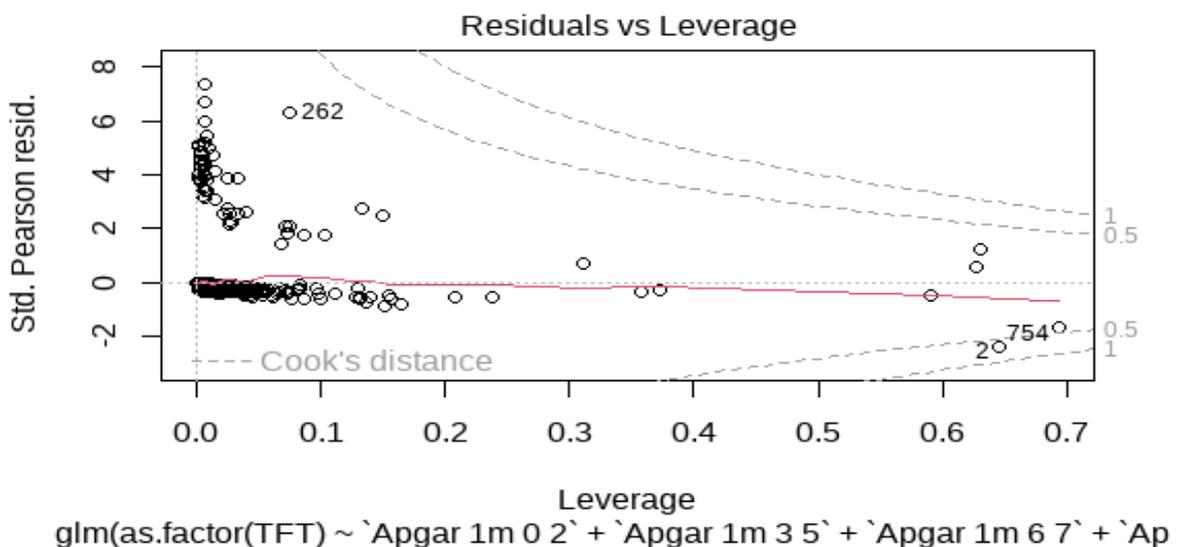
elevada, ficando com apenas 5,2%, totalizando 3 municípios acertados dos 57. Apesar disso, o modelo logístico ajustado apresentou uma taxa de acerto de 94,2%, mostrando uma boa qualidade preditiva.

#### 4.4.2 Avaliação de pontos influentes

O tratamento de *outliers* é realizado com o objetivo de lidar com valores atípicos ou extremos que estão fora do padrão esperado em um conjunto de dados. Os *outliers* podem surgir devido a erros de medição, falhas no processo de coleta de dados ou representar casos verdadeiramente raros ou incomuns. Além disso, podem ter um impacto desproporcional nos resultados de uma análise estatística ou em modelos de aprendizado de máquina. Eles podem distorcer as estimativas de média, variância e correlação, levando a conclusões incorretas ou viesadas. Ao tratar os *outliers*, podemos obter estimativas mais precisas e representativas das características do conjunto de dados.

Sendo assim, para averiguar a presença e a influência de *outliers*, foi primeiramente gerado um gráfico dos resíduos padronizados, mostrado na Figura 4.19. Onde se o gráfico apresentar pontos que estão fugindo do esperado, ou seja, pontos além da linha 0,5 do gráfico, ele será um possível *outliers* ou ponto influente. Para o modelo, apenas o ponto 2, ficou um pouco acima da linha, trazendo a interpretação de um possível ponto de influência no modelo.

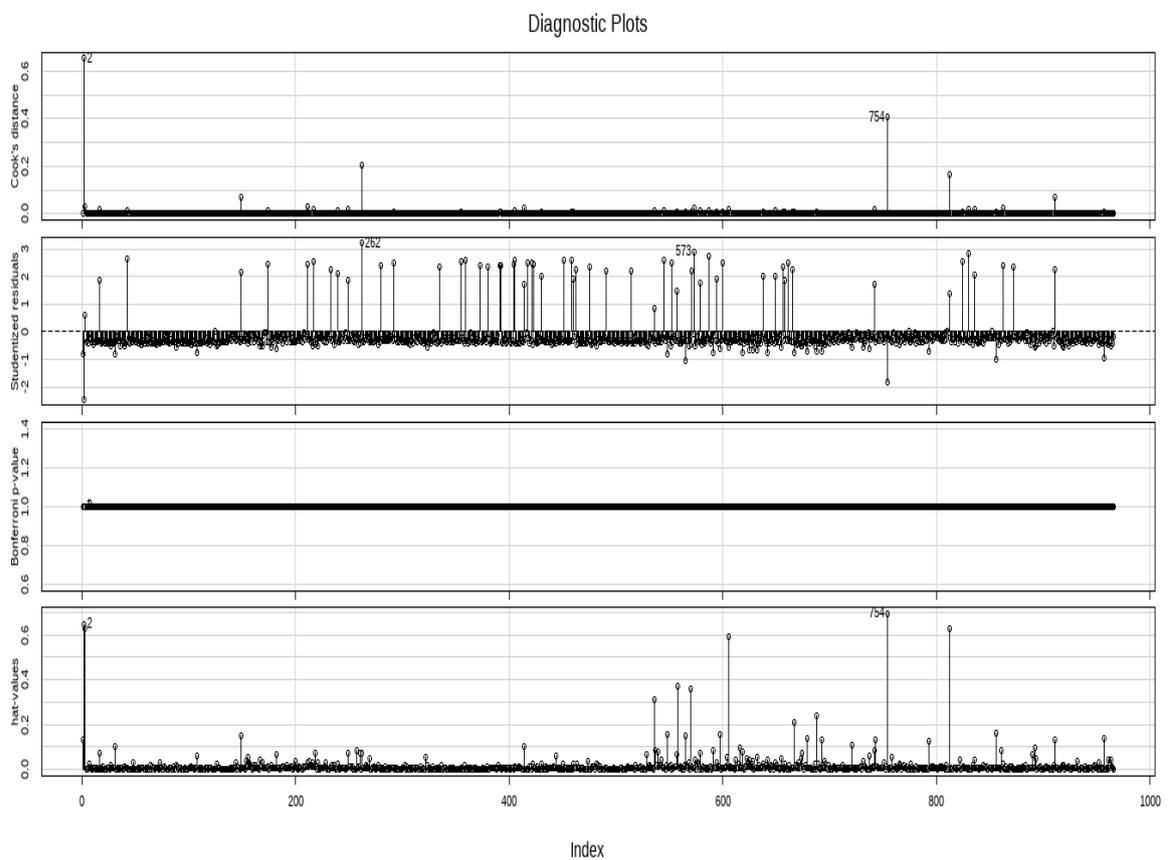
Figura 4.19: Avaliação de pontos influentes, 2021.



Fonte: Elaboração própria.

Para checar esse possível ponto, foi criado outro gráfico através da função `IndexPlot` do R, presente na Figura 4.20, onde nos traz um gráfico com algumas informações gerais a respeito do banco. Como a distância de Cook, gráfico dos resíduos e entre outros. Vamos nos ater ao segundo gráfico onde pontos que ficaram fora do intervalo  $[-3, 3]$ , possivelmente são pontos influentes. Sendo assim, vemos novamente a presença do ponto 2, confirmando a discrepância desses em relação ao demais. Porém, quando testado a retirada dele do modelo, não foi visto uma melhora na predição e na qualidade da regressão. Sendo assim, foi optado pela permanência dele no modelo.

Figura 4.20: Avaliação de pontos influentes II, 2021.



Fonte: Elaboração própria.

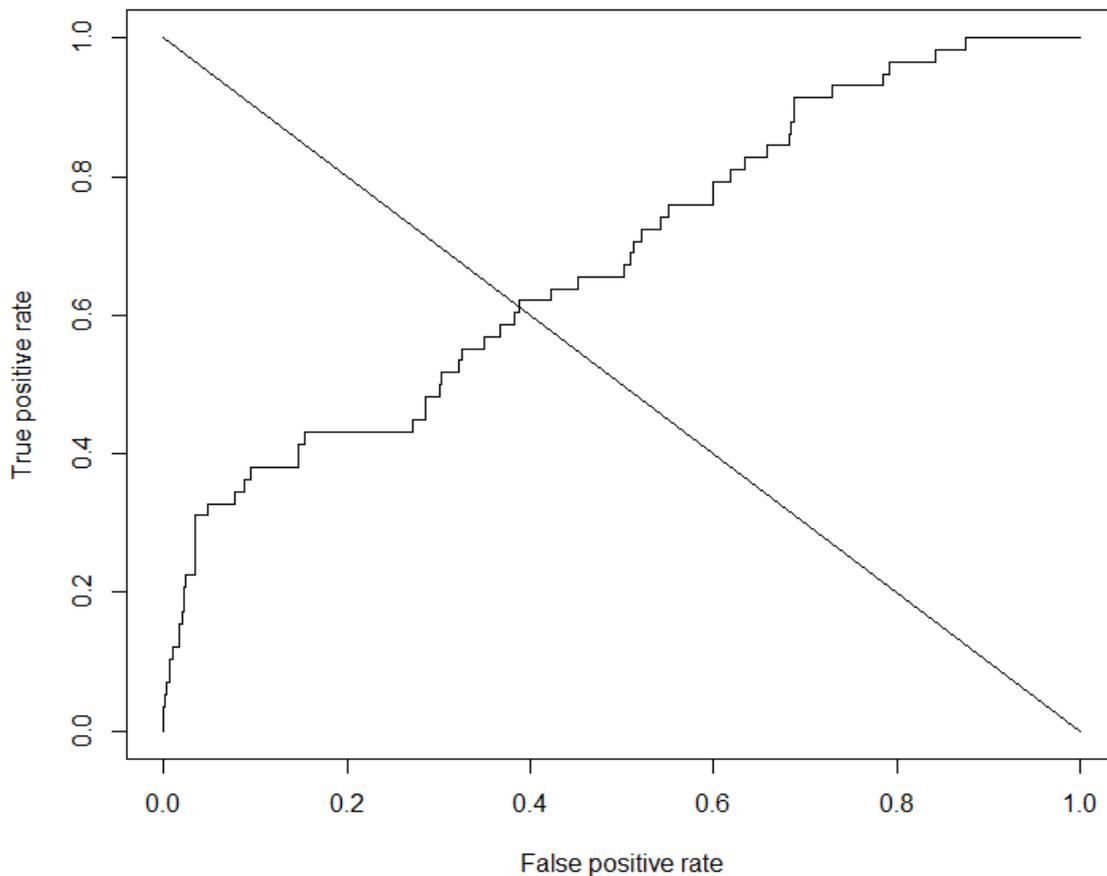
### 4.4.3 Curva ROC

A curva ROC (*Receiver Operating Characteristic*) é uma representação gráfica da performance de um modelo de classificação binária em diferentes níveis de threshold (limiar de decisão) para a classificação dos dados. A curva é criada plotando a taxa de verdadeiros positivos (TPR) no eixo y em relação à taxa de falsos positivos (FPR) no eixo x. Onde a

forma de interpretação é feita com base em sua forma e localização. Quanto mais próxima a curva estiver do canto superior esquerdo do gráfico, melhor será o desempenho do modelo, indicando uma maior capacidade de distinguir entre as classes positiva e negativa.

Para o modelo, foi gerado o gráfico contido na Figura 4.21, e para a escolha do melhor ponto de corte iremos nos basear nas taxas de verdadeiro positivo e falso positivo que o mesmo produz. Sendo assim, foi possível ver por meio do gráfico e da linha construída, que o ponto de equilíbrio está aparentemente perto do valor 0,4 da taxa do falso positivo, e 0,6 da taxa do verdadeiro positivo. Obtendo uma AUC (*Area Under the Curve*) de 0,68, e uma especificidade de 0,905, e uma sensibilidade de 0,379. Logo, sabendo que a área tem como máximo 1, podemos dizer que o valor encontrado, 0,68, trata-se de um modelo razoável, de forma que a taxa de *true* positive cresce muito mais rápido que a taxa de *false* positive.

Figura 4.21: Curva ROC, 2021.



Fonte: Elaboração própria.

## CAPÍTULO 5

### CONSIDERAÇÕES FINAIS

O estudo do Semiárido brasileiro e de seus fatores demográficos durante e pré-pandemia, são assuntos bem escassos na literatura, principalmente em relação à fecundidade das mulheres durante todo o período fértil do Semiárido brasileiro. Sendo assim, esse trabalho visou trazer respostas sobre o impacto da pandemia da COVID-19 sobre a qualidade das informações das declarações e nascidos vivos sobre essa a região, que há décadas vinha enfrentando sérios problemas na construção de indicadores de saúde fidedignos devido à fragilidade dos seus dados. Além disso, buscou-se identificar as variáveis materno-infantis que estão associadas aos níveis de fecundidade do Semiárido brasileiro em 2021.

Para o ano do referente estudo, 2021, período pandêmico, o estudo apontou certa ausência dos impactos da pandemia sobre a cobertura das informações das declarações dos nascidos vivos no Semiárido brasileiro. Desde 2010 a região vinha apresentado taxas de coberturas relativamente boas, a maioria acima de 90%. Fato associado a avanços na infraestrutura de comunicação, programas governamentais de inclusão digital e sexual e entre outros.

Em relação à completude das informações, é importante destacar a presença de percentuais de incompletude acima de 9% em algumas variáveis da Declaração de Nascido Vivo dos recém-nascidos em todos os estados do Semiárido. Dos 1.262 municípios da região, 296 apresentaram problemas mais graves na completude de suas informações, o que levou à exclusão desses municípios do estudo.

Por outro lado, em relação ao preenchimento das variáveis no ano de 2019 (pré-pandemia), algumas variáveis enfrentaram problemas significativos em sua completude.

A variável raça/cor da mãe registrou um total de 18.910 casos em que a informação foi ignorada, representando 4,88% do total geral. Além disso, o estado civil da mãe e as semanas de gestação também apresentaram 11.576 casos ignorados, correspondendo a um percentual de 2,99%. Além dessas variáveis, observou-se um total de 85.917 células vazias nas demais declarações, considerando um parâmetro geral.

Já no ano de 2021 (pandêmico), ocorreram algumas mudanças em relação ao cenário da completude de algumas variáveis. Primeiramente, ocorreu um aumento de vazios nas variáveis tipo de gravidez, parto e dos números de consultas pré-natal. Isso se deve, principalmente, a diminuição dos acompanhamentos pré-natal motivada pelas diversas interrupções dos serviços de saúde da mulher provocadas pela pandemia.

Apesar do aumento nessas variáveis, observa-se de forma geral uma diminuição no total de declarações vazias no ano de 2021 em comparação a 2019. Em 2019, foram identificadas 85.917 declarações vazias, enquanto em 2021 foram registradas 64.843 declarações, representando uma queda de 24,5%. Isso corresponde a uma redução de 21.074 declarações vazias. Esses dados indicam que, embora tenha havido um aumento nas variáveis mencionadas no parágrafo anterior, os impactos da pandemia da COVID-19 sobre as demais variáveis das declarações de nascidos vivos do ano de 2021 foram mínimos.

Ao analisar o total geral de declarações nos dois anos, constatou-se uma queda de aproximadamente 2,38%. Essa redução pode ser explicada justamente pelo adiamento da maternidade por parte das mulheres em 2021, devido ao período atípico que estava ocorrendo. Essa situação refletiu na diminuição da quantidade de nascidos vivos no referente ano.

Em relação ao preenchimento das variáveis das declarações de Nascidos Vivos, observou-se que a variável raça/cor foi a mais problemática em termos de completude. Isso pode estar relacionado à falta de conscientização das mulheres e dos profissionais de saúde sobre a importância da coleta precisa de dados sobre raça/cor, uma vez que isso auxilia na compreensão e abordagem das desigualdades raciais, visando promover a equidade para todos.

No que diz respeito à saúde materna, este trabalho identificou algumas fragilidades que requerem uma atenção mais cuidadosa por parte dos órgãos governamentais. A primeira delas é em relação ao tipo de parto, onde cerca de 56% dos partos realizados pelas mulheres do Semiárido no ano de 2021 foram cesáreos, enquanto 44% foram partos vaginais. Isso revela um cenário não muito favorável para as mulheres, uma vez que o ideal é ter uma maior frequência de partos vaginais, o que é mais benéfico para a saúde dessas mulheres.

Além disso, em relação à quantidade de consultas pré-natal, observamos que 75% das mulheres tiveram uma boa quantidade de consultas, tendo feito 7 ou mais consultas durante a gravidez, seguida por mulheres que tiveram de 4 a 6 consultas. Em relação a idades dessas mães, vemos que essas mulheres tiveram seus filhos com idades relativamente jovens, ou seja, 66% das mães tinham idades de 15 a 29 anos. Mostrando assim, mães bem jovens quando relacionadas a outras regiões do país, e a maioria dentro da idade escolar, representando 78% do total geral.

Esse estudo também trouxe reflexões a respeito da taxa de fecundidade total das mulheres desses municípios do Semiárido brasileiro. Como essa é uma estatística que mede o número médio de filhos que uma mulher teria ao longo de sua vida reprodutiva. Ela é amplamente utilizada para analisar os padrões de reprodução e o potencial de crescimento populacional de um determinado grupo ou país. Sendo assim, constatou que todos os estados do respectivo estudo, se encontra abaixo do nível de reposição populacional (2,1 filhos por mulher) estabelecido mundialmente. Em relação ao Semiárido, observamos o mesmo padrão, tendo como maiores TFTs o estado do Maranhão e o estado de Alagoas. Já em relação aos municípios, dos 1.262 municípios, 1.186 ficaram abaixo do nível de reposição, e apenas 76 acima. Olhando de um parâmetro mais geral, foi constatado uma TFT de 1,6 filhos por mulher na região do Semiárido brasileiro. Ou seja, sinalizando uma preocupação da região, pois baixos índices de reprodução acaba afetando o desenvolvimento econômico, social e cultural da região.

Sendo assim, foi empregado o Modelo de Regressão Logística para trazer uma resposta sobre a fecundidade das mães do Semiárido brasileiro e as variáveis das declarações de nascidos vivos. Após a aplicação do modelo, as variáveis maternas que conseguiram explicar e trazer uma relação com TFT, foi o Apgar 1 ° minuto, escolaridades, estado civil e idade das respectivas mães. Dessas, o modelo apontou que a probabilidade dos recém nascidos das mães do Semiárido brasileiro de apresentar um Apgar 1 ° minuto de 3 a 5,6 a 7 ou 8 a 10, tem probabilidades menores de pertencer a municípios com taxa de fecundidade total acima do nível de reposição quando comparado com o Apgar 1 ° minuto de 0-2.

Além disso, em relação à escolaridade, vemos que as mães dos recém nascidos que estão dentro da idade escolar tem menos chance de pertencerem a municípios com uma TFT acima do nível de reposição de 2,1, quando comparado a mães fora da idade escolar. Em relação ao estado civil das mães, observamos que mães com um estado civil viúva ou separada, tem mais chances de pertencer a municípios acima do nível de reposição quando comparado com os demais. Já em relação às idades, vemos que mães de 35 a 44 anos tem menos chances de pertencer a municípios acima do nível de reposição quando comparado com as demais idades.

Apesar dos resultados obtidos, o modelo utilizado neste estudo enfrentou algumas dificuldades na estimação e no balanceamento dos dados devido ao ponto de corte escolhido, que foi a taxa de reposição. Sendo assim, sugere-se que futuras pesquisas considerem a adoção de um novo ponto de corte para classificar as taxas de fecundidade total dos municípios, além da inclusão de novas variáveis e a utilização de técnicas alternativas de correção da fecundidade. Além disso, é recomendado o uso de modelos de regressão mais robustos, capazes de explicar de forma mais precisa a Taxa de Fecundidade Total da região, minimizando vieses e aumentando o poder preditivo. Resultados mais robustos e precisos contribuirão para uma melhor compreensão das questões relacionadas à saúde materna nessa região, que carece de informações na literatura, apesar de sua importância econômica, social e cultural para o país.

## REFERÊNCIAS BIBLIOGRÁFICAS

AASSVE, A. et al. **The COVID-19 pandemic and human fertility**. *Science*, v. 369, n. 6502, p. 370–371, 2020.

AGRANONIK, M. **Técnicas de Diagnóstico Aplicadas ao Modelo de Regressão Logística**. p. 33–35, 2005.

ALVES, J. E. D. **Bônus demográfico no Brasil: do nascimento tardio à morte precoce pela Covid-19**. *Revista Brasileira de Estudos de População*, Associação Brasileira de Estudos Populacionais, v. 37, n. 1, 2020.

APGAR. **Use and abuse of the Apgar score**. *Pediatrics*, v. 78, n. 6, p. 1148, 1986.

AQUINO, M. R. de. **Educação para a convivência com o Semiárido e direitos humanos: experiências educativas do Instituto Regional da Pequena Agropecuária Apropriada - IRPAA**. p. 87–90, 2010.

ARBACHE, J. **Demographic Transformation and International Competitiveness of the Brazilian Economy**. Social Science Electronic Publishing. Brasília: SSRN, 2011.

ASA. **SEMIÁRIDO - É no semiárido que a vida pulsa!** 2022. Disponível em: <<https://www.asabrasil.org.br/semiarido>>. Acesso em: 11 de Setembro 2022.

BANT, A. **Em um ano de pandemia, Fundo de População da ONU estima que interrupções em serviços levaram a 1,4 milhão de gravidezes não intencionais**. <https://brazil.unfpa.org/pt-br/news>, p. 6, 2022.

BARBOSA, G. G. et al. **Parto cesáreo: quem o deseja? Em quais circunstâncias?** *Cad. Saúde Pública*, v. 19, n. 6, p. 1611–1615, 2003.

CALDWELL, J. C. **Towards a restatement of demography transition**. *Population Council*, Canberra: Australian National University, 1977.

CASTRO, C.; BRAGA, A. **Supervised learning with imbalanced data sets: an overview**. *Controle e Automação Sociedade Brasileira de Automatica, Campinas*, n. 5, p. 440–441, 2011.

CAVENAGHI, S. M. E. **Notas sobre os diferenciais educacionais e econômicos da fecundidade no Brasil.** *Revista Brasileira de Estudos de População*, v. 31, n. 2, p. 471–482, 2016.

COUTINHO et al. **Considerações sobre a pandemia de Covid-19 e seus efeitos sobre a fecundidade e a saúde sexual e reprodutiva das brasileiras.** *Revista Brasileira de Estudos Populacionais*, v. 37, n. 1, p. 1–9, 2020.

COX, D.; SNELL, E. **Analysis of Binary Data. 2nd Edition.** Chapman and Hall/CRC, London, 1989.

EASTERLIN, R. A. **An economic framework for fertility analysis. Studies in family planning.** *Population Council*, v.6, n. 3, p. p.54–63, 1975.

ELÍDIO, V. **O envelhecimento, a transição epidemiológica, da população brasileira, e impacto nas internações no âmbito do SUS.** . n. 2, p. 147–148, 2019.

FAWCET, T. **Introduction to ROC analysis.** v. 27, p. 861–874, 2006.

FUSCO, W.; MORVAN, M. **Mapeando a fecundidade nordestina: 2000-2010.** *Open edition Journals*, n. 33, 2017.

G1. **Envelhecimento da população no Canadá causa problemas à economia.** 2022. Disponível em: <<https://g1.globo.com/Noticias/Mundo/0,AA1588849-5602,00-ENVELHECIMENTO+DA+POPULACAO+NO+CANADA+CAUSA+PROBLEMAS+A+ECONOMIA.html>>. Acesso em: 19 de Novembro 2022.

GIANNINI. **Taxa de fecundidade no Brasil é baixa e está em queda acelerada.** 2022. Disponível em: <<https://noticias.r7.com/saude/taxa-de-fecundidade-no-brasil-e-baixa-e-esta-em-queda-acelerada.>>. Acesso em: 10 de Novembro 2022.

GOBBI L. **Urbanização brasileira.** 2022. Disponível em: <<http://educacao.globo.com/geografia/assunto/urbanizacao/urbanizacao-brasileira.html>>. Acesso em: 19 de Novembro 2022.

GONÇALVES et al. **A transição da fecundidade no Brasil ao longo do século XX -uma perspectiva regional.** *Revista Brasileira de Ginecologia e Obstetrícia*, v. 36, p. 1–34, 2019.

IBGE. **Classificação e caracterização dos espaços rurais e urbanos do Brasil : uma primeira aproximação.** *Coordenação de Geografia. Rio de Janeiro: IBGE*, p. 84, 2022.

IDOETA, P. A. **"Covid: 300 mil bebês deixaram de nascer no Brasil por pandemia, com adiamentos e mais divórcios".** 2022. Disponível em: <<https://www.bbc.com/portuguese>>. Acesso em: 14 de Novembro 2022.

LEMONTE, A. **Diagnóstico em Regressão Normal Linear: princípios e aplicação.** *Rev. Bras. Biom.*, v. 26, p. 11–13, 2008.

LESLIE V. **ET AL APGAR Score.** 2023. Disponível em: <<https://www.ncbi.nlm.nih.gov/books/NBK470569/>>. Acesso em: 13 de Março 2023.

LINO, R. et al. **Tendência da incompletude das estatísticas vitais no período neonatal, estado do Rio de Janeiro, 1999-2014.** *Epidemiologia e Serviços de Saúde*, v. 28, 2019.

MAIA, L. T. S. et al. **Uso de linkage para melhoria da completude do SIM e do Sinasc nas capitais brasileiras.** *Revista de Saúde Pública*, v. 51, p. 112, 2017.

MARTINS, V. et al. **Mortalidade de mulheres em idade fértil de 2002 a 2011 em São Luís, MARANHÃO .** *Open Journal Systems*, v. 15, n. 1, 2014.

MINISTÉRIO DA SAÚDE. **Sistemas de informação em saúde.** 2021. Disponível em: <<https://www.gov.br/saude/pt-br/composicao/svsa/vigilancia-de-doencas-cronicas-nao-transmissiveis/sistemas-de-informacao-em-saude>>. Acesso em: 10 de Novembro 2022.

MINISTÉRIO DA SAÚDE. **Declaração de nascidos vivos, manual de instruções para preenchimento.** 2023. Disponível em: <<https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/publicacoes-svs/vigilancia/declaracao-de-nascido-vivo-manual-de-instrucoes-para-preenchimento>>. Acesso em: 26 de Fevereiro de 2023.

Ministério da Saúde. **O que é a Covid-19?** 2022. Disponível em: <<https://www.gov.br/saude/pt-br/coronavirus/o-que-e-o-coronavirus>>. Acesso em: 10 de Setembro 2022.

MONTGOMERY, D. C.; PECK, E.; VIKING, G. G. **Introduction to linear regression analysis.** *HTU John Wiley Sons.s. Inc. UTH, New York, NY*, 2001.

MOULTRIE, T. A. et al. **Tools for demographic estimation.** *Paris: International Union for the Scientific Study of Population*, 2013.

NELDER, J.; WEDDERBURN, R. W. M. **Generalized Linear Models.** *Journal of the Royal Statistical Society. Series A*, v. 135, n. 3, p. 370–384, 1972.

NERI et al. **Efeitos macroeconômicos do Programa Bolsa Família: uma análise comparativa das transferências sociais. Programa Bolsa Família: uma década de inclusão e cidadania.** *Brasília: Ipea*, v. 1, p. 193–2067, 2013.

OJIMA, R.; FUSCO, W. **Migrações e nordestinos pelo Brasil: uma breve contextualização.** *Editora Edgard Blücher*, <http://dx.doi.org/10.5151/BlucherOA-ojimafusco-04>, p. 11–26, 2015.

OLIVEIRA, A. C. et al. **Estado nutricional materno e sua associação com o peso ao nascer em gestações de alto risco.** *Ciência Saúde Coletiva*, v. 23, p. 2373–2382, 2018.

OLIVEIRA, E. **Modelos lineares e aditivos generalizados para posição, escala e forma na análise de estoques pesqueiros de lagostas (*Panulirus spp*) e de espadarte (*Xiphias gladius*).** p. 22–14, 2018.

OMS. **Histórico da pandemia de COVID-19.** 2022. Disponível em: <<https://www.paho.org/pt/covid19/historico-da-pandemia-covid-19>>. Acesso em: 14 de Novembro 2022.

ONU. *COVID-19 tem impactos “devastadores” na saúde das mulheres, afirma diretora da OPAS*. 2022. Disponível em: <<https://brasil.un.org/pt-br/128799-covid-19-tem-impactos-devastadores-na-saude-das-mulheres-afirma-diretora-da-opas>>. Acesso em: 10 de Setembro 2022.

OPAS. *COVID-19 tem impactos “devastadores” sobre as mulheres, afirma diretora da OPAS*. 2022. Disponível em: <<https://www.paho.org/pt/noticias/26-5-2021-covid-19-tem-impactos-devastadores-sobre-mulheres-afirma-diretora-da-opas#:~:text=Etienne.>>. Acesso em: 10 de Setembro 2022.

PAES, N. A. *Demografia estatística dos eventos vitais: com exemplos baseados na experiência brasileira*. [S.l.]: João Pessoa, Editora do CCTA, 2018.

PAGNO. *Brasil vive explosão de casos de dengue em 2022; ano pode terminar com recorde de mortes pela doença*. 2022. Disponível em: <<https://g1.globo.com/saude/noticia/2022/12/14/brasil-vive-explosao-de-casos-de-dengue>>. Acesso em: 10 de Dezembro 2022.

PEREIRA, J. B. H. *Modelagem da Fecundidade materno-adolescente no semiárido brasileiro*. Repositorio UFPB, p. 38–39, 2020.

PREGIBON, D. *On the Logistic Regression Diagnostics*. *Annals of statistics*, v. 9, n. 4, p. 704–724, 1981.

ROMERO; CUNHA, C. *Avaliação da qualidade das variáveis epidemiológicas e demográficas do Sistema de Informações sobre Nascidos Vivos, 2002*. *Scientific Electronic Library Online*, v. 23, n. 3, p. 700–716, 2007.

SANTOS, R. *O Registro Civil das Pessoas Naturais*. <<http://www.reinaldoveloso.not.br>>, 2006.

SANTOS, T. *INFLUÊNCIA DA PANDEMIA DA COVID-19 SOBRE O CICLO MENSTRUAL E SAÚDE MENTAL DE MULHERES*. p. 11–12, 2022.

SILVA; PAES, N.; CESAR, C. *Efeitos dos programas governamentais e da fecundidade sobre mortalidade infantil do semiárido brasileiro*. *Saúde em Debate*, v. 42, p. 139–147, 2018.

SILVESTRIN, S. et al. *Avaliação da incompletude da variável escolaridade materna nos registros das Declarações de Nascidos Vivos nas capitais brasileiras - 1996 a 2013*. *Cadernos de Saúde Pública*, v. 34, 2018.

SINASC. *Departamento de Análise Epidemiológica e Vigilância de Doenças Não Transmissíveis*. Disponível em: <<https://svs.aids.gov.br/daent/cgiae/sinasc/apresentacao/>>, 2022.

SINASC. *Departamento de Análise Epidemiológica e Vigilância de Doenças Não Transmissíveis SECRETARIA DE VIGILÂNCIA EM SAÚDE*. 2023. Disponível em: <<https://svs.aids.gov.br/daent/centrais-de-conteudos/dados-abertos/sinasc/>>. Acesso em: 10 de Fevereiro 2023.

SOBRAL, M.; ROCHA, L. *Teste de Durbin Watson: Aplicação para Variável data do Evento*. p. 13–15, 2020.

SOUSA. *Explosão demográfica é o crescimento elevado e acelerado da população de uma determinada localidade ou do mundo todo. Pode ser provocada por diferentes motivações*. 2022. Disponível em: <<https://mundoeducacao.uol.com.br/geografia/explosao-demografica.htm>>. Acesso em: 19 de Novembro 2022.

SUDENE. **Delimitação do semiárido - 2021**. <https://www.gov.br/sudene/pt-br/centrais-de-conteudo/02semiariadorelatorionv.pdf>, p. 07–17, 2021.

SZWARCWALD, C. L. et al. **Avaliação das informações do sistema de informação sobre Nascidos Vivos (SINASC)**. *Brasil. Cadernos de Saúde Pública*, v. 35, 2019.

UNFPA . *Fecundidade e dinâmica da população brasileira*. 2022. Disponível em: <[https://brazil.unfpa.org/sites/default/files/pub-pdf/swop\\_brasil\\_web.pdf](https://brazil.unfpa.org/sites/default/files/pub-pdf/swop_brasil_web.pdf)>. Acesso em: 10 de Novembro 2022.

UNICEF. *Declaração Universal dos Direitos Humanos*. 1948. Disponível em: <<https://www.unicef.org/brazil/declaracao-universal-dos-direitos-humanos>>. Acesso em: 10 de Novembro 2022.

VASCONCELOS, A.; GOMES, M. **Transição demográfica: a experiência brasileira**. *Epidemiologia e Serviços de Saúde*. 2012, v. 21, n. 4, p. 10, 2012.

WONG, L. L. R. **A projeção da fecundidade - um exercício aplicado ao Brasil para o período 1991-2020**. *Anais [...]*, p. p.1–21, 2016.

ANEXO - Declaração de Nascidos Vivos

**Automação** FONE/FAX: (51) 3594.4334 • e-mail: roberto@automacoes.com.br • 072757.CDR • AM VISITE NOSSA Home Page: www.automacoes.com.br

**República Federativa do Brasil**  
**Ministério da Saúde**  
 1ª VIA - SECRETARIA DE SAÚDE

**Declaração de Nascido Vivo** **00-00000000-0**

**1** Nome do Recém-nascido (RN) **Número do Cartão Nacional de Saúde do RN**  
 000 0000 0000 000

**2** Data **3** Sexo **4** Raça cor do Recém-nascido  
 Hora  M - Masculino  I - Ignorado  1  Branca  3  Amarela  5  Indígena  
 F - Feminino  2  Preta  4  Parda

**5** Índice de Apgar **6** Detectada alguma anomalia congênita?  
 1º minuto 5º minuto Causa alternativa, usar o código anormalia congênita para descrevê-las  
 1  Sim  2  Não  9  Ignorado

**7** Local da ocorrência **8** Estabelecimento **9** Código CNES  
 1  Hospital  3  Domicílio  5  Aldeia Indígena  9  Ignorado

**10** CEP **11** Bairro/Distrito **12** Município de ocorrência **13** UF

**14** Nome da Mãe **15** Cartão SUS

**16** Escolaridade (última série concluída) **17** Ocupação habitual **18** Código CBO 2002  
 Nível  0 Sem escolaridade  3 Médio (antigo 2º grau)  Ignorado Série  1 Fundamental I (1ª a 4ª série)  4 Superior incompleto  2 Fundamental II (5ª a 8ª série)  5 Superior completo

**19** Data nascimento da Mãe **20** Idade (anos) **21** Naturalidade da Mãe **22** Situação conjugal **23** Raça / Cor da Mãe  
 1 Solteira  4 Separada judicialmente/  
 2 Casada  5 União estável  1 Branca  4 Parda  
 3 Viúva  9 Ignorada  2 Preta  5 Indígena

**24** Logradouro **25** CEP **26** Bairro/Distrito **27** UF  
 Número Complemento

**28** Nome do Pai **29** Idade do Pai

**30** Histórico gestacional **31** Nº gestações anteriores **32** Nº de partos vaginais **33** Nº de cesáreas **34** Nº de nascidos vivos **35** Nº de perdas fetais / abortos

**36** Gestação atual **37** Parto **38** Tipo de parto **39** Cesárea ocorreu antes do trabalho de parto iniciar? **40** Nascimento assistido por  
 Idade Gestacional **41** Data da Última Menstruação (DUM) **42** Número de consultas em que iniciou o pré-natal **43** Mês de gestação em que iniciou o pré-natal **44** Tipo de gravidez **45** Apresentação **46** O Trabalho de parto foi induzido? **47** Tipo de parto **48** Cesárea ocorreu antes do trabalho de parto iniciar? **49** Nascimento assistido por  
 1 Única  1 Cefálica  1 Sim  1 Vaginal  1 Sim  1 Médico  
 2 Dupla  2 Péloco ou Podocóca  2 Não  2 Cesáreo  2 Não  2 Enfermeira/Obstetriz  
 3 Tripla ou mais  3 Transversa  3 Não se aplica  3 Parteira  
 9 Ignorado  9 Ignorado

**41** Descrever todas as anomalias congênicas observadas

**42** Data do preenchimento **43** Nome do responsável pelo preenchimento **44** Função **45** Nascimento assistido por  
 1 Médico  2 Enfermeiro  3 Parteira  4 Func. Cartório  
 5 Outros (descrever)

**46** Tipo documento **47** Nº do documento **48** Órgão emissor

**49** Cartório **50** Registro **51** Data **52** UF  
 Código

**53** Município

**ATENÇÃO: ESTE DOCUMENTO NÃO SUBSTITUI A CERTIDÃO DE NASCIMENTO**

---

# IMPLEMENTAÇÃO COMPUTACIONAL COM SOFTWARE ESTATÍSTICO - R

Listing 5.1: Código Utilizado

```
#Carregando banco

attach(Banco)

names(Banco)
Banco$TFT<-as.factor(Banco$TFT)
na.omit(Banco)

#Obtendo o modelo
#Retirada de forma regressiva,
#Maior pvalor para ate todos
#ficarem abaixo de 0.05

#modelo final
modelo8=glm(as.factor(TFT)~'Apgar 1m 0 2'+ 'Apgar 1m 3 5'+
+'Apgar 1m 6 7'+ 'Apgar 1m 8 10'+ 'Escolaridade fora da idade E'+
+'Escolaridade dentro da idade E'+ 'Estado Civil solteira'+
```

```

‘Estado Civil casada uni o e’+‘Estado Civil viuva separada’
+‘Idade 15 19’+‘Idade 20 24’+‘Idade 25 29’+
‘Idade 30 34’+‘Idade 35 39’+‘Idade 40 44’,binomial(link="logit")
,Banco)

```

```
s8=summary(modelo8)
```

```
s8
```

```
#Observano possivel pontos de alavancagem
```

```
plot(modelo8,which = 5)
```

```
install.packages("MASS")
```

```
library("MASS")
```

```
summary(stdres(modelo8))#possiveis pontos de alavancagem
```

```
#ouliers
```

```
influenceIndexPlot(modelo8)
```

```
#####BONDADE DE AJUSTE#####
```

```
#Desvio escalonado do ajuste para validacao do modelo
```

```
desvio = s8$deviance/s8$dispersion
```

```
q.quadr = qchisq(0.95, s8$df.residual)
```

```
q.quadr
```

```
desvio<q.quadr
```

```
#####Analise de diagnostico#####
```

```
b=modelo8
```

```
library("car")
```

```
library("boot")
```

```
library("nortest")
```

```
fit = fitted(b)
```

```
devres = glm.diag(b)$rd
```

```
#####Residuos#####
```

```
install.packages("lmtest")
```

```
library(lmtest)
```

```
residuos<-residuals(modelo8)
```

```
plot(residuos ,TFT)
```

```
par(mfrow = c(5, 3))
```

```
plot('Apgar 1m 0 2',residuos)
```

```
plot('Apgar 1m 3 5',residuos)
```

```
plot('Apgar 1m 6 7',residuos)
```

```
plot('Apgar 1m 8 10',residuos)
```

```
par(mfrow = c(2, 2))
```

```
plot('Escolaridade dentro da idade E',residuos)
```

```
plot('Escolaridade fora da idade E',residuos)
```

```
plot('Estado Civil casada uni o e',residuos)
```

```
plot('Estado Civil solteira ',residuos)
```

```
plot('Estado Civil viuva separada ',residuos)
```

```
par(mfrow = c(3, 2))
```

```
plot('Idade 15 19',residuos)
```

```
plot('Idade 20 24',residuos)
```

```
plot('Idade 25 29',residuos)
```

```
plot('Idade 30 34',residuos)
```

```
plot('Idade 35 39',residuos)
```

```
plot('Idade 40 44',residuos)
```

```
#####Distancia de cook#####
```

```
cooks_dist <- cooks.distance(modelo8)
```

```
# Plotar o gr fico de dist ncia de Cook
```

```
plot(cooks_dist , pch = 19, cex = 1.2, ylab =
```

```
"Dist ncia_de_Cook", xlab = "Observa o",
```

```
main = "Gr fico_de_Dist ncia_de_Cook")
```

```
#####R2#####
```

```
#Pseudo R2
```

```
install.packages("DescTools")  
library("DescTools")  
PseudoR2(modelo8, which = "Nagelkerke")
```

```
#OUTLIER TEST
```

```
library("car")  
car::outlierTest(modelo8)
```

```
#####Desempenho do modelo#####
```

```
pred<-predict(modelo8, Banco, type = "response")  
result<-as.factor(ifelse(pred > 0.05, 0, 1))  
table(result)
```

```
#desempenho do modelo
```

```
install.packages("caret")  
library("caret")  
confusionMatrix(result, Banco$'TFT CLA', positive = "1")
```

```
#####razao de chances OR
```

```
exp(cbind(OR = coef(modelo8), confint.default(modelo8)))
```

```
#####CURVA ROCK#####
```

```
fit = fitted(modelo8)  
library(ROCR)  
library(pROC)  
library(ggplot2)
```

```
install.packages("ggplot2")
```

```
#Considere em seu modelo o sucesso como Y=1 e o fracasso como Y=0.
```

```
pred <- prediction(fit, Banco$TFT)  
perf <- performance(pred, "tpr", "fpr")  
View(perf)  
area <- performance(pred, "auc")
```

```

plot(perf)
lines(c(0,1), c(1,0))

#area abaixo da curva
roc_data <- roc(TFT, fit)

# Calcule a AUC-ROC
auc <- auc(roc_data)

# Encontre o ponto timo (threshold)
com base na maior sensibilidade e especificidade
optimal_point <- coords(roc_data, "best")

# Plote o gr fico ROC com os pontos timos
plot(roc_data, print.thres = "best", print.auc = TRUE)

#####ODDS RATION#####
install.packages("sjPlot")
library("sjPlot")

#AVALIA A PREDI O COM OS VALORES JA EXISTENTES
marginalModelPlots(b)
library(effects)
plot(allEffects(modelo8))

```